JACC: ADVANCES VOL. 4, NO. 10, 2025

ORIGINAL RESEARCH

Deep Neural Network Algorithm Using the Electrocardiogram for Detection of Obstructive Sleep Apnea

Naima Covassin, PhD,[a] Kan Liu, PhD,[a,*] Jan Bukartyk, MS,[a] Paul C. Timm, MBA,[b] Paul A. Friedman, MD,[a] Erik K. St. Louis, MD, MS,[c] Zachi I. Attia, PhD,[a] Virend K. Somers, MD, PhD[a]

ABSTRACT

BACKGROUND Although highly prevalent, obstructive sleep apnea (OSA) remains largely underdiagnosed, thus justifying the need for high-performing screening tools.

OBJECTIVES The authors sought to develop a machine learning-powered algorithm to identify OSA from the 12-lead electrocardiogram (ECG), a routine clinical test.

METHODS A retrospective population of 11,299 patients who completed sleep evaluation and underwent 12-lead ECG at Mayo Clinic were included. OSA was defined as an apnea-hypopnea index ≥5. A deep convolutional neural network model was constructed to detect OSA from the ECG (artificial intelligence [AI]-ECG). Predictive performance of the algorithm in the total sample and separately in males and females was evaluated using the receiver-operating characteristic curve with area under the curve (AUC).

RESULTS The population consisted of 7,170 patients with OSA and 4,129 controls (53.7% males, median [Q1-Q3] of age 58 [47-68] years). The AUC of the AI-ECG model for identification of OSA in the test sample was 0.80 (95% CI: 0.77-0.83), with accuracy, sensitivity, and specificity of 73.7%, 77.0%, and 68.6%, respectively. The model showed better discriminatory performance in females (AUC: 0.82; 95% CI: 0.79-0.86) than in males (AUC: 0.73; 95% CI: 0.68-0.78; $P < 0.001$). Sensitivity analyses showed that the predictive abilities of the model were robust across different time intervals and even when including ECG recordings manifesting cardiac abnormalities.

CONCLUSIONS Our AI-ECG model demonstrated good diagnostic performance as an ECG-based screening tool for OSA in a clinical population, particularly among females. Incorporating this algorithm in medical practice may enable widespread low-cost screening for OSA, optimizing early diagnosis and therapy. (JACC Adv. 2025;4:102139)

From the [a]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota, USA; [b]Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota, USA; and the [c]Division of Pulmonary and Critical Care Medicine, Department of Neurology and Medicine, Mayo Center for Sleep Medicine, Mayo Clinic, Rochester, Minnesota, USA. *These authors contributed equally to this work.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received May 29, 2025; revised manuscript received August 1, 2025, accepted August 12, 2025.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2025.102139

ABBREVIATIONS AND ACRONYMS

AHI = apnea-hypopnea index
AI = artificial intelligence
AUC = area under the curve
ECG = electrocardiogram
LR = likelihood ratio
NPV = negative predictive value
OSA = obstructive sleep apnea
PPV = positive predictive value
PSG = polysomnography
ROC = receiver operating characteristic

Obstructive sleep apnea (OSA) is a highly prevalent sleep disorder, estimated to affect 17% of women and 34% of men in the United States[1] and 936 million adults globally.[2] OSA is a major health hazard, associated with increased risk of several chronic conditions, especially cardiovascular diseases and premature mortality.[3] Nevertheless, despite substantial adverse health effects and related socioeconomic impact, OSA remains largely underrecognized and thus undertreated.

The cornerstone for OSA diagnosis, namely in-laboratory polysomnography (PSG), is burdensome, labor-intensive, and costly. More convenient and inexpensive diagnostic home-based sleep apnea tests are becoming mainstream, but their broad application for screening purposes remains impractical. Several questionnaires and clinical score algorithms have also been developed to identify suspected OSA, with various levels of predictive performance.[4,5] However, while low-burden, these methods typically require access to the patient's medical history, ad hoc data acquisition, including physical measurements, and/or rely on self-reported symptoms. Hence, screening strategies that are objective, automated, and leverage already available information are desirable.

As a result of technological advancements and enhanced computational capabilities, an unprecedented surge in artificial intelligence (AI)-based applications has emerged and has been transforming several health care fields, including sleep medicine. By extracting features from raw PSG signals, AI-enabled technology and machine learning algorithms have demonstrated remarkable performances for automated sleep staging and detection of abnormal events during sleep, including disordered breathing events.[6] To this end, machine learning techniques have been applied to standard signals recorded during PSG to identify OSA, including electroencephalography,[7] respiratory channels,[8,9] and electrocardiography.[10-13] However, these approaches still require extended, overnight recordings.

The twelve-lead electrocardiogram (ECG) is among the most common, fundamental testing performed in individuals utilizing health care services. More than 40 million ECGs are conducted annually at clinic office visits[14] and more than 30 million ECGs are obtained in emergency departments.[15] Since ECG abnormalities associated with OSA can be noted also during wakefulness,[16,17] AI-guided analysis of the standard resting 12-lead ECG may unveil OSA signatures and thus have applicability for OSA screening.

Here, we tested the hypothesis that a deep learning algorithm, implemented using a deep convolutional neural network applied to standard 12-lead ECG (AI-ECG), can be used as a screening tool for OSA in adults. As OSA manifests with distinct sex-dependent characteristics,[18] we also evaluated the diagnostic performance of the model separately in males and females.

