A Novel Artificial Neural Network Based Sleep-Disordered Breathing Screening Tool

A Novel Artificial Neural Network Based Sleep Disordered Breathing Screening Tool

Ao Li, M.S.[1]; Stuart F. Quan, M.D.[2,6]; Graciela E. Silva, Ph.D., M.P.H.[3]; Michelle M. Perfect, Ph.D.[4]; Janet M. Roveda, Ph.D.[1,5];

[1]Department of Electrical and Computer Engineering, College of Engineering, University of Arizona, Tucson, AZ; [2]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [3]Biobehavioral Health Science Division, College of Nursing, University of Arizona, Tucson, AZ; [4]Disability and Psychoeducational Studies, College of Education, University of Arizona, Tucson, AZ; [5]Department of Biomedical Engineering, College of Engineering, University of Arizona, Tucson, AZ; [6]Asthma and Airway Disease Research Center, College of Medicine, University of Arizona, Tucson, AZ

Corresponding author's full address:

Ao Li, M.S.

Department of Electrical and Computer Engineering

University of Arizona

Tucson, AZ, 85719

Corresponding author's current Email: aoli1@email.arizona.edu

Institution where work was performed: University of Arizona

All authors have seen and approved the manuscript.

Ao Li, Stuart F. Quan, and Janet M. Roveda declare they filed a patent for the artificial neural network based sleep disordered breathing screening tool. All other authors have nothing to declare.

The manuscript reports are not on a clinical trial.

Number of tables: 5

Number of figures: 3

Manuscript word count:

Abstract

Study Objectives: This study evaluated a novel artificial neural network (ANN) based sleep disordered breathing (SDB) screening tool incorporating nocturnal pulse oximetry with demographic, anatomic and clinical data. The tool was compatible with 6 categories of apnea hypopnea index (AHI) with 4% oxyhemoglobin desaturation threshold, ≥ 5/hour, 10/hour, 15/hour, 20/hour, 25/hour, and 30/hour.

Methods: Using a general population dataset, the training set included 2,280 subjects, while the test set included 470 subjects. The input of this tool was a set of 22 variables. The tool had six multilayer perceptron (MLP) neural network models for each AHI threshold. Several criteria were explored to evaluate the accuracy of the tool: area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and 95% confidence intervals (CI).

Results: The AUCs were 0.904, 0.912, 0.913, 0.926, 0.930, and 0.954 respectively, with models of AHI ≥ 5/hour, 10/hour, 15/hour, 20/hour, 25/hour, and 30/hour thresholds. The sensitivities of all MLP neural network models were higher than 95%. The AHI ≥ 30/hour model had the maximum sensitivity: 98.31% (95% CI: 95.01% - 100%).

Conclusions: The results of this study suggested that the ANN based SDB screening tool can be used to identify the presence or absence of SDB. Future validation should be performed in other populations to determine the practicability of this screening tool in sleep clinics and other at risk populations.

Keywords: Sleep disordered breathing, screening, artificial neural network, general population.

Brief Summary

Current Knowledge/Study Rationale: All previous ANN screening tools were developed using clinical population datasets with relatively small numbers of subjects. This study introduces an ANN based screening tool developed by using a large general population database and incorporates clinical, anatomic and pulse oximetry input data to accurately screen for the presence or absence of SDB.

Study Impact: This study is one of the first studies to combine ANN models with easy to implement physiologic monitoring to predict the presence of SDB. Therefore, it has the potential for widespread clinical use and could result in a decrease in health care costs by reducing the need for both overnight laboratory-based polysomnograms (PSG) and home sleep studies (HST).

Introduction

Sleep disordered breathing (SDB) is a potentially remedial risk factor for hypertension, diabetes, stroke, coronary artery disease, and heart failure.[1] In one study, in a heart failure population, the prevalence of SDB was 76%.[2] Various SDB screening tools based on questionnaires and anthropometric data, such as the Berlin,[3] STOP,[4] STOP-BANG,[4] NoSAS,[5] and 4-variable tool have been developed over the past twenty years.[6] However, the accuracy of these existing screening tools to diagnose SDB is relatively low.[7] Therefore, confirmation requires a diagnostic study, either an overnight laboratory-based polysomnogram (PSG) or a home sleep study (HST). An overnight PSG is expensive, complex, and inconvenient. Although HST is less expensive,[8] false negative studies can occur and SDB severity tends to be underestimated. Consequently, many potential SDB patients never receive a diagnosis. One approach to address this deficiency would be a more accurate, convenient method of facilitating SDB screening.

