國立臺灣大學電機資訊學院資訊工程學系

博士論文

Department of Computer Science and Information Engineering,

National Taiwan University

Doctoral Dissertation

以支持向量機預測睡眠呼吸中止症

Support Vector Machine Prediction of Obstructive Sleep Apnea

黃文圻

Wen-Chi Huang

指導教授：賴飛羆 博士

Advisor: Feipei Lai, Ph.D.

中華民國 109 年 1 月

January, 2020

doi:10.6342/NTU202000475

國立臺灣大學博士學位論文
口試委員會審定書

以支持向量機預測睡眠呼吸中止症

Support Vector Machine Prediction of Obstructive Sleep Apnea

本論文係黃文圻君（學號 D03922025）在國立臺灣大學資訊工程學系完成之博士學位論文，於民國 109 年 1 月 7 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

賴飛羆

（指導教授）

李[illegible]　　汪大暉

周迺寬　　周承復

吳經閔　　歐陽彥

系　主　任　莊永裕

doi:10.6342/NTU202000173

中文摘要

睡眠多項生理檢查 (PSG) 是現今阻塞型睡眠呼吸中止症 (OSA) 的標準診斷工具，然而進行 PSG 時，受測者需要配戴多種感測器，於睡眠檢查室進行一整夜的檢查，既耗時又不舒適。除此之外，非睡眠專科醫師在門診時，多會將所有疑似有睡眠障礙的病人，轉診至睡眠科，或安排 PSG 進行診斷，皆可能造成 OSA 診斷效率下降。為解決上述問題，已有多項相關研究使用非連續量測資訊來設計 OSA 的快速篩檢工具，以達到更有效率的 PSG 排程。然而在以往的研究中，多有靈敏度 (sensitivity) 高，特異度 (specificity) 低的問題，且在資料收集時，部份資料定義不夠明確，造成使用上的困難。本研究使用個人基本資料、身體量測資訊、共病症及睡眠症狀等 32 項候選特徵，為非睡眠專科醫師設計一項 OSA 的快速篩檢工具。本研究比較了八種方法的分類效能，並選擇其中結果最佳的支持向量機 (SVM) 來進行特徵篩選與最佳化。本研究提出了二個階段的特徵篩選，再使用 SVM 各自對三種不同嚴重程度的 OSA [呼吸紊亂指數 (AHI) ≥5/hr、AHI ≥15/hr、AHI ≥30/hr] 進行二元分類器的訓練。特徵篩選的過程中發現，為達到較佳的分類效果 (AUROC ≥0.80)，當預測三種嚴重程度的 OSA 時，分別需要 2、6 及 6 個特徵。在 6,875 人的資料中，使用 2 個特徵進行預測、針對 AHI ≥5/hr 的 OSA 患者，分類器效能（精準度、靈敏度、特異度）達到 (75.7%、76.4%、72.2%)，使用 6 個特徵進行 AHI ≥15/hr 的預測效能為 (73.7%、75.1%、70.4%)，使用 6 個特徵進行 AHI ≥30/hr 的預測效能為 (73.0%、74.9%、70.0%)。本研究最終採用 6 個特徵做為快篩系統的輸入，並將訓練好的三個分類器，結合 RedCap 來提供網頁化的問卷，可即時回饋問卷填寫者 OSA 預測的結果，並提供醫師資料收集及閱覽的功能。

關鍵字：呼吸中止症，支持向量機，機器學習，特徵篩選，預測系統

doi:10.6342/NTU202000475

ABSTRACT

Polysomnography (PSG) is the gold standard for diagnosis of obstructive sleep apnea (OSA), but it is costly and access is often limited. Besides, the non-sleep specialist physician (NSSP) usually transfers most of suspected patients with sleep disorder to department of sleep for PSG. This situation lets OSA diagnosis more inefficient.

To solve above problems, there were several studies used discretely objective and subjective information to develop OSA screening tools which provide diagnosis support and prioritize PSG. However, several OSA prediction models from recent studies had two issues: one is sensitivity range was higher than specificity range, and another is feature definitions as model input may be not clear enough. The first issue may lead more non-OSA patients to do PSG test, and the second may cause the difficulty of applying the prediction model. This study proposed an OSA screening tool for NSSP by 32 features which includes patient basic information, anthropometrics, comorbidities, and sleep habitual information. After comparing the performance of 8 algorithms, the support vector machine (SVM) had the best performance and was chose to be optimized with feature selection.

The proposed method of this study applied a two stages of feature selection and using support vector machine to train classifier for OSA prediction. There were three classifiers trained for three apnea-hypopnea-index (AHI) cutoff (5/hr, 15/hr, and 30/hr). This study discovered that the classifier required more features with larger AHI cutoff to reach AUROC ≥0.80. Three AHI cutoffs required 2, 6, and 6 features, respectively. Three classifiers were trained and tested with 6,875 subject data. With 2, 6, and 6 features as input to predict three AHI cutoffs (5/hr, 15/hr, and 30/hr), the performance (accuracy, sensitivity, and specificity) achieved (74.24%, 74.14%, and 74.71%), (72.64%, 75.18%, and 68.73%), and (70.28%, 70.26%, and 70.30%), respectively. Finally, 6 features were selected and used in a web-based

doi:10.6342/NTU202000475

system which integrated trained SVM models. The web-based system is capable of giving user OSA risk as feedback in real-time, and it also lets medical staff collect and review the input data and outcome results.

Keywords: Obstructive Sleep Apnea, Support Vector Machine, Machine Learning, Feature Selection, Prediction System

CONTENTS

doi:10.6342/NTU202000475

LIST OF FIGURES

doi:10.6342/NTU202000475

LIST OF TABLES

doi:10.6342/NTU202000475

Chapter 1. Introduction

1.1 Background and Significance

Obstructive sleep apnea (OSA) is characterized by repeated episodes of upper airway obstruction that results in cessation of airflow during sleep [1]. OSA is a common disease with a prevalence of 9-38% in the general population [2]. Risk factors for OSA included age, male, obesity, smoking, anomalies of craniofacial features, and menopause in women [3]. Symptoms suggestive of OSA included habitual snore, witnessed apnea, choking or gasping at sleep, frequent awakening, nocturia, unrefreshed sleep, and daytime sleepiness [3]. Early diagnosis of OSA is essential because untreated OSA may increase the probability of developing cardiovascular diseases, metabolic disorders, and neurocognitive dysfunctions [4]. The overnight polysomnography (PSG) is the gold standard for the diagnosis of OSA, and the OSA severity is commonly determined by apnea-hypopnea index (AHI) with cutoff ≥5/hr for the presence of OSA, ≥15/hr for the presence of moderate-severe OSA, and ≥30/hr for the presence of severe OSA [2]. However, PSG is costly and the access is often limited. As a result, prioritizing patients with high risk for moderate-severe OSA for PSG can be crucial for many sleep laboratories.

A recent meta-analysis showed that care of patients with OSA by non-sleep specialist physician (NSSP) and sleep-specialist physician brought similar outcomes in terms of mortality, quality of life, adherence, and symptom score. Since most NSSPs in the included studies had extensive training in sleep medicine, the results may be inferior in the NSSP who were less seasoned or inadequately trained [5]. Hence, the development of a screening model based on clinical features commonly collected at clinic visits to predict the likelihood of OSA

doi:10.6342/NTU202000475

would be extremely practical for NSSP. Such a model can also help NSSP to prioritize patients with high pre-test probability of OSA for PSG [6].

To build an OSA screening model, there were several studies used discretely objective and subjective information to develop OSA screening tools which provide diagnosis support and prioritize PSG. However, several OSA prediction models from recent studies had two issues: one is sensitivity range was higher than specificity range, and another is feature definitions as model input may not clear enough. The first issue may lead more non-OSA patients to do PSG test, and the second may cause the difficulty of applying the prediction model.

With above reasons, this study proposed a method to have the balance between sensitivity and specificity, and the proposed method could address which features are important to predict OSA with clear definitions. Additionally, the model should be easy to use for NSSPs.

1.2 Literature Review

Prediction models reported in the literature were mostly built using clinical features (Table 1) including demographics (age, gender, smoking, alcohol consumption), co-morbidities, anthropometrics, OSA symptoms, physical findings [7], and physiologic measurements (e.g. blood pressure, overnight pulse oximetry, and pulmonary function) [8] collected from either sleep lab or community-based population. Among prediction models proposed so far, the sensitivity to predict AHI ≥5/hr ranged from 66% to 100% while the specificity ranged from 30.8% to 76.2%, and the sensitivity predict AHI ≥15/hr ranged from 60.3% to 92.7% while specificity ranged from 33.3% to 90.7% (Table 2). The wide-range discriminative ability of models could be attributed to the model complexity, number of participants, prevalence of OSA, and imbalance between different OSA severity proportion. Moreover, most prediction models for OSA tend to have a higher sensitivity with a lower specificity to promote early diagnosis (Table 1). These models can potentially cause a high false-positive rate and lead to over-prescription of PSG. Nevertheless, some models were established based on the data of which patients with comorbidities were excluded [9], where the generalizability of clinical implication would be constrained.

It is also crucial to validate the model efficacy in subgroups categorized with different features. For example, male patients often have fat distributed to the upper body and a higher percentage of snoring than female patients [10, 11]. Elder patients with OSA may be less susceptible to adverse effects of OSA like sleepiness, impaired quality of life, and mortality compared to middle-aged patients [12, 13]. It is also known that the Asian patients have higher AHI compared to body mass index (BMI)-matched Caucasians due to narrower craniofacial features [14]. Therefore, it may be more efficient to build a whole new model for Asian population to predict OSA with local dataset.

doi:10.6342/NTU202000475

Table 1. The Algorithm and selected features of related OSA-prediction models

Author year	**Algorithm**	**Feature No.**	**Feature detail**
Kirby 1999 [7]	GRNN	23	Age, gender, frequent awakening, witnessed apnea, observed chocking, excessive daytime sleepiness, ESS, HT, alcohol consumption, smoking (pack-year), height, weight, BMI, SBP ≥ 140, DBP ≥ 90, tonsillar enlargement, soft palate enlargement, crowding of the oral pharynx, sum of the clinical score for the binary categorical values
Rowley 2000 [15]	LR	4-6	Witnessed apnea, HT, BMI, age, gender, snoring, gasping, neck
Rodsutti 2004 [3]	LR	5	Age, sex, BMI, snoring, witnessed apnea
Rodsutti 2004 [16]	LR	4	Gender, BMI, snoring index, chocking index
Sharma 2006 [17]	LR	4	Gender, BP, BMI, snoring
Takegami 2009 [18]	Boosting	10	Neck circumference, BMI, age, snoring frequency, waist circumference, snoring loudness, gender, SOL, response to "What is the chance that you would doze off or fall asleep while sitting and reading?", and presence or absence of a heart attack.
Bouloukaki 2011 [7]	LR	4	Gender, EDS, neck circumference, and BMI
Caffo 2010 [19]	LR	1. 5 2. 4	1. Age, BMI, waist circumference, gender 2. Age, waist circumference, ESS score, and minimum oxygen saturation (SaO2)
Bouloukaki 2011 [20]	LR	5	Neck circumference, BMI, snoring, age, and gender
Zou 2013 [9]	LR	4	Gender, age, BMI, and snoring frequency
Ustun 2016 [11]	SLIM	1. 5 2. 1	1. Age, HTN, BMI, and gender 2. Age, BMI, DM, HTN, smoker, and gender
Marti-Soler 2016 [21]	SVM	4	BMI, neck circumference, waist circumference, and age
Shah 2016 [22]	LR	14	BMI, ASA score, age, dyslipidemia, chronic pulmonary disease, liver disease, HTN, CHF, pulmonary HTN, AF, DM, CAD, and hemiplegia/paraplegia
Ustun 2016 [23]	LR	5	Neck circumference, BMI, snoring, age, and gender
Traxdorf 2017 [15]	LR	5	ESS score, age, gasping, cardiovascular risk factors (e.g. CHF, CAD, myocardial infarction, AF, stroke), and witnessed apneas
Liu 2017 [24]	LR	2	Neck circumference, and age

Abbreviation: GRNN, generalized regression neural network; LR, logistic regression; SLIM, supersparse linear integer models.