METHODS

STUDY POPULATION. We identified consecutive patients who completed clinically indicated PSG at the Mayo Clinic Center for Sleep Medicine from November 16, 2009 through December 5, 2018 (N = 36,655) (Supplemental Figure 1). We excluded patients who were <18 years old (n = 4,372), patients who underwent nondiagnostic PSG (n = 2,237), and inadequate studies (ie, <120 min of total recording time or <60 min of total sleep time, n = 985). We did not include patients who did not provide authorization for research use of their medical records (Minnesota Statute 144.295, n = 1,682). Because most of the patients were referred for sleep evaluation due to a high pretest probability of sleep disturbances, a large proportion of those who underwent PSG was found to have OSA (n = 21,205, 77.4%). Hence, to enrich the sample with controls, we included patients who underwent nocturnal oximetry testing for suspected OSA from January 2, 2004, to December 31, 2014 (N = 86,501) and in whom this test was negative for probable OSA (see Sleep Apnea Assessment below).

From this pooled sample, we then selected patients who underwent 12-lead ECG testing at Mayo Clinic hospitals and clinics in the 3 months prior to PSG/oximetry assessment. If multiple, valid ECG tests were available for a patient within this interval, only the recording closest to the date of OSA assessment was used for analysis. Those with poor-quality ECG recordings (n = 2,864) were discarded prior to analysis. We also excluded tests indicating ECG abnormalities (n = 982) (Supplemental Table 1). The final sample for the primary analysis consisted of 11,299 patients. Patients' demographic and clinical characteristics were abstracted from the electronic health records. The Mayo Clinic Institutional Review Board approved the study.

SLEEP APNEA ASSESSMENT. Patients underwent in-laboratory, attended PSG using the Nicvue system (Nicolet, Inc) as described previously.[19] The majority

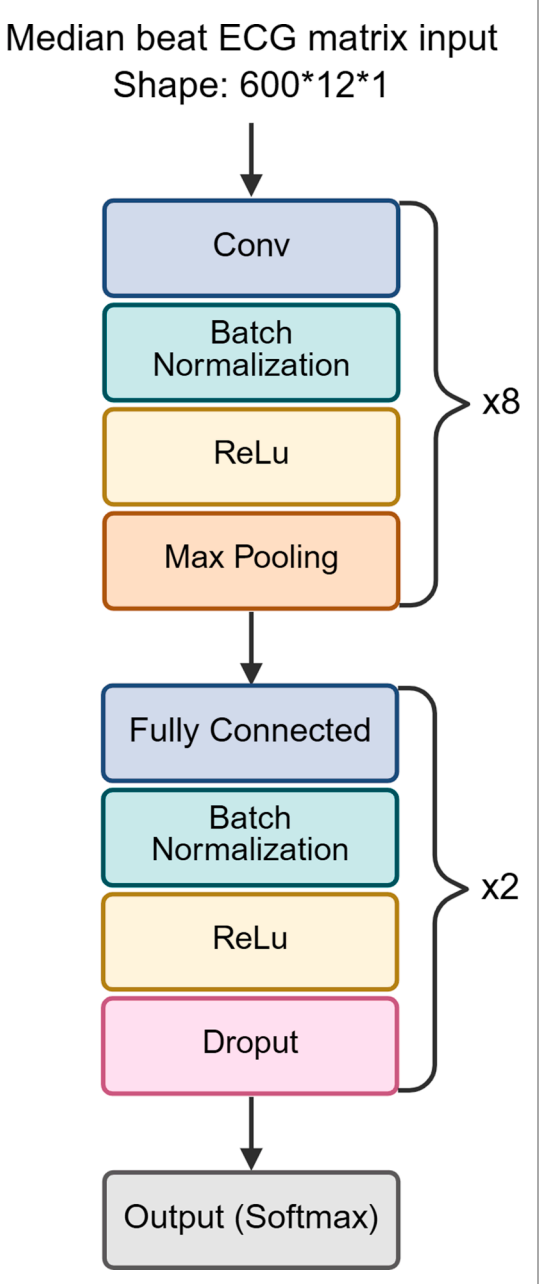
of the PSGs (80.7%) were split-night studies. Sleep studies were scored by registered PSG technicians following recommendations,[20] and confirmed by board-certified sleep physicians. Apnea was defined as a ≥90% decrease in airflow lasting at least 10 seconds. Hypopnea was defined as a ≥30% reduction in airflow lasting at least 10 seconds and with an associated ≥4% drop in peripheral oxygen saturation (Hypopnea Scoring Criteria 1B).[21] The apnea-hypopnea index (AHI) was calculated as the total number of apneas and hypopneas per hour of sleep. Presence of OSA was defined as AHI ≥5 per hour, with OSA severity defined as mild when AHI = 5 to 14.9 per hour, moderate when AHI = 15 to 29.9 per hour, and severe with AHI ≥30 per hour.

Data from portable overnight oximetry (Nonin Medical Inc) were obtained from an independent cohort of patients who were referred for suspected OSA but who did not complete PSG at our institution, ensuring lack of overlap between the oximetry no-OSA cohort and the PSG-proven OSA and no-OSA cohorts. Patients with mean nocturnal SpO_2 ≥95% and 5 or less desaturation events ≥4% on oximetry were regarded as no-OSA and pooled with patients who exhibited AHI <5 at PSG as controls in the study.

AI MODEL DEVELOPMENT. We created a deep convolutional neural network (Figure 1) using the Keras Framework with Tensorflow backend (Google). The network was composed of 6 convolutional layers, each of which was followed by a batch normalization layer and a max pooling layer, and a nonlinear "Relu" activation function. Following the last convolutional layer, the data were fed to a fully connected network with 2 hidden layers with dropout layers and an output layer with "Softmax" activation function. A dropout ratio of 50% was applied along with the batch size of 32 at the epoch count of 30. The median beat of the 10-second, 12-lead ECG was input at either 500 Hz or 250 Hz. ECGs with 250 Hz original sampling were upsampled to 500 Hz using the "Resample" function of the SCIPY python package. Therefore, the 12-lead median beat ECG contained a total of 600 data points for each lead. In addition, we also developed a model using the awake 10-second rhythm ECG. For this, the sampling rate was 500 Hz and the total data points for each lead were 5,000.

