https://doi.org/10.5664/jcsm.9630

JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Obstructive sleep apnea predicts 10-year cardiovascular disease–related mortality in the Sleep Heart Health Study: a machine learning approach

Ao Li, PhD[1,2]; Janet M. Roveda, PhD[1,2,3]; Linda S. Powers, PhD[1,2,3]; Stuart F. Quan, MD[4,5]

[1]Department of Electrical and Computer Engineering, College of Engineering, University of Arizona, Tucson, Arizona; [2]The BIO5 Institute, University of Arizona, Tucson, Arizona; [3]Department of Biomedical Engineering, College of Engineering, University of Arizona, Tucson, Arizona; [4]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [5]Asthma and Airway Disease Research Center, College of Medicine, University of Arizona, Tucson, Arizona

Study Objectives: Obstructive sleep apnea (OSA) is considered to be an important risk factor for the development of cardiovascular disease (CVD). This study aimed to develop and evaluate a machine learning approach with a set of features for assessing the 10-year CVD mortality risk of the OSA population.

Methods: This study included 2,464 patients with OSA who met study inclusion criteria and were selected from the Sleep Heart Health Study. We evaluated the importance of potential features by mutual information. The top 9 features were selected to develop a random forest model.

Results: We evaluated the model performance on a test set (n = 493) using the area under the receiver operating curve with 95% confidence interval and confusion matrix. A random forest model awarded the highest area under the receiver operating curve of 0.84 (95% confidence interval: 0.78–0.89). The specificity and sensitivity were 73.94% and 81.82%, respectively. Sixty-three years old was a threshold for increased risk of 10-year CVD mortality. Persons with severe OSA had higher risk than those with mild OSA.

Conclusions: This study demonstrated that a random forest model can provide a quick assessment of the risk of 10-year CVD mortality. Our model may be more informative for patients with OSA in determining their future CVD mortality risk.

Keywords: apnea-hypopnea index, cardiovascular mortality, machine learning, obstructive sleep apnea

Citation: Li A, Roveda JM, Powers LS, Quan SF. Obstructive sleep apnea predicts 10-year cardiovascular disease–related mortality in the Sleep Heart Health Study: a machine learning approach. *J Clin Sleep Med*. 2022;18(2):497–504.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Obstructive sleep apnea is a prevalent disease and associated with increased fatal and nonfatal cardiovascular disease (CVD). Early prediction is important to prevention of CVD morality for patients with obstructive sleep apnea.

Study Impact: This study is one of the first to combine random forest modeling and obstructive sleep apnea severity with other conventional CVD risk factors to predict the risk of 10-year CVD mortality. This approach has the potential to provide more personalized prediction of long-term CVD risk to patients with obstructive sleep apnea.

INTRODUCTION

Cardiovascular disease (CVD) is a class of prevalent diseases, including myocardial infarction, angina pectoris, heart failure, coronary revascularization procedure, and stroke. In the United States, a recent study found CVD is the cause of 655,000 deaths each year.[1] Preventing CVD disease and related mortality centers on identification of persons at greatest risk and subsequent treatment and mitigation of various risk factors.

Obstructive sleep apnea (OSA) is a highly prevalent disease and frequently coexists with CVD. In the general population, approximately 22% (range, 9–37%) in men and 17% (range, 4–50%) in women have OSA.[2] Studies found OSA is positively associated with increased fatal and nonfatal cardiovascular events,[3–5] and is an important risk factor for the development of stroke.[6] Information that would identify those at greatest risk for subsequent CVD mortality, defined as death due to cardiovascular or cerebrovascular disease, in an OSA population would be useful to clinicians.

The commonly used evaluation tools of 10-year CVD risk (eg, Framingham Risk Score[7]), however, do not include sleep factors as predictors. Our hypothesis is that we can use machine learning models with traditional risk factors and sleep factors to provide personalized prediction of 10-year CVD mortality risk for middle-aged adults with OSA. Therefore, in this study, we developed an approach that uses mutual information and random forest modeling to assess 10-year risk of CVD mortality for patients with OSA.

METHODS

Dataset

The Sleep Heart Health Study (SHHS) dataset was used to develop several machine learning models for assessing the 10-year risk of CVD mortality.[8–10] The SHHS is an ideal resource to be utilized for this purpose because it was a prospective multicenter cohort study designed to investigate the