doi:10.6342/NTU202000475

Table 2. Summary of related OSA-prediction models

Author year	**Source**	**Subject No.**	**Age (y/o)**	**Male (%)**	**AHI cutoff (/hr)**	**Preva (%)**	**AUROC**	**Sen (%)**	**Spec (%)**	**PPV (%)**	**NPV (%)**
Kirby 1999 [7]	S	150	51.0	66.0	≥10	65.00	0.94	98.90	80.00	88.10	98.00
Rowley 2000 [15]	S	370	46.5	51.6	1. ≥10	1. 67.02	1. 0.67-0.74	1. 76-96	1. 13-54	1 .66-77	N/A
					2. ≥20	2. 48.65	2. 0.70-0.76	2. 33-39	2. 87-93	2. 72-85	
Rodsutti 2004 [16]	S	243	51.0	63.0	≥5	72.02	0.79	100.00	30.88	80.66	84.50
Sharma 2006 [17]	S	104	NA	77.9	≥15	48.08	NA	82.00	90.70	89.13	84.50
Takegami 2009 [18]	S	308	43.8	99.0	1. ≥15*	1. 22.40	1. 0.78	1. 73.90	1 .66.10	1 .38.62	1. 89.77
					2. ≥30	2. 6.81	2. 0.85	2. 95.20	2. 61.00	2. 93.09	2. 85.62
Caffo 2010 [19]	C	1,383	65.0	47.3	≥7*	37.80	0.75	66.00	70.00	57.00	77.21
Bouloukaki 2011 [20]	S	2,690	50.7	79.0	≥15	79.00	0.78	70.00	73.10	81.00	62.00
Zou 2013 [9]	S	784	41.0	83.2	≥5	83.80	1. 0.84	1. 86.91	1. 74.80	1. 94.69	1. 52.49
							2. 0.98	2. 94.22	2. 85.83	2. 97.17	2. 74.15
Marti-Soler 2016 [21]	C	1,042	42.0	45.0	>20	11.52	0.74	85.00	77.00	33.00	98.00
Shah 2016 [22]	C	12,158	48.1	39.4	≥15	9.02	0.83	77.00	75.00	23.39	97.05
Ustun 2016 [23]	S	1,922	50.2	58.7	>5 or >10*	76.90	1. 0.77	1. 82.80	1. 56.20	1. 86.29	1. 49.53
							2. 0.79	2. 83.20	2. 58.90	2. 87.08	2. 51.29
Liu 2017 [24]	S	1,154	47.4	70.7	1. ≥15	1. 61.87	N/A	1. 68.35	1. 82.55	1. 86.40	1. 61.65
					2. ≥30	2. 44.04		2. 68.32	2. 79.85	2. 64.99	2. 77.45
Shin 2017 [25]	SW	108,781	54.4	44.4	>5	2.08	0.82	72.30	76.20	6.06	99.23
Tan 2017 [26]	C	242	48.3	50.4	1. ≥15	1. 28.10	1. 0.70	1. 60.30	1. 79.70	1. 53.90	1. 83.70
					2. ≥20	2. 20.20	2. 0.74	2. 69.40	2. 78.20	2. 44.70	2. 91.00
					3. ≥25	3. 14.50	3. 0.73	3. 71.40	3. 75.40	3. 32.90	3. 94.00
					4. ≥30	4 10.70	4. 0.71	4. 69.20	4. 73.10	4. 23.70	4. 95.20
Traxdorf 2017 [27]	S	100	48.1	76.0	1. ≥5	1. 70.00	N/A	1. 94.30	1. 50.00	1. 81.50	1. 78.90
					2. ≥15	2. 55.50		2. 92.70	2. 33.30	2. 62.90	2. 78.90
					3. ≥30	3. 26.00		3. 92.30	3. 22.90	3. 29.60	3. 89.50
Duarte 2018 [28]	S	2,035	44.0	53.2	1. ≥5	1. 76.40	1. 0.78	1. 83.10	1. 58.20	1. 86.50	1. 51.60
					2. ≥15	2. 54.70	2. 0.76	2. 88.70	2. 45.30	2. 66.20	2. 76.80
					3. ≥30	3. 35.80	3. 0.75	3. 91.50	3. 36.80	3. 44.60	3. 88.60

* respiratory disturbance index.

Abbreviations: S, sleep clinic; C, community, SW; surgical ward; Preva, prevalence.

doi:10.6342/NTU202000475

1.3 Machine Learning

Machine learning has been found to be a potential means in addressing these problems by its massive parallelism, self-organization, adaptive learning capability, and robustness. Support Vector Machine (SVM) has been increasingly applied in medical healthcare during the past few years since it can provide systematized architecture for analyzing and extracting important information from complex data [29]. Hence, SVM-based machine learning model may be promising for the prediction of OSA. In this study, to prove that SVM is the most appropriate method of OSA prediction, classification performance of five machine learning algorithms and eight approaches were estimated by AUROC. The eight approaches included: SVM with radial basis function (RBF) kernel, SVM with polynomial (Poly) kernel, logistic regression (LR), neural network (NN), random forest (RF), RF stacking with LR, gradient boosting tree (GBT), and GBT stacking with LR.

1.4 Aim of this Study

The present study aimed to propose an easy-to-use and accurate model to identify patients with OSA at three AHI cutoffs (≥5/hr, ≥15/hr, ≥30/hr). We developed a data-mining driven SVM prediction model using a large-scale sleep-lab database with features routinely collected at clinic visits. The model discriminative ability was also tested in the subgroups categorized with gender (men versus women) and age (<65 versus ≥65 y/o). The model discriminative ability was also compared with that of logistic regression, Berlin Questionnaire, NoSAS Score, and Supersparse Linear Integer Models (SLIM) scoring system. Finally, the model was integrated into a web-based questionnaire for NSSP practically.

doi:10.6342/NTU202000475

1.5 Organization of Thesis

This thesis shows how to build an SVM model for OSA prediction with two stages of feature selection and compare the proposed model with other approaches. The organization of this thesis is as following. Chapter 2 introduces the dataset collection, data definition, machine learning algorithms, feature selection, and SVM optimization processes. We first describe the data source, inclusive rules, exclusive rules, and feature definition. Then we introduce several popular machine learning algorithms for classification. These machine learning algorithms were tested with the targeted dataset, and find out which one had the best performance. The last section in Chapter 2 is focusing on the proposed feature selection method and the SVM optimization.

Chapter 3 provides the distribution and statistics results of each feature and outcome. Next, the performance of each algorithms are presented by receiver operating characteristic (ROC) curve. In this chapter, the feature selection results and the SVM performance after optimization are also addressed. The subgroups of the original dataset were tested to show the model discriminative ability in the last section.

In Chapter 4, we discuss pros and cons of the proposed method to predict OSA. We compared the proposed method by dataset, performance, selected features and feature definition with other related works. We reveals the limitation of the proposed method in the third section. Finally, we addressed the future work and conclusion.

doi:10.6342/NTU202000475

Chapter 2. Method

This chapter will introduce the dataset, algorithm selection, feature selection and the model estimating method. First, the data collection was addressed to introduce the data profile about data source, including and excluding criterion, outcome definition, and feature definition. Second, to realize which machine learning method is the most appropriate for OSA prediction, eight approaches were estimated. The SVM was chosen because it had the maximum AUROC, and therefore, we designed an optimization process to select features and reach the promising performance.

2.1 Dataset and Polysomnography

The dataset developed from information prospectively collected from 7,830 adult patients who underwent initial overnight PSG for the first time in the Center of Sleep Disorder of National Taiwan University Hospital between Jan. 2009 and Dec. 2016. For data-mining, only patients who had any following conditions were excluded: non-Chinese (n = 11), total recording time <240 min (n = 7), and missing data (n = 936). A total of 6,875 patients, with 5,223 men and 1,652 women (5,985 <65 y/o and 890 ≥65 y/o) were included (Tables 4, 5).

Thirty-two clinical features including demographics, anthropometrics, co-morbidities, self-reported habitual sleep patterns, and OSA symptoms were collected through self--administered questionnaires and medical records (Table 3). The demographics included age, gender, smoking, alcohol consumption, and hypnotic use defined as taking hypnotics ≥1 time/week over the past month. Anthropometrics included BMI, neck circumference, and waist circumference. Sleep history and OSA symptoms were collected with a self-administered questionnaire (description and definition listed in APPENDIX A). Sleep history included unrefreshed sleep, subjective sleepiness, frequency of awakening, awakening ≥3 times/night during sleep, minutes of sleep onset latency (SOL) and hours of sleep duration over the past month. In addition, the SOL <30 min and the sleep duration categorized as <6 hr/day, 6-8 hr/day, and ≥8 hr/day) were added. Subjective sleepiness was assessed by the Epworth Sleepiness Scale (ESS) with excessive daytime sleepiness (EDS) defined as ESS ≥10 [30]. The OSA symptoms included snore, witnessed apnea, frequency of nocturia, witnessed leg jerks at sleep, morning headache, nocturia ≥2 times/night and dry throat at wake up. All of 32 feature definitions were addressed in APPENDIX A.

Overnight PSG (Embla N7000, Medcare Flaga, Reykjavik, Iceland) was performed as previously reported [31]. Sleep stages and respiratory events were scored according to the 2007 AASM scoring rule [32]. Apnea was defined as ≥90% decrease in airflow for ≥10 seconds while hypopnea was ≥30% decrease in airflow ≥10 seconds associated with ≥4% reduction in arterial oxygen saturation. The PSG parameters collected included sleep efficiency, percentage of slow-wave sleep (% SWS) and % REM, AHI, oxygen desaturation index (ODI), percentage of total sleep time with SpO2 <90% (%TST-SpO2 <90%), and arousal index (AI).

doi:10.6342/NTU202000475

Table 3. Included 32 input features and categories

Demographics	**Anthropometric**	**Co-morbidities**	**Sleep history**	**Symptoms suggestive of OSA**
Age	BMI	Hypertension	SOL (min)	Snore
Gender	Neck circumference	Diabetes	SOL < 30 min	Witnessed apnea
Alcohol consumption	Waist circumference	CAD	Sleep duration, < 6, 6-8, and ≥ 8 hr	Freq. of nocturia (time/night)
Current smoking		CHF	Unrefreshed sleep	Nocturia ≥ 2 times/night
Hypnotics		CVA	Freq. of awakening at sleep (time/night)	Witnessed leg jerks at sleep
		CKD	Awakening at sleep ≥3 times/night	Morning headache
		COPD	ESS	Dry throat at wake up
		Asthma	EDS	
		Hypothyroidism		

Abbreviations: BMI, body mass index; circum., circumference; SOL, sleep onset latency; CAD, coronary artery disease; CHF, congestive heart failure; CVA, cerebrovascular accident; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; EDS, excessive daytime sleepiness

doi:10.6342/NTU202000475

2.2 Machine Learning Algorithm Selection

To find out the most appropriate method of OSA prediction, classification performance of five machine learning algorithms and eight approaches were estimated by AUROC. Eight approaches included: SVM with radial basis function (RBF) kernel, SVM with polynomial (Poly) kernel, logistic regression (LR), neural network (NN), random forest (RF), RF stacking with LR, gradient boosting tree (GBT), and GBT stacking with LR. The following contents in this section will introduce each algorithm and stacking process. These eight approaches were implemented and tested with collected dataset. The input features number was 32 totally. The AUROC of each approach was estimated by 5-fold cross-validation.

Support Vector Machine

The support vector machines (SVM) have become a common used method to solve difficult classification problems in a wide range of real application domains. There are two key advantages of using SVM: one is SVM had good generalization performance even in case of high-dimensional dataset [33]; another is SVM could find non-linear solutions efficiently by the kernel trick. There are two types of SVM method, one is linear and another is non-linear. In the case of OSA prediction, we found that the linear model could not solve the classification problem appropriately. Therefore, the following content will focus on introducing non-linear SVM model. Before starting to introduce the kernel function, please refer the notation of SVM decision function from [34].

The kernel trick is the key of non-linear SVM. What kernel function does is mapping the original vectors to a higher dimensional space, and the new mapping is linearly separable. There are two kernel function were tested as candidate, one is radius basis function (RBF) kernel as equation (1), and another is polynomial (Poly) kernel as equation (2).

doi:10.6342/NTU202000475

$$k_{\sigma}^{RBF}(\mathrm{x},\mathrm{x}') = \exp\left(-\frac{1}{\sigma}\|\mathrm{x}-\mathrm{x}'\|^2\right) \tag{1}$$

$$k_{d,K}^{Poly}(\mathrm{x},\mathrm{x}') = (\langle \mathrm{x},\mathrm{x}' \rangle + K)^d \tag{2}$$

The RBF kernel is also called Gaussian kernel. In equation (1), where $\sigma > 0$ is a parameter that controls the width of the Gaussian. It represents the degree d of the polynomial kernel in controlling the flexibility of the resulting classifier. The RBF kernel is zero if the squared distance $\|\mathrm{x}-\mathrm{x}'\|^2$ is much larger than σ. Basically, when σ is large, a given data point *x* has a nonzero kernel value relative to any example in the set of examples. Therefore, the whole set of support vectors affects the value of the discriminant function at *x*, leading to a smooth decision boundary [34]. In the other hand, with smaller σ, the kernel becomes more local, forming to greater curvature of the decision curve / surface, which means with too small σ may cause over-fitting problem. So, to prevent over-fitting, we restrict the minimum σ as 0.1 in this study.

In equation (2), where K is often chosen to be zero (homogeneous) or one (inhomogeneous). The feature space for the inhomogeneous kernel consists of all monomials with degree up to d [35]. And yet, its computation time is linear in the dimensionality of the input space. The kernel with d =1 and K =0, denoted by *k* linear, is the linear kernel leading to a linear discriminant function. The degree of the Poly kernel controls the flexibility of the resulting classifier. The lowest degree polynomial is the linear kernel, which is not sufficient when a nonlinear relationship between features exists.

