BMJ Open

Facio-cervical Measurements Based Machine Learning Model Predicts Obstructive Sleep Apnea in Chinese Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-048482
Article Type:	Original research
Date Submitted by the Author:	30-Dec-2020
Complete List of Authors:	zhang, liu; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Yan, Ya Ru; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Shi Qi; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Hong Peng; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Lin, Ying Ni; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Ning; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Sun, Xian Wen; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Ding, Yong Jie; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Library Li, Chuan Xiang; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Qing Yun; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory Medicine
Keywords:	PUBLIC HEALTH, SLEEP MEDICINE, Adult anaesthesia < ANAESTHETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Facio-cervical Measurements Based Machine Learning Model Predicts Obstructive Sleep Apnea in Chinese Population

Liu Zhang*[1,2], Ya Ru Yan*[1,2], Shi Qi Li *[1,2], Hong Peng Li [1,2], Ying Ni Lin [1,2], Ning Li [1,2], Xian Wen Sun [1,2], Yong Jie Ding[1,2], Chuan Xiang Li[1,2], Qing Yun Li[1,2]

[1]Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

* co-first authors

Corresponding author full contact details:

Name: Qing Yun Li, M.D, Ph.D.

Address: Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine

Post code: 200025

City: Shanghai

Country: China

Email: liqingyun68@hotmail.com

All authors have seen and approved the manuscript

Financial support: This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082).

No conflicts of interest are to be reported.

ABSTRACT:

Objectives Obstructive sleep apnea (OSA) has received much attention as a risk factor for perioperative complications and 68.5% OSA patients remain undiagnosed before surgery. Facio-cervical characteristics may predict OSA for Asians due to smaller upper airways compared to Caucasians. Thus, our study aimed to explore a machine-learning model to screen for moderate-severe OSA based on facio-cervical and anthropometric measurements.

Design A cross-sectional study.

Setting Data were collected from Shanghai Jiao Tong University School of Medicine affiliated Ruijin Hospital between February 2019 and August 2020.

Participants A total of 481 Chinese participants were included in the study.

Primary and secondary outcome (1) Identification moderate-severe OSA with apnea-hypopnoea index (AHI)≥15 events·h^{-1}. (2) Verification the machine learning model.

Results The SABIHC2 model (***S***ex-***A***ge-***B***ody mass index-maximum ***I***nterincisal distance-ratio of ***H***eight to thyro-sternum distance-neck ***C***ircumference-waist ***C***ircumference) was set up. The SABIHC2 model could predict moderate-severe OSA with area ender the curve (AUC)=0.787, sensitivity of 87.4% and specificity of 70.0%, and performed better than the STOP-BANG questionnaire, which showed AUC=0.597, sensitivity of 49.5% and specificity of 78.0%. Especially for asymptomatic patients (ESS < 10), the SABIHC2 model demonstrated better predictive ability compared to the STOP-BANG questionnaire, with sensitivity (0.915 *vs.* 0.404), specificity (0.704 *vs.* 0.788) and AUC (0.809 *vs.* 0.571).

Conclusion The SABIHC2 machine learning model provides a simple and accurate assessment of moderate-severe OSA, especially for those without significant daytime sleepiness.

Keywords: obstructive sleep apnea, machine learning, anthropometric, screening

Strengths and limitations of this study

- This is the first study to propose and validate facio-cervical measurements, which have been widely used in predicting difficult intubation, are important risk factors for screening moderate-severe OSA in Chinese subjects.
- Our SABIHC2 model, which is based on facio-cervical and anthropometric measurements, can be an effective and objective screening tool for OSA, independent of symptoms and comorbidities.
- The results obtained are hard to generalise for other populations, because the study cohort was based on Chinese population.
- The sample size was relatively small, and it was a single-centre study, which may affect the validation of the machine-learning algorithm model.

INTRODUCTION

Obstructive sleep apnea (OSA) is a common breathing sleep disorder that affects about 936 million adults globally,[1] of which about 80% are undiagnosed patients.[2] In Chinese population, 175 million adults have mild to severe obstructive sleep apnea and 65 million adults have moderate to severe obstructive sleep apnea.[1] In addition to cardiovascular injury and metabolic syndrome, OSA has recently been regarded as a risk factor for perioperative complications, including hypoxaemia, pneumonia, pulmonary embolism, unplanned transfer to the intensive care unit, and even death, especially for those who accepted abdominal or vascular surgery.[3] 68.5% of OSA patients were not diagnosed before surgery.[4] Guidelines have been recommended to develop a local protocol for screening possible OSA patients before elective surgery.[5,6] The most common screening scales, including STOP-BANG, Berlin questionnaire, and OSA50, are mainly based on symptoms and comorbidities,[7] which might lead to missed diagnosis for those without significant daytime sleepiness. Several factors contribute to the pathogenesis of OSA, including obesity, facio-cervical anatomy, and alteration in pharyngeal muscle function, etc.[8] Asians have relatively smaller upper airways compared to Caucasians.[9,10] Facio-cervical characteristics may predict OSA for Asians, such as thyromental distance (TMD), thyro-sternum distance (TSD), maximum interincisal distance (MID), and Mallampati test score, which have been widely used in predicting difficult intubation.[11] However, no studies assessed whether facio-cervical characteristics are suitable for predicting OSA. Thus, it is necessary to evaluate the relationship between facio-cervical characteristics and OSA.

During the past two decades, machine learning models have provided simple but effective approaches for improving diagnostic accuracy.[12] Support vector model (SVM) is a well-known classification technique and has achieved great success in bioinformatics applications. Thus, the study aimed to build a new model via SVM and monitor its effectiveness in predicting OSA.

METHODS

Study design

Participants with suspected OSA (snoring, witnessed apnea, or excessive daytime sleepiness, etc.) were enrolled in the study from Ruijin Hospital, Shanghai Jiao Tong University School of Medicine

between February 2019 and August 2020. All participants underwent anthropometric measurements and subsequent overnight polysomnography (PSG). Participants were divided into a deal group (apnea-hypopnoea index, AHI<15 events·h^{-1}) and a no-deal group (AHI≥15 events·h^{-1}). For proper set-up of the machine learning model, all participants were randomly partitioned into a training set and a validation set according to a ratio of 7:3. Exclusion criteria: (1) with severe cognitive impairment, (2) with severe heart failure, respiratory failure and other serious acute or chronic diseases, (3) neuromuscular diseases. The study was approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine at Ruijin Hospital (Protocol #2018-107). Written informed consent was obtained from all participants. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Clinical characteristics and anthropometric measurements

Age, sex, height, weight, BMI, neck circumference (NC), waist circumference (WC), Epworth Sleepiness Scale (ESS), and STOP-BANG data were recorded. Facio-cervical measurements including Mallampati score, MID, TMD, and TSD were measured. The Mallampati score was evaluated when participants were asked to sit upright and open their mouths as wide as possible. In grade I, the entire uvula, faucial pillar, and soft palate are visible. In grade II, part of the uvula and palate are visible. In grade III, the soft palate is visible, but the uvula is obscured.[13] MID was recorded by asking the participant to sit upright and open the mouth as wide as possible.[14] The TMD was measured as the straight distance between the thyroid notch and the lower border of the mental prominence, while the head was fully extended, and the mouth closed (Figure 1).[15] TSD was measured as the distance between the thyroid notch and the upper border of the sternum (Figure 1). The distance was rounded to the nearest 0.5 cm. The ratios of height to TMD (H/TMD) and height to TSD (H/TSD) were calculated.[15]

PSG

All participants underwent overnight PSG. No coffee, tea, cola-containing products or sedative-hypnotics were taken before sleep. The PSG monitoring included electroencephalography, electrooculography, chin electromyography, electrocardiography, measurements of ribcage and

abdominal movements, and airflow pressure and thermistor (Alice 5, Philips Respironics, USA), with the addition of oxygen saturation (Nonin, Herrsching, Germany). Sleep data were scored manually by registered polysomnographer of the United States according to American Academy of Sleep Medicine Recommendations (AASM).[16] The definition of arousal was based on the AASM guidelines. The diagnosis of moderate-severe OSA was defined by the presence of an obstructive AHI≥15/h according to AASM guidelines.

Model construction

Covariates

Differences in variables between the deal group and the no-deal group were assessed using t-test or chi-square test. Variables with $p<0.05$ were chosen as covariates to set up the machine-learning model. The effective of covariates were further evaluated by crude odds ratios (ORs) and adjusted OR by controlling for BMI. Statistical analysis was conducted using IBM SPSS software (version 24.0, IL, USA).

Model construction via SVM

SVM is a representative machine-learning algorithm for classification, which could be viewed as a nonlinear regression model. We used the SVM to capture the potential hyperplane that maximises the margin between the deal group and the no-deal group. The SVM has the following format:

$$C = \Sigma_{i=1}^{n} \alpha_i k(x_i, x) + b$$

Where c is the output of the model based on the new data x, which could be regarded as the classifier (0 or 1), x_i (i=1,…,n) is the training dataset, α_i and b are parameters determined by the algorithm.[17] We used the Gaussian radial basis function as the kernel function, with γ of 0.33 and the box constraint set to 10.

Based on the results of the statistical analysis, sex, age, NC, WC, BMI, MID, and H/TSD were set as independent variables ($p<0.05$). For the problem of multicollinearity, each participant was normalised and the principal component analysed before evaluating the SVM model. The SVM model, namely SABIHC, i.e. ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 C*s)** was estimated by the training dataset

(n=337, 70%); this model was then applied to predict OSA in the testing dataset (n=144, 30%). The flow chart including screening, randomisation and algorithm is presented in Figure 2.

The screening performance was assessed by means of sensitivity, specificity and area under the curve (AUC), negative likelihood ratio (-LR) and positive likelihood ratio (+LR) and accuracy. Visual optimal hyperplanes for the Support Vector Machine are shown in Figure 3. Finally, we compared the difference in predictive power between the STOP-BANG questionnaire and the SVM model.

Analyses were performed using the Scikit-learn package (0.20.3) based on Python (3.5.1).

Table 1 Subject demographics, crude and adjusted associations between morphometric variables and OSA

	AHI<15 (n=171)	AHI>15 (n=310)	P*	crude OR (95% CI)	adjusted OR for BMI (95% CI)
General characteristics					
Male (%)	93(54.39%)	232(74.84%)	<0.001	0.401(0.270-0.595)	0.463(0.308-0.697)
Age	44.05±13.80	49.11±12.08	<0.001	1.032(1.016-1.048)	1.037(1.021-1.054)
NC	37.63±3.74	39.83±4.20	<0.001	1.146(1.090-1.205)	1.121(0.054-1.193)
WC	90.58±11.93	97.39.76±10.20	<0.001	1.109(1.041-1.081)	1.074(1.041-1.107)
BMI	25.18±4.16	26.85±4.04	<0.001	1.109(1.056-1.166)	NA
Facio-cervical measurements					
MID	4.76±0.79	4.94±0.92	0.028	1.275(1.025-1.585)	1.298(1.037-1.626)
Mallampati test=1	77(45.02%)	118(38.06%)			
Mallampati test=2	54(31.58%)	88(28.39%)	0.120	0.589(0.371-0.938)	0.679(0.421-1.095)
Mallampati test=3	40(23.40%)	104(33.55%)		0.627(0.381-1.031)	0.713(0.428-1.186)
H/TMD	5.43±0.66	5.34±0.78	0.185	1.849(0.659-1.095)	0.856(0.661-1.108)
H/TSD	5.80±0.75	5.18±0.74	<0.001	0.292(0.212-0.404)	0.282(0.205-0.393)

Sex, age, body mass index, neck circumference, waist circumference, maximum interincisal distance, and height to thyro-sternum distance ratio were chosen to build machine-learning model, due to the significant difference between the two groups (p<0.05).

Data are presented as mean± SD or n (%). *: t test or chi-square test as appropriate. NA: not available.

Abbreviation: NC; neck circumference; WC: waist circumference; BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance.

RESULTS

A total of 512 participants were recruited for the study at first, among which 31 patients were excluded due to short total sleep time (<200 min, 13 cases) or missing data (18 cases). Finally, 481 participants (325 males and 156 females; aged between 14-77 y) were enrolled in our study (Figure 2). The participants were divided into the no-deal (AHI<15 events·h^{-1}, n=171, mean AHI 42.4 ± 20.7 events·h^{-1}) and the deal group (AHI≥15 events·h^{-1}, n=310, mean AHI 7.01 ± 4.47 events·h^{-1}).

Increased BMI, NC, WC, older age, and higher percentage of males were found in deal group (p<0.001). Meanwhile, they had higher MAI (32.5 ±13.2 vs 16.4 ± 8.3) and lower LSpO2 (72.1 ± 13.3 vs 87.13 ± 6.3) than those in the no-deal group (p<0.001) (Table 1).

Associations Between Facio-cervical Characteristics and OSA

Greater MID (4.94 vs 4.76, p=0.028) and smaller H/TSD (5.18 vs 5.80, p<0.001) were found in the deal group compared to no-deal group. MID (OR=1.275, adjusted OR=1.298) and H/TSD (OR=0.292, adjusted OR=0.282) are independent risk factors for OSA, even after controlling for BMI. No significant inter-group differences were found in Mallampati test scores and H/TMD (p>0.05) (Table 1).

Set-up and Predictive Accuracy of the SABIHC2 model

Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were chosen as covariates as a significant difference was found between the two groups. Of all participants, 337 patients (70%) were randomised to the training dataset and 144 patients (30%) were randomised to the testing dataset. We set up the SABIHC2 model based on training dataset (n=337), whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 Cs***). To predict moderate-severe OSA, the accuracy of SABIHC2 model was 90.1%, of which 92.58% for the deal group (blue dots in purple area), and 72.0% for no-deal group (green dots in green area; Figure 3).

Then, the model was verified using the testing dataset (n=144) to predict moderate-severe OSA. The AUC was 0.874 (95% CI, 0.702 to 0.872), with a sensitivity of 87.4% and specificity of 70.0%. Corresponding LR+ and LR- were 2.91 and 0.18, respectively (Table 2, Figure 4A).

Table 2 Performance of SABIHC2 model

		Sensitivity	Specificity	AUC	95% CI	+LR	-LR
SABIHC2	training dataset	0.949	0.727	0.838	0.787-0.889	3.480	0.070
	testing dataset	0.874	0.700	0.787	0.702-0.872	2.913	0.180
STOP-BANG	testing dataset	0.684	0.500	0.597	0.499-0.694	0.632	1.368

Abbreviations: SABIHC2 = Sex-age-BMI-MID-H/TSD-NC-WC. +LR: positive likelihood ratio; -LR: negative likelihood ratio

Discriminative Ability of the SABIHC2 model and STOP-BANG questionnaire

To compare the predictive ability for moderate-severe OSA between the SABIHC2 model and STOP-BANG questionnaire, we calculated the ROC curve of STOP-BANG for moderate-severe OSA on the testing dataset. The AUC based on testing dataset was 0.597 (95% CI, 0.499 to 0.694), with a sensitivity of 68.4% and specificity of 50.0%. (Table 2, Figure 4A).

As the STOP-BANG questions are based on the symptoms and comorbidities and may results in missed diagnoses, we further compared whether there are differences in the predictive capacity of SABIHC2 model and STOP-BANG questionnaire both for the symptomatic (ESS ≥ 10) and asymptomatic patients (ESS < 10) within the testing dataset. For asymptomatic patients, the SABIHC2 model demonstrated better predictive ability than STOP-BANG questionnaire [AUC (0.809 *vs.* 0.571)], with sensitivity (0.915 *vs.* 0.404), and specificity (0.704 *vs.* 0.788). Similarly, SABIHC2 model had higher predictive power than STOP-BANG questionnaire in patients experiencing sleepiness (ESS ≥ 10) [AUC (0.764 *vs.* 0.606)], with sensitivity (0.833 *vs.* 0.813), and specificity (0.696 *vs.* 0.478) (Figure 4B).

DISCUSSION

In the current study, we measured a broad range of anthropometric variables and assessed the association with moderate-severe OSA. Sex, age, BMI, MID, H/TSD, NC, and WC were significant determining factors. We also developed a SABIHC2 predictive model, which provides a simple and accurately assessment of OSA.

Facio-cervical measurements are essential risk factors for OSA. The majority of previous studies focused on the soft tissue of the upper airway. A significant correlation between the oropharyngeal soft tissue (Mallampati test and tonsillar enlargement, etc.) and AHI was reported.[18,19] However, Mallampati test was not a significant parameter in our study. It is possible that the association between soft tissue and OSA is weak in Asian populations.[10] No previous studies explored the relationship between moderate-severe OSA and facio-cervical measurements, such as MID, H/TMD, and H/TSD, which have been frequently used to evaluate difficult intubation.[11] Herein, we found that MID and H/TSD were strongly associated with moderate-severe OSA, and H/TSD showed a remarkable predictive power for OSA, even controlling for BMI. The result may suggest that H/TSD is a potential predictor for moderate-severe OSA in Chinese subjects.

As indicated previously, the overall scarcity and labour- and financially-onerous nature of PSG has prompted exploration of other suitable screening approaches, ranging from questionnaires to simplified multichannel recording. The STOP-BANG is one of the most widely used questionnaires. Some limitations of the STOP-BANG questionnaire should be considered. Firstly, it is based on simplified categories (0 or 1) and this might reduce its screening accuracy.[20] Secondly, the evaluation is based on symptoms and comorbidities, which may lead to omission of diagnosis for snorers without significant daytime sleepiness or hypertension.[21] Thus, we developed a new screening tool based on facio-cervical measurements by machine learning.