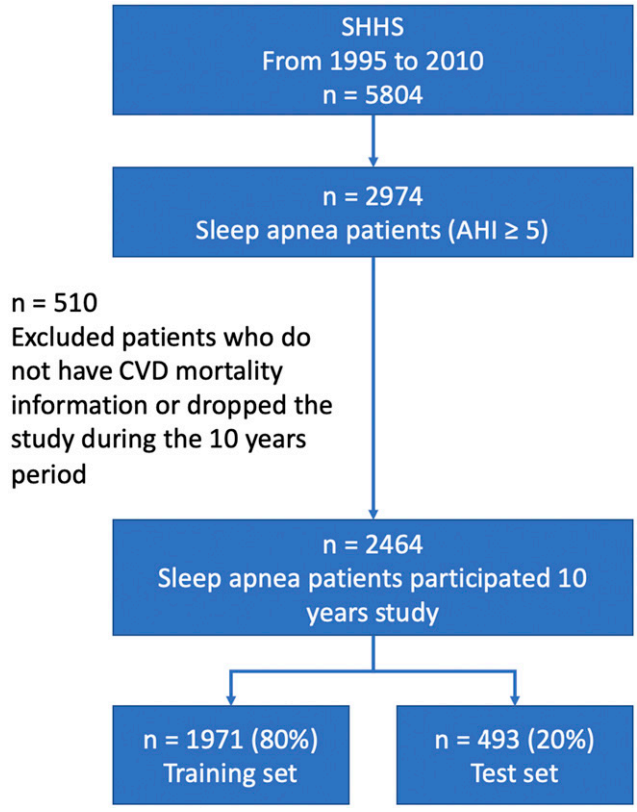
Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

longitudinal impact of OSA on CVD in the United States. A complete description of the SHHS has been previously published.[8] Briefly, 6,441participants, 40 years of age and older were recruited starting in 1995 from several ongoing "parent" cardiovascular and respiratory disease cohorts that were initially assembled between 1976 and 1995. The publicly accessible database (**Figure 1**) included 1,280 variables and 5,804 participants, with 2,765 males (47.6%) and 3,039 females (52.4%).[9,10]

Polysomnography

Home polysomnography was recorded using the Compumedics P Series System (Compumedics Ltd., Abbotsford, Australia). The montage included a single-lead electrocardiogram, heart rate (using a bipolar electrocardiogram lead), 2 electroencephalogram derivations (C4/A1 and C3/A2), 2 electrooculogram derivations, chin electromyography, thoracic and abdominal respiratory inductance plethysmography, airflow (detected by a nasal-oral thermocouple [Protec, Woodinville, WA]), finger pulse oximetry (Nonin, Minneapolis, MN), and body position (using a mercury gauge sensor), and ambient light (on/off, by a light sensor secured to the recording garment).[8] At the time of the home visit, each participant's medications were documented. In addition, the SHHS Sleep Habits Questionnaire and Health Interview were administered; these questionnaires included the Epworth Sleepiness Scale and the Medical Outcomes Study SF-36 as well as items which ascertained the presence of several health conditions.[11,12] Blood pressure was measured manually in triplicate in a seated position after 5 minutes of rest.[8] Body weight was obtained using a digital scale. Comprehensive descriptions of the polysomnography and quality-assurance procedures have been previously published.[13] The vital status and days since the baseline examination were recorded during the follow-up years until 2010. Adjudication of cardiovascular events and mortal status for SHHS was performed locally by each of the study's parent cohorts as previously described.[14]

Figure 1—Flow of study participants.



AHI = apnea-hypopnea index, CVD = cardiovascular disease, SHHS = Sleep Heart Health Study.

Scoring

Sleep was scored according to criteria described by Rechtschaffen and Kales.[15] Intrascorer and interscorer reliabilities were high.[13] The apnea-hypopnea index (AHI) was calculated for each participant as the sum of all apneas and hypopneas divided by the total sleep time. An apnea was defined as a complete or almost complete cessation of airflow, as measured by the amplitude of the thermocouple signal, lasting at least 10 seconds. Hypopneas were identified if the amplitude of a measure of flow or volume was reduced discernibly (at least 30% lower than baseline breathing) for at least 10 seconds, if the event did not meet the criteria for apnea, and the event was associated with a ≥ 4% oxygen desaturation from baseline.

Preprocessing

As shown in **Figure 1**, 2,974 participants had an AHI ≥ 5 events/h at the baseline examination of SHHS and were considered to have OSA; 510 were excluded because of the absence of 10-year mortality information. Therefore, the resulting analytic dataset included a total of 2,464 participants.

Selection of features to be used in model development was based on previous studies documenting their association with CVD.[16–23] These included demographics (age and sex), anthropometric (body mass index was calculated as weight [kg]/height [m^2]), social behavioral indices (smoking history [lifetime over 20 packs of cigarettes] and alcohol usage), plasma lipids (cholesterol, high density lipoprotein, and triglycerides), several diseases (hypertension, diabetes, and depression), spirometry (forced vital capacity [FVC] and forced expiratory volume [FEV_1]), and preexisting CVD history. We also included several sleep factors (insomnia, total sleep time, time in bed, and AHI). Because of the small number of individual CVD events, we defined the prevalence of CVD as the presence of myocardial infarction, angina pectoris, heart failure, coronary bypass surgery, stroke, angioplasty, or other heart/cardiac surgery history before the baseline study. Depression was defined as feeling "blue" or "down" for at least "a good bit of the time" for the previous 4 weeks before the baseline study or current use of antidepression medicine. We defined insomnia symptoms as often or almost always having "trouble falling asleep," "waking up during the middle of the night and having difficulty getting back to sleep," or "waking up too early in the morning and being unable to get back to sleep." Hypertension was defined as blood pressure greater than 140/90 mm Hg based on

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

the average of the second and third blood pressure readings or current treatment with hypertension medication. Sleepiness was stratified into 4 levels (normal: 0–10, mild: 11–14, moderate: 15–17, severe: 18 or higher) based on participants' Epworth Sleepiness Scale score.