Artificial neural networks (ANN) are increasingly used in biomedical fields such as classification of biologic specimens, prediction of pharmacokinetics of drugs and in the diagnosis and prognosis of diseases.[9] For example, they have been used in the fields of cardiology to predict the presence of coronary artery and congenital heart disease,[10,11] and in pulmonology to classify pulmonary nodules.[12] Some attempts have been made to use ANN to diagnose SDB based solely on anthropometric, demographic and historical clinical data with modest success. In this paper, we propose the inclusion of nocturnal physiologic data to potentially enhance accuracy. Pulse oximetry is a physiologic signal that is widely available. It has been considered as a screening tool for SDB, but is insufficient by itself to confirm a diagnosis of sleep apnea with an accuracy for an AHI > 15/hour of 86% and 80% in a high and low risk populations respectively.[13] We hypothesized that use of an ANN in combination with pulse oximetry would result in a more accurate screening tool for SDB. Therefore, in this study, we developed and tested a novel ANN

based SDB screening tool using a large general population database, the Sleep Heart Health Study (SHHS).

Methods

Data set

In this research, the Sleep Heart Health Study database (SHHS) was used to develop the neural network based screening tool.[14-16] It is an ideal resource to be utilized for this purpose because of its database of 6,441 subjects with polysomnograms and associated anthropometric and medical history data. A complete description of the SHHS has been previously published.[14-16] Only the baseline examination cycle between November 1, 1995 and January 31, 1998 was used as the study dataset. The dataset included 1,280 variables and 5,804 subjects, with 2,765 males (47.6%) and 3,039 females (52.4%). Six hundred two American Indian subjects were excluded because consent was withdrawn. We manually selected 22 SDB related variables from the 1,280 variables as the candidate variables for the screening tool. A total of 1,866 subjects were missing responses for some of the 22 SDB related variables in the baseline examination cycle. Additionally, 879 subjects responded *do not know* for frequency of snoring question. Therefore, we removed these subjects from our final dataset. We also removed the subjects who had poor pulse oximeter signal quality or a short PSG duration (< 5 hours). The resulting dataset thus included 2,850 subjects. Shown in Table 1 is the demographics and other relevant variables of the resulting dataset. The body mass index (BMI) was computed as weight in kilograms over height in meters squared. Neck circumference (cm) was measured at the medial point. Frequency of snoring was defined as 0=not snoring, 1=1 night/week, 2=1 or 2 nights/week, 3=3 to 5 nights/week, and 4=6 or 7nights/week. Fall asleep while in a car, fall asleep while sitting inactive in a public place, and fall asleep while sitting and talking were defined as 1=no chance, 2=slight chance, 3=moderate chance, and 4=high chance.

The resulting dataset was further randomly separated into two datasets, training set (80%) and test set (20%). Table 2 shows the number of positive and negative subjects in the training set and the test set, and their respective prevalence at each AHI threshold. The AHI mean of the overall training set is 11.0 ± 13.7. The AHI mean of the test set is 12.3 ± 15.8. Table 3 compares the characteristics of the final used variables in the training and test sets. There were no significant differences between the two sets.

Development of the Models

We classified SDB position and negative subjects by 6 thresholds of AHI with 4% oxyhemoglobin desaturation. The 22 manually selected variables were used as the candidate features. We created 6 MLP neural network models correspond to the 6 levels of AHI threshold: AHI ≥ 5/hour, 10/hour, 15/hour, 20/hour, 25/hour, and 30/hour. We normalized the features by a min-max normalization strategy to (0, 1), and used the extremely randomized trees algorithm to select input features of the MLP neural network models.[17,18]

A neural network is a type of mathematical model with optimizable parameters. The multilayer perceptron (MLP), a specific type of neural network model (Figure 1), was used in this study. In this paper, we used the training set to teach each of six neural network models, . Each neural network model corresponds to one of the 6 levels of AHI threshold. The backpropagation algorithm in conjunction with the limited memory version of the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm (L-BFGS) was used to train the neural network models.[19] The training process is an optimization process: it optimized the parameters in neural network models to reduce the output error rate. The output of the neural network model is a value between 0 and 1. (See appendix for more details)