Logistic Regression

Logistic Regression (LR) is another algorithm which could be used for classification, it is based on the concept of probability. Let Y denote the binary response variable of interest and

X_1, …, X_p the random variables considered as explaining variables, termed features in this paper. The logistic regression model links the conditional probability P ($Y = 1$ | X_1, ..., X_p) to X_1, …, X_p through

$$P\left(Y = 1\middle|X_1, \ldots, \mathrm{X}_\mathrm{p}\right) = \frac{\exp\left(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p\right)}{1+\exp\left(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p\right)}\ , \qquad (3)$$

where β_0, β_1, …, β_p are regression coefficients, which are estimated by maximum-likelihood from the considered dataset. The probability that $Y = 1$ for a new instance is then estimated by replacing the β's by their estimated counterparts and the X's by their realizations for the considered new instance in equation (3). The new instance is then assigned to class $Y = 1$ if P ($Y = 1$) > c, where c is a fixed threshold, and to class $Y = 0$ otherwise. The commonly used threshold $c = 0.5$, which is also used in this study, generates a so-called Bayes classifier [36]. As for all model-based methods, the prediction performance of LR depends on whether the data follow the assumed model. In this study, the LR will still be a comparing method after algorithm selection. Because the LR was the most used method in recent studies, and it is robust in clinic.

Artificial Neural network

Artificial NNs are formed with at least three-layer neuron structures, which are the input, hidden (middle) and output layers. The input layers collect numerical information data with feature sets and activation values. Input values are propagated through the interconnected neurons to the hidden layer. In the hidden layer, the input neurons are summed in order to compute weighted sum of the input neurons; and summed neurons are further combined to produce results in the output layer using an activation (or transfer) function [37, 38]. Both neurons and connection contain adjustable weights during the learning process. The summed neurons will transform mathematically in the output layer if the activation function threshold is exceeded.

A number of times the training functions are used to update the connection weights in the process of feeding the input values and terminating with output values in ANN is called an Epoch [38]. This is where the inputs of artificial neurons are multiplied by weights, and the resultant of these summation are fed to the output layer through an activation function [39]. The frequently used of activation functions include linear, sigmoid and hyperbolic tangent functions. The training terminates when the maximum epoch value and/or the validation checks are reached. The resultant trained data is fed into the test data in order to examine the ANN's performance [40].

The most common learning rule of ANNs is back-propagation (BP), which is a supervised learning approach and can be used for training the deep neural networks [40]. BP adjusts the weights of neurons through the calculated errors and enables the network to learn from the training process. Typical problem solving of ANNs include three archetypes of learning, i.e. supervised learning, unsupervised learning and reinforcement learning [39]. To improve the performance of classification, 5 hidden layers were used for NN's training during model selection stage in this study.

Random Forest

The random forest (RF) is an "ensemble learning" technique consisting of the assemble of a large number of decision trees, resulting in variance reduction compared to the single decision tree. In this study, the Leo Breiman's version of RF was considered [41], while acknowledging that other variants develop, for example RF based on conditional inference

doi:10.6342/NTU202000475

trees which address the problem of variable selection bias and perform better in some cases, or extremely randomized trees.

In the original version of RF [41], each tree of the RF is built based on a bootstrap sample drawn randomly from the original dataset using the conditional inference tree (CART) method and the decrease Gini impurity as the splitting criterion [41]. When building each tree, at each split, only a given number *mtry* of randomly selected features are considered as candidates for splitting. RF is usually considered a black-box algorithm, as gaining insight on a RF prediction rule is hard due to the large number of trees. In this study, one splitting criteria of the DGI and information gain were chosen depend on which performance is better.

Gradient Boosting Tree

Gradient boosting tree (GBT) was originally called gradient boosting machine, which was designed by Friedman (2001). The learning procedure of GBT consecutively fits new models to provide a more accurate estimate of the response variable. The principle idea behind this algorithm is to construct the new base-learners to be maximally correlated with the negative gradient of the loss function, associated with the whole ensemble. The loss functions applied can be arbitrary, but to give a better intuition, if the error function is the classic squared-error loss, the learning procedure would result in consecutive error-fitting. In general, the choice of the loss function is up to the researcher, with both a rich variety of loss functions derived so far and with the possibility of implementing one's own task-specific loss [42].

The high flexibility of GBT allows high customizability to any particular data-driven task. It addresses a lot of freedom into the model design so that making the choice of the most appropriate loss function a matter of trial and error. A particular GBT can be designed with

doi:10.6342/NTU202000475

different base-learner models on board. In this study, the decision tree algorithm was applied as base-learner [42].

Model Stacking

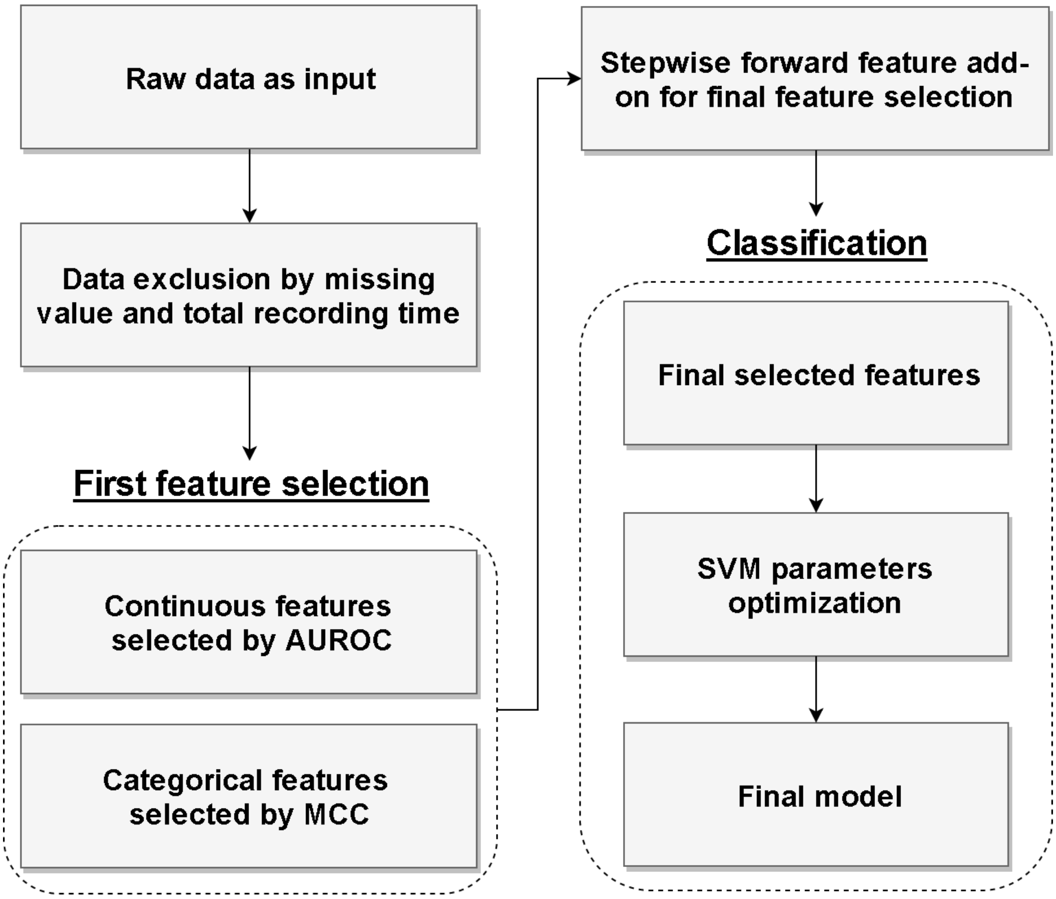
Stacking is a technique which usually is used to build a combining method for multiple classifiers ensemble. In the past, stacking showed success in data science competition, the Netflix competition for example, which was an open competition on using historical ratings of users to predict new films ratings. There were many teams with top rank employed stacking to combine classifiers. In particular, the winning team [43] applied stacking to combine hundreds of models, which accomplished the top performance. The stacking method with K-fold cross validation was addressed in Figure 1 [44]. There were two stacking approaches established, one was RF (first-level) + LR, and another was GBT (first-level) + LR.

```
**Input**: Training data $D = \{\mathbf{x}_\mathrm{i}, y_i\}_{i=1}^{m} (\mathbf{x}_\mathrm{i} \in \mathbb{R}^n, y_i \in \Upsilon)$
**Output**: An ensemble classifier H
Step 1: Adopt cross validation approach in preparing a training set for second-level classifier
Randomly split $D$ into $K$ equal-size subsets: $D = \{D_1, D_2, \dots, D_K\}$
**for** $k \leftarrow 1$ to $K$ **do**
    Step 1.1: Learn first-level classifiers
    **for** $t \leftarrow 1$ to $T$ **do**
        Learn a classifier $h_{kt}$ from $D \setminus D_k$
    **end for**
    Step 1.2: Construct a training set for second-level classifier
    **for** $x_i \in D_k$ **do**
        Get a record $\{\mathbf{x}_i', y_i\}$, where $\mathbf{x}_i' = \{h_{k1}(\mathbf{x}_\mathrm{i}), h_{k2}(\mathbf{x}_\mathrm{i}), \dots, h_{kT}(\mathbf{x}_\mathrm{i})\}$
    **end for**
**end for**
Step 2: Learn a second-level classifier
Learn a new classifier $h'$ from the collection of $\{\mathbf{x}_i', y_i\}$
Step 3: Re-learn first-level classifiers
**for** $t \leftarrow 1$ to $T$ **do**
    Learn a classifier $h_t$ based on $D$
**end for**
**return** $H(\mathbf{x}) = h'\{h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_T(\mathbf{x})\}$
```

Figure 1. Stacking with K-fold cross validation

2.3 Feature Selection and Support Vector Machine Optimization

After modeling methods selection, the SVM with RBF kernel was chosen for the next stage of optimization. Figure 2 illustrates the flowchart of SVM prediction model development (APPENDIX B shows the detail of model development). The training procedure of the proposed prediction model includes data input, data exclusion, feature selection, and OSA classification.

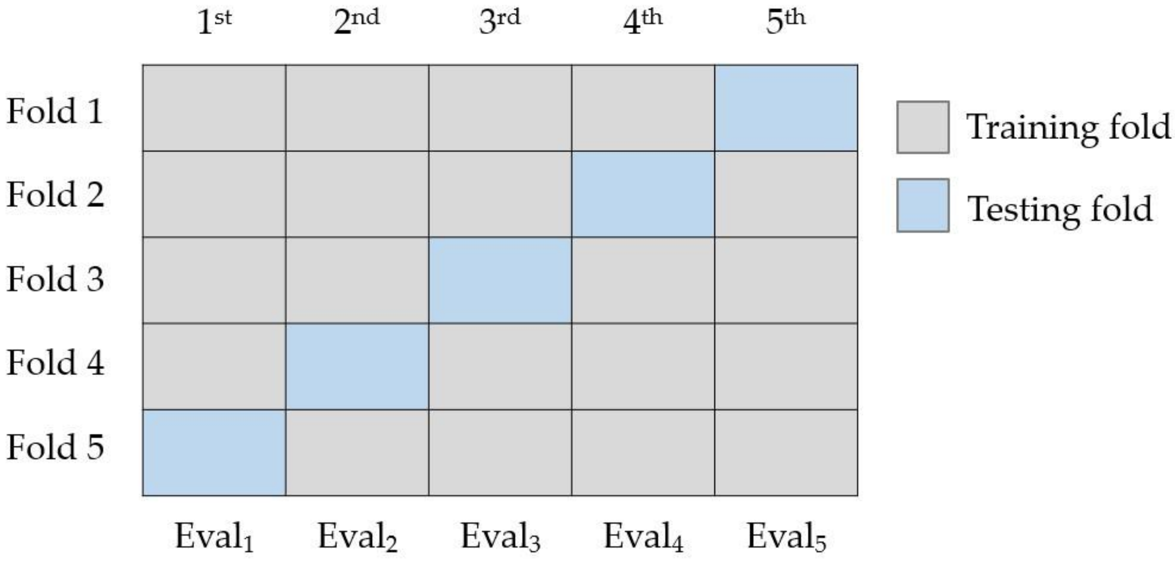

Raw data as input
Data exclusion by missing value and total recording time
First feature selection
Continuous features selected by AUROC
Categorical features selected by MCC
Stepwise forward feature add-on for final feature selection
Classification
Final selected features
SVM parameters optimization
Final model

Non-Chinese (n = 11), total recording time < 240 min (n = 7), and any missing data (n = 936) were excluded. In the first feature selection stage, we observed that the top half of AUROC and MCC feature ranks were similar in all three different AHI cutoffs suggesting the robustness of these features. Therefore, only features with AUROC or MCC higher than median remained in the model. During the last feature selection, the fewest features were selected to keep AUROC ≥ 0.80. Abbreviations: AUROC, area under receiver operating characteristic curve; MCC, Matthews correlation coefficient;

Figure 2. The flow chart of developing SVM-based prediction model

doi:10.6342/NTU202000475

The training procedure conducted in this study was based only on the training dataset to prevent overfitting. Subsequently, a comprehensive blind validation using the testing dataset was conducted during the testing stage. In the proposed method, we applied the cross-validation approach in this study to test the effectiveness of the selected features and the machine learning model. Cross validation (CV) is a re-sampling procedure used to hold out part of the available data as a testing set for model evaluation when data are limited. To perform CV, we put aside a portion of the data not used in model training for testing/validation as Figure 3.



$$CV\ result\ of\ 5\ testing\ fold = \frac{1}{5}\sum_{i=1}^{5} Eval_i$$

First, the whole dataset was separated into 5 folds randomly. Second, in the first iteration, fold 5 was an isolated testing fold, and fold 1-4 were the training folds for feature selection and model optimization. Third, after 5 iterations, the CV result is the average of the testing results from all iterations.