Based on their adjudicated mortality outcomes, each participant was labeled as "Negative/Positive" for 10-year CVD mortality and was used as ground truth for the model outcomes. We randomly separated the 2,464 participants into 2 sets: a training set (n = 1971) and a test set (n = 493). Approximately 24% of participants had at least 1 missing value that was considered missing at random. However, in both training and test sets, age, sex, hypertension, and all sleep factors (eg, AHI) did not include missing values. The overall missing rate was less than 2%. Therefore, we used the mean values calculated by the training set to impute the continuous variables in the training set and test set.[24,25] For handling missing categorical values, we used a dummy method to encode the categorical features (eg, diabetes [positive, negative, unknown] was encoded to diabetes positive, diabetes negative, and diabetes unknown features).

We used the same test set (n = 493) with additional variables (eg, systolic blood pressure) to evaluate the Framingham Risk Score. Approximately 8% of the participants included at least 1 missing value. The Framingham study provided mean values and negative labels (0) were used to impute continuous variables and categorical variables, respectively.[7]

Model development

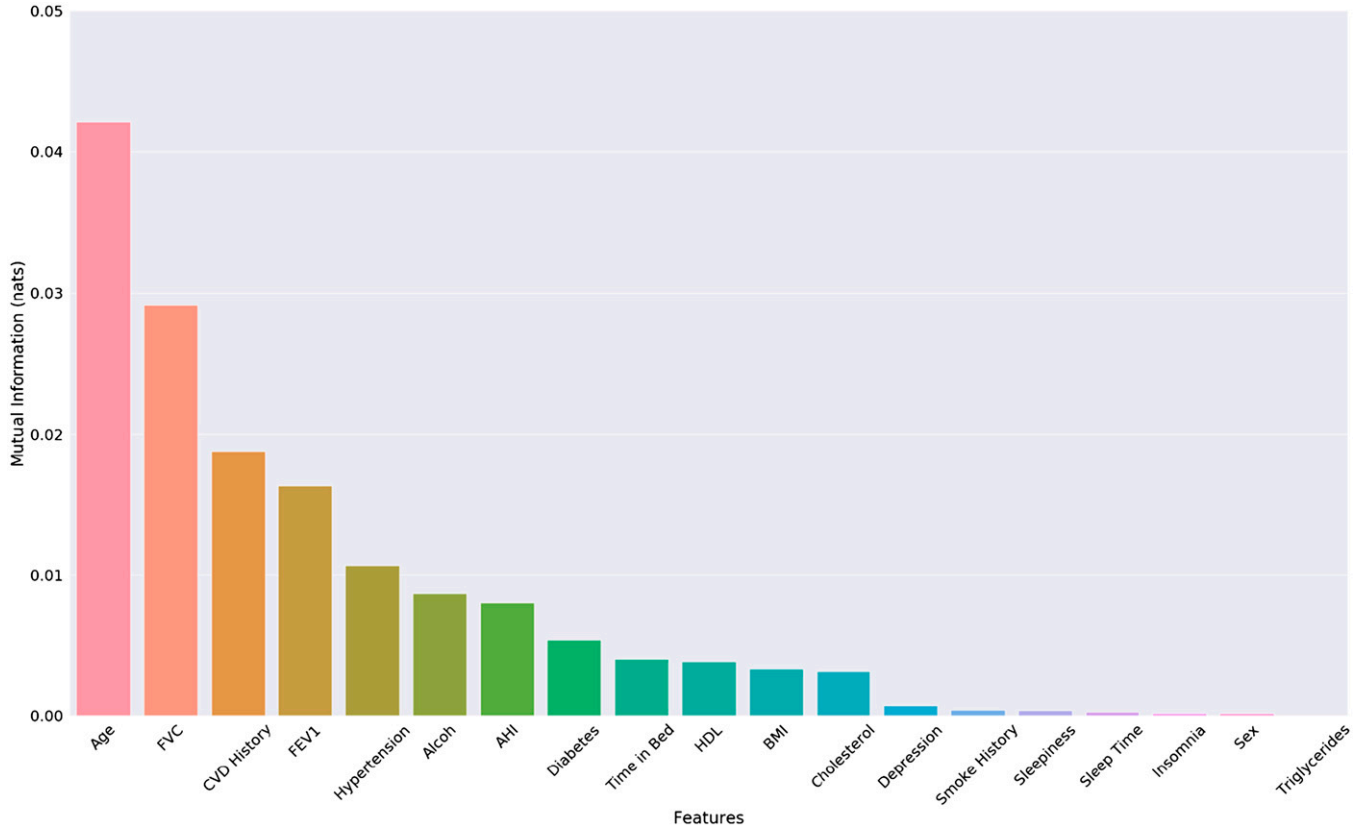
We developed and evaluated a random forest model for assessing the 10-year CVD mortality risk. Random forest is a type of ensemble learning algorithm. It builds multiple decision trees by randomly choosing subsamples from the training set and uses a weighted averaging method to improve their predictive accuracy where decision trees aim to learn decision rules from

Table 1—Description of training and test sets.