Evaluation and Statistical Analysis

The performance of the screening tool was evaluated by using the test set to calculate the tool's AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value

(NPV), and 95% confidence interval (CI). The receiver operating curve is a curve of true positive rate (sensitivity) vs false positive rate (1 – specificity). The confidence intervals were calculated by the normal approximation interval formula (as in (1)). Here, p is the probability, n is the sample size, and z is the z-value of 95% confidence interval.

$$CI = p \pm z * \sqrt{\frac{p(1-p)}{n}} \qquad (1)$$

Results

Evaluation of the screening tool is displayed in Table 4 and Figure 2. The AUCs were 0.904, 0.912, 0.913, 0.926, 0.930, and 0.954 when SDB was defined at the AHI threshold ≥ 5/hour, 10/hour, 15/hour, 20/hour, 25/hour, 30/hour, respectively. The result showed that the AHI ≥ 30/hour threshold model had the highest AUC although there were no large differences among the AUCs of each model. As shown in Table 4, we selected the operating points with high AUC and sensitivities for each model. The maximum sensitivity was 98.31% when the threshold was AHI ≥ 30/hour. The minimal sensitivity was 95.12% when the threshold was AHI ≥ 5/hour. The specificities of all MLP neural network models were greater than 60%. The minimum specificity was 62.81% when the threshold is AHI ≥ 5/hour. The maximum specificity was 72.0% when the threshold was AHI ≥ 20/hour. The AHI ≥ 30/hour model had the highest 99.73% NPV based on an 8.52% prevalence of positive subjects in the dataset. The AHI ≥ 5/hour model had the highest PPV (77.61%) based on 55.93% prevalence of positive subjects.

Discussion

In this study, we used ANN modelling of demographic, anthropometric, clinical and pulse oximetry data to develop a tool which can be used for screening individuals for the presence of SDB. We found that that at commonly used thresholds of AHI, the sensitivity, negative predictive value and AUC were greater than 90%. This suggests that addition of pulse oximetry in an MLP neural network model can be a useful screening tool for SDB in a general population.

In our study, we found 90-99% sensitivity, negative predictive value and AUC for AHI thresholds ranging from 5 to 30/hour using our MLP neural network models. These results exceed those published using other commonly used screening instruments when tested in the same SHHS dataset and in comparison to other studies. The Stop-Bang questionnaire consists of 8 Yes/No items.[4] It has been tested in the SHHS dataset and has 87.0% sensitivity and 43.4% specificity for moderate-to-severe (15≤AHI≤30) subjects, and 70.4% sensitivity and 59.5% specificity for severe (30≤AHI) subjects.[7] The 4 variable tool includes age, blood pressure (BP), body mass index (BMI), and snoring as input data.[6] It has 24.7% sensitivity and 93.2% specificity for moderate-to-severe patients, and 41.5% sensitivity and 93.2% specificity for severe patients when tested in the SHHS dataset.[7] The Berlin questionnaire is another commonly used instrument. In a recent review, it was reported to have a 69-93% sensitivity and 19-54% specificity using an AHI threshold of 30/hour with a 4% oxygen desaturation requirement.[20] As with the STOP-BANG and 4-variable tool, these validation statistics indicate a number of patients will be misclassified. From a clinical perspective, screen positive patients using these latter instruments will still need a confirmatory PSG or HST, and screen negative patients deemed to be at high-risk will also need further testing. Furthermore, in most validation studies, the test dataset consists of patients recruited from sleep clinics or those with a high suspicion of SDB, and results may not be applicable to a more diverse population.