MODEL TRAINING AND DATA SPLIT. We trained our AI-ECG model on a cluster with Nvidia GeForce Tesla V100 32Gb graphic processing unit. We randomly partitioned ECGs by patient into training (80%), internal validation (10%), and testing (10%) data sets without overlap of patients. The internal validation data set was used to find the optimal model during training and evaluate the optimal threshold for dichotomizing the neural network output from a probability to a binary decision for accurate calculations. The trained model was then applied directly without any further processing to the test data set. While the outcome used to train the model is binary,

FIGURE 1 Convolutional Neural Network Model Architecture for the Median Beat AI-ECG



Created with Biorender.com. ECG = electrocardiogram.

TABLE 1 Patient Characteristics

	All (N = 11,299)	OSA (n = 7,170)	No OSA (n = 4,129)	*P* Value
Age, y	58 (47-68)	62 (53-71)	49 (37-61)	<0.001
Male	6,072 (53.7)	4,556 (63.5)	1,516 (36.7)	<0.001
White	10,133 (92.9)	6,453 (92.5)	3,680 (93.6)	0.045
Body mass index, kg/m^2	31 (26.9-36)	32.5 (28.7-37.4)	27.3 (23.4-31.8)	<0.001
Hypertension	4,980 (44.1)	3,864 (53.9)	1,116 (27)	<0.001
Diabetes	2,054 (18.2)	1,675 (23.4)	379 (9.2)	<0.001
Ischemic heart disease	2,537 (22.5)	2,024 (28.2)	513 (12.4)	<0.001
Chronic kidney disease	827 (7.3)	688 (9.6)	139 (3.4)	<0.001
Chronic obstructive pulmonary disease	740 (6.5)	473 (6.6)	267 (6.5)	0.79
Heart failure	1,053 (9.3)	863 (12)	190 (4.6)	<0.001
Stroke	431 (3.8)	330 (4.6)	101 (2.4)	<0.001
AHI, event/hour	13 (6-31)	18 (9-37)	2 (1-3)	<0.001
OSA severity				
Mild (AHI 5-14.9)	3,040 (26.9)	3,040 (42.4)	-	-
Moderate (AHI 15-29.9)	1,804 (16)	1,804 (25.2)	-	
Severe (AHI ≥30)	2,326 (20.6)	2,326 (32.4)	-	
Training	9,067 (80.2)	5,784 (80.7)	3,283 (79.5)	0.13
Validation	1,081 (9.6)	687 (9.6)	394 (9.5)	
Testing	1,151 (10.2)	699 (9.7)	452 (10.9)	

Values are median (25th-75th percentiles) or n (%) and compared using the Mann-Whitney *U* test or chi-squared test, respectively.

AHI = apnea-hypopnea index; OSA = obstructive sleep apnea. AHI statistics are calculated only for patients who underwent polysomnography.

the network output is a continuous probability between 0 and 1 that represents the probability of the binary outcome (either case or control).

DATA ANALYSES. Categorical variables are reported as numbers and percentages, and continuous variables are reported as median and 25th-75th percentiles (Q1-Q3) due to skewed distributions. Chi-squared and Mann-Whitney tests were conducted to compare patient characteristics as appropriate.

The performance of the model in discriminating between OSA and non-OSA was assessed by the receiver-operating characteristic curve. We calculated the area under the curve (AUC) with 95% CIs, accuracy, sensitivity, specificity, positive and negative predictive values (NPV/PPV), positive and negative likelihood ratios (LR+/LR-), and F1 score. The optimal probability threshold for correct classification was determined using the Youden index. We calculated the proportion of patients in the sample who would be accurately classified as OSA or controls, as well as false positives and false negatives, using the proposed model. The algorithm's predictive ability was also compared between males and females using *t*-tests. We then tested temporal influences on the performance of the model by using ECGs collected at progressively increasing time intervals before sleep apnea assessment. We conducted a sensitivity analysis to investigate the impact of including in the analysis ECG recordings showing abnormal electrocardiographic findings, thus rejecting only poor-quality recordings. Predictions obtained using an algorithm based on rhythm ECG rather than median ECG were also explored. Last, we explored the model's performance using a 3-lead ECG configuration (leads I, II, and V_5). Statistical analysis was performed using Python version 3.10.11 (Python Software Foundation) and SPSS Statistics version 28.0.1.1 (IBM Corp). A two-sided *P* value <0.05 was regarded as significant for all tests.