Figure 3. Illustration of procedures of the 5-fold cross validation

To optimize the discriminative ability using the fewest features, continuous and categorical features were selected by single-feature SVM Area under the Receiver Operating Characteristic (AUROC) and Matthews Correlation Coefficient (MCC), respectively, during the feature selection [45]. The MCC calculation followed equations (4) and (5). The *TP* indicates the number true positive, the *TN* indicates the number of true negative, the *FP* indicates the number of false positive, and the *FN* means false negative number. The notation cov(*X*, *Y*) represents the covariance function [46]. The MCC values range between -1 and 1, which indicates the most negative correlated to the most positive, and the zero is non-correlated.

$$\text{MCC (Binary case)} = \frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(FP+FN)(TN+FP)(TN+FN)}}\ , \qquad (4)$$

$$\text{MCC (Multi-classes case)} = \frac{\text{cov}(X,Y)}{\sqrt{\text{cov}(X,X)\cdot\text{cov}(Y,Y)}} \qquad (5)$$

Due to different mathematical characteristics between continuous and categorical features, this study developed a two-stage feature selection procedure to prevent the selection relying on a single type of feature set in the proposed model. In the first stage, only the continuous and categorical features with the top half of AUROC (Table 5) and MCC (Figure 5), respectively, were reserved to reduce the interference from the redundant features, in which these features may be robustly related to different AHI cutoffs. Subsequently, forward stepwise feature selection (FSFS) was exploited in the second half of feature selection.

The feature set selected by FSFS was increased stepwise based on the greedy approach [47]. Specifically, the features with the maximum AUROC or MCC in the first stage were regarded as the feature candidates in the second stage. Afterward, each feature from the feature candidates was randomly integrated then used to train a new SVM for evaluating classification

doi:10.6342/NTU202000475

performance. During each iteration in the second stage, the add-on feature with the superior SVM performance was reserved for updating the selected feature. Accordingly, the updated selected set was used to evaluate the next incoming feature candidate.

The whole training procedure of FSFS was iterated until the stopping criteria (AUROC ≥ 0.8). To achieve significant clinical application, the selected feature set based on FSFS was aimed to achieve target criteria in AUROC in three AHI cutoffs during the selection procedure. Eventually, the selected feature set after the two-stage feature selection was used to establish the prediction model for OSA recognition based on SVM [48]. The posterior probability of the SVM was used to determine the class of the incoming datum [49], either OSA or non-OSA.

To further optimize the classification result, the Youden's index was employed to find the optimal threshold of SVM posterior probability to determine categories. In addition, the 5-fold CV was randomly repeated 5 times to verify the model reliability. The average AUROC of the 5-fold CV for each of the three AHI cutoffs was calculated. The prediction model was trained by sleep-lab-based dataset with three AHI cutoff 5/hr, 15/hr, and 30/hr, respectively, which means the models based on three AHI cutoffs were fairly trained and validated. To further evaluate the model robustness, the learning curves of three AHI cutoffs were depicted as Figure 7.

2.4 Data Analysis

The discriminative ability of the proposed SVM model was evaluated using average of 5-fold CV of AUROC, F1-score, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive likelihood ratio (LR+), and negative likelihood ratio (LR-) for each of the three AHI cutoffs. The performance was expressed as mean [95% confidence interval (CI)]. The cut-off of AUROC value was identified using the

doi:10.6342/NTU202000475

Youden’s index as equation (4). The SVM model was tested in subgroups categorized by gender (men versus woman) and age (< 65 y/o versus ≥ 65 y/o) to identify subgroups for which the model worked best. The performance of the SVM model was compared to logistic regression, Berlin Questionnaire (BQ), NoSAS score, and SLIM scoring system.

For logistic regression, 67% of participants were randomly selected as training set, while the remaining 33% of the participants were selected as testing set. Logistic models with forward selection was used to identify suitable factors to establish the prediction model for AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr in the training test. Each parameter has to be significant at the 0.0001 level to remain in the model. All the remaining variables were listed with odds ratio (OR) and 95% CI. The predictability of AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr was assessed by the AUROC. The feature selection results of LR were shown as APPENDIX C. In addition, sensitivity, specificity, PPV, and NPV were calculated by using the cut-off of Youden’s index.

The clinical features were compared between patients with and without OSA and among four subgroups at three AHI cutoffs. Continuous variables were expressed as mean ± standard deviation (SD) and categorical variables were expressed as percentage. Independent samples t-test and Chi-Square test were applied as appropriate in comparison of OSA datasets and non-OSA datasets as well as in the subgroups. A two-tailed P-value <0.05 was considered statistically significant. All statistical analyses were conducted by Python (Python Software Foundation. Python Language Reference, version 3.6.1. Available at http://www.python.org), and SAS Version 9.3 (SAS Institute, Cary, NC).

doi:10.6342/NTU202000475

Chapter 3. Preliminary Results

The clinical features of patients are listed in Tables 4 and 5. The mean age was 47.8 y/o and 76% were men. The mean AHI was 29.6/hr with a prevalence of 82.5%, 61.3%, and 40.6% at AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr cuto3ffs, respectively. Compared to patients without OSA, those with OSA were older, more obese, sleepier, and had higher percentage of men, history of smoking and alcohol consumption, comorbidities and OSA symptoms as well as shorter SOL. The patients without OSA had longer SOL, higher percentage of witnessed leg jerks in sleep and morning headache than those with OSA. The habitual SOL is weakly correlated with SOL recorded by polysomnography (Person correlation, $\gamma = 0.202$, $P < 0.001$).

With 32 features as input, 8 approaches were used to train a model, respectively. The comparison of 8 approaches of algorithms and stacking technique showed that the SVM with RBF kernel had the best AUROC compared to the others. And the stacking technique could not improve the performance by simply ensemble LR with GBT or RF. Figure 4 illustrates the ROC curves and AUROC values of each approach with 3 different AHI cutoffs. It was significant that with higher AHI cutoff, the performance was getting worse. Even with 32 features as input, it is difficult to reach AUROC ≥8.0 with AHI cutoff 30/hr.

doi:10.6342/NTU202000475

Table 4. Comparison of clinical features between patients with and without OSA at three AHI cutoffs

Feature name	**Overall** **(N = 6,875)**	**AHI < 5** **(N = 1,206)**	**AHI ≥ 5** **(N = 5,669)**	**AHI < 15** **(N = 2,664)**	**AHI ≥ 15** **(N = 4,211)**	**AHI < 30** **(N = 4,084)**	**AHI ≥ 30** **(N = 2,791)**
Age (y/o)	47.8±14.5	40.6±15.1	49.4±13.9	44.1±15.0	50.2±13.7	46±14.8	50.47±13.8
Man, n (%)	5,223 (76.0)	660 (54.8)	4,563 (80.5)	1,673 (62.8)	3,550 (84.3)	2,798 (68.5)	2,425 (86.9)
BMI (kg/m^2)	27.0±5.0	23.5±3.5	27.7±4.9	24.6±3.9	28.4±5.0	25.3±4	29.3±5.3
Neck circumference (cm)	37.7±4.1	34.6±3.4	38.3±4.0	35.6±3.7	39.0±3.9	36.3±3.8	39.7±3.8
Waist circumference (cm)	91.4±13.0	80.9±10.2	93.6±12.5	84.4±10.9	95.8±12.3	86.8±11.2	98±12.7
Current Smoker, n (%)	1,104 (16.1)	135 (11.2)	969 (17.1)	334 (12.5)	770 (18.3)	520 (12.7)	584 (20.9)
Alcohol consumption, n (%)	688 (10.0)	80 (6.6)	608 (10.7)	182 (6.8)	506 (12.0)	328 (8)	360 (12.9)
Hypnotic, n (%)	634 (9.2)	151 (12.5)	483 (8.5)	314 (11.8)	320 (7.6)	435 (10.7)	199 (7.1)
Comorbidity							
Hypertension, n (%)	2,021 (29.4)	126 (10.4)	1,895 (33.4)	435 (16.3)	1,586 (37.7)	835 (20.4)	1,186 (42.5)
Diabetes, n (%)	580 (8.4)	31 (2.6)	549 (9.7)	117 (4.4)	463 (11)	231 (5.7)	349 (12.5)
CAD, n (%)	248 (3.6)	16 (1.3)	232 (4.1)	62 (2.3)	186 (4.4)	119 (2.9)	129 (4.6)
CHF, n (%)	101 (1.5)	6 (0.5)	95 (1.7) †	22 (0.8)	79 (1.9)	37 (0.9)	64 (2.3)
CVA, n (%)	124 (1.8)	5 (0.4)	119 (2.1)	31 (1.2)	93 (2.2) †	54 (1.3)	70 (2.5)
CKD, n (%)	62 (0.9)	7 (0.6)	55 (1.0) $^{\&}$	13 (0.5)	49 (1.2) †	18 (0.4)	44 (1.6)
COPD, n (%)	67 (1.0)	14 (1.2)	53 (0.9) $^{\&}$	22 (0.8)	45 (1.1) $^{\&}$	37 (0.9)	30 (1.1) $^{\&}$
Asthma, n (%)	490 (7.1)	105 (8.7)	385 (6.8) †	210 (7.9)	280 (6.6) $^{\&}$	321 (7.9)	169 (6.1) †
Hypothyroidism, n (%)	156 (2.3)	25 (2.1)	131 (2.3) $^{\&}$	63 (2.4)	93 (2.2) $^{\&}$	103 (2.5)	53 (1.9) $^{\&}$

The data were presented as mean ± standard deviation or number (percentage)

Abbreviations: AHI, apnea-hypopnea index; BMI, body mass index; ESS, Epworth sleepiness scale; EDS, excessive daytime sleepiness; CAD, coronary artery disease; CHF, congestive heart failure; CVA, cerebrovascular disease; COPD, chronic obstructive pulmonary disease

The comparisons between non-OSA and OSA participants were analyzed with the independent t-test and Chi-square test. All P values were <0.001, except for variable marked with $^{\&}$ and †, of which the P-values were > 0.05 and < 0.05, respectively.

Table 5. Comparison of clinical features between patients with and without OSA at three AHI cutoffs