There have been other efforts to apply ANN modelling of clinical data to predict the presence of SDB. El-Solh et al.[21] developed a neural network model using 12 clinical input variables to predict AHI values. In the 80 subjects used to test their predictive model, they found comparable AUCs at AHI thresholds of 10, 15 and 20 /hour.[21] Kirby et al.[22] introduced a generalized regression neural network (GRNN) model to predict AHI values. There were 150 subjects used to test their GRNN model which had 23 input variables. This model acquired high sensitivity 98.9% when AHI ≥ 10 was applied to define obstructive sleep apnea (OSA).[22] Teferra

et al.[23] used 9 input variables in an ANN model which had only 74% sensitivity and 78% specificity to predict SDB at an AHI threshold of 15 /hour. In a recent study by Karamanli et al.[24] an ANN model was developed using 4 input variables and correctly classified 86.6% of subjects. However, all previous ANN research efforts used clinical population datasets with relatively small numbers of subjects to develop ANN models. The largest test dataset was 150 subjects,[22] and therefore may not have had enough subjects to adequately validate the neural network models. Furthermore, unlike our study, not all commonly employed thresholds of AHI were evaluated.

In this study, our novel ANN based screening tool was developed and tested using a general population dataset. The AUCs of all MLP neural network models were over 0.9. The sensitivities of all MLP neural network models were over 95%. The test results validate that the screening tool has high performance and its high negative predictive value of 97.61% at an AHI threshold of 15/hour indicates that it can be used in the general population to exclude the presence of moderate to severe OSA. This is clinically relevant because a recent comprehensive review concludes that it is unclear whether mild OSA is associated with an increase in cardiovascular or cerebrovascular events.[25]

Our study is not the first one to incorporate physiologic data in an ANN model to predict the presence of SDB. In a study by Lweesy et al.[26] features of the electrocardiogram (ECG) were used with >90% accuracy in classifying a small number of subjects with symptoms of OSA. Although it is possible to record ambulatory ECG signals, correct placement of the leads is important and prone to error. Use of pulse oximetry is easier for a lay person. Nevertheless, addition or substitution of other physiologic signals in our ANN model could produce better or comparable results.

This study does have some limitations. First, although the SHHS database is derived from the general population, it is oversampled with snorers and is limited to subjects over the age of 40 years.[14] Second, in clinical practice, some patients may not identify their own sleep problems

in questionnaires. Therefore, data from the SHHS population may not have the same predictive accuracy of responses from patients regarding sleep problems. Third, some patients deny sleep problems and will not either voluntarily report them. Both of these latter limitations may reduce the performance of the screening tool in clinical use. Second, despite the large number of subjects in SHHS, there were a relatively small number of subjects with high values of AHI. This reduces the reliability of the evaluation results in the high threshold models. Finally, subjects with missing data were excluded from the analysis. We believe that this was non-differential, and thus did not bias the results.

Despite these limitations, our study has important strengths. We used a large database of well-characterized subjects. It is one of the first studies to incorporate easy to implement physiologic monitoring to ANN modelling to predict the presence of SDB. Thus, it has the potential to be implemented in primary care physicians' offices to screen populations at high risk for SDB such as those with obesity, snoring, diabetes and heart failure, and thus decrease the need for referral to a sleep physician. Used by sleep physicians, results from the tool may be sufficient in some patients to determine whether or not a patient has SDB. Thus the tool could result in a decrease in health care costs by reducing the need for both PSG and HST.

Conclusion

In summary, we have developed ANN models that incorporate clinical, anatomic and pulse oximetry input data to accurately screen for the presence or absence of SDB. This tool may have utility in identifying patients with SDB. Future studies should be done in other populations to determine the feasibility of applying this screening tool in clinics and other at risk populations.