RESULTS

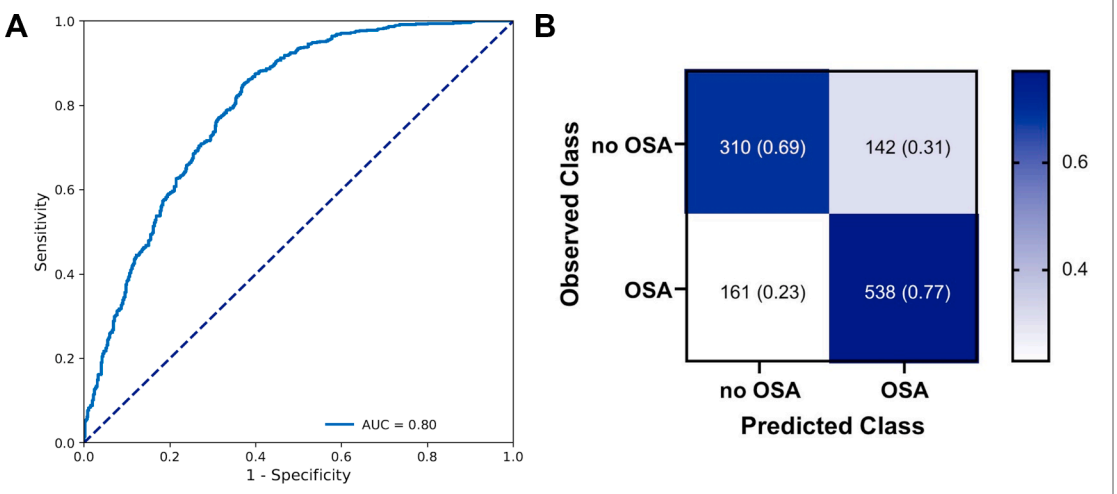
STUDY POPULATION. The sample for the primary analysis consisted of 11,299 patients, of whom 7,170 (63.5%) were OSA patients and 4,129 were free of OSA (36.5%). Among those with OSA, 42.4% of the sample had mild OSA, 25.2% had moderate OSA, and 32.4% had severe OSA. Compared to those without OSA, patients with OSA were older, had higher body mass index, were more likely to be male, non-White and to have medical comorbidities. Detailed characteristics of case and control groups as well as training, validation, and testing sets are reported in **Table 1** and Supplemental Table 2, respectively.

MODEL PREDICTIVE PERFORMANCE. In the overall population, the AUC of the median beat AI-ECG model for detection of OSA on testing was 0.80 (95% CI: 0.77-0.83) (**Figure 2A**, **Table 2**), with corresponding sensitivity and specificity of 77.0% and 65.8%, respectively. The optimal AHI prediction threshold was 0.72. PPV and NPV were 79.1% and 68.5%, respectively, and the F1 score was 78.0%. LR+ was 2.45 and LR- was 0.34.

Overall, the algorithm correctly classified 848 out of 1,151 (73.7%) individuals included in the test sample (**Figure 2B**). Of 161 OSA cases incorrectly identified as controls (14.0% of the test sample, 23% of the OSA group), 76 were mild OSA (27% of all mild cases), 41 were moderate OSA (23.2% of all moderate cases), and 44 were severe OSA (18.3% of all severe cases). The predicted probabilities across different OSA severities were similar, as illustrated in **Figure 3**.

SEX DIFFERENCES. Characteristics of male and female patients included in the sample are summarized in Supplemental Table 3. Lower prevalence of OSA and milder sleep disorder severity were noted in females, along with younger age and lower frequency of concurrent diseases compared with males. In sex-stratified analysis, we observed significantly greater AUC for OSA classification in females (0.82; 95% CI:

FIGURE 2 AI ECG Model Performance for OSA Detection



(A) Receiver-Operating Characteristic (ROC) curve and corresponding AUC (95% CI) values of median beat AI ECG model for OSA prediction. (B) Confusion matrix for patient classification of OSA Using AI ECG. Data are reported as count (proportions). Darker colors indicate higher proportions. OSA = obstructive sleep apnea; other abbreviation as in Figure 1.

0.79-0.86) than in males (0.73; 95% CI: 0.68-0.78) (Figure 4, Table 2). Sensitivity and specificity were 67.9% and 79.1% in females and 81.9% and 52.5% in males, respectively. LR+ was also higher in females than in males (3.25 vs 1.72). Confusion matrices are displayed in Supplemental Figure 2.

ADDITIONAL ANALYSES. To investigate the robustness of the model, we assessed its performance using ECG recordings obtained up to 1 year, 3 years, and 5 years prior to the sleep evaluation (Supplemental Table 4). Results showed the AUC for OSA detection progressively decreased, albeit modestly, with increasing time intervals (1 year: 0.77; 95% CI: 0.75-0.80; 3 years: 0.74; 95% CI: 0.72-0.77; 5 years: 0.74; 95% CI: 0.71-0.76) (Supplemental Table 5). Sensitivities ranged from 69.9% to 72.5% and specificities ranged from 64.4% to 67.8%.

Including ECG recordings with evidence of abnormal electrocardiographic findings (such as acute myocardial infarction, supraventricular tachycardia, or atrial fibrillation) had relatively little effect on the discriminatory power of the algorithm, resulting in an AUC of 0.76 (95% CI: 0.73-0.79) (Supplemental Table 6). Lastly, the AI-ECG model using rhythm ECG instead of median beat ECG showed a similar though nominally worse diagnostic power, with AUC for OSA of 0.76 (95% CI: 0.73-0.79) (Supplemental Table 7). Accuracy, sensitivity, and specificity were 71.9%, 78.9%, and 60.5%, respectively. Similar results were obtained when using a 3-lead ECG configuration (AUC: 0.75; 95% CI: 0.72-0.78) (Supplemental Table 7).

DISCUSSION

In the present study, we developed a deep learning-enabled algorithm to detect OSA from clinically indicated 12-lead ECGs and showed that this approach can be used as an economical, valuable, and practical screening for OSA (**Central Illustration**).