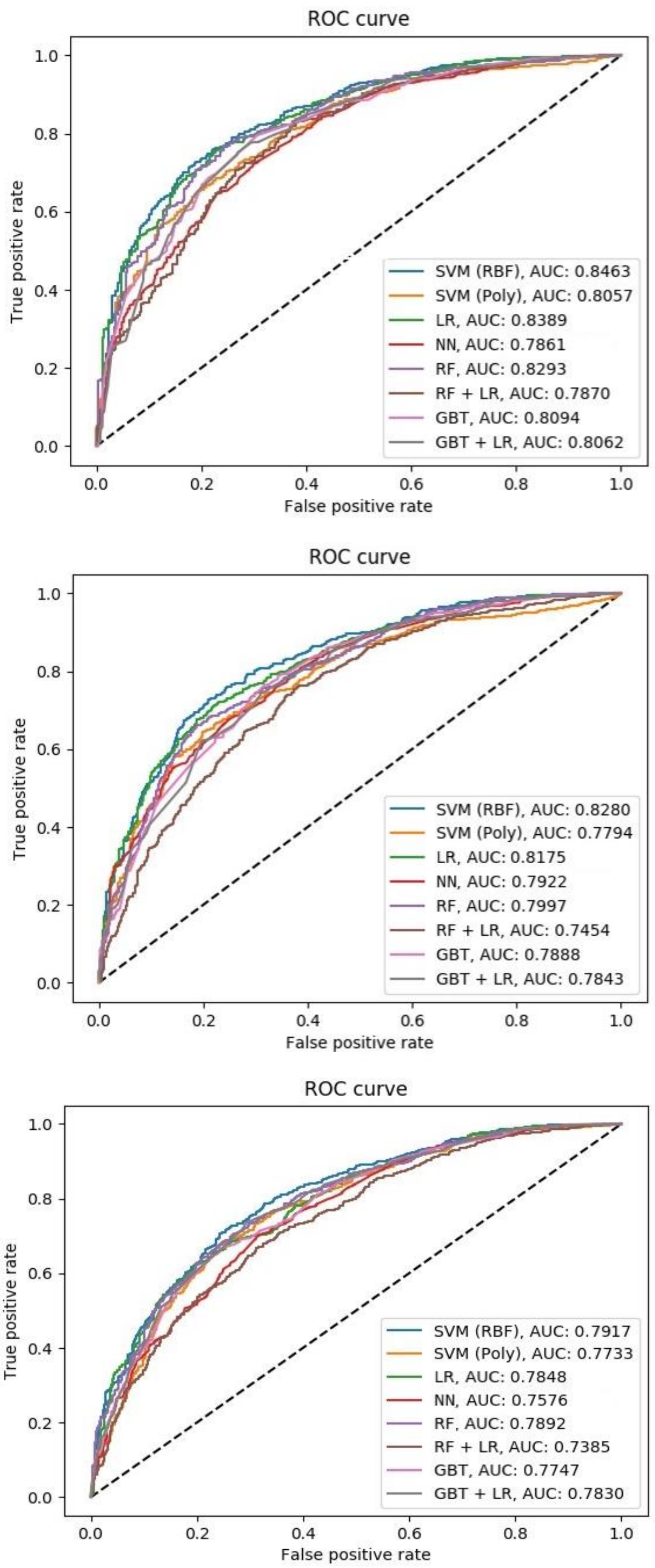
Feature name	**Overall** **(N = 6,875)**	**AHI < 5** **(N = 1,206)**	**AHI ≥ 5** **(N = 5,669)**	**AHI < 15** **(N = 2,664)**	**AHI ≥ 15** **(N = 4,211)**	**AHI < 30** **(N = 4,084)**	**AHI ≥ 30** **(N = 2,791)**
Habitual sleep pattern							
Habitual SOL (min)	20.9±22.5	25±29.9	20±20.5	23.7±26.7	19.1±19.3	22.6±24.2	18.4±19.6
Habitual SOL <30 min, n (%)	4,794 (69.7)	770 (63.8)	4,024 (71.0)	1,744 (65.5)	3,050 (72.4)	2,752 (67.4)	2,042 (73.2)
Habitual sleep duration (hr)	6.6±3.3	6.6±2.7	6.5±3.5	6.5±2.6	6.6±3.7†	6.5±2.8	6.6±4&
Unrefreshed sleep, n (%)	3,685 (53.6)	770 (63.8)	2,915 (51.4)	1,613 (60.5)	2,072 (49.2)	2,325 (56.9)	1,360 (48.7)
Freq. of awakening in sleep (time/night)	0.5±1.5	0.5±1.5	0.6±1.5&	0.5±1.5	0.6±1.5&	0.5±1.5	0.6±1.5
Awakening at sleep ≥3 times/night	1,504 (21.9)	237 (19.7)	1,267 (22.3) †	526 (19.7)	978 (23.2) †	802 (19.6)	702 (25.2)
ESS	10.3±4.9	10.0±4.9	10.4±4.9†	9.8±4.8	10.6±4.9	9.9±4.7	11±5
EDS, n (%)	3,764 (54.7)	628 (52.1)	3,136 (55.3) †	1,355 (50.9)	2,409 (57.2)	2,098 (51.4)	1,666 (59.7)
Symptom suggestive of OSA							
Snoring, n (%)	5,480 (79.7)	753 (62.4)	4,727 (83.4)	1,912 (71.8)	3,568 (84.7)	3,099 (75.9)	2,381 (85.3)
Witnessed apnea, n (%)	1,066 (15.5)	79 (6.6)	987 (17.4)	221 (8.3)	845 (20.1)	417 (10.2)	649 (23.3)
Freq. of nocturia (times/night)	1.1±1.2	0.9±1.1	1.2±1.2	1±1.1	1.3±1.2	1±1.1	1.3±1.3
Nocturia ≥2 times/night, n (%)	2,352 (34.2)	308 (25.5)	2,044 (36.1)	746 (28.0)	1,606 (38.1)	1,203 (29.5)	1,149 (41.2)
Witnessed leg jerks in sleep, n (%)	3,278 (47.7)	603 (50)	2,675 (47.2) &	1,303 (48.9)	1,975 (46.9) &	1,974 (48.3)	1,304 (46.7) &
Morning headache, n (%)	799 (11.6)	192 (15.9)	607 (10.7)	351 (13.1)	448 (10.6) †	513 (12.6)	286 (10.2) †
Dry throat at waking up, n (%)	3,856 (56.1)	577 (47.8)	3,279 (57.8)	1,324 (49.7)	2,532 (60.1)	2,132 (52.2)	1,724 (61.8)
AHI (/hr)	29.6±26.0	1.9±1.5	35.5±24.9	6.1±4.5	44.5±22.8	11.6±8.7	56±19.5

The data were presented as mean ± standard deviation or number (percentage)

Abbreviations: SOL, sleep onset latency; AHI, apnea-hypopnea index

The comparisons between non-OSA and OSA participants were analyzed with the independent t-test and Chi-square test. All P values were <0.001, except for variable marked with & and †, of which the P-values were > 0.05 and < 0.05, respectively.

doi:10.6342/NTU202000475



From top to bottom, the AHI cutoff is 5/hr, 15/hr, and 30/hr, respectively

Figure 4. ROC curves and AUROC comparison of 8 machine learning approaches

doi:10.6342/NTU202000475

3.1 Feature Selection

The MCCs of categorical features for different AHI cutoffs are listed in Figure 5. Twelve categorical features with top half of MCC value were selected for each AHI cutoff. The results of AUROC evaluation with SVMs trained by each single continuous feature are listed in Table 6. Four continuous features with top half of AUROC in predicting OSA were waist, neck circumference, BMI and age. In total, 16 features were selected.

In the final feature selection, 2, 6, and 6 features were selected with FSFS for AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr, respectively (Table 7, Figure 6) where the detailed iterations are listed in Table S5. The learning curve showed no evidence of overfitting (Figure 7). In addition to waist circumference and age, snoring, neck circumference, witnessed apnea, and SOL < 30 min were selected for AHI ≥ 15/hr and AHI ≥ 30/hr (Table 7). For logistic regression, 7, 10, and 10 features were selected for AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr, respectively (Table 6). Five features selected in the SVM model including waist circumference, age, neck circumference, snoring, and witnessed apnea were also selected in the LR. The SOL was selected instead of SOL < 30 min. Additional selected features in LR includes BMI, dry throat, gender, hypnotic, and hypertension (Table 7).

doi:10.6342/NTU202000475

Table 6. AUROC of single continuous feature at SVM model for three AHI cutoffs.

Order	**AHI ≥ 5 /hr**		**AHI ≥ 15 /hr**		**AHI ≥ 30 /hr**	
	Feature	**AUROC**	**Feature**	**AUROC**	**Feature**	**AUROC**
1	Waist	0.794	Waist	0.763	Waist	0.754
2	BMI	0.772	Neck	0.747	Neck	0.742
3	Neck	0.769	BMI	0.739	BMI	0.736
4	Age	0.665	Age	0.561	Age	0.584
Features not selected						
5	SOL	0.525	Freq. of nocturia	0.520	Freq. of nocturia	0.510
6	Freq. of nocturia	0.513	ESS	0.516	ESS	0.518
7	Freq. of awakening in sleep	0.513	Freq. of awakening in sleep	0.494	Freq. of awakening in sleep	0.503
8	ESS	0.495	SOL	0.488	SOL	0.493

doi:10.6342/NTU202000475

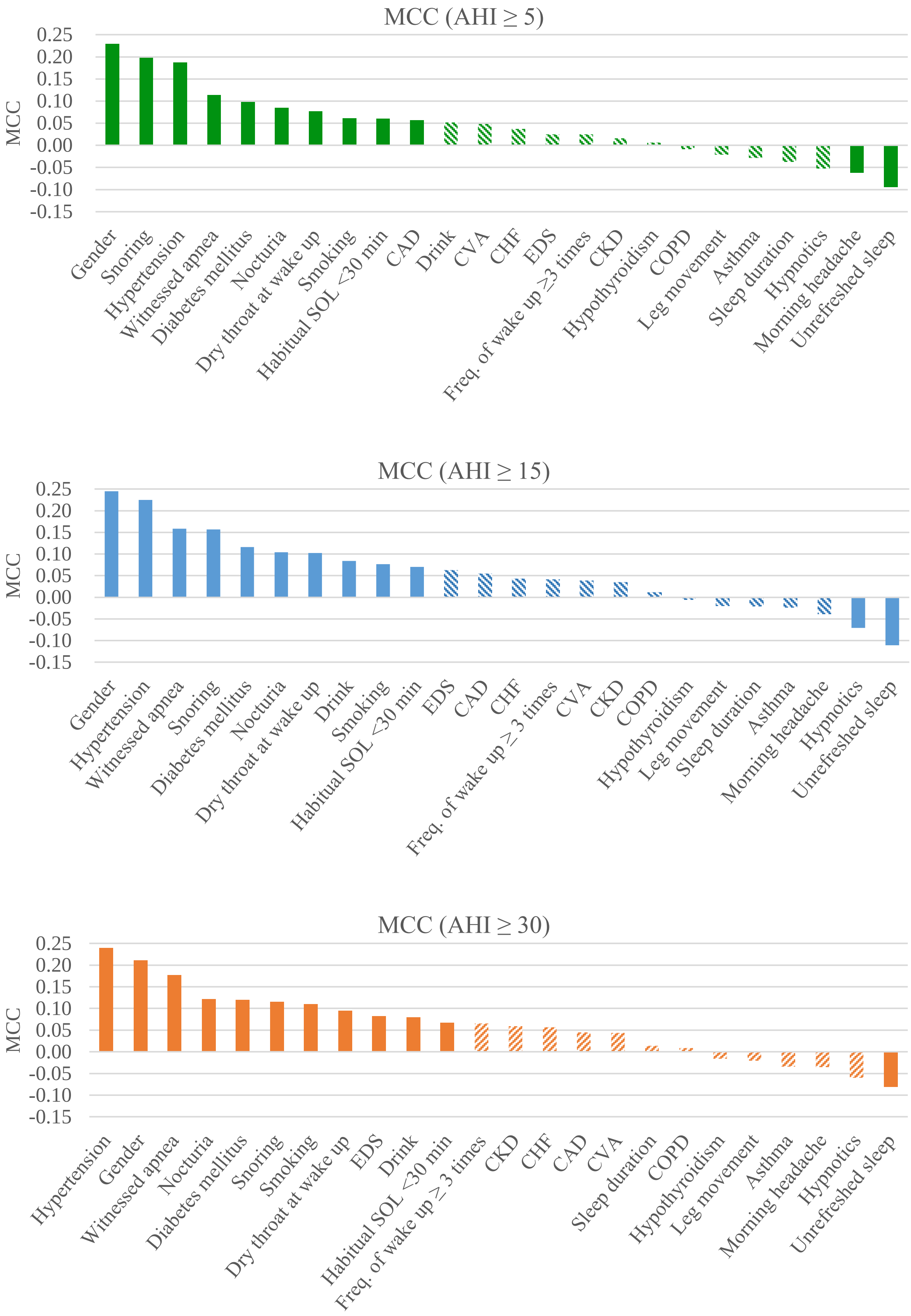


Figure 5. MCC values of categorical features correlated with three AHI cutoffs

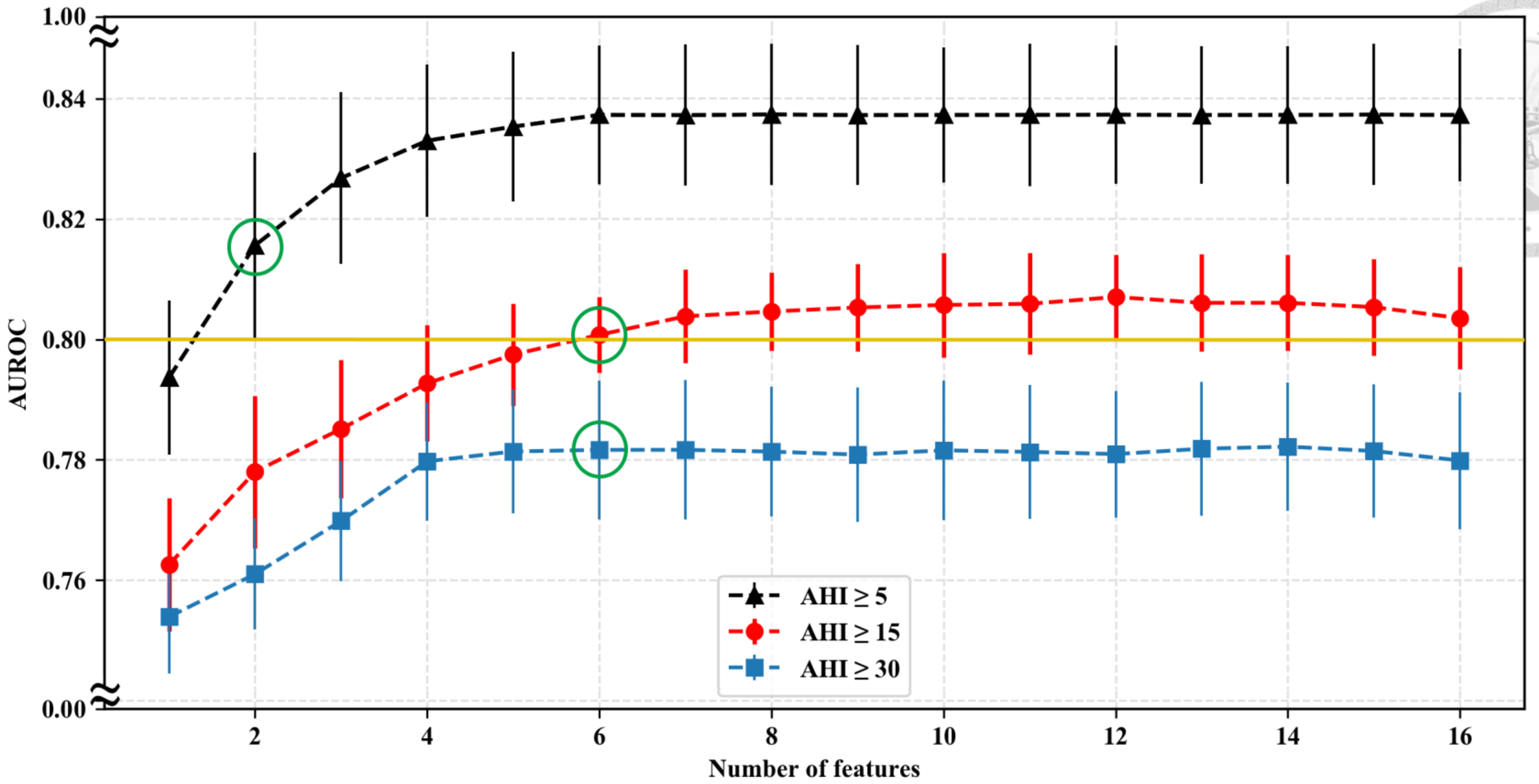
doi:10.6342/NTU202000475

Table 7. The features selected with forward stepwise feature selection of SVM model and logistic regression for three AHI cutoffs.