Abbreviations

AHI, Apnea hypopnea index

ANN, Artificial neural network

AUC, Area under the receiver operating characteristic curve

BMI, Body Mass Index

BP, Blood pressure

CHS, Cardiovascular Health Study

CI, Confidence intervals

FHS, Framingham Heart Study

GRNN, Generalized regression neural network

H&E, Health and Environ

HST, Home sleep study

L-BFGS, limited memory BFGS

Logistic, Logistic function

MinO2Sa, Minimal oxygen saturation in sleep

MLP, Multilayer perceptron

NC, Neck Circumference

NPV, Negative predictive value

OSA, Obstructive sleep apnea

O2Sa75, Percent of sleep time oxygen saturation below 75%

O2Sa80, Percent of sleep time oxygen saturation below 80%

O2Sa85, Percent of sleep time oxygen saturation below 85%

O2Sa90, Percent of sleep time oxygen saturation below 90%

O2Sa95, Percent of sleep time oxygen saturation below 95%

PSG, Polysomnogram

PPV, Positive predictive value

ROC, Receiver operating characteristic

SDB, Sleep disordered breathing

SHHS, Sleep Heart Health Study

SHS, Strong Heart Study

SInPub, Fall asleep while sitting inactive in a public place

Snore, Frequency of Snoring

Tanh, Hyperbolic tangent function

TES, Tucson Epidemiologic Study of Airways Obstructive Diseases

Acknowledgements

The database used in the study was developed using the following National Heart, Lung and Blood Institute cooperative agreements: U01HL53940 (University of Washington), U01HL53941 (Boston University), U01HL53938 (University of Arizona), U01HL53916 (University of California, Davis), U01HL53934 (University of Minnesota), U01HL53931 (New York University), U01HL53937 and U01HL64360 (Johns Hopkins University), U01HL63463 (Case Western Reserve University), and U01HL63429 (Missouri Breaks Research).

Sleep Heart Health Study (SHHS) acknowledges the Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Cornell/Mt. Sinai Worksite and Hypertension Studies, the Strong Heart Study (SHS), the Tucson Epidemiologic Study of Airways Obstructive Diseases (TES) and the Tucson Health and Environment Study (H&E) for allowing their cohort members to be part of the SHHS and for permitting data acquired by them to be used in the study. SHHS is particularly grateful to the members of these cohorts who agreed to participate in SHHS as well. SHHS further recognizes

all of the investigators and staff who have contributed to its success. A list of SHHS investigators, staff and their participating institutions is available on the SHHS website, www.jhucct.com/shhs.

This material is based upon work partially supported by the National Science Foundation under Grant No. 1433185. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Reference

1. Redline S, Quan SF. Sleep Apnea: A Common Mechanism for the Deadly Triad—Cardiovascular Disease, Diabetes, and Cancer? *Am J Respir Crit Care Med*. 2012;186:123-124.

2. Oldenburg O, Lamp B, Faber L, Teschler H, Horstkotte D, Töpfer V. Sleep-disordered breathing in patients with symptomatic heart failure: a contemporary study of prevalence in and characteristics of 700 patients. *Eur J Heart Fail*. 2007;9:251-257.

3. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire To Identify Patients at Risk for the Sleep Apnea Syndrome. *Ann Intern Med*. 1999;131:485.

4. Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology*. 2008;108:812-821.

5. Marti-Soler H, Hirotsu C, Marques-Vidal P, et al. The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. *Lancet Respir Med*. 2016;4:742-748.

6. Takegami, M., Hayashino, Y., Chin, K., Sokejima, S., Kadotani, H., Akashiba, T., ... & Fukuhara, S. Simple Four-Variable Screening Tool for Identifcation of Patients with Sleep-Disordered Breathing. *Sleep*. 2009.

7. Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. *J Clin Sleep Med*. 2011;7:467-472.

8. Barker J, Jones SF. Obstructive Sleep Apnea, An Issue of Sleep Medicine Clinics,. Elsevier Health Sciences; 2013.

9. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed*. 2013;11:47-58.

10. Li H, Luo M, Zheng J, et al. An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. *Medicine* . 2017;96:e6090.

11. Isma'eel HA, Cremer PC, Khalaf S, et al. Artificial neural network modeling enhances risk stratification and can reduce downstream testing for patients with suspected acute coronary syndromes, negative cardiac biomarkers, and normal ECGs. *Int J Cardiovasc Imaging*. 2015;32:687-696.

12. Li W, Cao P, Zhao D, Wang J. Pulmonary Nodule Classification with Deep Convolutional Neural Networks on Computed Tomography Images. *Comput Math Methods Med*. 2016;2016:6215085.

13. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline. *J Clin Sleep Med*. 2017;13:479-504.

14. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;20:1077-1085.

15. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep*. 1998;21:759-767.

16. Lind BK, Goodwin JL, Hill JG, Ali T, Redline S, Quan SF. Recruitment of Healthy Adults into a Study of Overnight Sleep Monitoring in the Home: Experience of the Sleep Heart Health Study. *Sleep Breathing*. 2003;7:13-24.