Our AI-ECG model showed a sensitivity of 77.0% and a specificity of 68.6% for OSA detection. Compared to our method, meta-analytical evidence[5]

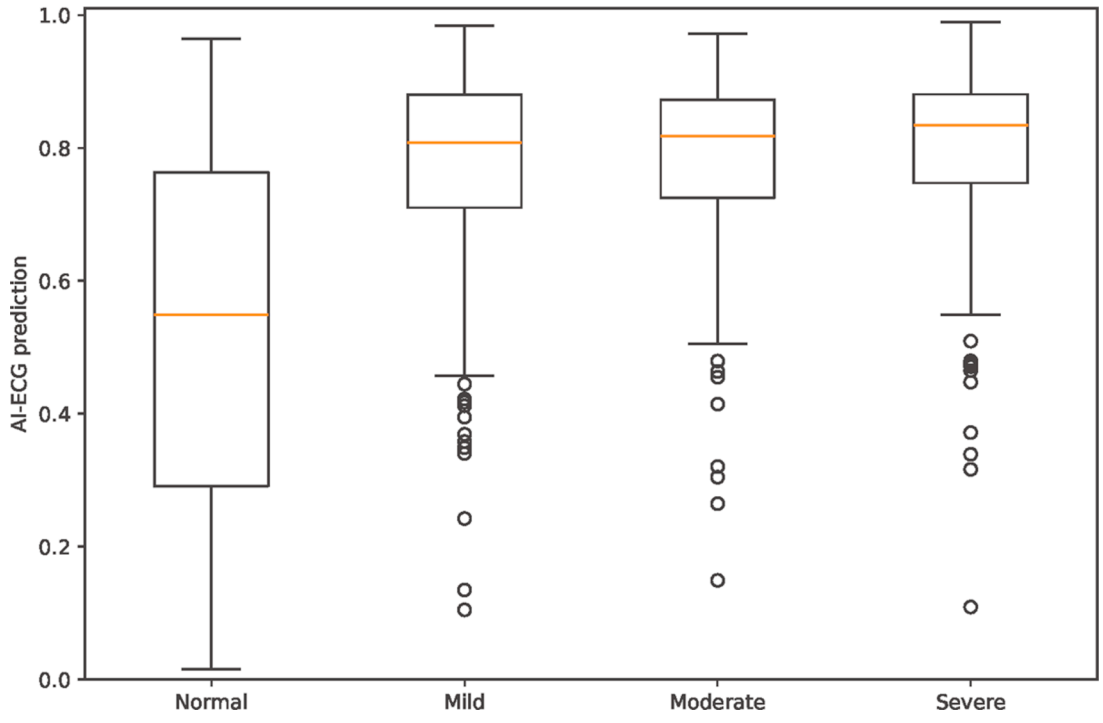
TABLE 2 Performance of the Median Beat AI-ECG Model for OSA Prediction in the Total Sample and Separately in Males and Females

	AUC (95% CI)	Accuracy %	Sensitivity %	Specificity %	PPV, %	NPV, %	F1, %	LR+	LR-
Total sample	0.80 (0.77-0.83)	73.7	77.0	68.6	79.1	65.8	78.0	2.45	0.34
Males	0.73 (0.68-0.78)	73.6	81.9	52.5	81.4	53.4	81.6	1.72	0.34
Females	0.82 (0.79-0.86)[a]	73.8	67.9	79.1	74.6	73.2	71.1	3.25	0.41

[a] $P < 0.001$ for sex comparison.

AI = artificial intelligence; AUC = area under the curve; LR = likelihood ratio; NPV = negative predictive value; PPV = positive predictive value; other abbreviation as in Table 1.

FIGURE 3 Box Plot of AI ECG-Predicted Probabilities of Different OSA Severity



Patients with a probability >0.5 are predicted to have OSA according to the AI ECG model. Data are median and 25th-75th percentiles. Circles indicate outliers. Abbreviations as in Figures 1 and 2.

shows higher sensitivity for the Berlin questionnaire[22] and the STOP-BANG questionnaire[23] in sleep clinic populations, but at the expenses of much lower specificity, suggesting that these instruments may result in unwarranted sleep studies. A similar comparison can be drawn for recently developed clinical scores, such as the NoSAS.[24] Of note, our model outperforms diagnostic capabilities of the Epworth Sleepiness Scale,[25] another instrument for OSA screening which is frequently used despite its modest-to-low sensitivity. Future studies exploring screening diagnostic performance of combined subjective sleep instruments together with AI-ECG should be considered since a combined approach leveraging both symptoms of sleep disturbance and objective AI-ECG data might further improve both sensitivity and specificity for OSA detection.

OSA is a common sleep disorder,[1,2] and its prevalence will presumably continue to rise as a consequence of the aging population[26] and the upward North American population trends in obesity risk.[27] Because of its detrimental impact on health, it is paramount to develop cost-effective screening strategies that clinicians can adopt promptly, without added burden, to identify patients at high risk of OSA. Screening questionnaires and scores, such as those mentioned above, use various combinations of demographics, anthropometrics, clinical characteristics, and/or symptoms to predict OSA risk. Typically, this information is self-reported and thus subject to bias. While taking direct measurements (eg, height and weight, neck circumference) upon screening may enhance accuracy, this gain should be considered against the resulting extended time for screening completion, and being possibly time-consuming for providers who would be collecting these measures. Other machine learning-guided predictive models use data mining of electronic health records to abstract OSA-related variables and identify high-risk patients.[28,29] However, this approach is conditional upon having such data available in the clinical records: thus it may be inapplicable to new patients.