Method	**Order**	**AHI ≥ 5/hr**	**AHI ≥ 15/hr**	**AHI ≥ 30/hr**
SVM	1	Waist circumference	Waist circumference	Waist circumference
	2	Age	Age	Witnessed apnea
	3		Neck circumference	Age
	4		Snoring	Neck circumference
	5		Witnessed apnea	Snoring
	6		SOL < 30 min	SOL < 30 min
Logistic regression	N/A	Snoring	Snoring	Witnessed apnea
		Gender	Witnessed apnea	Gender
		Age	Dry throat	Snoring
		Neck circumference	Gender	Hypertension
		SOL	Hypnotic	Dry throat
		BMI	Age	Waist circumference
		Waist circumference	Waist circumference	Age
			Neck circumference	Neck circumference
			SOL	BMI
			BMI	SOL

In SVM model, the minimal feature set was selected to achieve the target criteria in the AUROC. When AHI cutoffs were 5/hr and 15/hr, the target AUROC was set as 0.8. While AHI cutoff was 30/hr, the experiment showed that the maximum AUROC was 0.78, so we selected minimum features to achieve the performance.

Abbreviations: AHI, apnea-hypopnea index; BMI, body mass index; SOL, sleep onset latency; N/A, not applicable



This figure illustrates the relationship between AUROC of prediction model and corresponding numbers of features in the stepwise forward feature selection. The results show that the fewest numbers of features to achieve AUROC ≥ 0.80 were 2, 6 and 6 for AHI ≥ 5, 15 and 30/hr, respectively. The solid dot and bar indicated mean and standard deviation, respectively. The green circles indicated selected feature numbers with specific AHI cutoff.

Figure 6. The second stage of feature selection with FSFS

doi:10.6342/NTU202000475

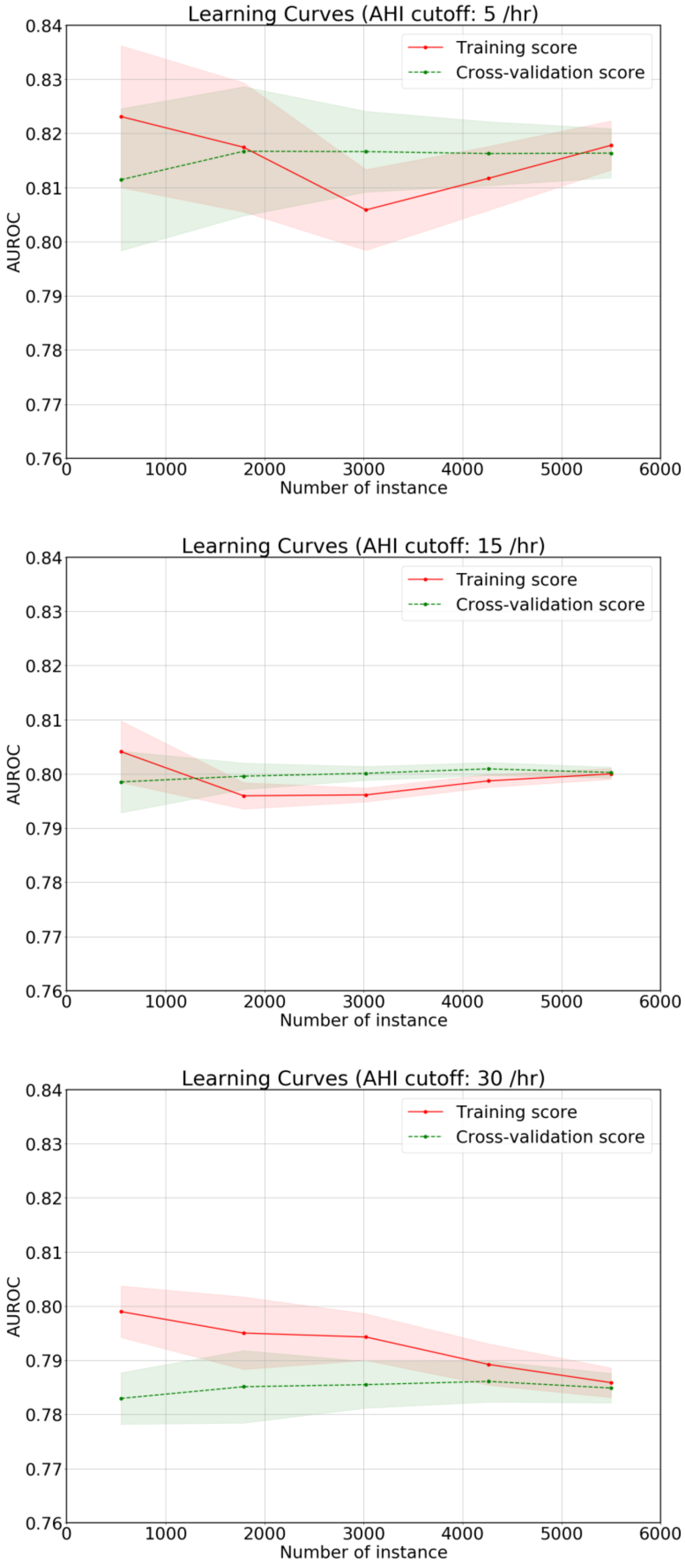


Figure 7. Learning curves of SVM with three AHI cutoffs

doi:10.6342/NTU202000475

3.2 Model Discriminative Ability

The discriminative ability of the SVM model and logistic regression, BQ, NoSAS score, and SLIM scoring system for three AHI cutoffs are shown in Table 8. The performance of SVM model remains good consistently across three AHI criteria. The AUROC was 0.82, 0.80, and 0.78 for AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr, respectively, while the accuracy was 74.24%, 72.68%, and 70.28%, respectively. The sensitivity was 74.14%, 75.18%, and 70.26%, for AHI ≥ 5/hr, ≥ 15/hr, and ≥ 30/hr, respectively, while the specificity was 74.71%, 68.73%, and 70.30% respectively. Compared to logistic regression, the SVM model had similar AUROC and accuracy across three AHI cutoffs. Moreover, at higher AHI cutoffs, the SVM model upheld good sensitivity and NPV without losing specificity and PPV. Compared to the BQ, the SVM model had higher AUROC, accuracy, specificity, PPV and NPV across three AHI criteria while it had higher AUROC, accuracy, sensitivity, and NPV compared to NoSAS score. Compared to SLIM scoring system, the SVM model had higher AUROC, accuracy, and sensitivity across three AHI criteria.

The discriminative ability of SVM model in four subgroups are shown in Tables 9 and 10. The AUROC and accuracy were similar between male and female while AUROC, accuracy, specificity, PPV and NPV were higher in < 65 y/o than ≥ 65 y/o subgroup. Moreover, the discriminative ability was best for male < 65 y/o and modest for female ≥ 65 y/o. To make this proposed prediction model available to researchers and clinicians, we have built an easy-to-use website (http://howareyou.csie.ntu.edu.tw), which provides OSA probability prediction based on our machine learning model.

doi:10.6342/NTU202000475

Table 8. The performance of SVM, logistic regression, Berlin Questionnaire, NoSAS score, and SLIM scoring system at three AHI cutoffs.

Model	**AHI cutoff (/hr)**	**Feature no.**	**AUROC**	**F1-factor**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**	**LR+**	**LR-**
SVM	≥5	2	0.82 (0.79－0.85)	0.83 (0.81－0.85)	74.24 (71.59－76.89)	74.14 (71.33－76.95)	74.71 (70.88－78.54)	93.23 (92.17－94.29)	38.15 (34.79－41.52)	2.96 (2.52－3.41)	0.35 (0.30-0.40)
	≥15	6	0.80 (0.79－0.81)	0.77 (0.74－0.80)	72.68 (70.52－74.84)	75.18 (67.61－82.76)	68.73 (61.72－75.75)	79.32 (76.84－81.80)	64.03 (59.75－68.31)	2.45 (2.04－2.87)	0.36 (0.29-0.43)
	≥30	6	0.78 (0.77－0.80)	0.66 (0.61－0.70)	70.28 (68.68－71.88)	70.26 (60.21－80.31)	70.3 (64.18－76.43)	61.93 (59.21－64.35)	77.86 (73.68－82.03)	2.39 (2.14－2.64)	0.42 (0.32-0.52)
LR	≥5	7	0.84 (0.83－0.86)	-	73.77	94.41	37.87	72.55	79.56	1.52	0.15
	≥15	10	0.81 (0.80－0.82)	-	72.14	79.94	62.69	72.21	72.03	2.14	0.32
	≥30	10	0.79 (0.78－0.81)	-	72.83	65.01	78.77	69.94	74.77	3.06	0.44
BQ	≥5	-	0.54 (0.52－0.56)	-	67.58	74.95	32.91	84.01	21.89	1.11	0.76
	≥15	-	0.53 (0.52－0.55)	-	58.39	76.09	30.41	63.34	44.58	1.09	0.79
	≥30	-	0.53 (0.51－0.54)	-	48.09	76.68	28.55	42.31	64.17	1.07	0.81
NoSAS score	≥5	4	0.70 (0.68－0.71)	-	57.25	50.62	88.39	95.31	27.58	4.36	0.56
	≥15	4	0.68 (0.67－0.70)	-	66.01	57.99	78.67	81.13	54.23	2.72	0.53
	≥30	4	0.68 (0.67－0.69)	-	68.3	64.88	70.64	60.16	74.64	2.2	0.5
SLIM (10 size)	≥5	10	0.69 (0.67－0.70)	0.63	54.68	47.1	90.3	95.8	26.64	4.86	0.59
	≥15	10	0.68 (0.67－0.69)	0.65	64.77	54.33	81.27	82.1	52.96	2.9	0.56
	≥30	10	0.68 (0.67－0.70)	0.62	69.4	62.24	74.29	62.33	74.22	2.42	0.51

Table 9. The performance of SVM model in subgroups including men, women, < 65 y/o, and ≥ 65 y/o.