17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. October 2011:2825-2830.

18. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63:3-42.

19. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program*. 1989;45:503-528.

20. Senaratna CV, Perret JL, Matheson MC, et al. Validity of the Berlin questionnaire in detecting obstructive sleep apnea: A systematic review and meta-analysis. *Sleep Med Rev*. April 2017.

21. El-Solh AA, Jeffrey Mador M, Ten-Brock E, Shucard DW, Abul-Khoudoud M, Grant BJB. Validity of Neural Network in Sleep Apnea. *Sleep*. 1999;22:105-111.

22. Kirby SD, Eng P, Danter W, et al. Neural network prediction of obstructive sleep apnea from clinical criteria. *Chest*. 1999;116:409-415.

23. Teferra RA, Grant BJB, Mindel JW, et al. Cost minimization using an artificial neural network sleep apnea prediction tool for sleep studies. *Ann Am Thorac Soc*. 2014;11:1064-1074.

24. Karamanli H, Yalcinoz T, Yalcinoz MA, Yalcinoz T. A prediction model based on artificial neural networks for the diagnosis of obstructive sleep apnea. *Sleep Breath*. 2016;20:509-514.

25. Chowdhuri S, Quan SF, Almeida F, et al. An Official American Thoracic Society Research Statement: Impact of Mild Obstructive Sleep Apnea in Adults. *Am J Respir Crit Care Med*. 2016;193:e37-54.

26. Lweesy K, Fraiwan L, Khasawneh N, Dickhaus H. New automated detection method of OSA based on artificial neural networks using P-wave shape and time changes. *J Med Syst*. 2011;35:723-734.

27. Mosteller F. Data Analysis, Including Statistics.; 1968.

Figure Titles and Captions

Figure 1—Multilayer perceptron neural network

Figure 2—Receiver operating characteristic (ROC) curves of each apnea hypopnea index (AHI) threshold model

Figure 3—Development Process

Table 1—Demographics and Related Variables of the Sleep Heart Health Study Dataset

	Mean ± SD (min-max)
Age (years)	62.1 ± 10.1 (39-90)
Body Mass Index (kg/m^2)	28.51 ± 5.0 (18-50)
Neck Circumference (cm)	38.4 ± 4.2 (27-59)
Percent of sleep time oxygen saturation below 70% (%)	0.03 ± 0.6 (0-24)
Percent of sleep time oxygen saturation below 75% (%)	0.07 ± 1.2 (0-46.4)
Percent of sleep time oxygen saturation below 80% (%)	0.19 ± 2.1 (0-66.3)
Percent of sleep time oxygen saturation below 85% (%)	0.57 ± 3.8 (0- 82.0)
Percent of sleep time oxygen saturation below 90% (%)	3.45 ± 9.9 (0-99.6)
Percent of sleep time oxygen saturation below 95% (%)	44.9 ± 34.4 (0 -100)
Minimum oxygen saturation in sleep (%)	85.3 ± 6.1 (24 - 97)
Frequency of Snoring	2.6 ± 1.2 (0 - 4)

Fall asleep while in a car	1.2 ± 0.5 (1 - 4)
Fall asleep while sitting inactive in a public place	1.8 ± 0.9 (1 - 4)
Fall asleep while sitting and talking	1.2 ± 0.5 (1 - 4)
AHI	11.26 ± 14.2 (0 - 199.5)
	Percentage
Female (%)	44.5%
Male (%)	55.5%
Heart Failure (%)	1.26%
Heart Attack (%)	6.56%
Stroke (%)	2.96%
Hypertension (%)	67.9%
Diastolic Blood Pressure (> 90)(%)	6.14%
Systolic Blood Pressure (> 170)(%)	1.86%
Diabetes (%)	6.42%

Table 2—Prevalence and Number of SDB Positive and Negative Subjects

AHI ≥	Training Positive	Training Negative	Test Positive	Test Negative	Prevalence
5	1266	1014	328	242	55.93%
10	814	1466	216	354	36.14%
15	523	1757	157	413	23.86%
20	366	1914	109	461	16.67%
25	259	2021	80	490	11.89%
30	184	2096	59	511	8.52%