	Training Set (n = 1,971)	Test Set (n = 493)	*P*
Risk factors			
Age (years)	67 ± 10	66 ± 10	.45
Male (%)	58	57	.69
Race (%)			.86
White	89	88	
Black	7	8	
Others	4	4	
Body mass index (kg/m^2)	30 ± 5	29 ± 5	.62
Total cholesterol	207 ± 39	209 ± 36	.31
High density lipoprotein (mg/dL)	48 ± 15	49 ± 15	.39
Triglycerides (mg/dL)	160 ± 113	159 ± 112	.89
Alcohol usage (drinks/day)	3 ± 6	3 ± 6	.33
Forced expiratory volume (liters)	3 ± 1	3 ± 1	.60
Forced vital capacity (liters)	4 ± 1	4 ± 1	.55
Apnea-hypopnea index (events/h)	18 ± 15	18 ± 14	.43
Sleep time (minutes)	586 ± 105	577 ± 114	.09
Time in bed (minutes)	439 ± 58	435 ± 62	.18
Insomnia (%)	30	30	.98
Sleepiness (%)			.31
Normal	73	76	
Mild	18	17	
Moderate	6	5	
Severe	3	2	
Hypertension (%)	47	49	.35
A history of cardiovascular diseases (%)	21	21	.73
Diabetes (%)	10	9	.33
Depression (%)	5	7	.28
Smoking history (> 20 packs in whole lifetime) (%)	56	56	.91
Ground truth			
10-year cardiovascular disease mortality (%)	9	11	.19

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

Figure 2—Feature importance.



AHI = apnea-hypopnea index, BMI = body mass index, CVD = cardiovascular disease, FEV1 = forced expiratory volume, FVC = forced vital capacity, HDL = high-density lipoprotein, nat = natural unit of information.

the training data.[26] Random forest can catch both linear and nonlinear relationships between input features and the target variable.

We used mutual information technology with 5-fold stratified cross validation to evaluate the importance of the features. Mutual information measures the amount of information that 1 random variable contains about the target variable.[27,28] High mutual information means a large reduction in the uncertainty of the target variable, when the values of a random variable are provided. Zero mutual information means the 2 variables are independent. The top 50% of features (9 features) with the highest mutual information were selected as input to train the random forest model. In development, we set the weight of the negative class to 11 to overcome a class imbalance problem. Grid search with stratified 5-fold cross validation was used to find the optimal values of hyperparameters (criterion, max depth, max samples, max features, and number of estimators), where the stratified cross validation was used to preserve the percentage of samples of each class for each fold. After finding optimal hyperparameters and evaluating the performance of a model by 5-fold cross validation, we trained the model with the best found hyperparameters on the entire training set (n = 1971) and evaluated the performance on the hold-out test set (n = 493). The random forest model was implemented by Python package Scikit-learn v0.23.[29]

Statistical analysis

We used mean and standard deviation as well as percentages to provide an overall description of the training and test sets. Chi-square and 2-tailed *t* tests were used to measure the difference between the training and test sets. We used the area under the receiver operating curve (AUC) as the metric to evaluate performance. Delong's method was used to compute the 95% confidence interval of AUC.[30,31] A bootstrap method with 1,000 iterations and 50% samples was used to compute the 95% confidence interval of adjusted odds ratios. We considered that $P < .05$ indicated statistical significance in our analyses. Analyses were performed using Python package Scikit-learn v0.23 and Scipy v1.4.[29,32]