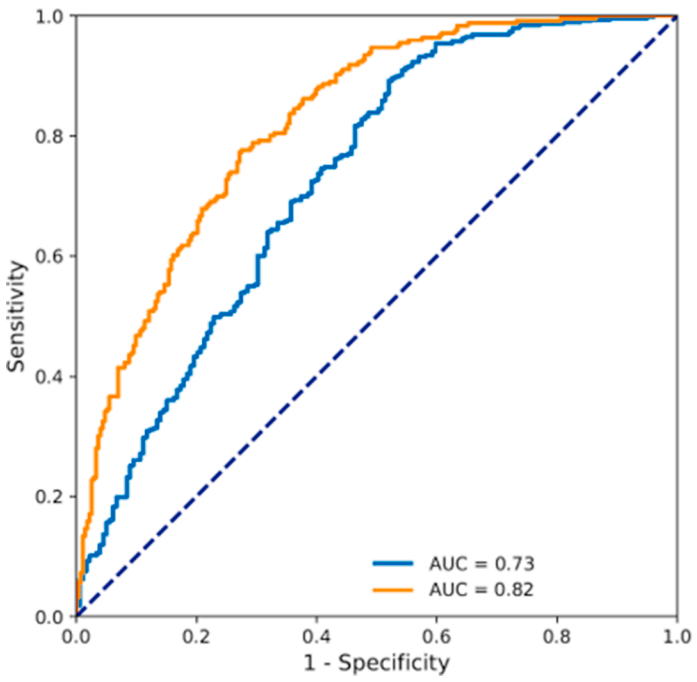
Our ECG-based screening has several advantages over these other methods. By requiring only the 12-lead ECG signal to generate predictions, it is not

affected by potential sources of error ensuing from subjective reporting of clinical information or physical measurements. Since it is entirely automated, it places no additional burden on clinicians nor on patients. This screening tool can be deployed into clinical practice, integrated with the health record system and alert providers of OSA risk during routine ECG testing, yielding actionable information. Patients who are identified as being at high risk can be triaged and referred to sleep medicine for diagnostic confirmation and treatment. In turn, prompt disease management could significantly improve functioning and outcomes in OSA patients. Of note, our exploratory analysis showed promising performance of our model even when using 3-lead ECG, extending its applicability to other clinical settings. However, ad hoc training of the current algorithm on clinically relevant reduced ECG configurations would likely improve its discriminative power and be required prior to generalization.

It is important to point out that our model was similarly predictive of OSA when we included patients with prior myocardial infarction, arrhythmias, or implantable cardiac devices as noted on the ECG, suggesting robust performance of our algorithm. Because OSA is frequently comorbid with heart disease and contributes to unfavorable prognosis in these patients,[3] OSA screening of patients who present with cardiac symptoms, rhythm disturbances, or other cardiac abnormalities evident on 12-lead ECG can be streamlined using our approach.

It is also worth noting that the discriminative abilities of our model appear to be superior in females relative to males, with significantly higher AUC for OSA detection in females. Higher NPV also suggests better ability to rule out elevated risk of OSA in females. While sensitivity of available OSA screening questionnaires and scores is generally poor in women,[30,31] our AI ECG model showed acceptable sensitivity among females. This could be attributed to the focus of traditional screening tools on symptoms and physical attributes of OSA that may be more closely associated with this sleep disorder in males than in females, such as obesity and sleepiness.[32,33] It is now well accepted that the clinical presentation of OSA is sex-dependent, with women more likely than men to present with somewhat atypical symptoms such as fatigue, insomnia, and morning headaches.[18] To our knowledge, no prior studies have reported on sex differences in the performance of AI-based algorithms applied to ECG for OSA detection. We speculate that the OSA features captured by our method from raw ECG signals are better predictors in female subjects, suggesting that the effects of OSA on the heart may be more manifest in females rather than males. This inference would be consistent with emerging evidence that OSA may be more deleterious in females, despite the condition being far more severe in males.[34,35]

FIGURE 4 Receiver-Operating Characteristic (ROC) Curve and Corresponding AUC (95% CI) Values of Median Beat AI ECG Model for Prediction of OSA in Males (blue line) and Females (orange line)