Considering non-linear relationships and data interaction effects,[22] the machine learning algorithm (SVM) was used to find the complex linkages between variables. Distinct from previous machine learning models that were based on age, sex, NC, WC, and BMI,[23,24] we included the structural facio-cervical measurements to construct the SABIHC2 model. This showed much better performance than the STOP-BANG in screening for OSA.

The results confirmed that the SABIHC2 model performed better than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for asymptomatic patients (ESS < 10), demonstrating that it may be particularly useful for this cohort of patient. With increasing spread of smartphones, the SABIHC2 algorithm could be easily used in pre-operational screening for OSA via mobile application.

To our knowledge, this is the first study examining the predictive power of facio-cervical measurements, which may become a new index to predict OSA besides sex, age, and BMI. However, several limitations should be mentioned. First, the study cohort was limited to the Chinese population. The results obtained are hard to generalise for other populations. Second, the cross-sectional design of our study makes it difficult to establish the causal relationship between variables. Third, sample size was relatively small, and it was a single-centre study, which may affect the validation of the machine-learning algorithm model.

In conclusion, we confirmed that facio-cervical measurements are important risk factors for screening moderate-severe OSA, and developed a validated machine learning algorithm that is based on facio-cervical and anthropometric measurements, called the SABIHC2 predictive model. The model is more effective than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for those without significant daytime symptoms.

REFERENCES

1.Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698.

2. Young T, Evans L, Finn L, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep.* 1997;20(9):705-706.

3.Kaw R, Pasupuleti V, Walker E, et al. Postoperative complications in patients with obstructive sleep apnea. *Chest.* 2012; 141: 436-41.

4.Singh M, Liao P, Kobah S, et al. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. *Br J Anaesth*. 2013 Apr;110(4):629-36.

5.American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology.* 2014;120(2):268-286. Available: https://pubmed.ncbi.nlm.nih.gov/24346178/

6. Dalesio NM, Hendrix CW, McMichael DH, et al. Effects of Obesity and Leptin Deficiency on Morphine Pharmacokinetics in a Mouse Model. *Anesth Analg.* 2016 Dec;123(6):1611-1617.

7.Kee K, Dixon J, Shaw J, et al. Comparison of Commonly Used Questionnaires to Identify Obstructive Sleep Apnea in a High-Risk Population. *J Clin Sleep Med.* 2018; 14: 2057-64.

8.Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers.* 2015;7(1):15015.

9.Sutherland K, Lee RW, Cistulli PA. Obesity and craniofacial structure as risk factors for obstructive sleep apnoea: impact of ethnicity. *Respirology*. 2012; 17: 213-22.

10.Schorr F, Kayamori F, Hirata RP, et al. Different Craniofacial Characteristics Predict Upper Airway Collapsibility in Japanese-Brazilian and White Men. *Chest*. 2016; 149: 737-46.

11.Roth D, Pace NL, Lee A, et al Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018; 5: CD008874. Available: http:// doi:10.1111/ijn.12805

12.Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016; 375: 1216-9.

13.Menon SM, Sampangiramaiah S, Mathew M. Cross Sectional Observational Study Performed to See for Relation of Mallampati Score and Extended Mallampati Score with Body Mass Index. *J Clin Diagn Res* 2017; 11: Ug01-ug03.

14.Den Herder C, Schmeck J, Appelboom DJ, et al.Risks of general anaesthesia in people with obstructive sleep apnoea. *BMJ* 2004; 329: 955-9.

15.Schmitt HJ, Kirmse M, Radespiel-Troger M. Ratio of patient's height to thyromental distance improves prediction of difficult laryngoscopy. *Anaesth Intensive Care*. 2002; 30: 763-5.

16.The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep.* 1999; 22: 667-89.

17.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. Available: http://doi:10.1371/journal.pone.0176991

18.Schellenberg JB, Maislin G, Schwab RJ. Physical findings and the risk for obstructive sleep apnea. The importance of oropharyngeal structures. *Am J Respir Crit Care Med.* 2000; 162: 740-8.

19.Schwab RJ, Leinwand SE, Bearn CB, et al Digital Morphometrics: A New Upper Airway Phenotyping Paradigm in OSA. *Chest* 2017; 152: 330-42.

20.Chung F, Abdullah HR, Liao P. STOP-BANG Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. Chest. 2016;149(3):631-638.

21.Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. *Eur Respir J.* 2014; 44:1600–7.

22.Linnen DT, Escobar GJ, Hu X, et al.Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. *J Hosp Med.* 2019; 14:161–9.

23.Mencar C, Gallo C, Mantero M, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics J.* 2019;1460458218824725. Available: http://doi:10.1177/1460458218824725

24.Bozkurt S, Bostanci A, Turhan M. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf Med.* 2017; 56:308–18.

Author affiliations

[1]Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Acknowledgements The authors thank Peizhe Liu for his assistance with the SVM algorithm.

Author contributions Liu Zhang and Ya Ru Yan drafted the manuscript, had full access to all the data, and takes responsibility for the integrity of the data and the accuracy of the data analysis. Shi Qi Li contributed to data analysis and interpretation. Qing Yun Li and Shi Qi Li contributed to study conception and design. Hong Peng Li, Ying Ni Lin, Ning Li, Xian Wen Sun, Yong Jie Ding contributed to data collection, critical revision. Hong Peng Li and Qing Yun Li contributed to final approval of the manuscript.

Funding This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082). No conflicts of interest are to be reported

Competing interests None declared.

Patient and Public Involvement statement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Ethics approval The study was approved by the ethics committee of the Shanghai Jiao Tong University School of Medicine at Ruijin Hospital, with the following reference number: 2018-107.

Data availability statement Data are available upon reasonable request.

FIGURE LEGENDS:

Figure 1. TMD and TSD measurements.

Abbreviations: TMD: the straight distance between the thyroid notch and the lower border of the mental prominence; TSD: the distance between the thyroid notch and the upper border of the sternum

Figure 2. Flow chart showing screening, randomization and algorithm.

481 participants were enrolled in our study after excluding 49 patients. Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were potential risk factors for OSA due to the significant difference between the two groups. We chose the following parameters to setup the SABIHC2 model based on training dataset, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s). Then, the model was verified on testing dataset.

Abbreviations: BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance; NC: neck circumference; WC: waist circumference;

Figure 3. Optimal hyperplane for Support Vector Machine (SVM). The green area means the SVM predicting AHI ＜15 events·h^{-1}, the purple means the SVM predicts the AHI≥15 events·h^{-1}. Green dots represent no-deal groups (AHI＜15 events·h^{-1}), blue dots represent deal groups(AHI≥15 events·h^{-1}). The accuracy of SABIHC2 model was 90.1%, of which 92.58 % for the deal group (blue dots in purple area), and 72.0 % for no-deal group (green dots in green area). The boundary was obtained from the SVM classifier and the figure was created using Python.

Abbreviations: SVM: support vector machine; SABIHC2 model: ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s).

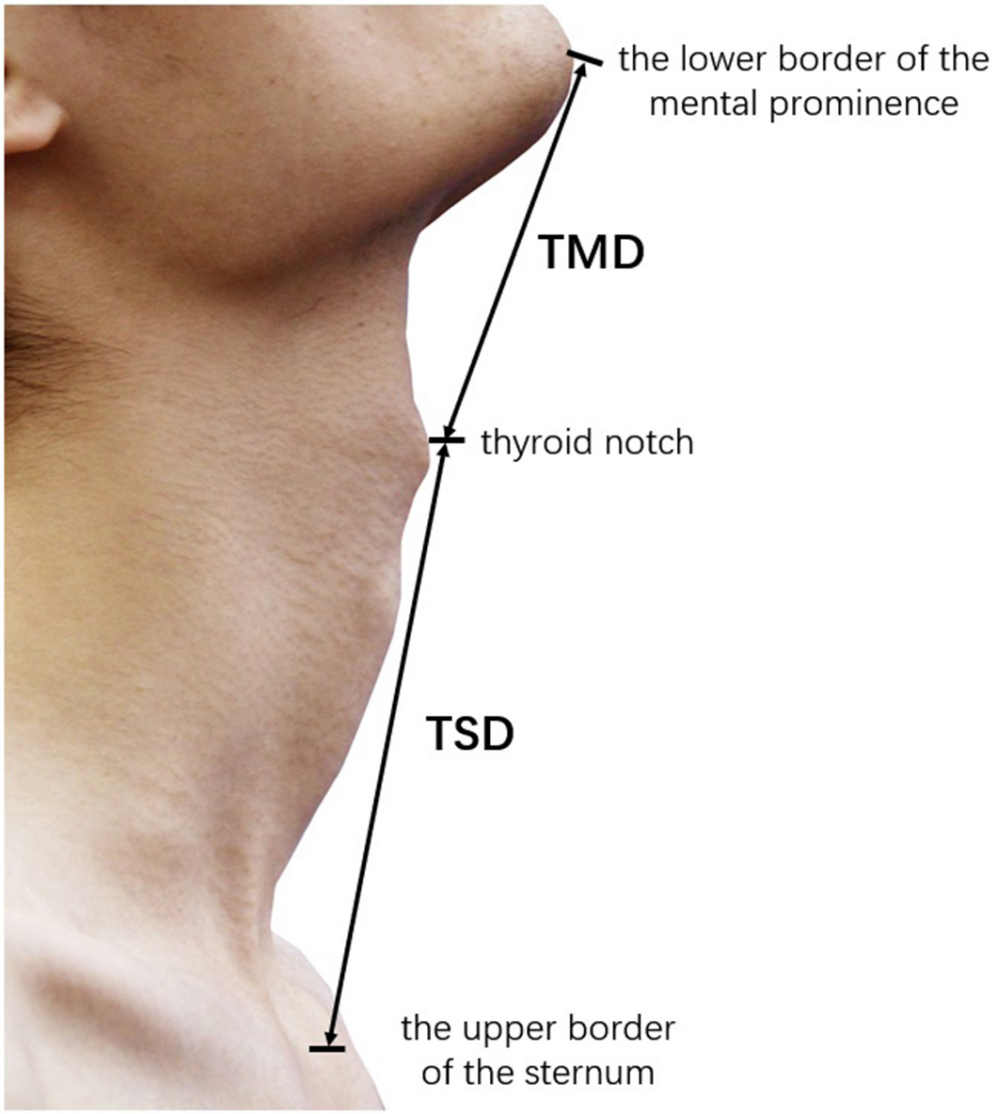
Figure 4. ROC curve of cross-validation.

A: The ROC curve of SABIHC2 model and STOP-BANG. The AUC of the SABIHC2 model based on the training database (n=336, 69.9%, brown line) was 0.949 (95% CI, 0.787 to 0.889), with a sensitivity of 94.9% and specificity of 72.7%. About testing database (n=145 30.1%, red line), the AUC was 0.874 (95% CI, 0.702 to 0.872), with a sensitivity of 87.4% and specificity of 70.0%. The ROC curve of STOP-BANG questionnaire (blue line) on the testing database was 0.597 (95% CI, 0.499 to 0.694). STOP-BANG>2 showed the sensitivity and specificity of 83.2% and 22.0%, respectively, while STOP-BANG>4 with the sensitivity of 49.5% and specificity of 78.0%.

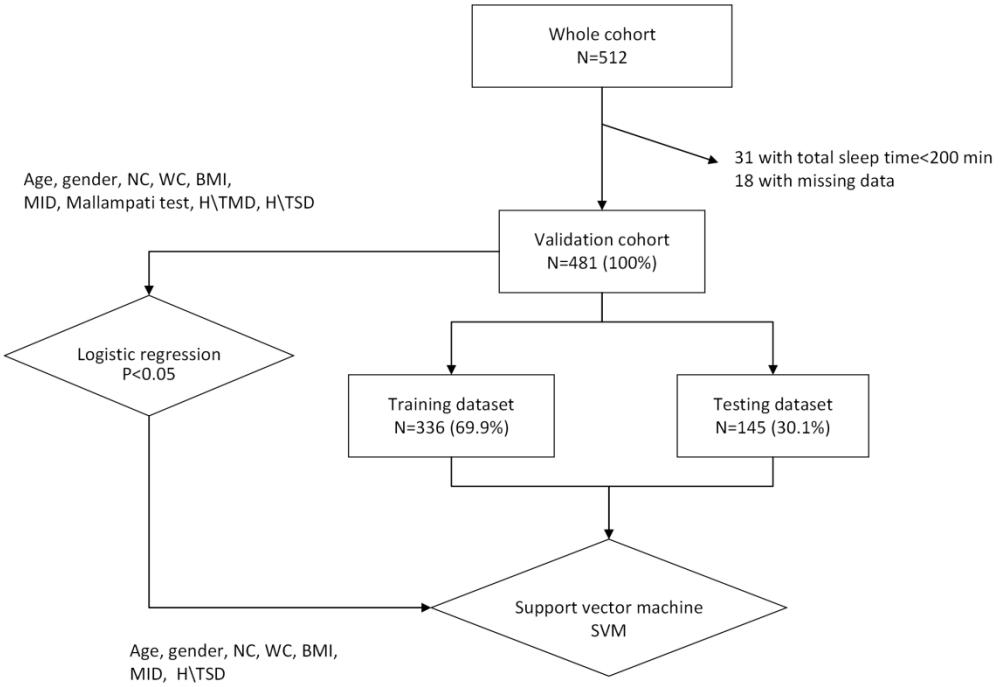
Blue: SVM based on training dataset; red: SVM based on testing dataset; brown: STOP-BANG based on testing dataset.

B: The ROC curve of SABIHC2 model and STOP-BANG based on asymptomatic patients (ESS < 10) and sleepiness patients (ESS ≥ 10). In asymptomatic patients, SABIHC2 model (red line) remarkably demonstrated better predictive ability than STOP-BANG questionnaire (dark blue line), with sensitivity (0.915 vs 0.404), specificity (0.704 vs 0.788) and AUC (0.809 vs 0.571). Similarly, SABIHC2 model (orange line) had higher predictive power than STOP-BANG questionnaire (soft blue line) in sleepiness patients (ESS ≥ 10), with sensitivity (0.833vs 0.813), specificity (0.696 vs 0.478) and AUC (0.764 vs 0.606)

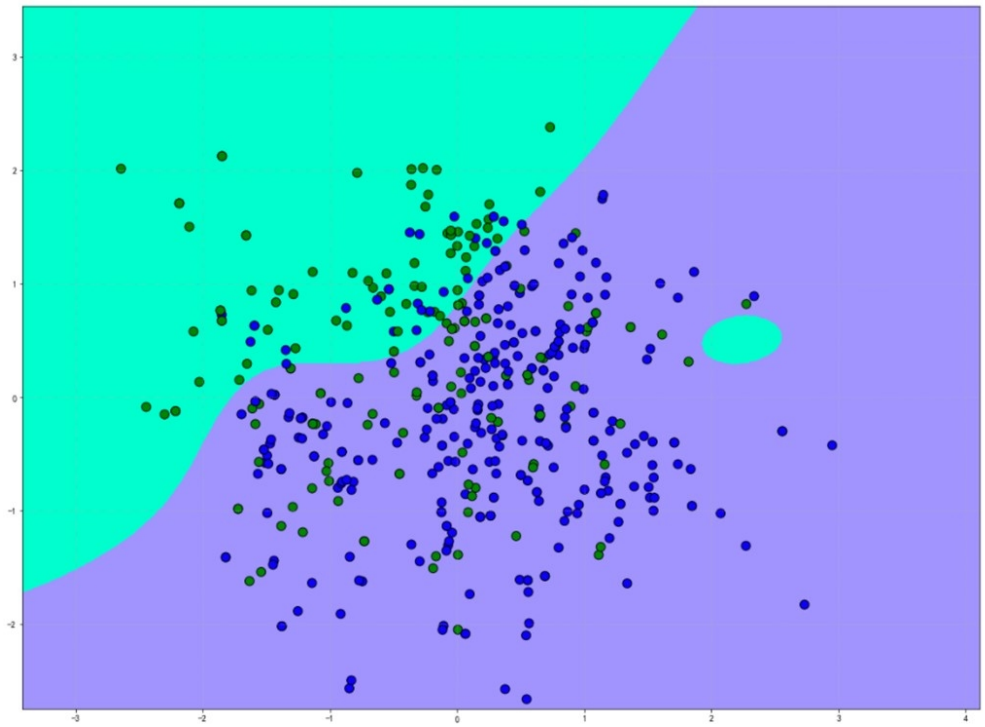
Abbreviations: AUC: the area under ROC curve; SVM: support vector machine; CIs: confidence intervals.