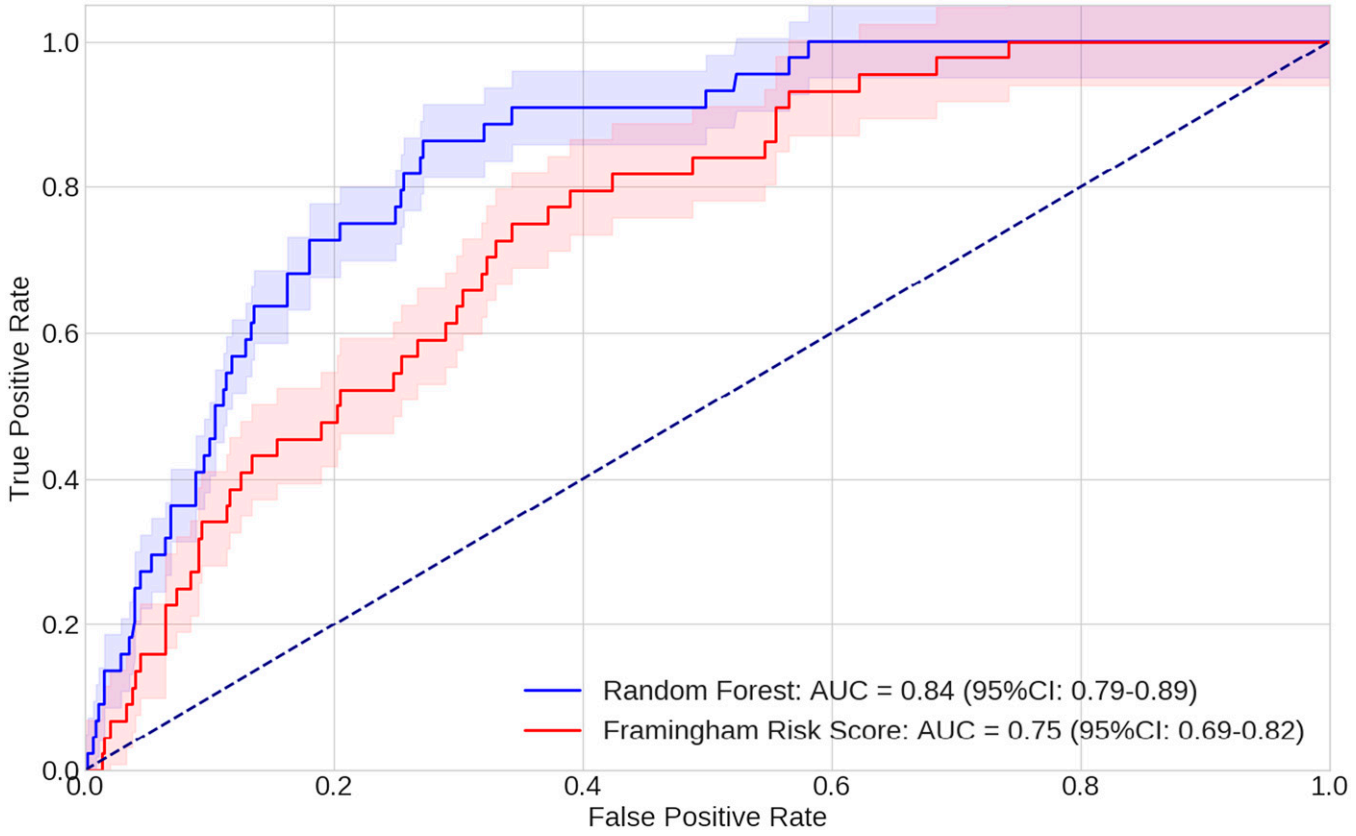
RESULTS

Table 1 describes the demographic, anthropometric and clinical characteristics of datasets. We did not find statistically significant differences between the training set and test set.

We used mutual information technology to evaluate the importance of the 19 features (**Figure 2**). The age feature had the highest mutual information. **Figure 2** also shows that the FVC and FEV_1 have higher mutual information than most of

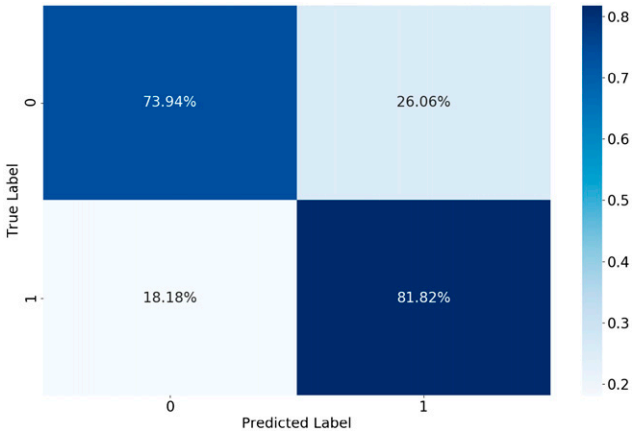
Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

Figure 3—Receiver operating characteristics analysis.



The blue dot line (45-degree diagonal) indicates the performance of a random classifier. The shadows indicate 95% CI. AUC = area under the receiver operating characteristic curve, CI = confidence interval.

Figure 4—Confusion matrix.



0.5 was selected as the probability threshold. The confusion matrix was normalized over the true label to show prediction performance as percentages. The diagonal elements represent the percentage of participants for which the predicted label is equal to the true label. The off-diagonal elements represent the percentage of misclassified participants. The color bar represents percentage; darker blue means higher percentage.

the other features. We also found that the dependency between 10-year CVD mortality and AHI was lower than for FEV_1 and FVC but higher than for diabetes, high density lipoprotein, body mass index, and cholesterol.