	AHI cutoff	No. of ≥ AHI cutoff	AUROC	F1 score	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-
Male	≥5	4,653	0.8 (0.77—0.83)	0.82 (0.80—0.83)	71.82 (69.59—74.04)	71.21 (69.04—73.37)	76.06 (73.04—79.08)	95.35 (94.67—96.04)	27.69 (25.55—29.84)	3 (2.59—3.42)	0.38 (0.34—0.42)
	≥15	3,550	0.77 (0.77—0.78)	0.76 (0.70—0.81)	69.86 (65.74—73.98)	69.3 (59.97—78.63)	71.08 (63.99—78.16)	83.7 (81.83—85.58)	52.62 (48.17—57.07)	2.44 (2.07—2.82)	0.43 (0.35—0.51)
	≥30	2,425	0.76 (0.74—0.78)	0.66 (0.59—0.73)	69.65 (67.17—72.13)	64.58 (52.25—76.91)	74.06 (67.47—80.64)	68.5 (66.53—70.48)	71.09 (66.21—75.98)	2.52 (2.29—2.74)	0.47 (0.36—0.59)
Female	≥5	1,106	0.78 (0.73—0.83)	0.79 (0.74—0.84)	72.76 (67.28—78.24)	77.75 (70.55—84.96)	62.64 (56.94—68.35)	80.79 (77.74—83.85)	58.64 (50.32—66.96)	2.11 (1.65—2.58)	0.36 (0.23—0.48)
	≥15	661	0.79 (0.75—0.83)	0.67 (0.65—0.69)	67.92 (63.79—72.06)	81.84 (79.15—84.53)	58.64 (50.16—67.12)	57.13 (52.74—61.53)	82.87 (81.88—83.86)	2.02 (1.65—2.39)	0.31 (0.29—0.33)
	≥30	366	0.79 (0.75—0.82)	0.52 (0.47—0.57)	68.17 (61.43—74.91)	77.04 (74.41—79.67)	65.64 (56.66—74.62)	39.47 (33.20—45.75)	90.91 (89.86—91.96)	2.33 (1.68—2.98)	0.35 (0.31—0.40)
<65 y/o	≥5	4,870	0.82 (0.79—0.85)	0.83 (0.80—0.85)	74.52 (71.62—77.41)	74.68 (71.42—77.93)	73.81 (69.80—77.82)	92.56 (91.41—93.71)	40.14 (36.42—43.86)	2.88 (2.46—3.31)	0.34 (0.29—0.40)
	≥15	3,577	0.81 (0.80—0.82)	0.77 (0.74—0.81)	73.43 (71.47—75.39)	76.46 (68.81—84.10)	68.94 (62.42—75.46)	78.66 (76.51—80.82)	66.74 (62.27—71.22)	2.5 (2.15—2.85)	0.34 (0.27—0.41)
	≥30	2,347	0.79 (0.78—0.81)	0.66 (0.62—0.70)	70.96 (69.22—72.70)	71.83 (62.05—81.61)	70.4 (63.93—76.87)	61.22 (58.19—64.25)	79.75 (75.68—83.82)	2.46 (2.13—2.79)	0.4 (0.29—0.50)
≥65 y/o	≥5	799	0.7 (0.64—0.76)	0.78 (0.76—0.79)	65.84 (63.68—68.00)	65.83 (63.37—68.30)	65.73 (52.01—79.45)	94.47 (92.42—96.52)	17.94 (14.85—21.04)	2.19 (0.88—3.50)	0.53 (0.43—0.63)
	≥15	634	0.69 (0.67—0.72)	0.72 (0.66—0.79)	64.94 (59.99—69.88)	65.75 (54.68—76.82)	62.92 (50.00—75.84)	81.77 (77.98—85.57)	42.92 (39.06—46.79)	1.87 (1.33—2.40)	0.54 (0.46—0.63)
	≥30	444	0.7 (0.68—0.72)	0.62 (0.54—0.69)	64.16 (61.20—67.12)	58.52 (45.19—71.85)	69.74 (60.90—78.58)	66.07 (63.15—68.99)	63.23 (58.75—67.71)	1.97 (1.71—2.23)	0.59 (0.47—0.71)

Table 10. The performance of SVM model in subgroups including men, women, with < 65 y/o, and ≥ 65 y/o.

	AHI cutoff	No. of ≥ AHI cutoff	AUROC	F1 score	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-
Male <65 y/o	≥5	3,972	0.81 (0.78－0.84)	0.82 (0.80－0.84)	72.47 (70.56－74.37)	71.8 (69.64－73.97)	76.81 (72.96－80.67)	95.29 (94.52－96.06)	29.46 (27.70－31.22)	3.14 (2.63－3.65)	0.37 (0.33－0.40)
	≥15	3,075	0.78 (0.77－0.79)	0.76 (0.72－0.81)	70.74 (67.41－74.07)	70.18 (61.74－78.62)	71.89 (64.65－79.14)	83.77 (81.54－86.01)	54.5 (50.56－58.44)	2.56 (2.08－3.05)	0.41 (0.34－0.48)
	≥30	2,077	0.77 (0.75－0.79)	0.66 (0.61－0.72)	70.15 (68.40－71.91)	65.96 (54.65－77.26)	73.64 (67.30－79.97)	67.68 (65.67－69.69)	72.62 (68.15－77.10)	2.53 (2.29－2.78)	0.46 (0.35－0.56)
Male ≥65 y/o	≥5	591	0.66 (0.50－0.82)	0.77 (0.71－0.84)	64.99 (56.26－73.71)	65.65 (58.02－73.28)	57.09 (30.16－84.02)	94.53 (90.98－98.08)	13.31 (5.67－20.95)	1.82 (0.85－2.79)	0.81 (-0.07－1.70)
	≥15	475	0.69 (0.64－0.75)	0.72 (0.59－0.85)	64.67 (52.79－76.55)	64.21 (46.97－81.45)	66.08 (57.94－74.23)	84.07 (81.07－87.07)	40.99 (30.06－51.92)	1.91 (1.45－2.37)	0.54 (0.31－0.76)
	≥30	348	0.68 (0.65－0.72)	0.6 (0.51－0.70)	61.58 (57.45－65.71)	54.82 (40.88－68.77)	69.49 (61.48－77.50)	67.95 (66.43－69.47)	57.03 (52.80－61.27)	1.8 (1.68－1.92)	0.64 (0.52－0.77)
Female <65 y/o	≥5	898	0.78 (0.72－0.84)	0.78 (0.74－0.83)	72.32 (67.63－77.01)	78.52 (72.21－84.82)	61.35 (55.96－66.73)	78.23 (75.33－81.13)	62.05 (54.74－69.35)	2.06 (1.67－2.45)	0.35 (0.24－0.46)
	≥15	502	0.79 (0.76－0.83)	0.66 (0.62－0.70)	67.7 (62.00－73.39)	85.67 (81.39－89.94)	57.7 (48.59－66.82)	53.22 (48.26－58.19)	87.83 (84.22－91.43)	2.07 (1.68－2.45)	0.25 (0.17－0.33)
	≥30	270	0.79 (0.76－0.83)	0.49 (0.43－0.55)	66.83 (59.86－73.80)	81.11 (76.69－85.54)	63.44 (55.04－71.83)	35.01 (28.68－41.34)	93.33 (91.53－95.14)	2.3 (1.61－2.99)	0.3 (0.21－0.39)
Female ≥65 y/o	≥5	208	0.71 (0.57－0.85)	0.78 (0.72－0.85)	68.88 (62.49－75.26)	68.31 (56.76－79.85)	71.07 (47.64－94.50)	93.19 (89.36－97.02)	29.79 (28.03－31.56)	3.14 (1.14－5.13)	0.44 (0.39－0.49)
	≥15	159	0.68 (0.61－0.74)	0.66 (0.59－0.73)	60.3 (54.38－66.22)	60.32 (50.11－70.54)	60.39 (50.49－70.30)	73.41 (68.65－78.17)	46.04 (39.42－52.66)	1.56 (1.25－1.86)	0.66 (0.49－0.84)
	≥30	96	0.7 (0.65－0.76)	0.59 (0.49－0.69)	66.88 (56.89－76.87)	60.47 (47.36－73.59)	70.95 (55.26－86.63)	58.64 (42.90－74.38)	73.92 (65.99－81.84)	2.66 (0.42－4.91)	0.57 (0.34－0.79)

Chapter 4. Discussion

4.1 Preliminary Findings

This study proposed an SVM model driven by data mining that uses a large-scale data set based on the sleep laboratory to predict OSA with three different AHI limits. The characteristics selected in the model were as few as 2, 6 and 6 for AHI ≥ 5/hr, ≥ 15/hr, and ≥30/hr, respectively, and all were collected in the clinics. Compared to the logistic regression, the SVM model had a non-inferior discriminative capacity, balanced sensitivity and specificity, and with fewer features. The discriminative capacity of the SVM model was better than BQ, NoSAS score and SLIM scoring system. The SVM model worked best for men < 65 y/o.

4.2 Comparison with Prior Work

Compared to other related studies, a major strength of the proposed model is that the SVM prediction is built using a large-scale dataset from sleep clinics with very few exclusions which enhance the representativeness of the dataset and minimize the selection bias of small samples (Table 2) [9, 15-21, 23, 24, 26-28, 50]. Moreover, all 32 features are information routinely collected at the clinic visits and are not physiological parameters derived from overnight pulse oximetry or pulmonary function test. Unlike certain model [7] that includes physical findings of oral cavity which may be difficult to measure precisely [16], this study did not include such features for the model development. Similarly, there was a concern that single office blood pressure may not be representative so it was not included as an input feature.

The six selected features, with the exception of SOL < 30 min, were often selected in the prediction models reported in the literature while SOL < 30 min has never been described to predict AHI ≥ 15/hr or AHI ≥ 30/hr (Table 2). The selection of SOL < 30 min as a feature in

doi:10.6342/NTU202000475

the SVM model was a surprise but was echoed by the selection of SOL in logistic regression. Despite that SOL was an input feature in the Boosting model proposed by Caffo et al., it was not selected for their modeling [19]. The difference may be due to the fact that the participants in the Caffo study were from a community population and were older with more females.

Similar to the findings of other studies, neither ESS nor EDS was selected as a model feature, which reflects that there is a high prevalence of EDS in OSA patients regardless of AHI. It may be related to the clinical practice that patients with EDS are more prone to be referred for sleep study and sleepiness is not necessarily concordant with OSA severity [16].

Our results are comparable to another model validated in the Chinese population. Liu et al. [24] developed an SVM-based model to predict OSA with three anthropometric features, i.e., waist size, neck size, and BMI. That study included age, BQ, and anthropometrics as features. The predictability for AHI ≥ 15/hr and AHI ≥ 30/hr in Liu’s model is highest in females 50 y/o. Compared to the discriminative ability for AHI ≥ 15/hr in the model reported by Liu et al., our model has higher AUROC and accuracy in elderly males while lower in elderly females. The difference in performance is likely due to the inclusion of OSA symptoms such as snoring and witnessed apnea. Moreover, the age cutoff in our study for subgroup analysis is 65 y/o as opposed to 50 y/o in the study by Liu. This study arbitrarily chose 65 y/o as cutoff for subgroup analysis. Elderly OSA patients often have poorer association between AHI and body habitus variables (neck size, BMI, and waist-to-hip ratio), a lower percentage of habitual snoring, and a longer SOL compared to younger patients [5], which may contribute to the poorer performance of our model in the elderly as the proposed model was built with anthropometrics and OSA symptoms.

doi:10.6342/NTU202000475

4.3 Limitations

There are some limitations in the present study. First, several features highly associated with OSA, such as hyperlipidemia and atrial fibrillation, were not included in the database and, therefore, were not used for the development of the model. Second, the dataset was created from the information collected from patients referred to our sleep laboratory for the study of sleep in which the prevalence of OSA is high. The result may not be applicable to the general population where the prevalence of OSA is much lower. Third, all participants are Chinese and the accuracy of this model in other ethnic groups remains unclear. The validity of this model needs to be confirmed in multi-ethnic community populations to address the meaning and implications. Fourth, the use of the AHI limit value as the sole objective of prediction is one of the limitations of our study. AHI is known for its flexible association with OSA-related outcomes [37], while factors such as EDS may have a better prediction of cardiovascular outcome than AHI [38]. In the future, other parameters such as morbidity should be considered as prediction objectives. Fifth, we do not compare our model with STOP-Bang [1] since STOP-Bang was not included as part of the routine questionnaires of the NTUH sleep laboratory until January 2017. An additional study that compares our SVM model with STOP-Bang would be justified. Finally, the proposed method only works when the SVM is the most or the nearly most appropriate algorithm for targeted data classification, and the feature selection method requires enough features which are truly related to the outcome, or the feature selection may be failed.

4.4 Future Work

To validate the proposed feature selection method is robust or not, other types of open data may be used for validation. Future studies and the development of the machine learning

doi:10.6342/NTU202000475

model algorithm should focus on validation in the sleep laboratory and community populations with multiple ethnicities for greater clinical application.

4.5 Conclusion

To solve the sensitivity and specificity imbalance problem during using clinical features to predict OSA and to develop an easy-to-use tool for non-sleep specialist physician, this study compared several popular machine learning models to predict OSA with three AHI cutoffs. We found that the SVM with RBF kernel had better performance. With the proposed feature selection method, the feature importance could be clarified and the selected features were matched as clinical experience. The feature selection method is effective to find out the most related features to predict OSA with different AHI cutoffs. The proposed SVM model provides a simple and precise modality for the early identification of patients with OSA. To understand which subgroup (gender and age) fits our model, we also used 8 subgroups of dataset to test trained models. The results showed that our model could fit well with male and < 65 y/o. Finally, the web-based questionnaire integrated with trained models is easy to use for NSSPs and patients.

doi:10.6342/NTU202000475

REFERENCES

1. Patil, S.P., et al., *Adult obstructive sleep apnea: pathophysiology and diagnosis.* Chest, 2007. **132**(1): p. 325-337.

2. Senaratna, C.V., et al., *Prevalence of obstructive sleep apnea in the general population: A systematic review.* Sleep Med Rev, 2017. **34**: p. 70-81.

3. Bradley, T.D. and J.S. Floras, *Obstructive sleep apnoea and its cardiovascular consequences.* Lancet, 2009. **373**(9657): p. 82-93.

4. Somers, V.K., et al., *Sleep apnea and cardiovascular disease: an American Heart Association/american College Of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council On Cardiovascular Nursing. In collaboration with the National Heart, Lung, and Blood Institute National Center on Sleep Disorders Research (National Institutes of Health).* Circulation, 2008. **118**(10): p. 1080-111.

5. Kunisaki, K.M., et al., *Provider Types and Outcomes in Obstructive Sleep Apnea Case Finding and Treatment: A Systematic Review.* Ann Intern Med, 2018. **168**(3): p. 195-202.