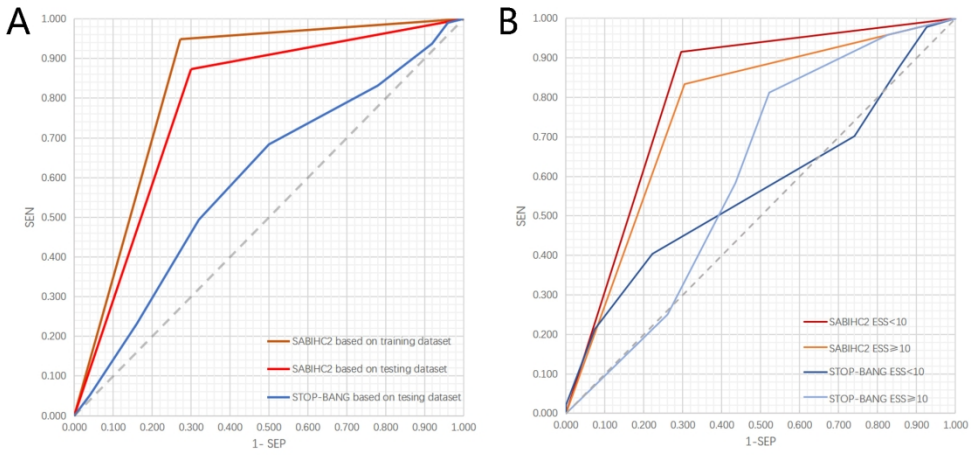
90x100mm (300 x 300 DPI)



189x131mm (300 x 300 DPI)



86x63mm (300 x 300 DPI)



189x100mm (300 x 300 DPI)

BMJ Open

Screening value of Facio-cervical Measurements Based Machine Learning Model on moderate-severe OSA in Chinese Population

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-048482.R1
Article Type:	Original research
Date Submitted by the Author:	12-Apr-2021
Complete List of Authors:	zhang, liu; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Yan, Ya Ru; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Shi Qi; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Hong Peng; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Lin, Ying Ni; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Ning; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Sun, Xian Wen; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Ding, Yong Jie; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Chuan Xiang; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital Li, Qing Yun; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Anaesthesia
Keywords:	PUBLIC HEALTH, SLEEP MEDICINE, Adult anaesthesia < ANAESTHETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Screening value of Facio-cervical Measurements Based Machine Learning Model on moderate-severe OSA in Chinese Population

Liu Zhang*[1,2], Ya Ru Yan*[1,2], Shi Qi Li *[1,2], Hong Peng Li [1,2], Ying Ni Lin [1,2], Ning Li [1,2], Xian Wen Sun [1,2], Yong Jie Ding[1,2], Chuan Xiang Li[1,2], Qing Yun Li[1,2]

[1]Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

* co-first authors

Corresponding author full contact details:

Name: Qing Yun Li, M.D, Ph.D.

Address: Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

Post code: 200025

City: Shanghai

Country: China

Email: liqingyun68@hotmail.com

All authors have seen and approved the manuscript

Financial support: This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082).

No conflicts of interest are to be reported.

ABSTRACT:

Objectives Obstructive sleep apnea (OSA) has received much attention as a risk factor for perioperative complications and 68.5% OSA patients remain undiagnosed before surgery. Facio-cervical characteristics may screen OSA for Asians due to smaller upper airways compared to Caucasians. Thus, our study aimed to explore a machine-learning model to screen moderate-severe OSA based on facio-cervical and anthropometric measurements.

Design A cross-sectional study.

Setting Data were collected from the Shanghai Jiao Tong University School of Medicine affiliated Ruijin Hospital between February 2019 and August 2020.

Participants A total of 481 Chinese participants were included in the study.

Primary and secondary outcome (1) Identification moderate-severe OSA with apnea-hypopnea index (AHI)15 events·h^{-1}. (2) Verification the machine learning model.

Results The SABIHC2 model (***S***ex-***A***ge-***B***ody mass index-maximum ***I***nterincisal distance-ratio of ***H***eight to thyro-sternum distance-neck ***C***ircumference-waist ***C***ircumference) was set up. The SABIHC2 model could screen moderate-severe OSA with area under the curve (AUC)=0.787, sensitivity of 0.874 and specificity of 0.700, and performed better than the STOP-BANG questionnaire, which showed AUC=0.597, sensitivity of 0.495 and specificity of 0.780. Especially for asymptomatic patients (ESS < 10), the SABIHC2 model demonstrated better predictive ability compared to the STOP-BANG questionnaire, with sensitivity (0.915 *vs.* 0.404), specificity (0.704 *vs.* 0.788) and AUC (0.809 *vs.* 0.571).

Conclusion The SABIHC2 machine learning model provides a simple and accurate assessment of moderate-severe OSA, especially for those without significant daytime sleepiness.

Keywords: obstructive sleep apnea, machine learning, anthropometric, screening

Strengths and limitations of this study

- The study aims to propose and validate the association between facio-cervical measurements and moderate-severe OSA in Chinese subjects.
- This is the first study to set up a machine learning model based on facio-cervical measurements to predict moderate-severe OSA in Chinese population.
- Instead of most questionnaires based on symptoms and comorbidities, the SABIHC2 model which is entirely based on measurements, avoids the confounding effect of symptoms and comorbidities.
- The cross-sectional design of the study makes it difficult to establish the causal relationship between variables.
- The study is a single-centre study with relatively small sample size, and further confirmation with a large-sample, multi-centre, prospective clinical trial is needed in the future.

INTRODUCTION

Obstructive sleep apnea (OSA) is a common breathing sleep disorder that affects about 936 million adults globally,[1] of which about 80% are undiagnosed patients.[2] In Chinese population, 175 million adults have mild to severe obstructive sleep apnea and 65 million adults have moderate to severe obstructive sleep apnea.[1] In addition to cardiovascular injury and metabolic syndrome, OSA has recently been regarded as a risk factor for perioperative complications, including hypoxaemia, pneumonia, pulmonary embolism, unplanned transfer to the intensive care unit, and even death, especially for those who received abdominal or vascular surgery.[3] The rates of postoperative cardiovascular events show a rise in moderate or severe OSA (25.1%) compared to no or mild OSA(16.8%).[4] 38% surgical patients had moderate or severe OSA, of which 68.5% were not diagnosed before surgery.[5] Guidelines have been recommended to develop a local protocol for screening possible OSA patients before elective surgery.[6,7] The most common screening scales, including STOP-BANG, Berlin questionnaire, and OSA50, are mainly based on symptoms and comorbidities,[8] which might lead to missed diagnosis for those without significant daytime sleepiness. Several factors contribute to the pathogenesis of OSA, including obesity, facio-cervical anatomy, and alteration in pharyngeal muscle function, etc.[9] Asians have relatively smaller upper airways compared to Caucasians.[10] Only in Asians, smaller upper airways are predictors of upper airway collapsibility, and an anatomic imbalance between tongue and mandible volume influenced upper airway collapsibility among Caucasians. [11] The above evidence prompts facio-cervical characteristics may predict OSA for Asians, such as thyromental distance (TMD), thyro-sternum distance (TSD), maximum interincisal distance (MID), and Mallampati test score, which have been widely used in predicting difficult intubation.[12] However, no studies assessed whether facio-cervical characteristics are suitable for predicting OSA. Thus, it is necessary to evaluate the relationship between facio-cervical characteristics and OSA.

During the past two decades, machine learning models have provided simple but effective approaches for improving diagnostic accuracy.[13] Support vector model (SVM) is a well-known classification technique and has achieved great success in bioinformatics applications. Thus, the study aimed to build a new model via SVM and monitor its effectiveness in screening OSA.

METHODS

Study design

Participants with suspected OSA (snoring, witnessed apnea, or excessive daytime sleepiness, etc.) were enrolled in the study from Ruijin Hospital, Shanghai Jiao Tong University School of Medicine between February 2019 and August 2020. All participants underwent anthropometric measurements and subsequent overnight polysomnography (PSG). Participants were grouped according to their AHI gain into the following: 1) no or mild OSA (subjects without moderate-to-severe OSA: AHI <15 events·h^{-1}), 2) moderate or severe OSA (moderate to severe OSA: AHI ≥ 15 events·h^{-1}). Exclusion criteria: (1) Patients showing complications with severe respiratory diseases, such as severe COPD, interstitial lung disease or acute asthma; (2) Patients showing complications with serious cardiovascular diseases such as acute myocardial infarction, acute heart failure or chronic congestive heart failure (Grade III and IV); (3) Patients with mental illnesses who could not cooperate with the examination; (4) Patients who receiving non-invasive positive pressure ventilation therapy; (5) Patients who might had other sleep disorders under clinical evaluation. For proper set-up of the machine learning model, all participants were randomly partitioned into a training set and a testing set according to a ratio of 7:3. The study was approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine at Ruijin Hospital (Protocol #2018-107). Written informed consent was obtained from all participants. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Clinical characteristics and anthropometric measurements

Age, sex, height, weight, BMI, neck circumference (NC), waist circumference (WC), Epworth Sleepiness Scale (ESS), and STOP-BANG questionnaire were recorded. The STOP-BANG questionnaire is a scoring model consisting of eight questions and its scores are based on Yes/No answers (score: 1/0). The eight questions included snoring, tiredness, observed apnea and high blood pressure, BMI, age, neck circumference and gender. Facio-cervical measurements including Mallampati score, MID, TMD, and TSD were measured. The Mallampati score was evaluated when participants were asked to sit upright and open their mouths as wide as possible. In grade I, the entire

uvula, faucial pillar, and soft palate are visible. In grade II, part of the uvula and palate are visible. In grade III, the soft palate is visible, but the uvula is obscured.[14] MID was recorded by asking the participant to sit upright and open the mouth as wide as possible.[15] The TMD was measured as the straight distance between the thyroid notch and the lower border of the mental prominence, while the head was fully extended, and the mouth closed.[16] TSD was measured as the distance between the thyroid notch and the upper border of the sternum. The distance was rounded to the nearest 0.5 cm. The ratios of height to TMD (H/TMD) and height to TSD (H/TSD) were calculated.[16]

PSG

All participants underwent overnight PSG. No coffee, tea, caffeine-containing products, or sedative-hypnotics were taken before sleep. The PSG monitoring included electroencephalography, electrooculography, chin electromyography, electrocardiography, measurements of thoracal and abdominal movements, and airflow pressure and thermistor (Alice 5, Philips Respironics, USA), with the addition of oxygen saturation (Nonin, Herrsching, Germany). Sleep recordings were scored according to the American Academy of Sleep Medicine (AASM) 2007 criteria.[17] The diagnosis of moderate or severe OSA was defined by the presence of an obstructive AHI≥15 events·h^{-1} according to AASM guidelines.

Statistical analysis

Differences in variables between moderate or severe OSA and no or mild OSA were assessed using t-test or chi-square test. Variables with $p<0.05$ were chosen as covariates to set up the machine-learning model. The effective of covariates were further evaluated by crude odds ratios (ORs) and adjusted OR by controlling for BMI. Statistical analysis was conducted using IBM SPSS software (version 24.0, IL, USA).

Model construction via SVM

SVM is a representative machine-learning algorithm for classification, which could be viewed as a nonlinear regression model. We used the SVM to capture the potential hyperplane that maximises the margin between moderate or severe OSA and no or mild OSA. The SVM has the following format:

$$C = \Sigma_{i=1}^{n} \alpha_i k(x_i, x) + b$$

Where c is the output of the model based on the new data x, which could be regarded as the classifier (0 or 1), x_i (i=1,…,n) is the training dataset, α_i and b are parameters determined by the algorithm.[18] We used the Gaussian radial basis function as the kernel function, with γ of 0.23 and the box constraint set to 20. Based on the results of the statistical analysis, sex, age, NC, WC, BMI, MID, and H/TSD were set as independent variables ($p<0.05$). We performed a significant principal component taking in to account the strong collinearity among parameters, such as BMI and WC. Five principal components were selected according to the accumulative variance contribution more than 90% and scree plot. [19]

The SVM model, namely SABIHC, i.e., ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s) was estimated by the training dataset (n=337, 70%); this model was then applied to predict OSA in the testing dataset (n=144, 30%). The flow chart including screening, randomisation and algorithm is presented in Figure 1.

The screening performance was assessed by means of sensitivity, specificity, and area under the curve (AUC), negative likelihood ratio (-LR) and positive likelihood ratio (+LR) and accuracy. Visual optimal hyperplanes for the Support Vector Machine are shown in Figure 2. Finally, we compared the difference in predictive power between the STOP-BANG questionnaire and the SVM model.

Analyses were performed using the Scikit-learn package (0.20.3) based on Python (3.5.1).

Table 1 Subject demographics, crude, and adjusted associations between morphometric variables and moderate-severe OSA.

	All patients (n=481)	AHI<15 (n=171)	AHI≥15 (n=310)	p-value*	crude OR (95% CI) #	p-value‡	adjusted OR for BMI (95% CI) #	p-value‡
General characteristics								
Male (%)	325(67.6%)	93(54.39%)	232(74.84%)	<0.001	0.401(0.270-0.595)	<0.001	0.463(0.308-0.697)	<0.001
Age	47.32±12.94	44.05±13.80	49.11±12.08	<0.001	1.032(1.016-1.048)	<0.001	1.037(1.021-1.054)	<0.001
NC	39.05±4.17	37.63±3.74	39.83±4.20	<0.001	1.146(1.090-1.205)	<0.001	1.121(0.054-1.193)	<0.001
WC	94.97±11.32	90.58±11.93	97.39±10.20	<0.001	1.109(1.041-1.081)	<0.001	1.074(1.041-1.107)	<0.001
BMI	26.26±4.16	25.18±4.16	26.85±4.04	<0.001	1.109(1.056-1.166)	<0.001	N/A	N/A
Facio-cervical measurements								
MID	4.88±0.88	4.76±0.79	4.94±0.92	0.028	1.275(1.025-1.585)	0.029	1.298(1.037-1.626)	0.023
Mallampati test=1	195(40.5%)	77(45.02%)	118(38.06%)		N/A	N/A	N/A	N/A
Mallampati test=2	142(29.5%)	54(31.58%)	88(28.39%)	0.12	0.589(0.371-0.938)†	0.026†	0.679(0.421-1.095) †	0.113†
Mallampati test=3	144(29.9%)	40(23.40%)	104(33.55%)		0.627(0.381-1.031)†	0.066†	0.713(0.428-1.186) †	0.192†
H/TMD	18.98±2.75	18.69±2.30	19.14±2.95	0.085	1.064(0.991-1.142)	0.086	1.071(0.998-1.150)	0.057
H/TSD	18.92±2.97	17.49±2.06	19.70±3.10	<0.001	1.448(1.311-1.599)	<0.001	1.458(1.417-1.614)	<0.001

Data are presented as mean± SD or n (%). *: t test or chi-square test as appropriate. ‡: logistic regression. #: Odds ratios are depicted for moderate or severe OSA relative to no or mild OSA. N/A: not applicable. †: Odds ratios of Mallampati test are depicted for Mallampati test=2 (or 3) relative to Mallampati test=1.

Abbreviation: NC; neck circumference; WC: waist circumference; BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance.

RESULTS

A total of 512 participants were recruited for the study at first, among which 31 patients were excluded due to short total sleep time (<200 min, 13 cases) or missing data (18 cases). Finally, 481 participants (325 males and 156 females; aged between 14-77 y) were enrolled in our study (Figure 1). The participants were divided into no or mild OSA (AHI<15 events·h^{-1}, n=171, mean AHI 7.01 ± 4.47 events·h^{-1}) and moderate or severe OSA (AHI≥15 events·h^{-1}, n=310, mean AHI 42.4 ± 20.7 events·h^{-1}). Increased BMI, NC, WC, older age, and higher percentage of males were found in moderate or severe OSA (p<0.001). Meanwhile, they had higher MAI (32.5 ±13.2 vs 16.4 ± 8.3) and lower LSpO2 (72.1 ± 13.3 vs 87.13 ± 6.3) than no or mild OSA (p<0.001) (Table 1).

Associations Between Facio-cervical Characteristics and OSA

Greater MID (4.94 vs 4.76, p=0.028) and H/TSD (19.70 vs 17.49, p<0.001) were found in moderate or severe group compared to no or mild OSA. MID (OR=1.275, adjusted OR=1.298) and H/TSD (OR=1.448, adjusted OR=1.458) are association to moderate-severe OSA, even after controlling for BMI. No significant inter-group differences were found in Mallampati test scores and H/TMD (p>0.05) (Table 1).

Set-up and Predictive Accuracy of the SABIHC2 model

Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were chosen as covariates because a significant difference was found between the two groups. Of all participants, 337 patients (70%) were randomised to the training dataset and 144 patients (30%) were randomised to the testing dataset. We set up the SABIHC2 model based on training dataset (n=337), whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s). The AUC was 0.829 (95% CI, 0.777 to 0.880), with a sensitivity of 0.930 and specificity of 0.727. Corresponding LR+ and LR- were 3.407 and 0.096, respectively (Table 2, Figure 3A). Then, the model was verified using the testing dataset (n=144) to screen moderate-severe OSA. The AUC was 0.787 (95% CI, 0.702 to 0.872), with a sensitivity of 0.874 and specificity of 0.700. Corresponding LR+ and LR- were 2.913 and 0.180, respectively (Table 2, Figure 3A).

The accuracy of SABIHC2 model based on all patients was 0.881, of which 0.928 for moderate or severe group (blue dots in purple area; Figure 2), and 0.719 for no or mild OSA (green dots in green area; Figure 2).

Table 2 Performance of SABIHC2 model

		Sensitivity	Specificity	AUC	95% CI	+LR	-LR
SABIHC2	training dataset	0.930	0.727	0.829	0.777-0.880	3.407	0.096
	testing dataset	0.874	0.700	0.787	0.702-0.872	2.913	0.180
STOP-BANG	testing dataset	0.684	0.500	0.597	0.499-0.694	0.632	1.368

Abbreviations: SABIHC2 = Sex-age-BMI-MID-H/TSD-NC-WC. +LR: positive likelihood ratio; -LR: negative likelihood ratio

Discriminative Ability of the SABIHC2 model and STOP-BANG questionnaire

To compare the predictive ability between the SABIHC2 model and STOP-BANG questionnaire, we calculated the ROC curve of STOP-BANG for moderate-severe OSA on the testing dataset. The AUC based on testing dataset was 0.597 (95% CI, 0.499 to 0.694), with a sensitivity of 0.684 and specificity of 0.500. (Table 2, Figure 3A).