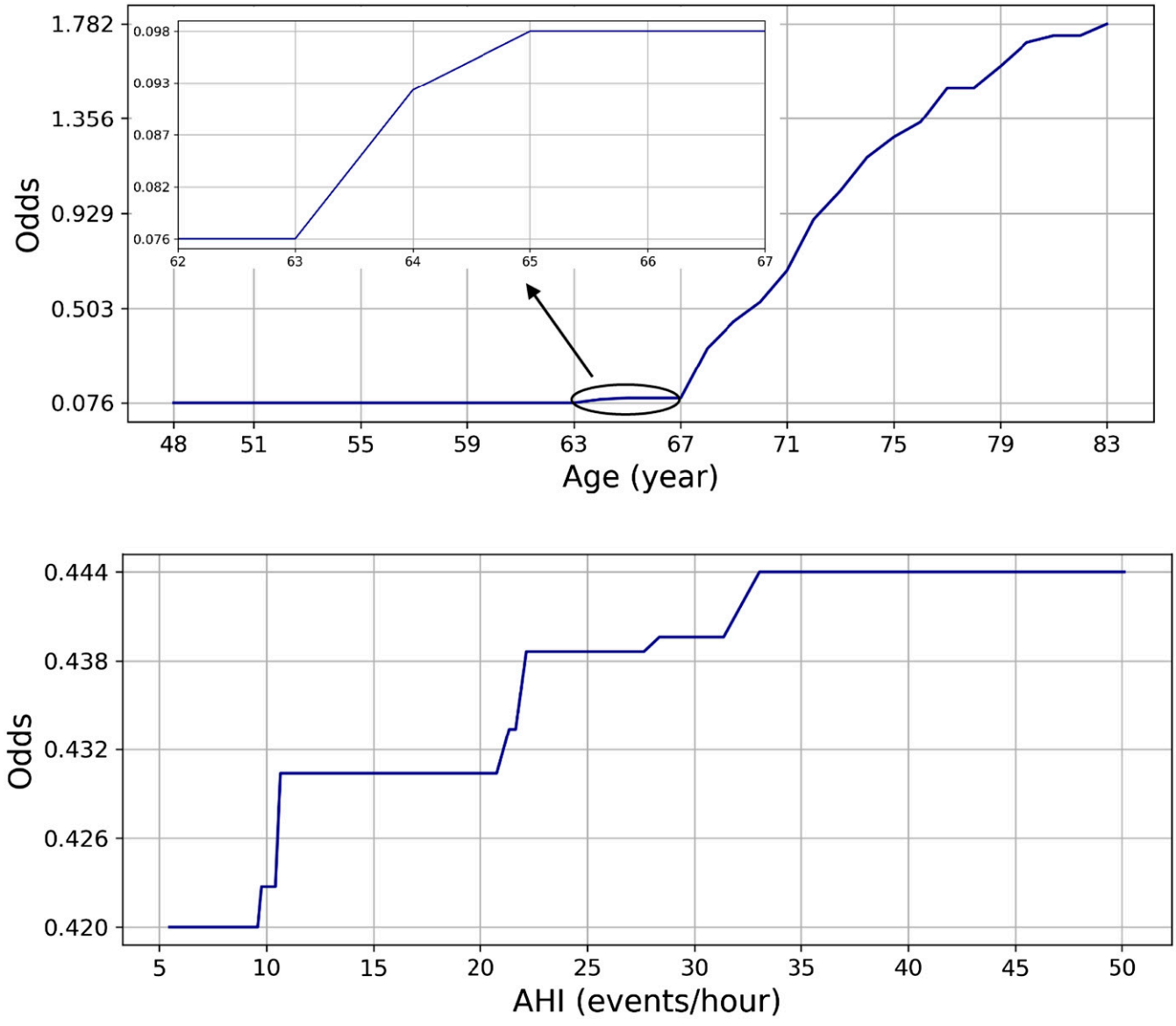
On the training set, the random forest model awarded an AUC of 0.84. On the test set (n = 493), the random forest model achieved higher AUC than the Framingham Risk Score (AUC: 0.84 vs 0.75) which we show for comparison (**Figure 3**).

In evaluation, we chose 0.5 as the threshold for classifying positive/negative 10-year CVD mortality. The confusion matrix of the ensemble random forest model is shown in **Figure 4**. It showed excellent discrimination ability. The sensitivity and specificity of the random forest model were 81.82% and 73.94%, respectively.

Figure 5 shows the impact of features on the output. As age and AHI increase, the probability of CVD mortality after 10 years increases as well. For example, 63 years is a threshold for increased risk of 10-year CVD mortality. The odds for CVD mortality after 10 years for a 67-year-old is 1.290-fold (95% confidence interval: 1.276–1.297) that of under 63-year-old. Similarly, persons with severe OSA (AHI = 35 events/h) have a 1.057-fold (95% confidence interval: 1.055–1.059) greater risk than those with mild OSA (AHI = 5 events/h).

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

Figure 5—Adjusted odds of CVD mortality, given features of the model.



The odds shown were adjusted for the remaining 8 features. Calculation of odds ratios:

- $\text{OR}\left(\frac{Age=67}{Age=63}\right) = \frac{\text{Odds(Age=67)}}{\text{Odds(Age=63)}} = \frac{0.098}{0.076} = 1.290$
- $\text{OR}\left(\frac{Age=73}{Age=63}\right) = \frac{\text{Odds(Age=73)}}{\text{Odds(Age=63)}} = \frac{1.031}{0.076} = 13.566$
- $\text{OR}\left(\frac{AHI=35}{AHI=5}\right) = \frac{\text{Odds(AHI=35)}}{\text{Odds(AHI=5)}} = \frac{0.444}{0.420} = 1.057$

AHI = apnea-hypopnea index, CVD = cardiovascular disease, OR = odds ratio.

DISCUSSION

In this study, we developed a random forest model on a population with OSA to assess its ability to calculate the risk of CVD mortality after 10 years and compared it with the Framingham Risk Score. The random forest model performed the best with a 0.84 AUC (95% confidence interval: 0.78–0.89). This study demonstrated that the random forest model provided an accurate decision model for 10-year mortality in patients with OSA.