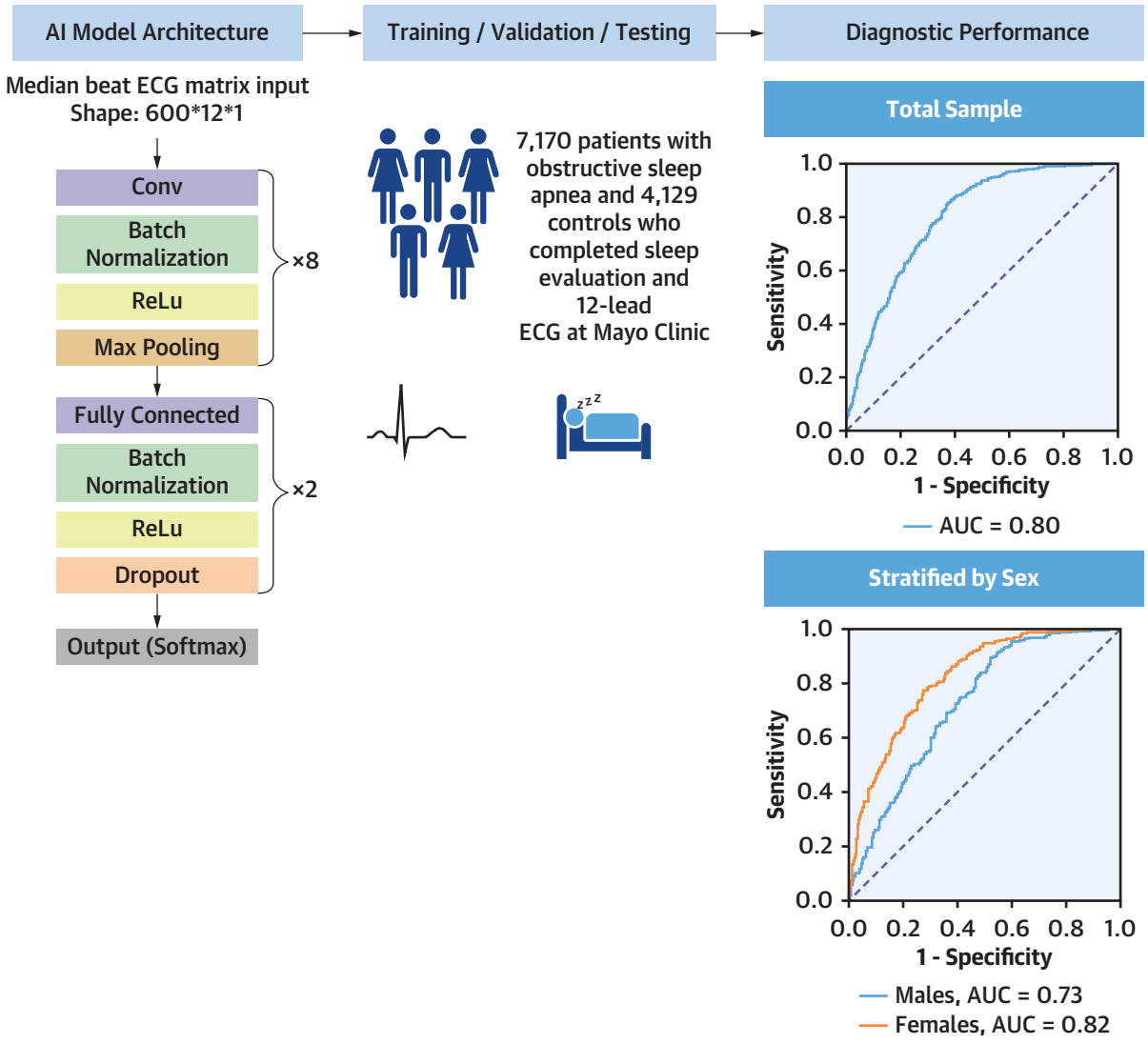


Abbreviations as in Figures 1 and 2.

Within the rapidly evolving AI landscape, deep learning has been dominating AI-enabled applications in health care. Artificial neural networks applied to ECG signals have demonstrated remarkable value as predictive tools for a multitude of cardiovascular diseases, including atrial fibrillation[36] and hypertrophic cardiomyopathy.[37] Neural networks have also successfully detected OSA signatures in the ECG in prior work.[10-13] Notably, those investigations used recordings obtained during sleep, when OSA episodes are overtly manifest and cause pronounced perturbations in the ECG. Because OSA-induced cardiac aberrations are evident also during wakefulness,[16,17] these features can be captured and analyzed to predict disease, as demonstrated by our findings.

Biologic plausibility of our approach and findings are supported by: first, that OSA is well established as a cause of cardiac disease; second, that the predictive

CENTRAL ILLUSTRATION **Convolutional Neural Network Model Development and Performance for OSA Detection**



Covassin N, et al. JACC Adv. 2025;4(10):102139.

Created with Biorender.com. Abbreviations as in Figures 1 and 2.

value of the ECG increases with closer proximity to the actual diagnostic sleep study; third, that predictive value is greater prior to rather than after the sleep study, the latter of which would include patients on therapy; and last, that the ECG is better at identifying OSA in females, who are emerging as potentially at greater cardiovascular risk from OSA than males.

Further potential applications of our AI-ECG screening modality for OSA include wearables. As utilization of digital health technologies rapidly increases, advances in solutions providing high-quality cardiac monitoring can be powered by our tool to alert users if at high risk of OSA. However, because those sensors are worn under ambulatory, real-life conditions, the predictive performance of the model in the context of contaminated data needs to be tested.

STUDY LIMITATIONS. Our findings should be interpreted in the context of some study limitations. Due to the retrospective nature of the study, clinical ECG

tests were not conducted concurrently with laboratory testing for OSA ascertainment. Changes in physiological features (eg, body mass index) over time could have affected the performance of the algorithm. To mitigate such effects and potential confounding of OSA treatment, we selected only ECGs conducted within 3 months prior to the sleep testing. However, in sensitivity analysis, we showed that the predictive abilities of our method remain similar across all considered time intervals, up to 5 years prior to OSA evaluation. Using a clinical sample who underwent evaluation for suspected OSA might have introduced a selection bias, as these patients had high pretest probability of the disease, thus potentially limiting the model's generalizability as a population-level screening instrument. Some performance metrics such as NPV and PPV are highly dependent on disease prevalence and would vary in populations with different prevalence of OSA. Accordingly, validation of our algorithm is needed in unselected primary care settings, where the prevalence of OSA would mirror that of the general population. Relatedly, to minimize the imbalance between numbers of cases and controls, we included, as non-OSA controls, patients who had a negative PSG (ie, AHI <5) as well as those who had a negative nocturnal oximetry. Although pooling patients assessed with 2 different diagnostic techniques might have caused misclassification, we used a very low ODI threshold to minimize the risk of inaccurately categorizing false negatives at the oximetry as controls. Further validation in a more racially diverse sample is also warranted.

CONCLUSIONS

Here we show that the standard, medically ubiquitous 12-lead ECG can serve as a convenient, cost-effective screening solution for OSA in clinical settings. Our ECG-based, machine learning-powered approach has scalability potential for broad OSA screening, hence facilitating earlier diagnosis and impactful treatment of this common and deleterious sleep disorder.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This research was supported by the Mayo Foundation. Dr St. Louis has received research funding from the American Academy of Sleep Medicine Foundation (unrelated to present work) and Spark, Inc; and has served on the Sleep Number Scientific Advisory Board. Dr Somers has served as a consultant for Lilly, Apnimed, Jazz Pharmaceuticals, Mineralys, and iRhythm; and is on the Sleep Number Scientific Advisory Board. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Virend K. Somers, Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota 55902, USA. E-mail: somers.virend@mayo.edu.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Despite its high and rising prevalence and adverse health implications, OSA remains largely underdiagnosed and thus undertreated. This can be partly ascribed to the burden, availability and costs associated with conventional OSA diagnostics. Accordingly, development of cost-effective, practical, and scalable screening solutions for OSA is warranted.