As the STOP-BANG questions are based on the symptoms and comorbidities and may results in missed diagnoses, we further compared whether there are differences in the predictive capacity of SABIHC2 model and STOP-BANG questionnaire both for the symptomatic (ESS ≥ 10) and asymptomatic patients (ESS < 10) within the testing dataset. For asymptomatic patients, the SABIHC2 model demonstrated better predictive ability than STOP-BANG questionnaire [AUC (0.810 *vs.* 0.571)], with sensitivity (0.910 *vs.* 0.404), and specificity (0.709 *vs.* 0.788). Similarly, SABIHC2 model had higher predictive power than STOP-BANG questionnaire in patients experiencing sleepiness (ESS ≥ 10) [AUC (0.762 *vs.* 0.606)], with sensitivity (0.830 *vs.* 0.813), and specificity (0.699 *vs.* 0.478) (Figure 3B).

DISCUSSION

In the current study, we measured a broad range of anthropometric variables and assessed the association with moderate-severe OSA. Sex, age, BMI, MID, H/TSD, NC, and WC were significant determining factors. We also developed the SABIHC2 predictive model, which provides a simple and accurately assessment of moderate-severe OSA.

Facio-cervical measurements are essential factors related to OSA. The majority of previous studies focused on the soft tissue of the upper airway. A significant correlation between the oropharyngeal soft tissue (Mallampati test and tonsillar enlargement, etc.) and AHI was reported.[20,21] However, Mallampati test was not a significant parameter in our study. It is possible that the association between soft tissue and OSA is weak in Asian populations.[11] No previous studies explored the relationship between moderate-severe OSA and facio-cervical measurements, such as MID, H/TMD, and H/TSD, which have been frequently used to evaluate difficult intubation. Herein, we found that MID and H/TSD were strongly associated with moderate-severe OSA, and H/TSD showed a significant correlation with moderate-severe OSA, even controlling for BMI. The result may suggest that H/TSD is a potential factor for moderate-severe OSA in Chinese subjects.

As indicated previously, the overall scarcity and labour- and financially-onerous nature of PSG has prompted exploration of other suitable screening approaches, ranging from questionnaires to simplified multichannel recording. The STOP-BANG is one of the most widely used questionnaires. Some limitations of the STOP-BANG questionnaire should be considered. Firstly, it is based on simplified categories (0 or 1) and this might reduce its screening accuracy.[22] Secondly, the evaluation is based on symptoms and comorbidities, which may lead to omission of diagnosis for snorers without significant daytime sleepiness or hypertension.[23] Thus, we developed a new screening tool based on facio-cervical measurements by machine learning.

Considering non-linear relationships and data interaction effects,[24] the machine learning algorithm (SVM) was used to find the complex linkages between variables. Distinct from previous machine learning models which were based on age, sex, NC, WC, and BMI,[25,26] we included the structural facio-cervical measurements to construct the SABIHC2 model. This showed much better performance than the STOP-BANG in screening for OSA.

The results confirmed that the SABIHC2 model performed better than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for asymptomatic patients (ESS < 10), which prompted that it may be particularly useful for this cohort of patient. As most models work, we are planning to develop a software or application in the future, to allow health care worker friendly installation and application.

To our knowledge, this is the first study examining the predictive power of facio-cervical measurements, which may become a new index to predict OSA besides sex, age, and BMI. However, several limitations should be mentioned. First, the study cohort was based on Chinese population. The results obtained are hard to generalise for other populations. Second, the cross-sectional design of the study makes it difficult to establish the causal relationship between variables. Third, sample size was relatively small, and it was a single-centre study, which may affect the validation of the machine-learning algorithm model.

In conclusion, we confirmed that facio-cervical measurements are association to moderate-severe OSA. The machine learning model called the SABIHC2 model was set up based on facio-cervical and anthropometric measurements. The model is more effective than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for those without significant daytime symptoms.

REFERENCES

1.Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698.

2. Young T, Evans L, Finn L, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep.* 1997;20(9):705-706.

3.Kaw R, Pasupuleti V, Walker E, et al. Postoperative complications in patients with obstructive sleep apnea. *Chest.* 2012; 141: 436-41.

4.Chan MTV, Wang CY, Seet E, et al. Postoperative Vascular Complications in Unrecognized Obstructive Sleep Apnea (POSA) Study Investigators. Association of Unrecognized Obstructive Sleep Apnea With Postoperative Cardiovascular Events in Patients Undergoing Major Noncardiac Surgery. *JAMA*. 2019 May 14;321(18):1788-1798.

5.Singh M, Liao P, Kobah S, et al. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. *Br J Anaesth*. 2013 Apr;110(4):629-36.

6.American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology.* 2014;120(2):268-286. Available: https://pubmed.ncbi.nlm.nih.gov/24346178/

7. Dalesio NM, Hendrix CW, McMichael DH, et al. Effects of Obesity and Leptin Deficiency on Morphine Pharmacokinetics in a Mouse Model. *Anesth Analg.* 2016 Dec;123(6):1611-1617.

8.Kee K, Dixon J, Shaw J, et al. Comparison of Commonly Used Questionnaires to Identify Obstructive Sleep Apnea in a High-Risk Population. *J Clin Sleep Med.* 2018; 14: 2057-64.

9.Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers.* 2015;7(1):15015.

10.Sutherland K, Lee RW, Cistulli PA. Obesity and craniofacial structure as risk factors for obstructive sleep apnoea: impact of ethnicity. *Respirology*. 2012; 17: 213-22.

11.Schorr F, Kayamori F, Hirata RP, et al. Different Craniofacial Characteristics Predict Upper Airway Collapsibility in Japanese-Brazilian and White Men. *Chest*. 2016; 149: 737-46.

12.Roth D, Pace NL, Lee A, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018; 5: CD008874. Available: http:// doi:10.1111/ijn.12805

13.Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016; 375: 1216-9.

14.Menon SM, Sampangiramaiah S, Mathew M. Cross Sectional Observational Study Performed to See for Relation of Mallampati Score and Extended Mallampati Score with Body Mass Index. *J Clin Diagn Res* 2017; 11: Ug01-ug03.

15.Den Herder C, Schmeck J, Appelboom DJ, et al. Risks of general anaesthesia in people with obstructive sleep apnoea. *BMJ* 2004; 329: 955-9.

16.Schmitt HJ, Kirmse M, Radespiel-Troger M. Ratio of patient's height to thyromental distance improves prediction of difficult laryngoscopy. *Anaesth Intensive Care*. 2002; 30: 763-5.

17.The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep.* 1999; 22: 667-89.

18.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.

19. Björklund M. Be careful with your principal components. Evolution. 2019 Oct;73(10):2151-2158.

20.Schellenberg JB, Maislin G, Schwab RJ. Physical findings and the risk for obstructive sleep apnea. The importance of oropharyngeal structures. *Am J Respir Crit Care Med.* 2000; 162: 740-8.

21.Schwab RJ, Leinwand SE, Bearn CB, et al. Digital Morphometrics: A New Upper Airway Phenotyping Paradigm in OSA. *Chest* 2017; 152: 330-42.

22.Chung F, Abdullah HR, Liao P. STOP-BANG Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. *Chest*. 2016;149(3):631-638.

23.Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. *Eur Respir J.* 2014; 44:1600–7.

24.Linnen DT, Escobar GJ, Hu X, et al. Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. *J Hosp Med.* 2019; 14:161–9.

25.Mencar C, Gallo C, Mantero M, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics J.* 2019;1460458218824725. Available: http:// doi:10.1177/1460458218824725

26.Bozkurt S, Bostanci A, Turhan M. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf Med.* 2017; 56:308–18.

Author affiliations

[1]Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Acknowledgements The authors thank Peizhe Liu for his assistance with the SVM algorithm.

Author contributions Liu Zhang and Ya Ru Yan drafted the manuscript, had full access to all the data, and takes responsibility for the integrity of the data and the accuracy of the data analysis. Shi Qi Li contributed to data analysis and interpretation. Qing Yun Li and Shi Qi Li contributed to study conception and design. Hong Peng Li, Ying Ni Lin, Ning Li, Xian Wen Sun, Yong Jie Ding contributed to data collection, critical revision. Chuan Xiang Li and Qing Yun Li contributed to final approval of the manuscript.

Funding This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082). No conflicts of interest are to be reported

Competing interests None declared.

Patient and Public Involvement statement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Ethics approval The study was approved by the ethics committee of the Shanghai Jiao Tong University School of Medicine at Ruijin Hospital, with the following reference number: 2018-107.

Data availability statement Data are available upon reasonable request.

FIGURE LEGENDS:

Figure 1. Flow chart showing screening, randomisation and algorithm.

481 participants were enrolled in our study after excluding 31 patients. Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were potential factors related to OSA due to the significant difference between the two groups ($p<0.05$). We chose the following parameters to setup the SABIHC2 model based on training dataset, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 C*s)**. Then, the model was verified on testing dataset.

Abbreviations: BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance; NC: neck circumference; WC: waist circumference;

Figure 2. Optimal hyperplane for Support Vector Machine (SVM). The green area means the SVM predicting AHI<15 events·h^{-1}, the purple means the SVM predicts the AHI≥15 events·h^{-1}. Green dots represent no or mild OSA (AHI<15 events·h^{-1}), blue dots represent moderate or severe OSA (AHI≥15 events·h^{-1}). The accuracy of SABIHC2 model based on all patients was 0.881, of which 0.928 for moderate or severe OSA (blue dots in purple area), and 0.719 for no or mild OSA (green dots in green area). The boundary was obtained from the SVM classifier and the figure was created using Python.

Abbreviations: SVM: support vector machine; SABIHC2 model: ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 C*s)**.

Figure 3. ROC curve of SABIHC2 model and STOP-BANG.

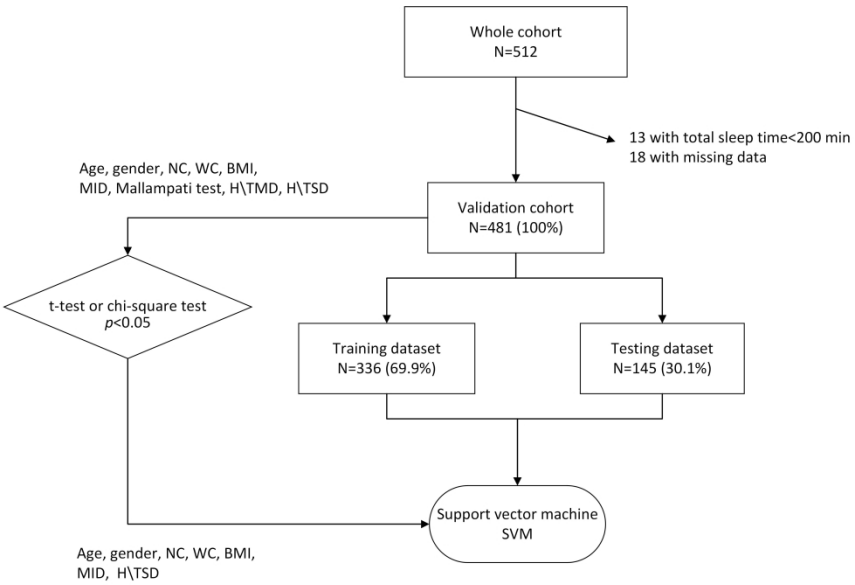
A: The ROC curve of SABIHC2 model and STOP-BANG. The AUC of the SABIHC2 model based on the training database (n=337, 70.1%, brown line) was 0.829 (95% CI, 0.777 to 0.880), with a sensitivity of 0.930 and specificity of 0.727. About testing database (n=144 29.9%, red line), the AUC was 0.787 (95% CI, 0.702 to 0.872), with a sensitivity of 0.874 and specificity of 0.700, respectively.

The ROC curve of STOP-BANG questionnaire (blue line) on the testing database was 0.597 (95% CI, 0.499 to 0.694). STOP-BANG>2 showed the sensitivity and specificity of 0.832 and 0.220, respectively, while STOP-BANG>4 with the sensitivity of 0.495 and specificity of 0.780.

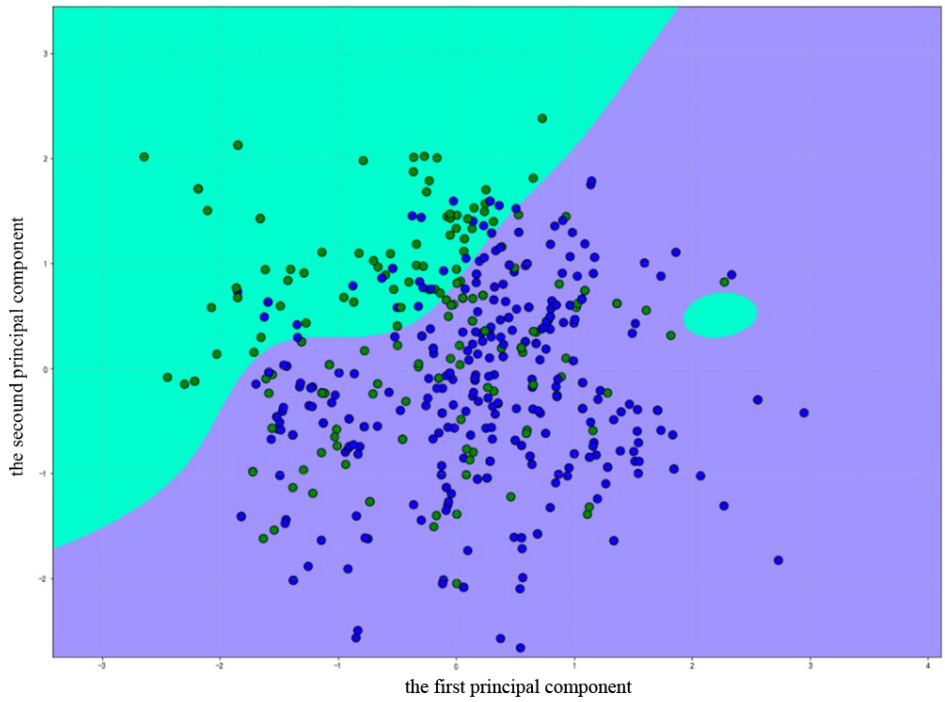
Brown: SVM based on training dataset; red: SVM based on testing dataset; blue: STOP-BANG based on testing dataset.

B: The ROC curve of SABIHC2 model and STOP-BANG based on asymptomatic patients (ESS < 10) and sleepiness patients (ESS ≥ 10). In asymptomatic patients, SABIHC2 model (red line) remarkably demonstrated better predictive ability than STOP-BANG questionnaire (dark blue line), (0.910 *vs.* 0.404), and specificity (0.709 *vs.* 0.788) and AUC (0.810 *vs.* 0.571). Similarly, SABIHC2 model (orange line) had higher predictive power than STOP-BANG questionnaire (soft blue line) in sleepiness patients (ESS ≥ 10), with sensitivity (0.830 *vs.* 0.813), and specificity (0.699 *vs.* 0.478) and AUC (0.762 *vs.* 0.606).

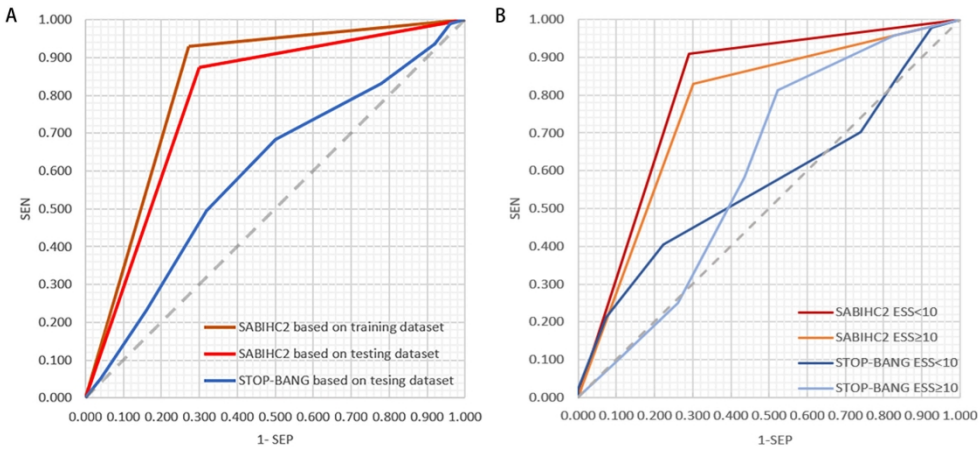
Abbreviations: AUC: the area under ROC curve; SVM: support vector machine; CIs: confidence intervals.



482x281mm (300 x 300 DPI)



90x65mm (300 x 300 DPI)



189x90mm (300 x 300 DPI)

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Moderate-Severe OSA Screening Based on the Machine Learning Model of the Chinese Population Facio-cervical Measurements Dataset

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-048482.R2
Article Type:	Original research
Date Submitted by the Author:	21-Jul-2021

Complete List of Authors:	zhang, liu; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Yan, Ya Ru; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Li, Shi Qi; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Li, Hong Peng; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Lin, Ying Ni; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Li, Ning; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Sun, Xian Wen; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Ding, Yong Jie; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine,; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Li, Chuan Xiang; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Institute of Respiratory Medicine Li, Qing Yun; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital,

	Institute of Respiratory Medicine
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Respiratory medicine
Keywords:	PUBLIC HEALTH, SLEEP MEDICINE, Adult anaesthesia < ANAESTHETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Moderate-Severe OSA Screening Based on the Machine Learning Model of the Chinese Population Facio-cervical Measurements Dataset

Liu Zhang*[1,2], Ya Ru Yan*[1,2], Shi Qi Li *[1,2], Hong Peng Li [1,2], Ying Ni Lin [1,2], Ning Li [1,2], Xian Wen Sun [1,2], Yong Jie Ding[1,2], Chuan Xiang Li[1,2], Qing Yun Li[1,2]

[1]Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

* co-first authors

Corresponding author full contact details:

Name: Qing Yun Li, M.D, Ph.D.

Address: Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

Post code: 200025

City: Shanghai

Country: China

Email: liqingyun68@hotmail.com

All authors have seen and approved the manuscript

Financial support: This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082).