Our finding that the random forest model performed better than use of the Framingham Risk Score is of particular interest. The Framingham Risk Score includes age, sex, smoking

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

history, lipid levels, blood pressure, and use of antihypertensive medications.[7] In contrast, the most important features in our approach had little overlap with the Framingham Risk Score with the exception of age and history of hypertension. Thus, it appears that in an OSA population, factors predicting 10-year CVD mortality are different than in a general adult population.

We found AHI is an important predictor for assessing the risk of 10-year CVD mortality. By applying mutual information analysis, we found it was more informative for predicting CVD mortality than diabetes, high density lipoprotein, and cholesterol, which are commonly used features, although it was less useful than FVC and FEV_1. Additionally, the random forest model learned to use AHI as an important feature to build decision trees and showed a high discrimination ability. We also found severe OSA patients have a higher risk than mild OSA patients on 10-year CVD mortality. These findings are consistent with previous research. Results from several large cohort studies also have found OSA is associated with CVD and CVD-related mortality.[33,34] Our study extends these previous reports by using a different analytic approach to confirm that OSA severity is an important pathophysiologic component leading to CVD mortality in a population with OSA. Furthermore, it suggests that in this population, OSA severity is more important than the presence of diabetes and hyperlipidemia.

Age contributed the highest predictive power. Several previous studies have proven that age is a primary risk factor of CVD and CVD mortality and adults aged over 65 years accounted for 82% of total CVD mortality in 2005.[35–37] Through the random forest model, we further found the risk of CVD mortality had a nonlinear association with age. Sixty-three years is an important threshold for increased risk of 10-year CVD mortality. Additionally, we also observed that the 10-year risk for CVD mortality for a 67-year-old and a 73-year-old are 1.3-fold and 13.6-fold that of those under 63 years old, respectively. The random forest model combined age with other features and showed higher discrimination ability. Our observations suggest that OSA severity and age in conjunction with other conventional CVD risk factors have the potential to provide a more personalized prediction of long-term CVD risk to patients with OSA.

Interestingly, we found that pulmonary function (FEV_1 and FVC) was an important predictor of mortality. Furthermore, the dependency between FEV_1 and FVC and 10-year CVD mortality was higher than all other risk factors, except age and history of previous CVD. Previous studies also have reported an inverse risk of all-cause mortality with better pulmonary function. In one study there was a linear negative association with FVC and nonlinear negative association with FEV_1.[38] Magnussen et al found that reduced FEV_1 and FVC were related to all-cause mortality in the general population.[39] Our findings reinforce the concept that the pulmonary and cardiovascular systems are intertwined and suggest that improvement in pulmonary function, if possible, may be more important than correction of other cardiac risk factors.

Several limitations of our study should be noted. First, the average age of the study population is over 65 years old. Further study is required to test the model on a younger population. Second, we classified patients with OSA by an AHI definition requiring at least a 4% desaturation for hypopneas. In the future, the models need to be evaluated on an OSA population who are identified by an AHI definition requiring a minimum 3% desaturation for hypopneas or arousal. Finally, there may have been some misclassification of preexisting CVD. Our definition combined several different outcomes and was dependent on adjudication of these outcomes by SHHS's parent cohorts. In some cases, these may have been self-reported.

Although there are limitations, our study has several strengths. We used the innovative mutual information procedure to evaluate the dependency between features and 10-year CVD mortality on a large well characterized OSA cohort. The random forest model outperformed the Framingham Risk Score and explored nonlinear associations between risk factors and CVD mortality. Importantly, all participants had polysomnography to document the severity of their OSA and CVD outcomes were objectively adjudicated.

In conclusion, we evaluated a machine learning model on a large OSA population and found that the model outperformed the Framingham Risk Score for prediction of 10-year CVD mortality and that OSA severity was a significant contributor to the prediction model. In an OSA population, use of our model may be more informative for patients in determining their future CVD mortality risk than more conventional risk assessments.

ABBREVIATIONS

AHI, apnea-hypopnea index
AUC, area under the receiver operating curve
CVD, cardiovascular disease
FEV_1, forced expiratory volume
FVC, forced vital capacity
OSA, obstructive sleep apnea
SHHS, Sleep Heart Health Study

REFERENCES

1. Virani SS, Alonso A, Benjamin EJ, et al; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation.* 2020;141(9):e139–e596.
2. Franklin KA, Lindberg E. Obstructive sleep apnea is a common disorder in the population-a review on the epidemiology of sleep apnea. *J Thorac Dis.* 2015;7(8):1311–1322.
3. Peker Y, Carlson J, Hedner J. Increased incidence of coronary artery disease in sleep apnoea: a long-term follow-up. *Eur Respir J.* 2006;28(3):596–602.
4. Arzt M, Young T, Finn L, Skatrud JB, Bradley TD. Association of sleep-disordered breathing and the occurrence of stroke. *Am J Respir Crit Care Med.* 2005;172(11):1447–1451.
5. Lüthje L, Andreas S. Obstructive sleep apnea and coronary artery disease. *Sleep Med Rev.* 2008;12(1):19–31.
6. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med.* 2005;353(19):2034–2041.
7. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.* 2008;117(6):743–753.
8. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep.* 1997;20(12):1077–1085.