Table 3—Key Variables in the Training Set and Test Set

	Training Set (Mean ± SD)	Test Set (Mean ± SD)	p-value*
Minimal oxygen saturation in sleep (%)	85.4 ± 5.86	85.0 ± 6.78	0.224
Percent of sleep time oxygen saturation below 75% (%)	0.07 ± 1.17	0.13 ± 1.43	0.355
Percent of sleep time oxygen saturation below 80% (%)	0.17 ± 1.92	0.29 ± 2.77	0.310
Percent of sleep time oxygen saturation below 85% (%)	0.53 ± 3.53	0.77 ± 4.69	0.238
Percent of sleep time oxygen saturation below 90% (%)	3.36 ± 9.72	3.85 ± 10.7	0.313
Percent of sleep time oxygen saturation below 95% (%)	44.5 ± 34.4	47.0 ± 34.2	0.117
Age (years)	62.0 ± 10.2	62.5 ± 9.89	0.313
Body Mass Index (kg/m^2)	28.5 ± 4.89	28.7 ± 5.22	0.376
Neck Circumference (cm)	38.4 ± 4.18	38.4 ± 4.25	0.969
Frequency of Snoring	2.63 ± 1.16	2.61 ± 1.19	0.711
Fall asleep while sitting	1.78 ± 0.85	1.75 ± 0.84	0.564

inactive in a public place			

* The two-tailed Student's t-test was used to assess the difference in the mean values. The statistical significance value was $p < 0.05$.

Table 4—Test Set Evaluation

AHI ≥	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
5	95.12% (92.79% - 97.45%)	62.81% (56.72% - 68.90%)	77.61% (73.54% - 81.69%)	90.78% (86.04% - 94.92%)	0.904
10	95.37% (92.57% - 98.17%)	68.93% (64.11% - 73.75%)	65.19% (59.94% - 70.44%)	96.01% (93.67% - 98.46%)	0.912
15	95.54% (92.31% - 98.77%)	69.25% (64.80% - 73.70%)	54.15% (48.28% - 60.02%)	97.61% (95.86% - 99.36%)	0.913
20	97.25% (94.18% - 100%)	72.02 % (67.92% - 76.11%)	45.10% (38.74% - 51.47%)	99.10% (98.10% - 100%)	0.926
25	96.25% (92.09% - 100%)	70.20% (66.16% - 74.25%)	34.53% (28.29% - 40.77%)	99.13% (98.16% - 100%)	0.930
30	98.31% (95.01% - 100%)	71.62 % (67.72% - 75.53%)	28.57% (22.36% - 34.79%)	99.73% (99.2% - 100%)	0.954

CI: Confidence intervals; PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under the receiver operating characteristic curve

Development of the Algorithm

The development process of the screening tool is schematically described in Figure S1. Initially, our team (with substantial input from SFQ and GES) empirically selected 22 candidate variables based on their known association with SDB from the 1280 variables as features. The 22 candidate variables included: the SaO_2% during sleep, age at time of study in years, based on start date of SHHS1 PSG recording; Gender as reported by Parent Cohort; Parent Cohort reported Diabetes Status; Neck circumference in centimeters. Questions included: How often do you snore? What is chance that you would doze off or fall asleep while in a car, while stopped for a few minutes in traffic? What is the chance that you would doze off or fall asleep while sitting inactive in a public place? What is the chance that you would doze off or fall asleep while sitting and talking to someone?

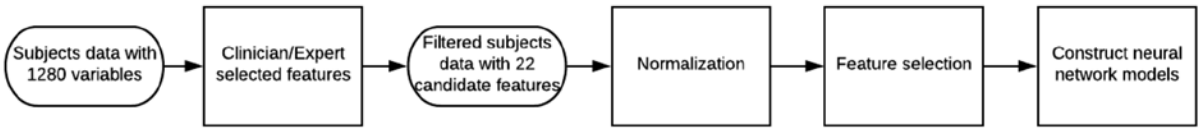

Figure S1—Development Process

Normalization

We used the Scikit-learn machine learning library to develop the screening tool.[17] The min-max normalization strategy was applied to normalize the 22 candidate features to the range [0, 1]. The following equations further clarify the normalization process:

$$X_{std} = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (2)$$

$$X_{scaled} = X_{std} * (\max - \min) + \min \qquad (3)$$

Here in (2), X is a 2D array storing all candidate features. The max(X) and min(X) are two 1D arrays with maximum values and minimum values of the features in the full dataset. In (3), the (max, min) is the normalized range of the candidate features. In this study, max is 1, min is 0.