No conflicts of interest are to be reported.

ABSTRACT:

Objectives Obstructive sleep apnea (OSA) has received much attention as a risk factor for perioperative complications and 68.5% of OSA patients remain undiagnosed before surgery. Facio-cervical characteristics may screen OSA for Asians due to smaller upper airways compared to Caucasians. Thus, our study aimed to explore a machine-learning model to screen moderate-severe OSA based on facio-cervical and anthropometric measurements.

Design A cross-sectional study.

Setting Data were collected from the Shanghai Jiao Tong University School of Medicine affiliated Ruijin Hospital between February 2019 and August 2020.

Participants A total of 481 Chinese participants were included in the study.

Primary and secondary outcome (1) Identification of moderate-severe OSA with apnea-hypopnea index (AHI)15 events·h^{-1}. (2) Verification of the machine learning model.

Results The SABIHC2 model (***S***ex-***A***ge-***B***ody mass index-maximum ***I***nterincisal distance-ratio of ***H***eight to thyro-sternum distance-neck ***C***ircumference-waist ***C***ircumference) was set up. The SABIHC2 model could screen moderate-severe OSA with an area under the curve (AUC)=0.832, the sensitivity of 0.916, and specificity of 0.749, and performed better than the STOP-BANG questionnaire, which showed AUC=0.631, the sensitivity of 0.487, and specificity of 0.772. Especially for asymptomatic patients (ESS < 10), the SABIHC2 model demonstrated better predictive ability compared to the STOP-BANG questionnaire, with AUC (0.824 *vs.* 0.530), sensitivity (0.892 *vs.* 0.348), and specificity (0.755 *vs.* 0.809).

Conclusion The SABIHC2 machine learning model provides a simple and accurate assessment of moderate-severe OSA in the Chinese population, especially for those without significant daytime sleepiness.

Keywords: obstructive sleep apnea, machine learning, anthropometric, screening

Strengths and limitations of this study

- This is the first study to assess the screening value of facio-cervical measurements in predicting moderate-severe OSA.
- Support vector model, a well-known machine learning method, was used to set up the model to screen moderate-severe OSA in the Chinese population.
- Instead of most questionnaires based on symptoms and comorbidities, the SABIHC2 model which is entirely based on measurements, avoids the confounding effect of symptoms and comorbidities.
- The cross-sectional design of the study makes it difficult to establish the causal relationship between variables.
- The study is a single-centre study with a relatively small sample size, and further confirmation with large-sample, multi-centre, prospective clinical trial is needed in the future.

INTRODUCTION

Obstructive sleep apnea (OSA) is a common breathing sleep disorder that affects about 936 million adults globally,[1] of which about 80% are estimated to be undiagnosed.[2] In the Chinese population, approximately 175 million adults have mild to severe obstructive sleep apnea, of whom about 65 million adults have moderate-severe obstructive sleep apnea.[1] In addition to cardiovascular injury and metabolic syndrome, OSA has recently been regarded as a risk factor for perioperative complications, including hypoxaemia, pneumonia, pulmonary embolism, unplanned transfer to the intensive care unit, and even death, especially for those who received abdominal or vascular surgery.[3] The rates of postoperative cardiovascular events show a rise in moderate-severe OSA (25.1%) compared to no or mild OSA(16.8%).[4] 38% of surgical patients had moderate-severe OSA, of which 68.5% were not diagnosed before surgery.[5] Guidelines have been recommended to develop a local protocol for screening possible OSA patients before elective surgery.[6,7] The most common screening scales, including STOP-BANG, Berlin questionnaire, and OSA50, are mainly based on symptoms and comorbidities,[8] which might lead to missed diagnoses for those without significant daytime sleepiness.

Several factors contribute to the pathogenesis of OSA, including obesity, facio-cervical anatomy, and alteration in pharyngeal muscle function, etc.[9] Asians have relatively smaller upper airways compared to Caucasians.[10] Only in Asians, smaller upper airways are predictors of upper airway collapsibility, and an anatomic imbalance between tongue and mandible volume influenced upper airway collapsibility among Caucasians. [11] The above evidence prompts facio-cervical characteristics that may predict OSA for Asians, such as thyromental distance (TMD), thyro-sternum distance (TSD), maximum interincisal distance (MID), and Mallampati test score, which have been widely used in predicting difficult intubation.[12] However, no studies assessed whether facio-cervical characteristics are suitable for predicting OSA. Thus, it is necessary to evaluate the relationship between facio-cervical characteristics and OSA.

During the past two decades, machine learning models have provided simple but effective approaches for improving diagnostic accuracy.[13] Support vector model (SVM) is a well-known classification

technique and has achieved great success in bioinformatics applications. Thus, the study aimed to build a new model via SVM and monitor its effectiveness in screening OSA.

METHODS

Study design

Participants with suspected OSA (snoring, witnessed apnea, or excessive daytime sleepiness, etc.) were enrolled in the study from Ruijin Hospital, Shanghai Jiao Tong University School of Medicine between February 2019 and August 2020. All participants underwent anthropometric measurements and subsequent overnight polysomnography (PSG). Participants were grouped according to their AHI gain into the following: 1) no or mild OSA (subjects without moderate-to-severe OSA: AHI <15 events·h^{-1}), 2) moderate-severe OSA (moderate to severe OSA: AHI ≥ 15 events·h^{-1}). Exclusion criteria: (1) Patients showing complications with severe respiratory diseases, such as severe COPD, interstitial lung disease, or acute asthma; (2) Patients showing complications with serious cardiovascular diseases such as acute myocardial infarction, acute heart failure, or chronic congestive heart failure (Grade III and IV); (3) Patients with mental illnesses who could not cooperate with the examination; (4) Patients who receiving non-invasive positive pressure ventilation therapy; (5) Patients who might have other sleep disorders under clinical evaluation.

The study was approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine at Ruijin Hospital (Protocol #2018-107). Written informed consent was obtained from all participants. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Clinical characteristics and anthropometric measurements

Age, sex, height, weight, BMI, neck circumference (NC), waist circumference (WC), Epworth Sleepiness Scale (ESS), and STOP-BANG questionnaire were recorded. The STOP-BANG questionnaire is a scoring model consisting of eight questions and its scores are based on Yes/No answers (score: 1/0). The eight questions included snoring, tiredness, observed apnea, high blood pressure, BMI, age, neck circumference, and gender. Facio-cervical measurements including Mallampati score, MID, TMD, and TSD were measured. The Mallampati score was evaluated when

participants were asked to sit upright and open their mouths as wide as possible. In grade I, the entire uvula, faucial pillar, and soft palate are visible. In grade II, part of the uvula and palate are visible. In grade III, the soft palate is visible, but the uvula is obscured.[14] MID was recorded by asking the participant to sit upright and open the mouth as wide as possible.[15] The TMD was measured as the straight distance between the thyroid notch and the lower border of the mental prominence, while the head was fully extended, and the mouth closed.[16] TSD was measured as the distance between the thyroid notch and the upper border of the sternum. The distance was rounded to the nearest 0.5 cm. The ratios of height to TMD (H/TMD) and height to TSD (H/TSD) were calculated.[16]

PSG

All participants underwent overnight PSG. No coffee, tea, caffeine-containing products, or sedative-hypnotics were taken before sleep. The PSG monitoring included electroencephalography, electrooculography, chin electromyography, electrocardiography, measurements of thoracal and abdominal movements, and airflow pressure and thermistor (Alice 5, Philips Respironics, USA), with the addition of oxygen saturation (Nonin, Herrsching, Germany). Sleep recordings were scored according to the American Academy of Sleep Medicine (AASM) 2007 criteria.[17] The diagnosis of moderate-severe OSA was defined by the presence of an obstructive AHI≥15 events·h^{-1} according to AASM guidelines.

Statistical analysis

Continuous variables with a normal distribution are presented as mean ± standard deviation, while values without a normal distribution are presented as median (25th to 75th percentiles). Categorical variables are presented as numbers and percentages. Independent samples t-test and Mann–Whitney U test were used to determine differential risk factors between the two groups. The chi-square test and Fisher's exact chi-square test were used to compare the categorical data, as appropriate. Logistic regression analyses were performed to calculate the odds ratio (OR). Receiver operating characteristic (ROC) analysis was used to determine the area under the curve (AUC) with a 95% confidence interval (CI), sensitivity, and specificity. All tests were two-sided and used a significance level of 0.05. All analyses were performed using the SPSS software (version 24.0; SPSS Inc. Chicago, IL, USA).

Model construction via SVM

SVM is a representative machine-learning algorithm for classification, which could be viewed as a nonlinear regression model. We used the SVM to capture the potential hyperplane that maximises the margin between moderate or severe OSA and no or mild OSA. The SVM has the following format:

$$C = \Sigma_{i=1}^{n} \alpha_i k(x_i, x) + b$$

Where c is the output of the model based on the new data x, which could be regarded as the classifier (0 or 1), x_i (i=1,…,n) is the training dataset, α_i and b are parameters determined by the algorithm.[18] We used the Gaussian radial basis function as the kernel function, with γ of 0.23 and the box constraint set to 20. Based on the results of the statistical analysis, sex, age, NC, WC, BMI, MID, and H/TSD were set as independent variables ($p<0.05$). We performed a significant principal component taking in to account the strong collinearity among parameters, such as BMI and WC. Five principal components were selected according to the accumulative variance contribution of more than 90% and scree plot. [19]

The performance of the SVM methods was evaluated by the ten-fold cross-validation. The participants were stratified sampling into ten subsamples: one formed the test dataset for verifying the effectiveness of the model, and the others formed the training dataset to predict moderate-severe OSA for the model. The SVM model, namely SABIHC, i.e., ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 Cs*)** was estimated by the training dataset and then applied to predict OSA in the testing dataset. The flow chart including screening, randomisation, and algorithm is presented in Figure 1.

The screening performance was assessed using sensitivity, specificity, area under the curve (AUC), negative likelihood ratio (-LR), positive likelihood ratio (+LR), and accuracy. Visual optimal hyperplanes for the Support Vector Machine are shown in Figure 2. Finally, we compared the difference in predictive power between the STOP-BANG questionnaire and the SVM model.

Analyses were performed using the Scikit-learn package (0.20.3) based on Python (3.5.1).

Table 1 Subject demographics, crude, and adjusted associations between morphometric variables and moderate-severe OSA.

	All patients (n=481)	AHI<15 (n=171)	AHI≥15 (n=310)	p-value*	crude OR (95% CI) #	p-value‡	adjusted OR for BMI (95% CI) #	p-value‡
General characteristics								
Male (%)	325(67.6%)	93(54.39%)	232(74.84%)	<0.001	0.401(0.270-0.595)	<0.001	0.463(0.308-0.697)	<0.001
Age	47.32±12.94	44.05±13.80	49.11±12.08	<0.001	1.032(1.016-1.048)	<0.001	1.037(1.021-1.054)	<0.001
NC	39.05±4.17	37.63±3.74	39.83±4.20	<0.001	1.146(1.090-1.205)	<0.001	1.121(0.054-1.193)	<0.001
WC	94.97±11.32	90.58±11.93	97.39±10.20	<0.001	1.109(1.041-1.081)	<0.001	1.074(1.041-1.107)	<0.001
BMI	26.26±4.16	25.18±4.16	26.85±4.04	<0.001	1.109(1.056-1.166)	<0.001	N/A	N/A
Facio-cervical measurements								
MID	4.88±0.88	4.76±0.79	4.94±0.92	0.028	1.275(1.025-1.585)	0.029	1.298(1.037-1.626)	0.023
Mallampati test=1	195(40.5%)	77(45.02%)	118(38.06%)		N/A	N/A	N/A	N/A
Mallampati test=2	142(29.5%)	54(31.58%)	88(28.39%)	0.12	0.589(0.371-0.938)†	0.026†	0.679(0.421-1.095) †	0.113†
Mallampati test=3	144(29.9%)	40(23.40%)	104(33.55%)		0.627(0.381-1.031)†	0.066†	0.713(0.428-1.186) †	0.192†
H/TMD	18.98±2.75	18.69±2.30	19.14±2.95	0.085	1.064(0.991-1.142)	0.086	1.071(0.998-1.150)	0.057
H/TSD	18.92±2.97	17.49±2.06	19.70±3.10	<0.001	1.448(1.311-1.599)	<0.001	1.458(1.417-1.614)	<0.001

Data are presented as mean± SD or n (%). *: t test or chi-square test as appropriate. ‡: logistic regression. #: Odds ratios are depicted for moderate-severe OSA relative to no or mild OSA. N/A: not applicable. †: Odds ratios of Mallampati test are depicted for Mallampati test=2 (or 3) relative to Mallampati test=1.

Abbreviation: NC; neck circumference; WC: waist circumference; BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance.

RESULTS

A total of 512 participants were recruited for the study at first, among which 31 patients were excluded due to short total sleep time (<200 min, 13 cases) or missing data (18 cases). Finally, 481 participants (325 males and 156 females; aged between 14-77 y) were enrolled in our study (Figure 1). The participants were divided into no or mild OSA (AHI<15 events·h^{-1}, n=171, mean AHI 7.01 ± 4.47 events·h^{-1}) and moderate-severe OSA (AHI≥15 events·h^{-1}, n=310, mean AHI 42.4 ± 20.7 events·h^{-1}). Increased BMI, NC, WC, older age and a higher percentage of males were found in moderate-severe OSA (p<0.001). Meanwhile, they had a higher micro-arousal index (MAI) (32.5 ±13.2 vs 16.4 ± 8.3 events·h^{-1}) and lower lowest pulse oxygen saturation ($LSpO_2$) (72.1 ± 13.3 vs 87.13 ± 6.3 %) than no or mild OSA (p<0.001) (Table 1).

Associations Between Facio-cervical Characteristics and OSA

Greater MID (4.94 vs 4.76, p=0.028) and H/TSD (19.70 vs 17.49, p<0.001) were found in the moderate-severe group compared to no or mild OSA. MID (OR=1.275, adjusted OR=1.298) and H/TSD (OR=1.448, adjusted OR=1.458) are associated with moderate-severe OSA, even after controlling for BMI. No significant inter-group differences were found in Mallampati test scores and H/TMD (p>0.05) (Table 1).

Set-up and Predictive Accuracy of the SABIHC2 model

Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were chosen as covariates because a significant difference was found between the two groups. We set up the SABIHC2 model, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s). The AUC was 0.832 (95% CI, 0.790 to 0.875), with a sensitivity of 0.916 and specificity of 0.749. Corresponding LR+ and LR- were 3.649 and 0.112, respectively (Table 2, Figure 3A). The accuracy of the SABIHC2 model was 0.857, of which 0.916 for the moderate-severe group (blue dots in the purple area; Figure 2), and 0.748 for no or mild OSA (green dots in the green area; Figure 2).

Table 2 Performance of SABIHC2 model

	Sensitivity	Specificity	AUC	95% CI	+LR	-LR
SABIHC2	0.916	0.749	0.832	0.790-0.875	3.649	0.112
STOP-BANG	0.487	0.772	0.631	0.581-0.682	2.136	0.665

Abbreviations: SABIHC2 = Sex-age-BMI-MID-H/TSD-NC-WC. +LR: positive likelihood ratio; -LR: negative likelihood ratio

Discriminative Ability of the SABIHC2 model and STOP-BANG questionnaire

To compare the predictive ability between the SABIHC2 model and the STOP-BANG questionnaire, we calculated the ROC curve of STOP-BANG for moderate-severe OSA. The AUC was 0.631 (95% CI, 0.581 to 0.682), with a sensitivity of 0.487 and specificity of 0.772. (Table 2, Figure 3A).

As the STOP-BANG questions are based on the symptoms and comorbidities and may result in missed diagnoses, we further compared whether there are differences in the predictive capacity of the SABIHC2 model and STOP-BANG questionnaire both for the symptomatic (ESS ≥ 10) and asymptomatic patients (ESS < 10). For asymptomatic patients, the SABIHC2 model demonstrated better predictive ability than the STOP-BANG questionnaire, with AUC (0.824 *vs.* 0.530), sensitivity (0.892 *vs.* 0.348), and specificity (0.755 *vs.* 0.809). Similarly, the SABIHC2 model had higher predictive power than the STOP-BANG questionnaire in patients experiencing sleepiness (ESS ≥ 10), with AUC (0.841 *vs.* 0.720), sensitivity (0.941 *vs.* 0.632), and specificity (0.740 *vs.* 0.727) (Figure 3B).

DISCUSSION

In the current study, we measured a broad range of anthropometric variables and assessed the association with moderate-severe OSA. Sex, age, BMI, MID, H/TSD, NC, and WC were significant determining factors. We also developed the SABIHC2 predictive model, which provides a simple and accurate assessment of moderate-severe OSA.

Facio-cervical measurements are essential factors related to OSA. The majority of previous studies focused on the soft tissue of the upper airway. A significant correlation between the oropharyngeal soft tissue (Mallampati test and tonsillar enlargement, etc.) and AHI was reported.[20,21] However, the Mallampati test was not a significant parameter in our study. It is possible that the association between soft tissue and OSA is weak in Asian populations.[11] No previous studies explored the relationship between moderate-severe OSA and facio-cervical measurements, such as MID, H/TMD, and H/TSD, which have been frequently used to evaluate difficult intubation. Herein, we found that MID and H/TSD were strongly associated with moderate-severe OSA, and H/TSD showed a significant correlation with moderate-severe OSA, even controlling for BMI. The result may suggest that H/TSD is a potential factor for moderate-severe OSA in Chinese subjects.