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.

9. Dean DA II, Goldberger AL, Mueller R, et al. Scaling up scientific discovery in sleep medicine: The National Sleep Research Resource. *Sleep.* 2016;39(5): 1151–1164.
10. Zhang G-Q, Cui L, Mueller R, et al. The National Sleep Research Resource: towards a sleep data commons. *J Am Med Inform Assoc.* 2018;25(10):1351–1358.
11. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540–545.
12. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30(6):473–483.
13. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep.* 1998;21(7):749–757.
14. Punjabi NM, Caffo BS, Goodwin JL, et al. Sleep-disordered breathing and mortality: a prospective cohort study. *Plos Med.* 2009;6(8):e1000132.
15. Rechtschaffen A. Kales A, eds. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Sleep*. Los Angeles: Brain Information Service/Brain Research Institute; 1968.
16. Djoussé L, Lee IM, Buring JE, Gaziano JM. Alcohol consumption and risk of cardiovascular disease and death in women: potential mediating mechanisms. *Circulation.* 2009;120(3):237–244.
17. Wang B, Zhou Y, Xiao L, et al. Association of lung function with cardiovascular risk: a cohort study. *Respir Res.* 2018;19(1):214.
18. Whittaker HR, Bloom C, Morgan A, Jarvis D, Kiddle SJ, Quint JK. Accelerated FEV 1 decline and risk of cardiovascular disease and mortality in a primary care population of COPD patients. *Eur Respir J.* 2021;57(3):2000918.
19. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. *Lancet.* 2014;384(9943):626–635.
20. Bradley SM, Rumsfeld JS. Depression and cardiovascular disease. *Trends Cardiovasc Med.* 2015;25(7):614–622.
21. Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J.* 1991;121(1 Pt 2):293–298.
22. Appelman Y, van Rijn BB, Ten Haaf ME, Boersma E, Peters SAE. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis.* 2015;241 (1):211–218.
23. Wald NJ, Simmonds M, Morris JK. Screening for future cardiovascular disease using age alone compared with multiple risk factors and age. *PLoS One.* 2011;6(5):e18742.
24. Josse J, Prost N, Scornet E, Varoquaux G. On the consistency of supervised learning with missing values; 2020. https://hal.archives-ouvertes.fr/hal-02024202/ document. Accessed September 12, 2021.
25. Pérez A, Dennis RJ, Gil JFA, Rondón MA, López A. Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. *Stat Med.* 2002;21(24):3885–3896.
26. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
27. Ross BC. Mutual information between discrete and continuous data sets. *PLoS One.* 2014;9(2):e87357.
28. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys E.* 2004;69:066138.
29. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
30. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett.* 2014;21(11):1389–1393.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–845.
32. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261–272.
33. Geovanini GR, Wang R, Weng J, et al; The Multi-Ethnic Study of Atherosclerosis. Association between obstructive sleep apnea and cardiovascular risk factors: variation by age, sex, and race. The Multi-Ethnic Study of Atherosclerosis. *Ann Am Thorac Soc.* 2018;15(8):970–977.
34. Shahar E, Whitney CW, Redline S, et al. Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. *Am J Respir Crit Care Med.* 2001;163(1):19–25.
35. Harman D. The free radical theory of aging. *Antioxid Redox Signal.* 2003;5(5):557–561.
36. Lloyd-Jones DM, Wilson PWF, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. *Am J Cardiol.* 2004;94(1):20–24.
37. Yazdanyar A, Newman AB. The burden of cardiovascular disease in the elderly: morbidity, mortality, and costs. *Clin Geriatr Med.* 2009;25(4):563–577, vii.
38. Neas LM, Schwartz J. Pulmonary function levels as predictors of mortality in a national sample of US adults. *Am J Epidemiol.* 1998;147(11):1011–1018.
39. Magnussen C, Ojeda FM, Rzayeva N, et al; Gutenberg Health Study Investigators. FEV1 and FVC predict all-cause mortality independent of cardiac function - Results from the population-based Gutenberg Health Study. *Int J Cardiol.* 2017;234:64–68.

ACKNOWLEDGMENTS

The Sleep Heart Health Study was supported by 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). A list of SHHS investigators, staff, and participating institutions is available on the SHHS website, http://jhuccs1.us/shhs/details/investigators.htm.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication April 9, 2021
Submitted in final revised form August 19, 2021
Accepted for publication August 19, 2021
Address correspondence to: Ao Li, PhD, Department of Electrical and Computer Engineering, The University of Arizona, 1230 E Speedway Blvd, Tucson, AZ, 85719; Email: aoli1@Email.arizona.edu

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at the University of Arizona. Dr. Li reports grants from the National Science Foundation, outside the submitted work. Dr. Quan is a consultant for Best Doctors, Bryte Foundation, and Whispersom, outside the submitted work. The other authors report no conflicts of interest.

Downloaded from jcsm.aasm.org by Kirsten Taylor on February 7, 2022. For personal use only. No other uses without permission.
Copyright 2022 American Academy of Sleep Medicine. All rights reserved.