6. Harding, S.M., *Prediction formulae for sleep-disordered breathing.* Curr Opin Pulm Med, 2001. **7**(6): p. 381-5.

7. Kirby, S.D., et al., *Neural network prediction of obstructive sleep apnea from clinical criteria.* Chest, 1999. **116**(2): p. 409-15.

8. Zerah-Lancner, F., et al., *Predictive value of pulmonary function parameters for sleep apnea syndrome.* Am J Respir Crit Care Med, 2000. **162**(6): p. 2208-12.

9. Zou, J., et al., *An Effective Model for Screening Obstructive Sleep Apnea: A Large-Scale Diagnostic Study.* PLOS ONE, 2013. **8**(12): p. e80704.

10. Appleton, S., et al., *Influence of Gender on Associations of Obstructive Sleep Apnea Symptoms with Chronic Conditions and Quality of Life.* International Journal of Environmental Research and Public Health, 2018. **15**(5): p. 930.

11. Lin, C.M., T.M. Davidson, and S. Ancoli-Israel, *Gender differences in obstructive sleep apnea and treatment implications.* Sleep medicine reviews, 2008. **12**(6): p. 481-496.

12. Launois, S.H., J.L. Pepin, and P. Levy, *Sleep apnea in the elderly: a specific entity?* Sleep Med Rev, 2007. **11**(2): p. 87-97.

13. Martinez-Garcia, M.A., et al., *Obstructive sleep apnea has little impact on quality of life in the elderly.* Sleep Med, 2009. **10**(1): p. 104-11.

14. Yamagishi, K., et al., *Cross-cultural comparison of the sleep-disordered breathing prevalence among Americans and Japanese.* Eur Respir J, 2010. **36**(2): p. 379-84.

15. Rowley, J.A., L.S. Aboussouan, and M.S. Badr, *The use of clinical prediction formulas in the evaluation of obstructive sleep apnea.* Sleep, 2000. **23**(7): p. 929-38.

16. Rodsutti, J., et al., *A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea.* Sleep, 2004. **27**(4): p. 694-9.

17. Sharma, S.K., et al., *Prediction of obstructive sleep apnea in patients presenting to a tertiary care center.* Sleep Breath, 2006. **10**(3): p. 147-54.

18. Takegami, M., et al., *Simple four-variable screening tool for identification of patients with sleep-disordered breathing.* Sleep, 2009. **32**(7): p. 939-48.

19. Caffo, B., et al., *A novel approach to prediction of mild obstructive sleep disordered breathing in a population-based sample: the Sleep Heart Health Study.* Sleep, 2010. **33**(12): p. 1641-8.

20. Bouloukaki, I., et al., *Prediction of obstructive sleep apnea syndrome in a large Greek population.* Sleep Breath, 2011. **15**(4): p. 657-64.

21. Marti-Soler, H., et al., *The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study.* Lancet Respir Med, 2016. **4**(9): p. 742-748.

22. Shah, N., et al., *Sex-Specific Prediction Models for Sleep Apnea From the Hispanic Community Health Study/Study of Latinos.* Chest, 2016. **149**(6): p. 1409-1418.

23. Ustun, B., et al., *Clinical Prediction Models for Sleep Apnea: The Importance of Medical History over Symptoms.* J Clin Sleep Med, 2016. **12**(2): p. 161-8.

24. Liu, W.T., et al., *Prediction of the severity of obstructive sleep apnea by anthropometric features via support vector machine.* PLoS One, 2017. **12**(5): p. e0176991.

25. Shin, C.H., et al., *Development and validation of a Score for Preoperative Prediction of Obstructive Sleep Apnea (SPOSA) and its perioperative outcomes.* BMC anesthesiology, 2017. **17**(1): p. 71-71.

26. Tan, A., et al., *Validation of NoSAS score for screening of sleep-disordered breathing in a multiethnic Asian population.* Sleep Breath, 2017. **21**(4): p. 1033-1038.

27. Traxdorf, M., et al., *The Erlangen Questionnaire: a new 5-item screening tool for obstructive sleep apnea in a sleep clinic population - A prospective, double blinded study.* Eur Rev Med Pharmacol Sci, 2017. **21**(16): p. 3690-3698.

28. Duarte, R.L.M., et al., *Simplifying the Screening of Obstructive Sleep Apnea With a 2-Item Model, No-Apnea: A Cross-Sectional Study.* Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine, 2018. **14**(7): p. 1097-1107.

29. Jothi, N., N.A.A. Rashid, and W. Husain, *Data Mining in Healthcare – A Review.* Procedia Computer Science, 2015. **72**: p. 306-313.

30. Chen, N.H., et al., *Validation of a Chinese version of the Epworth sleepiness scale.* Qual Life Res, 2002. **11**(8): p. 817-21.

31. Liu, H.-W., et al., *Combining MAD and CPAP as an effective strategy for treating patients with severe sleep apnea intolerant to high-pressure PAP and unresponsive to MAD.* PLOS ONE, 2017. **12**(10): p. e0187032.

32. Iber, C., S. Ancoli-Israel, and J.A.L. Chesson, *For the American Academy of Sleep Medicine the AASM Manual for the Scoring of Sleep and Associated Events.* 2007: p. 1-59.

doi:10.6342/NTU202000475

33. Burges, C.J.C., *A Tutorial on Support Vector Machines for Pattern Recognition.* Data Mining and Knowledge Discovery, 1998. **2**(2): p. 121-167.

34. Ben-Hur, A., et al., *Support vector machines and kernels for computational biology.* PLoS computational biology, 2008. **4**(10): p. e1000173-e1000173.

35. Schölkopf, B., *Learning with kernels : support vector machines, regularization, optimization, and beyond.* 2002.

36. Couronné, R., P. Probst, and A.-L. Boulesteix, *Random forest versus logistic regression: a large-scale benchmark experiment.* BMC Bioinformatics, 2018. **19**(1): p. 270.

37. Vandamme, J., N. Meskens, and J. Superby, *Predicting Academic Performance by Data Mining Methods.* Education Economics, 2007. **15**: p. 405-419.

38. Rashid, T.A. and H.A. Ahmad, *Lecturer performance system using neural network with Particle Swarm Optimization.* Computer Applications in Engineering Education, 2016. **24**(4): p. 629-638.

39. Kose, U. and A. Arslan, *Optimization of self-learning in Computer Engineering courses: An intelligent software system supported by Artificial Neural Network and Vortex Optimization Algorithm.* Computer Applications in Engineering Education, 2017. **25**(1): p. 142-156.

40. Lau, E.T., L. Sun, and Q. Yang, *Modelling, prediction and classification of student academic performance using artificial neural networks.* SN Applied Sciences, 2019. **1**(9): p. 982.

41. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.

42. Natekin, A. and A. Knoll, *Gradient boosting machines, a tutorial.* Frontiers in neurorobotics, 2013. **7**: p. 21.

43. Bell, R.M. and Y. Koren, *Lessons from the Netflix prize challenge*. Vol. 9. 2007: Association for Computing Machinery. 75–79.

44. Aggarwal, C.C., *Data Classification*. 2014: Taylor & Francis.

45. Boughorbel, S., F. Jarray, and M. El-Anbari, *Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric.* PLoS One, 2017. **12**(6): p. e0177678.

46. Jurman, G., S. Riccadonna, and C. Furlanello, *A Comparison of MCC and CEN Error Measures in Multi-Class Prediction.* PLOS ONE, 2012. **7**(8): p. e41882.

47. Alvarez, D., et al., *Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis.* Int J Neural Syst, 2013. **23**(5): p. 1350020.

48. Yuanyuan, S., et al. *The comparison of optimizing SVM by GA and grid search*. in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. 2017.

49. Karatzoglou, A., et al., *Support Vector Machines in R.* Journal of Statistical Software, 2006. **15**(9): p. 1-28.

50. Ting, H., et al., *Decision Tree Based Diagnostic System for Moderate to Severe Obstructive Sleep Apnea.* Journal of Medical Systems, 2014. **38**(9): p. 94.

APPENDIX

APPENDIX A. Description and definition of sleep pattern parameters and OSA symptoms

Parameter	**Description**	**Domain**	**Definition in SVM**
Habitual sleep pattern			
Habitual SOL (min)	During the past month, how long (in minutes) does it usually take you to fall asleep at night?		
Habitual SOL < 30 min			< 30 min = 1, ≥ 30 min = 0
Habitual sleep duration (hr)	During the past month, how many hours of actual sleep did you get at night?		
Category of habitual sleep duration			< 6 hr = -1, 6-8 hr = 0, ≥ 8 hr = 1
Unrefreshed sleep	Do you feel unrefreshed after waking up in the morning?	Yes, No	Yes = 1, No = 0
Freq. of wakening in sleep (times/night)	How many times do you wake up during the night?		
Awakening in sleep ≥ 3 times/night			≥ 3 times/night = 1, < 3 times/night = 0
EDS			ESS ≥10 = 1, ESS <10 = 0
Symptom suggestive of OSA			
Snoring	Do you snore?	Yes, No, don’t know	Yes = 1, No = 0, don’t know = 0
Witnessed apnea	How often in the past month have you been told to have long pauses between breaths while in sleep?	No, < 1tme/week, 1-2 times/week, 3 times/week	No = 0, < 1tme/week = 0, 1-2 times/week = 0, 3 times/week = 1
Freq. of nocturia (times/night)	How many times do you need to get out of bed to go to the bathroom during your sleep period?		
Nocturia ≥ 2 times/night			≥2 times/night = 1, < 2 times/night = 0
Witnessed leg jerks in sleep	How often in the past month have you been told to have had leg twitching or jerking while in sleep?	No, < 1tme/week, 1-2 times/week, 3 times/week	No = 0, < 1tme/week = 0, 1-2 times/week = 0, 3 times/week = 1
Morning headache	Do you experience headaches while waking up in the morning?	Yes, No	Yes = 1, No = 0
Dry throat at waking up	Do you experience dry throat at waking up?	Yes, No	Yes = 1, No = 0

doi:10.6342/NTU202000475

APPENDIX B. Details of SVM prediction model training and testing procedures

Task	Task name	Comment	Criterion	Cutoff
1	Data input		N/A	N/A
2	Data exclusion	After exclusion, 6,875 subjects left	PSG total recording time	< 240 min
			Not Chinese	N/A
			Any missing value in 32 features	N/A
3	Data splitting for 5-fold CV		Each fold's prevalence rate is nearly same	N/A
4	First feature selection (continuous feature)	Use 4 folds for these tasks	Median of single-feature-SVM's AUROC	≥ Median value
	First feature selection (categorical feature)		Median of MCC	≥ Median value
5	Forward stepwise feature selection		AUROC	≥ 0.8
6	SVM optimization		AUROC	Maximum and not overfitting
7	Testing on hold out fold		N/A	N/A
8	Averaging the results of five test folds		N/A	N/A
9	Plotting learning curve	To evaluate model for overfitting	No significant difference between training and testing curve	N/A
10	Repeating task 4-8 until five times	With different training and testing folds	N/A	N/A
11	Averaging results from task 10	Calculate mean and 95% confidence interval	N/A	N/A

doi:10.6342/NTU202000475

APPENDIX C. The result of multivariable logistic regression.

AHI ≥ 5/hr			**AHI ≥ 15/hr**			**AHI ≥ 30/hr**		
Variable	**OR (95% CI)**	**P-value**	**Variable**	**OR (95% CI)**	**P-value**	**Variable**	**OR (95% CI)**	**P-value**
Snoring	2.732 (2.231-3.345)	<.0001	Snoring	1.846 (1.543-2.208)	<.0001	Witnessed apnea	2.036 (1.684-2.460)	<.0001
Gender	1.640 (1.244-2.163)	0.0005	Witnessed apnea	1.873 (1.506-2.329)	<.0001	Gender	1.751 (1.376-2.227)	<.0001
Age	1.046 (1.039-1.053)	<.0001	Dry throat	1.451 (1.259-1.674)	<.0001	Snoring	1.471 (1.222-1.770)	<.0001
Neck	1.081 (1.030-1.134)	0.0014	Gender	1.640 (1.301-2.066)	<.0001	Hypertension	1.416 (1.202-1.667)	<.0001
SOL	0.992 (0.988-0.996)	<.0001	Hypnotic	0.635 (0.494-0.817)	0.0004	Dry throat	1.343 (1.168-1.544)	<.0001
BMI	1.131 (1.079-1.186)	<.0001	Age	1.040 (1.034-1.046)	<.0001	Waist	1.013 (0.999-1.027)	<.0001
Waist	1.029 (1.010-1.048)	0.0026	Waist	1.024 (1.009-1.039)	<.0001	Age	1.028 (1.022-1.034)	<.0001
			Neck	1.082 (1.043-1.122)	<.0001	Neck	1.078 (1.041-1.116)	<.0001
			SOL	0.991 (0.988-0.995)	<.0001	BMI	1.141 (1.103-1.181)	<.0001
			BMI	1.126 (1.086-1.168)	<.0001	SOL	0.988 (0.984-0.992)	<.0001

Abbreviation: AHI, apnea-hypopnea index; OR, odd ratio; SOL, sleep onset latency; BMI, body mass index.

doi:10.6342/NTU202000475