As indicated previously, the overall scarcity and labour- and financially-onerous nature of PSG has prompted exploration of other suitable screening approaches, ranging from questionnaires to simplified multichannel recording. The STOP-BANG is one of the most widely used questionnaires. Some limitations of the STOP-BANG questionnaire should be considered. Firstly, it is based on simplified categories (0 or 1) and this might reduce its screening accuracy.[22] Secondly, the evaluation is based on symptoms and comorbidities, which may lead to the omission of diagnosis for snorers without significant daytime sleepiness or hypertension.[23] Thus, we developed a new screening tool based on facio-cervical measurements by machine learning.

Considering non-linear relationships and data interaction effects,[24] the machine learning algorithm (SVM) was used to find the complex linkages between variables. Distinct from previous machine learning models which were based on age, sex, NC, WC, and BMI,[25,26] we included the structural facio-cervical measurements to construct the SABIHC2 model. This showed much better performance than the STOP-BANG in screening for OSA.

The results confirmed that the SABIHC2 model performed better than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for asymptomatic patients (ESS < 10), which prompted that it may be particularly useful for this cohort of patients. As most models work, we are planning to develop a software or application in the future, to allow health care worker-friendly installation and application.

To our knowledge, this is the first study examining the predictive power of facio-cervical measurements, which may become a new index to predict OSA besides sex, age, and BMI. However, several limitations should be mentioned. First, the study cohort was based on the Chinese population. The results obtained are hard to generalise for other populations. Second, the cross-sectional design of the study makes it difficult to establish the causal relationship between variables. Third, the sample size was relatively small, and it was a single-centre study, which may affect the validation of the machine-learning algorithm model.

In conclusion, we confirmed that facio-cervical measurements are associated with moderate-severe OSA in the Chinese population. The machine learning model called the SABIHC2 model was set up based on facio-cervical and anthropometric measurements. The model is more effective than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for those without significant daytime symptoms.

REFERENCES

1.Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698.

2. Young T, Evans L, Finn L, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep.* 1997;20(9):705-706.

3.Kaw R, Pasupuleti V, Walker E, et al. Postoperative complications in patients with obstructive sleep apnea. *Chest.* 2012; 141: 436-41.

4.Chan MTV, Wang CY, Seet E, et al. Postoperative Vascular Complications in Unrecognized Obstructive Sleep Apnea (POSA) Study Investigators. Association of Unrecognized Obstructive Sleep Apnea With Postoperative Cardiovascular Events in Patients Undergoing Major Noncardiac Surgery. *JAMA*. 2019 May 14;321(18):1788-1798.

5.Singh M, Liao P, Kobah S, et al. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. *Br J Anaesth*. 2013 Apr;110(4):629-36.

6.American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology.* 2014;120(2):268-286. Available: https://pubmed.ncbi.nlm.nih.gov/24346178/

7. Dalesio NM, Hendrix CW, McMichael DH, et al. Effects of Obesity and Leptin Deficiency on Morphine Pharmacokinetics in a Mouse Model. *Anesth Analg.* 2016 Dec;123(6):1611-1617.

8.Kee K, Dixon J, Shaw J, et al. Comparison of Commonly Used Questionnaires to Identify Obstructive Sleep Apnea in a High-Risk Population. *J Clin Sleep Med.* 2018; 14: 2057-64.

9.Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers.* 2015;7(1):15015.

10.Sutherland K, Lee RW, Cistulli PA. Obesity and craniofacial structure as risk factors for obstructive sleep apnoea: impact of ethnicity. *Respirology*. 2012; 17: 213-22.

11.Schorr F, Kayamori F, Hirata RP, et al. Different Craniofacial Characteristics Predict Upper Airway Collapsibility in Japanese-Brazilian and White Men. *Chest*. 2016; 149: 737-46.

12.Roth D, Pace NL, Lee A, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018; 5: CD008874. Available: http:// doi:10.1111/ijn.12805

13.Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016; 375: 1216-9.

14.Menon SM, Sampangiramaiah S, Mathew M. Cross Sectional Observational Study Performed to See for Relation of Mallampati Score and Extended Mallampati Score with Body Mass Index. *J Clin Diagn Res* 2017; 11: Ug01-ug03.

15.Den Herder C, Schmeck J, Appelboom DJ, et al. Risks of general anaesthesia in people with obstructive sleep apnoea. *BMJ* 2004; 329: 955-9.

16.Schmitt HJ, Kirmse M, Radespiel-Troger M. Ratio of patient's height to thyromental distance improves prediction of difficult laryngoscopy. *Anaesth Intensive Care*. 2002; 30: 763-5.

17.The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep.* 1999; 22: 667-89.

18.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.

19. Björklund M. Be careful with your principal components. Evolution. 2019 Oct;73(10):2151-2158.

20.Schellenberg JB, Maislin G, Schwab RJ. Physical findings and the risk for obstructive sleep apnea. The importance of oropharyngeal structures. *Am J Respir Crit Care Med.* 2000; 162: 740-8.

21.Schwab RJ, Leinwand SE, Bearn CB, et al. Digital Morphometrics: A New Upper Airway Phenotyping Paradigm in OSA. *Chest* 2017; 152: 330-42.

22.Chung F, Abdullah HR, Liao P. STOP-BANG Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. *Chest*. 2016;149(3):631-638.

23.Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. *Eur Respir J.* 2014; 44:1600–7.

24.Linnen DT, Escobar GJ, Hu X, et al. Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. *J Hosp Med.* 2019; 14:161–9.

25.Mencar C, Gallo C, Mantero M, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics J.* 2019;1460458218824725. Available: http:// doi:10.1177/1460458218824725

26.Bozkurt S, Bostanci A, Turhan M. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf Med.* 2017; 56:308–18.

Author affiliations

[1]Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Acknowledgements The authors thank Peizhe Liu for his assistance with the SVM algorithm.

Author contributions Liu Zhang and Ya Ru Yan drafted the manuscript, had full access to all the data, and takes responsibility for the integrity of the data and the accuracy of the data analysis. Shi Qi Li contributed to data analysis and interpretation. Qing Yun Li and Shi Qi Li contributed to study conception and design. Hong Peng Li, Ying Ni Lin, Ning Li, Xian Wen Sun, Yong Jie Ding contributed to data collection, critical revision. Chuan Xiang Li and Qing Yun Li contributed to final approval of the manuscript.

Funding This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082). No conflicts of interest are to be reported

Competing interests None declared.

Patient and Public Involvement statement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Ethics approval The study was approved by the ethics committee of the Shanghai Jiao Tong University School of Medicine at Ruijin Hospital, with the following reference number: 2018-107.

Data availability statement Data are available upon reasonable request.

FIGURE LEGENDS:

Figure 1. Flow chart showing screening, randomisation and algorithm.

481 participants were enrolled in our study after excluding 31 patients. Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were potential factors related to OSA due to the significant difference between the two groups ($p<0.05$). We chose the following parameters to setup the SABIHC2 model based on training dataset, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 C*s)**. Then, the model was verified on testing dataset.

Abbreviations: BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance; NC: neck circumference; WC: waist circumference;

Figure 2. Optimal hyperplane for Support Vector Machine (SVM). The green area means the SVM predicting AHI<15 events·h^{-1}, the purple means the SVM predicts the AHI≥15 events·h^{-1}. Green dots represent no or mild OSA (AHI<15 events·h^{-1}), blue dots represent moderate-severe OSA (AHI≥15 events·h^{-1}). The accuracy of SABIHC2 model based on all patients was 0.881, of which 0.928 for moderate-severe OSA (blue dots in purple area), and 0.719 for no or mild OSA (green dots in green area). The boundary was obtained from the SVM classifier and the figure was created using Python.

Abbreviations: SVM: support vector machine; SABIHC2 model: ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 C*s)**.

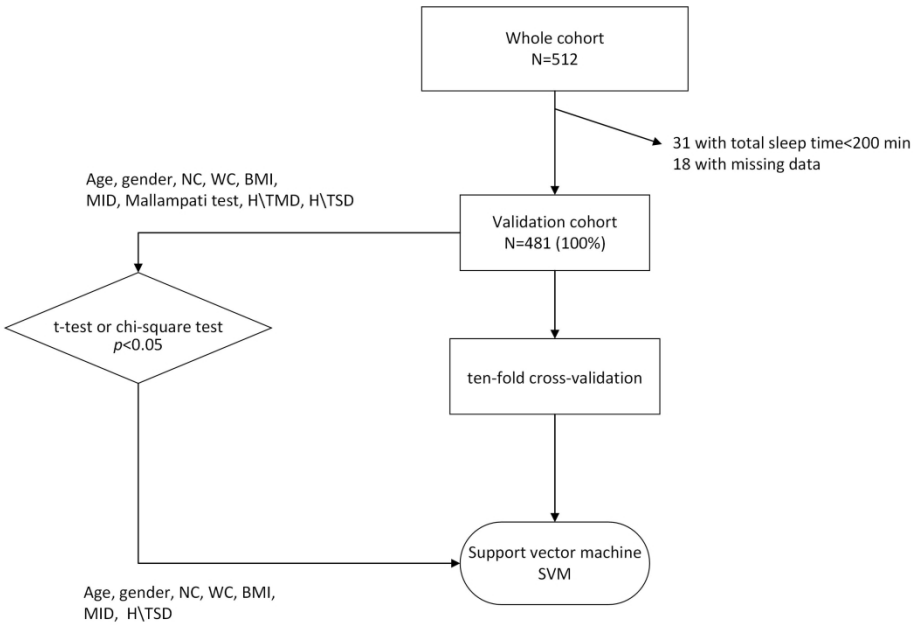
Figure 3. ROC curve of SABIHC2 model and STOP-BANG.

A: The ROC curve of SABIHC2 model and STOP-BANG. The AUC of the SABIHC2 model based on the training database (n=337, 70.1%, brown line) was 0.829 (95% CI, 0.777 to 0.880), with a sensitivity of 0.930 and specificity of 0.727. About testing database (n=144 29.9%, red line), the AUC

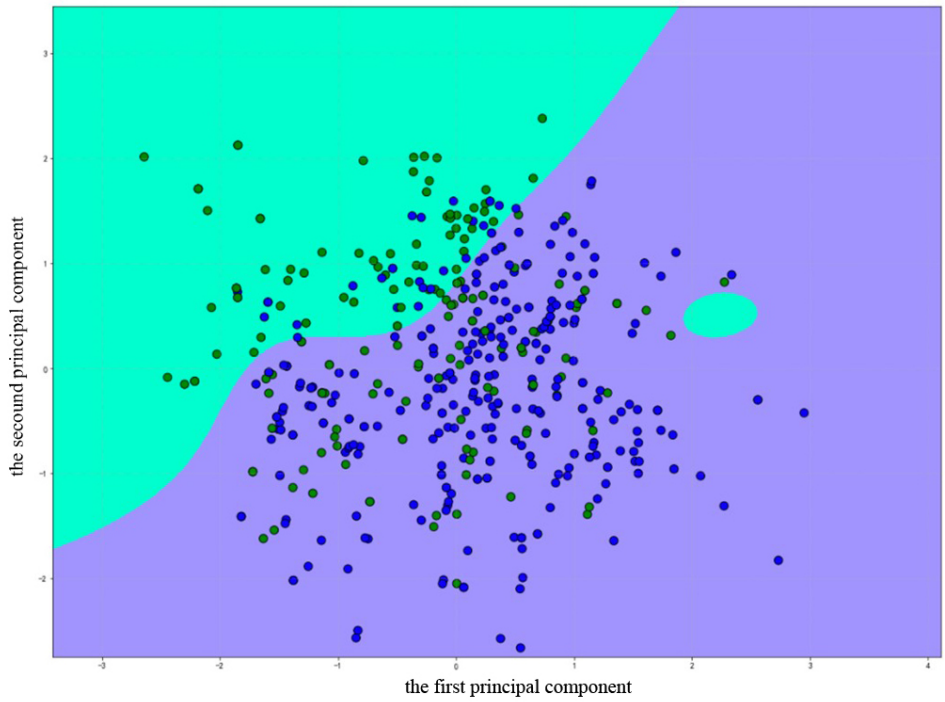
was 0.787 (95% CI, 0.702 to 0.872), with a sensitivity of 0.874 and specificity of 0.700, respectively. The ROC curve of STOP-BANG questionnaire (blue line) on the testing database was 0.597 (95% CI, 0.499 to 0.694). STOP-BANG>2 showed the sensitivity and specificity of 0.832 and 0.220, respectively, while STOP-BANG>4 with the sensitivity of 0.495 and specificity of 0.780. Brown: SVM based on training dataset; red: SVM based on testing dataset; blue: STOP-BANG based on testing dataset.

B: The ROC curve of SABIHC2 model and STOP-BANG based on asymptomatic patients (ESS < 10) and sleepiness patients (ESS ≥ 10). In asymptomatic patients, SABIHC2 model (red line) remarkably demonstrated better predictive ability than STOP-BANG questionnaire (dark blue line), with AUC (0.810 *vs.* 0.571), sensitivity (0.910 *vs.* 0.404), and specificity (0.709 *vs.* 0.788). Similarly, SABIHC2 model (orange line) had higher predictive power than STOP-BANG questionnaire (soft blue line) in sleepiness patients (ESS ≥ 10), with sensitivity (0.830 *vs.* 0.813), and specificity (0.699 *vs.* 0.478) and AUC (0.762 *vs.* 0.606).

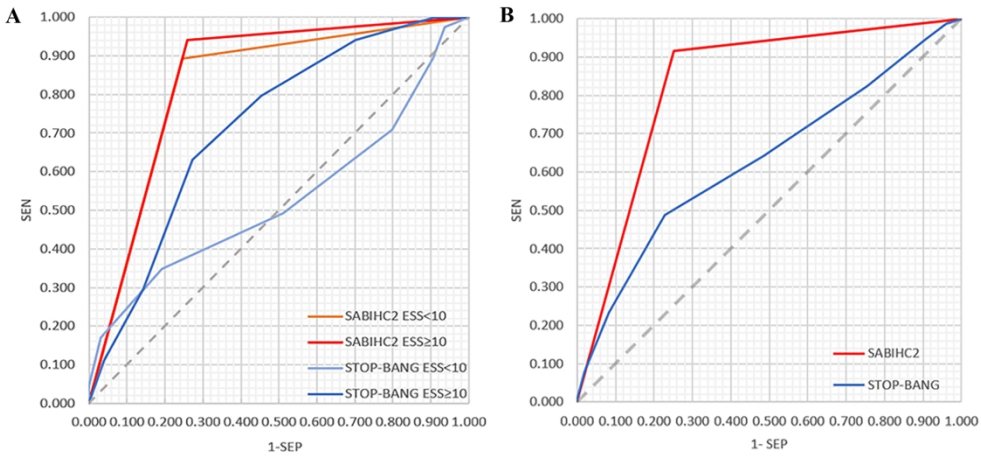
Abbreviations: AUC: the area under ROC curve; SVM: support vector machine; CIs: confidence intervals.



279x175mm (300 x 300 DPI)



90x65mm (300 x 300 DPI)



189x89mm (300 x 300 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 4
	4	Study objectives and hypotheses	Page 4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Not applicable
Participants	**6**	Eligibility criteria	Page 5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 5
	9	Whether participants formed a consecutive, random or convenience series	Page 5
Test methods	**10a**	Index test, in sufficient detail to allow replication	Page 5, 6
	10b	Reference standard, in sufficient detail to allow replication	Page 5, 6
	11	Rationale for choosing the reference standard (if alternatives exist)	Page 4, 5
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Not applicable
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Page 4, 5
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	No
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Yes
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	Not applicable
	15	How indeterminate index test or reference standard results were handled	Not applicable
	16	How missing data on the index test and reference standard were handled	Page 9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Page 6
	18	Intended sample size and how it was determined	Not applicable
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Figure 1
	20	Baseline demographic and clinical characteristics of participants	Page 8, Table 1
	21a	Distribution of severity of disease in those with the target condition	Page 8
	21b	Distribution of alternative diagnoses in those without the target condition	Not applicable
	22	Time interval and any clinical interventions between index test and reference standard	Not applicable
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Page 9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Page 8, 9, 10
	25	Any adverse events from performing the index test or the reference standard	Not applicable
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 12
	27	Implications for practice, including the intended use and clinical role of the index test	Page 12
OTHER INFORMATION			
	28	Registration number and name of registry	Page 16
	29	Where the full study protocol can be accessed	Page 16
	30	Sources of funding and other support; role of funders	Page 15

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Moderate-severe OSA screening based on support vector machine of the Chinese population facio-cervical measurements dataset: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-048482.R3
Article Type:	Original research
Date Submitted by the Author:	04-Aug-2021
Complete List of Authors:	zhang, liu; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Yan, Ya Ru; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Li, Shi Qi; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Li, Hong Peng; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Lin, Ying Ni; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Li, Ning; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Sun, Xian Wen; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Ding, Yong Jie; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine,; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Li, Chuan Xiang; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine Li, Qing Yun; Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Department of Respiratory and Critical Care Medicine; Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Medicine
Primary Subject	Respiratory medicine

Heading:	
Secondary Subject Heading:	Respiratory medicine
Keywords:	PUBLIC HEALTH, SLEEP MEDICINE, Adult anaesthesia < ANAESTHETICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Moderate-severe OSA screening based on support vector machine of the Chinese population facio-cervical measurements dataset: a cross-sectional study

Liu Zhang*[1,2], Ya Ru Yan*[1,2], Shi Qi Li *[1,2], Hong Peng Li [1,2], Ying Ni Lin [1,2], Ning Li [1,2], Xian Wen Sun [1,2], Yong Jie Ding[1,2], Chuan Xiang Li[1,2], Qing Yun Li[1,2]

[1]Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

* co-first authors

Corresponding author full contact details:

Name: Qing Yun Li, M.D, Ph.D.