Feature Selection

We employed the extremely randomized trees algorithm with model selection as the feature selection algorithm.[1,2] The extremely randomized trees algorithm is a tree-based ensemble method to build 10 total randomized trees.[2] The importance weights of each feature were computed by the feature selection algorithm. Features used at top of the trees have higher important weight.[1] The total weight is 1 of the 22 candidate features. The input features of each MLP neural network model were selected from the 22 candidate features based on their importance weights. The AHI threshold ≥ 5/hour had 7 input features. The AHI thresholds ≥ 10/hour, 15/hour, and 20/hour models had 8 input features each. The AHI threshold ≥ 25/hour model had 11 input features. The AHI threshold ≥ 30/hour model had 9 input features.

Construction of Neural Network Models

The MLP neural network had two layers: one hidden layer, and an output layer. The MLP neural network models were trained by the backpropagation learning method in conjunction with the well-known limited memory BFGS optimization algorithm (L-BFGS) with L2 regularization. The L-BFGS is a quasi-Newton method algorithm with limited computer memory.[3] During the optimization process, the parameters in neural network models were random initiated. The parameters of the model were optimized to reduce the output error. The output error was computed by the cross-entropy cost function. This process was repeated for all subjects in the training set over several iterations. After sufficient training, the model learned how to accurately compute the output result. In this study, we used two types of activation functions for the hidden

layer of different neural network models: the logistic function (Logistic, in (4)) and the hyperbolic tangent function (Tanh, in (5)). The activation function of output layer is the logistic function. We applied the grid search method to optimize the hyper-parameters of the MLP neural network models and used the AUC as the metric. In the optimization procedure, we used a 10-fold cross-validation to evaluate the AUC.[4] The optimized hyper-parameters, activation function, and input features for each model are listed in Table S1.

$$f_{logistic}(x) = \frac{1}{1+e^{-x}} \quad (4)$$

$$f_{tanh}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \quad (5)$$

Table S1—Hyper-parameters and Selected Features of Each Model

AHI ≥	Activation function of hidden layer	Hidden neurons	L2 regularization term	Features
5	Logistic	4	10^{-4}	Age, BMI, MinO_2Sa, NC, O_2Sa90, O_2Sa95, Snore
10	Tanh	3	10^{-2}	Age, BMI, MinO_2Sa, NC, O_2Sa85, O_2Sa90, O_2Sa95, Snore
15	Logistic	6	10^{-3}	Age, BMI, MinO_2Sa, NC, O_2Sa85, O_2Sa90, O_2Sa95, Snore
20	Logistic	4	10^{-2}	Age, BMI, MinO_2Sa, NC, O_2Sa85, O_2Sa90, O_2Sa95, Snore
25	Tanh	10	10^{-3}	Age, BMI, MinO_2Sa, NC, O_2Sa75, O_2Sa80, O_2Sa85, O_2Sa90, O_2Sa95, Snore, SInPub
30	Tanh	7	10^{-3}	Age, BMI, MinO_2Sa, NC, O_2Sa80, O_2Sa85, O_2Sa90, O_2Sa95, Snore

BMI: Body Mass Index; MinO_2Sa: Minimal oxygen saturation in sleep; NC: Neck circumference; O_2Sa90: Percent of sleep time oxygen saturation below 90%; O_2Sa95: Percent of sleep time oxygen saturation below 95%; Snore: Frequency of Snoring; O_2Sa85: Percent of sleep time oxygen saturation below 85%; O_2Sa75: Percent of sleep time oxygen saturation below 75%; O_2Sa80: Percent of sleep time oxygen saturation below 80%; SInPub: Fall asleep while sitting inactive in a public place

References

1. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011:2825-2830.
2. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63:3-42.
3. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program. 1989;45:503-528.
4. Mosteller F, Tukey JW. Data analysis, including statistics. In: Lindzey G, Aronson E, eds. Handbook of Social Psychology, Vol. 2. Addison–Wesley; 1968.