Address: Department of Respiratory and Critical Care Medicine, Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital, Shanghai 200025, China

Post code: 200025

City: Shanghai

Country: China

Email: liqingyun68@hotmail.com

All authors have seen and approved the manuscript

Financial support: This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082).

No conflicts of interest are to be reported.

ABSTRACT:

Objectives Obstructive sleep apnea (OSA) has received much attention as a risk factor for perioperative complications and 68.5% of OSA patients remain undiagnosed before surgery. Facio-cervical characteristics may screen OSA for Asians due to smaller upper airways compared to Caucasians. Thus, our study aimed to explore a machine-learning model to screen moderate-severe OSA based on facio-cervical and anthropometric measurements.

Design A cross-sectional study.

Setting Data were collected from the Shanghai Jiao Tong University School of Medicine affiliated Ruijin Hospital between February 2019 and August 2020.

Participants A total of 481 Chinese participants were included in the study.

Primary and secondary outcome (1) Identification of moderate-severe OSA with apnea-hypopnea index (AHI)15 events·h^{-1}. (2) Verification of the machine learning model.

Results The SABIHC2 model (***S***ex-***A***ge-***B***ody mass index-maximum ***I***nterincisal distance-ratio of ***H***eight to thyro-sternum distance-neck ***C***ircumference-waist ***C***ircumference) was set up. The SABIHC2 model could screen moderate-severe OSA with an area under the curve (AUC)=0.832, the sensitivity of 0.916, and specificity of 0.749, and performed better than the STOP-BANG questionnaire, which showed AUC=0.631, the sensitivity of 0.487, and specificity of 0.772. Especially for asymptomatic patients (ESS < 10), the SABIHC2 model demonstrated better predictive ability compared to the STOP-BANG questionnaire, with AUC (0.824 *vs.* 0.530), sensitivity (0.892 *vs.* 0.348), and specificity (0.755 *vs.* 0.809).

Conclusion The SABIHC2 machine learning model provides a simple and accurate assessment of moderate-severe OSA in the Chinese population, especially for those without significant daytime sleepiness.

Keywords: obstructive sleep apnea, machine learning, anthropometric, screening

Strengths and limitations of this study

- This is the first study to assess the screening value of facio-cervical measurements in predicting moderate-severe OSA.
- Support vector machine, a well-known machine learning method, was used to set up the model to screen moderate-severe OSA in the Chinese population.
- Instead of most questionnaires based on symptoms and comorbidities, the SABIHC2 model which is entirely based on measurements, avoids the confounding effect of symptoms and comorbidities.
- The cross-sectional design of the study makes it difficult to establish the causal relationship between variables.
- The study is a single-centre study with a relatively small sample size, and further confirmation with large-sample, multi-centre, prospective clinical trial is needed in the future.

INTRODUCTION

Obstructive sleep apnea (OSA) is a common breathing sleep disorder that affects about 936 million adults globally,[1] of which about 80% are estimated to be undiagnosed.[2] In the Chinese population, approximately 175 million adults have mild to severe obstructive sleep apnea, of whom about 65 million adults have moderate-severe obstructive sleep apnea.[1] In addition to cardiovascular injury and metabolic syndrome, OSA has recently been regarded as a risk factor for perioperative complications, including hypoxaemia, pneumonia, pulmonary embolism, unplanned transfer to the intensive care unit, and even death, especially for those who received abdominal or vascular surgery.[3] The rates of postoperative cardiovascular events show a rise in moderate-severe OSA (25.1%) compared to no or mild OSA(16.8%).[4] 38% of surgical patients had moderate-severe OSA, of which 68.5% were not diagnosed before surgery.[5] Guidelines have been recommended to develop a local protocol for screening possible OSA patients before elective surgery.[6,7] The most common screening scales, including STOP-BANG, Berlin questionnaire, and OSA50, are mainly based on symptoms and comorbidities,[8] which might lead to missed diagnoses for those without significant daytime sleepiness.

Several factors contribute to the pathogenesis of OSA, including obesity, facio-cervical anatomy, and alteration in pharyngeal muscle function, etc.[9] Asians have relatively smaller upper airways compared to Caucasians.[10] Only in Asians, smaller upper airways are predictors of upper airway collapsibility, and an anatomic imbalance between tongue and mandible volume influenced upper airway collapsibility among Caucasians. [11] The above evidence prompts facio-cervical characteristics that may predict OSA for Asians, such as thyromental distance (TMD), thyro-sternum distance (TSD), maximum interincisal distance (MID), and Mallampati test score, which have been widely used in predicting difficult intubation.[12] However, no studies assessed whether facio-cervical characteristics are suitable for predicting OSA. Thus, it is necessary to evaluate the relationship between facio-cervical characteristics and OSA.

During the past two decades, machine learning models have provided simple but effective approaches for improving diagnostic accuracy.[13] Support vector machine (SVM) is a well-known classification

technique and has achieved great success in bioinformatics applications. Thus, the study aimed to build a new model via SVM and monitor its effectiveness in screening OSA.

METHODS

Study design

Participants with suspected OSA (snoring, witnessed apnea, or excessive daytime sleepiness, etc.) were enrolled in the study from Ruijin Hospital, Shanghai Jiao Tong University School of Medicine between February 2019 and August 2020. All participants underwent anthropometric measurements and subsequent overnight polysomnography (PSG). Participants were grouped according to their AHI gain into the following: 1) no or mild OSA (subjects without moderate-to-severe OSA: AHI <15 events·h^{-1}), 2) moderate-severe OSA (moderate to severe OSA: AHI ≥ 15 events·h^{-1}). Exclusion criteria: (1) Patients showing complications with severe respiratory diseases, such as severe COPD, interstitial lung disease, or acute asthma; (2) Patients showing complications with serious cardiovascular diseases such as acute myocardial infarction, acute heart failure, or chronic congestive heart failure (Grade III and IV); (3) Patients with mental illnesses who could not cooperate with the examination; (4) Patients who receiving non-invasive positive pressure ventilation therapy; (5) Patients who might have other sleep disorders under clinical evaluation.

The study was approved by the Ethics Committee of Shanghai Jiao Tong University School of Medicine at Ruijin Hospital (Protocol #2018-107). Written informed consent was obtained from all participants. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Clinical characteristics and anthropometric measurements

Age, sex, height, weight, BMI, neck circumference (NC), waist circumference (WC), Epworth Sleepiness Scale (ESS), and STOP-BANG questionnaire were recorded. The STOP-BANG questionnaire is a scoring model consisting of eight questions and its scores are based on Yes/No answers (score: 1/0). The eight questions included snoring, tiredness, observed apnea, high blood pressure, BMI, age, neck circumference, and gender. Facio-cervical measurements including Mallampati score, MID, TMD, and TSD were measured. The Mallampati score was evaluated when

participants were asked to sit upright and open their mouths as wide as possible. In grade I, the entire uvula, faucial pillar, and soft palate are visible. In grade II, part of the uvula and palate are visible. In grade III, the soft palate is visible, but the uvula is obscured.[14] MID was recorded by asking the participant to sit upright and open the mouth as wide as possible.[15] The TMD was measured as the straight distance between the thyroid notch and the lower border of the mental prominence, while the head was fully extended, and the mouth closed.[16] TSD was measured as the distance between the thyroid notch and the upper border of the sternum. The distance was rounded to the nearest 0.5 cm. The ratios of height to TMD (H/TMD) and height to TSD (H/TSD) were calculated.[16]

PSG

All participants underwent overnight PSG. No coffee, tea, caffeine-containing products, or sedative-hypnotics were taken before sleep. The PSG monitoring included electroencephalography, electrooculography, chin electromyography, electrocardiography, measurements of thoracal and abdominal movements, and airflow pressure and thermistor (Alice 5, Philips Respironics, USA), with the addition of oxygen saturation (Nonin, Herrsching, Germany). Sleep recordings were scored according to the American Academy of Sleep Medicine (AASM) 2007 criteria.[17] The diagnosis of moderate-severe OSA was defined by the presence of an obstructive AHI≥15 events·h^{-1} according to AASM guidelines.

Statistical analysis

Continuous variables with a normal distribution are presented as mean ± standard deviation, while values without a normal distribution are presented as median (25th to 75th percentiles). Categorical variables are presented as numbers and percentages. Independent samples t-test and Mann–Whitney U test were used to determine differential risk factors between the two groups. The chi-square test and Fisher's exact chi-square test were used to compare the categorical data, as appropriate. Logistic regression analyses were performed to calculate the odds ratio (OR). Receiver operating characteristic (ROC) analysis was used to determine the area under the curve (AUC) with a 95% confidence interval (CI), sensitivity, and specificity. All tests were two-sided and used a significance level of 0.05. All analyses were performed using the SPSS software (version 24.0; SPSS Inc. Chicago, IL, USA).

Model construction via SVM

SVM is a representative machine-learning algorithm for classification, which could be viewed as a nonlinear regression model. We used the SVM to capture the potential hyperplane that maximises the margin between moderate or severe OSA and no or mild OSA. The SVM has the following format:

$$C = \Sigma_{i=1}^{n} \alpha_i k(x_i, x) + b$$

Where c is the output of the model based on the new data x, which could be regarded as the classifier (0 or 1), x_i (i=1,…,n) is the training dataset, α_i and b are parameters determined by the algorithm.[18] We used the Gaussian radial basis function as the kernel function, with γ of 0.23 and the box constraint set to 20. Based on the results of the statistical analysis, sex, age, NC, WC, BMI, MID, and H/TSD were set as independent variables ($p<0.05$). We performed a significant principal component taking into account the strong collinearity among parameters, such as BMI and WC. Five principal components were selected according to the accumulative variance contribution of more than 90% and scree plot. [19]

The performance of the SVM methods was evaluated by the ten-fold cross-validation. The participants were stratified sampling into ten subsamples: one formed the test dataset for verifying the effectiveness of the model, and the others formed the training dataset to predict moderate-severe OSA for the model. The SVM model, namely SABIHC, i.e., ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** **(*2 Cs*)** was estimated by the training dataset and then applied to predict OSA in the testing dataset. The flow chart including screening, randomisation, and algorithm is presented in Figure 1.

The screening performance was assessed using sensitivity, specificity, area under the curve (AUC), negative likelihood ratio (-LR), positive likelihood ratio (+LR), and accuracy. Visual optimal hyperplanes for the Support Vector Machine are shown in Figure 2. Finally, we compared the difference in predictive power between the STOP-BANG questionnaire and the SVM model.

Analyses were performed using the Scikit-learn package (0.20.3) based on Python (3.5.1).

Table 1 Subject demographics, crude, and adjusted associations between morphometric variables and moderate-severe OSA.

	All patients (n=481)	AHI<15 (n=171)	AHI≥15 (n=310)	p-value*	crude OR (95% CI) #	p-value‡	adjusted OR for BMI (95% CI) #	p-value‡
General characteristics								
Male (%)	325(67.6%)	93(54.39%)	232(74.84%)	<0.001	0.401(0.270-0.595)	<0.001	0.463(0.308-0.697)	<0.001
Age	47.32±12.94	44.05±13.80	49.11±12.08	<0.001	1.032(1.016-1.048)	<0.001	1.037(1.021-1.054)	<0.001
NC	39.05±4.17	37.63±3.74	39.83±4.20	<0.001	1.146(1.090-1.205)	<0.001	1.121(0.054-1.193)	<0.001
WC	94.97±11.32	90.58±11.93	97.39±10.20	<0.001	1.109(1.041-1.081)	<0.001	1.074(1.041-1.107)	<0.001
BMI	26.26±4.16	25.18±4.16	26.85±4.04	<0.001	1.109(1.056-1.166)	<0.001	N/A	N/A
Facio-cervical measurements								
MID	4.88±0.88	4.76±0.79	4.94±0.92	0.028	1.275(1.025-1.585)	0.029	1.298(1.037-1.626)	0.023
Mallampati test=1	195(40.5%)	77(45.02%)	118(38.06%)		N/A	N/A	N/A	N/A
Mallampati test=2	142(29.5%)	54(31.58%)	88(28.39%)	0.12	0.589(0.371-0.938)†	0.026†	0.679(0.421-1.095) †	0.113†
Mallampati test=3	144(29.9%)	40(23.40%)	104(33.55%)		0.627(0.381-1.031)†	0.066†	0.713(0.428-1.186) †	0.192†
H/TMD	18.98±2.75	18.69±2.30	19.14±2.95	0.085	1.064(0.991-1.142)	0.086	1.071(0.998-1.150)	0.057
H/TSD	18.92±2.97	17.49±2.06	19.70±3.10	<0.001	1.448(1.311-1.599)	<0.001	1.458(1.417-1.614)	<0.001

Data are presented as mean± SD or n (%). *: t test or chi-square test as appropriate. ‡: logistic regression. #: Odds ratios are depicted for moderate-severe OSA relative to no or mild OSA. N/A: not applicable. †: Odds ratios of Mallampati test are depicted for Mallampati test=2 (or 3) relative to Mallampati test=1.

Abbreviation: NC; neck circumference; WC: waist circumference; BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance.

RESULTS

A total of 512 participants were recruited for the study at first, among which 31 patients were excluded due to short total sleep time (<200 min, 13 cases) or missing data (18 cases). Finally, 481 participants (325 males and 156 females; aged between 14-77 y) were enrolled in our study (Figure 1). The participants were divided into no or mild OSA (AHI<15 events·h^{-1}, n=171, mean AHI 7.01 ± 4.47 events·h^{-1}) and moderate-severe OSA (AHI≥15 events·h^{-1}, n=310, mean AHI 42.4 ± 20.7 events·h^{-1}). Increased BMI, NC, WC, older age, and a higher percentage of males were found in moderate-severe OSA (p<0.001). Meanwhile, they had a higher micro-arousal index (MAI) (32.5 ±13.2 vs 16.4 ± 8.3 events·h^{-1}) and lower lowest pulse oxygen saturation ($LSpO_2$) (72.1 ± 13.3 vs 87.13 ± 6.3 %) than no or mild OSA (p<0.001) (Table 1).

Associations Between Facio-cervical Characteristics and OSA

Greater MID (4.94 vs 4.76, p=0.028) and H/TSD (19.70 vs 17.49, p<0.001) were found in the moderate-severe group compared to no or mild OSA. MID (OR=1.275, adjusted OR=1.298) and H/TSD (OR=1.448, adjusted OR=1.458) are associated with moderate-severe OSA, even after controlling for BMI. No significant inter-group differences were found in Mallampati test scores and H/TMD (p>0.05) (Table 1).

Set-up and Predictive Accuracy of the SABIHC2 model

Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were chosen as covariates because a significant difference was found between the two groups. We set up the SABIHC2 model, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C***s). The AUC was 0.832 (95% CI, 0.790 to 0.875), with a sensitivity of 0.916 and specificity of 0.749. Corresponding LR+ and LR- were 3.649 and 0.112, respectively (Table 2, Figure 3A). The accuracy of the SABIHC2 model was 0.857, of which 0.916 for the moderate-severe group (blue dots in the purple area; Figure 2), and 0.748 for no or mild OSA (green dots in the green area; Figure 2).

Table 2 Performance of SABIHC2 model

	Sensitivity	Specificity	AUC	95% CI	+LR	-LR
SABIHC2	0.916	0.749	0.832	0.790-0.875	3.649	0.112
STOP-BANG	0.487	0.772	0.631	0.581-0.682	2.136	0.665

Abbreviations: SABIHC2 = Sex-age-BMI-MID-H/TSD-NC-WC. +LR: positive likelihood ratio; -LR: negative likelihood ratio

Discriminative Ability of the SABIHC2 model and STOP-BANG questionnaire

To compare the predictive ability between the SABIHC2 model and the STOP-BANG questionnaire, we calculated the ROC curve of STOP-BANG for moderate-severe OSA. The AUC was 0.631 (95% CI, 0.581 to 0.682), with a sensitivity of 0.487 and specificity of 0.772. (Table 2, Figure 3A).

As the STOP-BANG questions are based on the symptoms and comorbidities and may result in missed diagnoses, we further compared whether there are differences in the predictive capacity of the SABIHC2 model and STOP-BANG questionnaire both for the symptomatic (ESS ≥ 10) and asymptomatic patients (ESS < 10). For asymptomatic patients, the SABIHC2 model demonstrated better predictive ability than the STOP-BANG questionnaire, with AUC (0.824 *vs.* 0.530), sensitivity (0.892 *vs.* 0.348), and specificity (0.755 *vs.* 0.809). Similarly, the SABIHC2 model had higher predictive power than the STOP-BANG questionnaire in patients experiencing sleepiness (ESS ≥ 10), with AUC (0.841 *vs.* 0.720), sensitivity (0.941 *vs.* 0.632), and specificity (0.740 *vs.* 0.727) (Figure 3B).

DISCUSSION

In the current study, we measured a broad range of anthropometric variables and assessed the association with moderate-severe OSA. Sex, age, BMI, MID, H/TSD, NC, and WC were significant determining factors. We also developed the SABIHC2 predictive model, which provides a simple and accurate assessment of moderate-severe OSA.

Facio-cervical measurements are essential factors related to OSA. The majority of previous studies focused on the soft tissue of the upper airway. A significant correlation between the oropharyngeal

soft tissue (Mallampati test and tonsillar enlargement, etc.) and AHI was reported.[20,21] However, the Mallampati test was not a significant parameter in our study. It is possible that the association between soft tissue and OSA is weak in Asian populations.[11] No previous studies explored the relationship between moderate-severe OSA and facio-cervical measurements, such as MID, H/TMD, and H/TSD, which have been frequently used to evaluate difficult intubation. Herein, we found that MID and H/TSD were strongly associated with moderate-severe OSA, and H/TSD showed a significant correlation with moderate-severe OSA, even controlling for BMI. The result may suggest that H/TSD is a potential factor for moderate-severe OSA in Chinese subjects.

As indicated previously, the overall scarcity and labour- and financially-onerous nature of PSG has prompted exploration of other suitable screening approaches, ranging from questionnaires to simplified multichannel recording. The STOP-BANG is one of the most widely used questionnaires. Some limitations of the STOP-BANG questionnaire should be considered. Firstly, it is based on simplified categories (0 or 1) and this might reduce its screening accuracy.[22] Secondly, the evaluation is based on symptoms and comorbidities, which may lead to the omission of diagnosis for snorers without significant daytime sleepiness or hypertension.[23] Thus, we developed a new screening tool based on facio-cervical measurements by machine learning.

Considering non-linear relationships and data interaction effects,[24] the machine learning algorithm (SVM) was used to find the complex linkages between variables. Distinct from previous machine learning models which were based on age, sex, NC, WC, and BMI,[25,26] we included the structural facio-cervical measurements to construct the SABIHC2 model. This showed much better performance than the STOP-BANG in screening for OSA.

The results confirmed that the SABIHC2 model performed better than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for asymptomatic patients (ESS < 10), which prompted that it may be particularly useful for this cohort of patients. As most models work, we are planning to develop a software or application in the future, to allow health care worker-friendly installation and application.

To our knowledge, this is the first study examining the predictive power of facio-cervical measurements, which may become a new index to predict OSA besides sex, age, and BMI. However,

several limitations should be mentioned. First, the study cohort was based on the Chinese population. The results obtained are hard to generalise for other populations. Second, the cross-sectional design of the study makes it difficult to establish the causal relationship between variables. Third, the sample size was relatively small, and it was a single-centre study, which may affect the validation of the machine-learning algorithm model.

In conclusion, we confirmed that facio-cervical measurements are associated with moderate-severe OSA in the Chinese population. The machine learning model called the SABIHC2 model was set up based on facio-cervical and anthropometric measurements. The model is more effective than the STOP-BANG questionnaire in predicting moderate-severe OSA, especially for those without significant daytime symptoms.

REFERENCES

1.Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698.

2. Young T, Evans L, Finn L, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep.* 1997;20(9):705-706.

3.Kaw R, Pasupuleti V, Walker E, et al. Postoperative complications in patients with obstructive sleep apnea. *Chest.* 2012; 141: 436-41.

4.Chan MTV, Wang CY, Seet E, et al. Postoperative Vascular Complications in Unrecognized Obstructive Sleep Apnea (POSA) Study Investigators. Association of Unrecognized Obstructive Sleep Apnea With Postoperative Cardiovascular Events in Patients Undergoing Major Noncardiac Surgery. *JAMA*. 2019 May 14;321(18):1788-1798.

5.Singh M, Liao P, Kobah S, et al. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. *Br J Anaesth*. 2013 Apr;110(4):629-36.

6.American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology.* 2014;120(2):268-286. Available: https://pubmed.ncbi.nlm.nih.gov/24346178/

7. Dalesio NM, Hendrix CW, McMichael DH, et al. Effects of Obesity and Leptin Deficiency on Morphine Pharmacokinetics in a Mouse Model. *Anesth Analg.* 2016 Dec;123(6):1611-1617.

8.Kee K, Dixon J, Shaw J, et al. Comparison of Commonly Used Questionnaires to Identify Obstructive Sleep Apnea in a High-Risk Population. *J Clin Sleep Med.* 2018; 14: 2057-64.

9.Lévy P, Kohler M, McNicholas WT, et al. Obstructive sleep apnoea syndrome. *Nat Rev Dis Primers.* 2015;7(1):15015.

10.Sutherland K, Lee RW, Cistulli PA. Obesity and craniofacial structure as risk factors for obstructive sleep apnoea: impact of ethnicity. *Respirology*. 2012; 17: 213-22.

11.Schorr F, Kayamori F, Hirata RP, et al. Different Craniofacial Characteristics Predict Upper Airway Collapsibility in Japanese-Brazilian and White Men. *Chest*. 2016; 149: 737-46.

12.Roth D, Pace NL, Lee A, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. *Cochrane Database Syst Rev* 2018; 5: CD008874. Available: http:// doi:10.1111/ijn.12805

13.Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016; 375: 1216-9.

14.Menon SM, Sampangiramaiah S, Mathew M. Cross Sectional Observational Study Performed to See for Relation of Mallampati Score and Extended Mallampati Score with Body Mass Index. *J Clin Diagn Res* 2017; 11: Ug01-ug03.

15.Den Herder C, Schmeck J, Appelboom DJ, et al. Risks of general anaesthesia in people with obstructive sleep apnoea. *BMJ* 2004; 329: 955-9.

16.Schmitt HJ, Kirmse M, Radespiel-Troger M. Ratio of patient's height to thyromental distance improves prediction of difficult laryngoscopy. *Anaesth Intensive Care*. 2002; 30: 763-5.

17.The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep.* 1999; 22: 667-89.

18.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.

19. Björklund M. Be careful with your principal components. Evolution. 2019 Oct;73(10):2151-2158.

20.Schellenberg JB, Maislin G, Schwab RJ. Physical findings and the risk for obstructive sleep apnea. The importance of oropharyngeal structures. *Am J Respir Crit Care Med.* 2000; 162: 740-8.

21.Schwab RJ, Leinwand SE, Bearn CB, et al. Digital Morphometrics: A New Upper Airway Phenotyping Paradigm in OSA. *Chest* 2017; 152: 330-42.

22.Chung F, Abdullah HR, Liao P. STOP-BANG Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. *Chest*. 2016;149(3):631-638.

23.Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. *Eur Respir J.* 2014; 44:1600–7.

24.Linnen DT, Escobar GJ, Hu X, et al. Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. *J Hosp Med.* 2019; 14:161–9.

25.Mencar C, Gallo C, Mantero M, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics J.* 2019;1460458218824725. Available: http:// doi:10.1177/1460458218824725

26.Bozkurt S, Bostanci A, Turhan M. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf Med.* 2017; 56:308–18.

Author affiliations

[1]Department of Respiratory and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

[2]Institute of Respiratory Medicine, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Acknowledgements The authors thank Peizhe Liu for his assistance with the SVM algorithm.

Author contributions Liu Zhang and Ya Ru Yan drafted the manuscript, had full access to all the data, and takes responsibility for the integrity of the data and the accuracy of the data analysis. Shi Qi Li contributed to data analysis and interpretation. Qing Yun Li and Shi Qi Li contributed to study conception and design. Hong Peng Li, Ying Ni Lin, Ning Li, Xian Wen Sun, Yong Jie Ding contributed to data collection, critical revision. Chuan Xiang Li and Qing Yun Li contributed to final approval of the manuscript.

Funding This work was supported by grants from National Key R&D Program of China (2018YFC1311900) and the National Natural Science Foundation of China (founding no. 81770084, 81700084, 81570082). No conflicts of interest are to be reported

Competing interests None declared.

Patient and Public Involvement statement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research

Ethics approval The study was approved by the ethics committee of the Shanghai Jiao Tong University School of Medicine at Ruijin Hospital, with the following reference number: 2018-107.

Data availability statement Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi: 10.5061/dryad.qnk98sfhg.

FIGURE LEGENDS:

Figure 1. Flow chart showing screening, randomisation and algorithm.

481 participants were enrolled in our study after excluding 31 patients. Age, sex, NC, WC, BMI, and facio-cervical measurements (MID, H/TSD) were potential factors related to OSA due to the significant difference between the two groups ($p<0.05$). We chose the following parameters to setup the SABIHC2 model based on training dataset, whose name refers to ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C*****s**). Then, the model was verified on testing dataset.

Abbreviations: BMI: body mass index; MID: maximum interincisal distance; H/TMD: ratio of height to thyromegaly distance; H/TSD: ratio of height to thyro-sternum distance; NC: neck circumference; WC: waist circumference;

Figure 2. Optimal hyperplane for Support Vector Machine (SVM). The green area means the SVM predicting AHI<15 events·h^{-1}, the purple means the SVM predicts the AHI≥15 events·h^{-1}. Green dots represent no or mild OSA (AHI<15 events·h^{-1}), blue dots represent moderate-severe OSA (AHI≥15 events·h^{-1}). The accuracy of SABIHC2 model based on all patients was 0.881, of which 0.928 for moderate-severe OSA (blue dots in purple area), and 0.719 for no or mild OSA (green dots in green area). The boundary was obtained from the SVM classifier and the figure was created using Python.

Abbreviations: SVM: support vector machine; SABIHC2 model: ***s***ex, ***a***ge, ***B***MI, M***I***D, ***H***/TSD, N***C***, and W***C*** (***2 C*****s**).

Figure 3. ROC curve of SABIHC2 model and STOP-BANG.

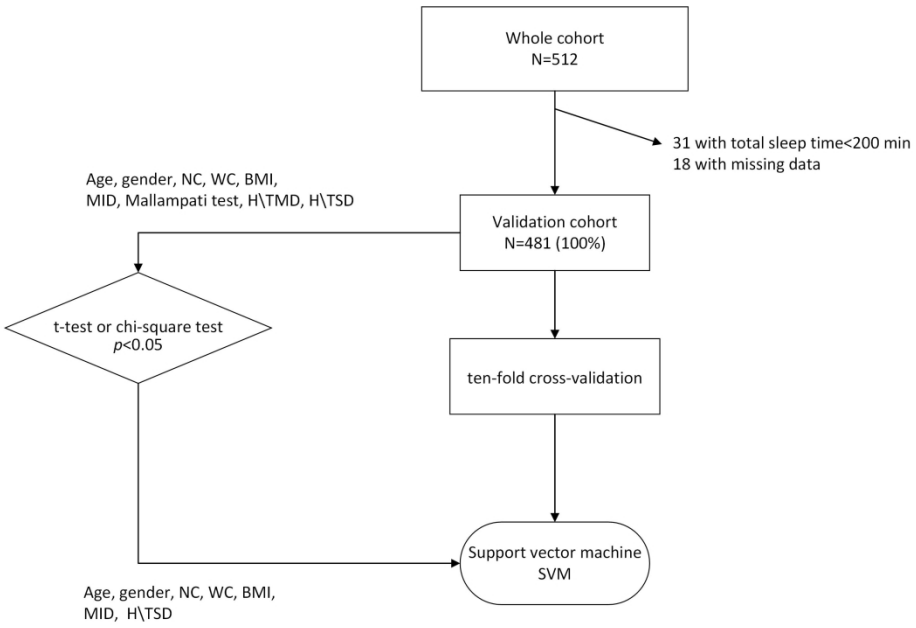
A: The ROC curve of SABIHC2 model and STOP-BANG. The AUC of the SABIHC2 model based on the training database (n=337, 70.1%, brown line) was 0.829 (95% CI, 0.777 to 0.880), with a

sensitivity of 0.930 and specificity of 0.727. About testing database (n=144 29.9%, red line), the AUC was 0.787 (95% CI, 0.702 to 0.872), with a sensitivity of 0.874 and specificity of 0.700, respectively. The ROC curve of STOP-BANG questionnaire (blue line) on the testing database was 0.597 (95% CI, 0.499 to 0.694). STOP-BANG>2 showed the sensitivity and specificity of 0.832 and 0.220, respectively, while STOP-BANG>4 with the sensitivity of 0.495 and specificity of 0.780.

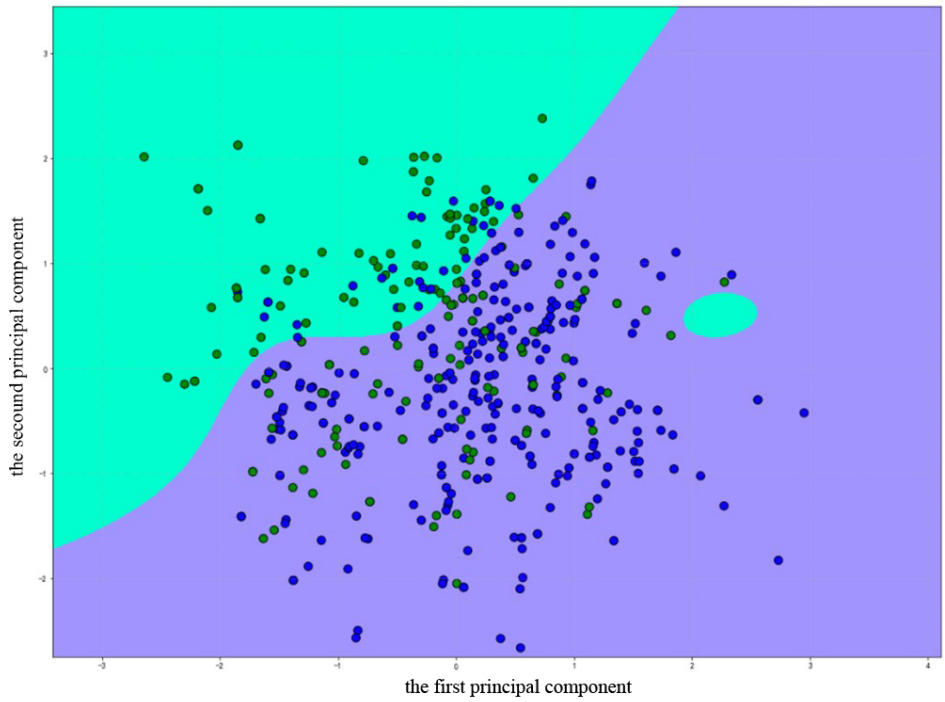
Brown: SVM based on training dataset; red: SVM based on testing dataset; blue: STOP-BANG based on testing dataset.

B: The ROC curve of SABIHC2 model and STOP-BANG based on asymptomatic patients (ESS < 10) and sleepiness patients (ESS ≥ 10). In asymptomatic patients, SABIHC2 model (red line) remarkably demonstrated better predictive ability than STOP-BANG questionnaire (dark blue line), with AUC (0.810 *vs.* 0.571), sensitivity (0.910 *vs.* 0.404), and specificity (0.709 *vs.* 0.788). Similarly, SABIHC2 model (orange line) had higher predictive power than STOP-BANG questionnaire (soft blue line) in sleepiness patients (ESS ≥ 10), with sensitivity (0.830 *vs.* 0.813), and specificity (0.699 *vs.* 0.478) and AUC (0.762 *vs.* 0.606).

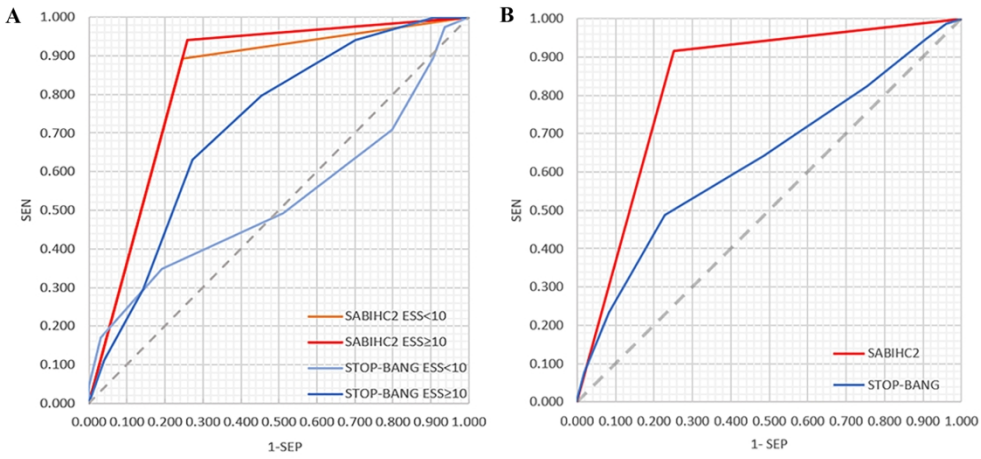
Abbreviations: AUC: the area under ROC curve; SVM: support vector machine; CIs: confidence intervals.



279x175mm (300 x 300 DPI)



90x65mm (300 x 300 DPI)



189x89mm (300 x 300 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	Page 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	Page 2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	Page 4
	4	Study objectives and hypotheses	Page 4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	Not applicable
Participants	**6**	Eligibility criteria	Page 5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	Page 5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	Page 5
	9	Whether participants formed a consecutive, random or convenience series	Page 5
Test methods	**10a**	Index test, in sufficient detail to allow replication	Page 5, 6
	10b	Reference standard, in sufficient detail to allow replication	Page 5, 6
	11	Rationale for choosing the reference standard (if alternatives exist)	Page 4, 5
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Not applicable
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Page 4, 5
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	No
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	Yes
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	Not applicable
	15	How indeterminate index test or reference standard results were handled	Not applicable
	16	How missing data on the index test and reference standard were handled	Page 9
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	Page 6
	18	Intended sample size and how it was determined	Not applicable
RESULTS			
Participants	**19**	Flow of participants, using a diagram	Figure 1
	20	Baseline demographic and clinical characteristics of participants	Page 8, Table 1
	21a	Distribution of severity of disease in those with the target condition	Page 8
	21b	Distribution of alternative diagnoses in those without the target condition	Not applicable
	22	Time interval and any clinical interventions between index test and reference standard	Not applicable
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Page 9
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	Page 8, 9, 10
	25	Any adverse events from performing the index test or the reference standard	Not applicable
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	Page 12
	27	Implications for practice, including the intended use and clinical role of the index test	Page 12
OTHER INFORMATION			
	28	Registration number and name of registry	Page 16
	29	Where the full study protocol can be accessed	Page 16
	30	Sources of funding and other support; role of funders	Page 15

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.