TRANSLATIONAL OUTLOOK 1: Our machine learning-guided algorithm applied to standard 12-lead ECG showed accuracy, sensitivity, and specificity for OSA detection of 73.7%, 77.0% and 68.6%, respectively—a diagnostic performance similar or even superior to that of traditional OSA screening questionnaires or scores. In addition, we found that the performance of the model was better in females and consistent across different time intervals and cardiac diagnoses.

TRANSLATIONAL OUTLOOK 2: This nonobtrusive, practical approach leverages a routine, inexpensive clinical test, does not require collection of any additional information, and may be easily incorporated into electronic health systems to streamline OSA screening and referral in clinical settings.

REFERENCES

1. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol*. 2013;177:1006-1014.

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

3. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive sleep apnea and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*. 2021;144:e56-e67.

4. Gamaldo C, Buenaver L, Chernyshev O, et al. Evaluation of clinical tools to screen and assess for obstructive sleep apnea. *J Clin Sleep Med*. 2018;14:1239-1244.

5. Chiu H-Y, Chen P-Y, Chuang L-P, et al. Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: a bivariate meta-analysis. *Sleep Med Rev*. 2017;36:57-70.

6. Bandyopadhyay A, Goldstein C. Clinical applications of artificial intelligence in sleep medicine: a sleep clinician's perspective. *Sleep Breath*. 2023;27:39-55.

7. Kang C, An S, Kim HJ, Devi M, Hwang S, Lee HW. Age-integrated artificial intelligence framework for sleep stage classification and obstructive sleep apnea screening. *Front Neurosci*. 2023;17:1059186.

8. Choi SH, Yoon H, Kim HS, et al. Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. *Comput Biol Med*. 2018;100:123-131.

9. Wang E, Koprinska I, Jeffries B. Sleep apnea prediction using deep learning. *IEEE J Biomed Health Inform*. 2023;27:5644-5654.

10. Chang H-Y, Yeh C-Y, Lee C-T, Lin C-C. A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram. *Sensors*. 2020;20:4157.

11. Erdenebayar U, Kim YJ, Park J-U, Joo EY, Lee K-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput Methods Programs Biomed*. 2019;180:105001.

12. Li K, Pan W, Li Y, Jiang Q, Liu G. A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal. *Neurocomputing*. 2018;294:94-101.

13. Urtnasan E, Park J-U, Joo E-Y, Lee K-J. Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network. *J Med Syst*. 2018;42:1-8.

14. Santo L, Kang K. National Hospital Ambulatory Medical Care Survey: 2019 National Summary Tables. National Ambulatory Medical Care Survey; 2023.

15. Cairns C, Kang K. National Hospital Ambulatory Medical Care Survey: 2019 Emergency Department Summary tables. National Ambulatory Medical Care Survey; 2022.

16. Corotto PS, Kang H, Massaro B, et al. Obstructive sleep apnea and electrocardiographic p-wave morphology. *Ann Noninvasive Electrocardiol*. 2019;24:e12639.

17. Shamsuzzaman A, Amin RS, van der Walt C, et al. Daytime cardiac repolarization in patients with obstructive sleep apnea. *Sleep Breath*. 2015;19:1135-1140.

18. Bonsignore MR, Saaresranta T, Riha RL. Sex differences in obstructive sleep apnoea. *Eur Respir Rev*. 2019;28.

19. Covassin N, Lu D, St Louis EK, et al. Sex-specific associations between daytime sleepiness, chronic diseases and mortality in obstructive sleep apnea. *Front Neurosci*. 2023;17:1210206.

20. Iber C, Ancoli-Israel S, Chesson AL, Quan SF. *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications*. Westchester, IL: American Academy of Sleep Medicine; 2007.

21. Berry R, Brooks R, Gamaldo C, et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: American Academy of Sleep Medicine; 2013. Version 2.0.2.

22. 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;131:485-491.

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

24. 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.

25. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14:540-545.

26. He W, Goodkind D, Kowal PR. *An aging world: 2015*. Washington, DC: United States Census Bureau; 2016.

27. Ward ZJ, Bleich SN, Cradock AL, et al. Projected US state-level prevalence of adult obesity and severe obesity. *N Engl J Med*. 2019;381:2440-2450.

28. Huang W-C, Lee P-L, Liu Y-T, Chiang AA, Lai F. Support vector machine prediction of obstructive sleep apnea in a large-scale Chinese clinical sample. *Sleep*. 2020;43:zsz295.

29. Ramesh J, Keeran N, Sagahyroon A, Aloul F. Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning. *Healthcare*. 2021;9:1450.

30. Bauters FA, Loof S, Hertegonne KB, Chirinos JA, De Buyzere ML, Rietzschel ER. Sex-specific sleep apnea screening questionnaires: closing the performance gap in women. *Sleep Med*. 2020;67:91-98.

31. Zhang Z, Yang D, Wang H, Liu X. Effects of age and sex on the performance of the NoSAS score as a screening tool for obstructive sleep apnea: a hospital-based retrospective study in China. *Sleep Breath*. 2021;25:1407-1417.

32. Cairns A, Poulos G, Bogan R. Sex differences in sleep apnea predictors and outcomes from home sleep apnea testing. *Nat Sci Sleep*. 2016;8:197-205.

33. Earl DE, Lakhani SS, Loriaux DB, Spector AR. Predictors of moderate to severe obstructive sleep apnea: identification of sex differences. *Sleep Breath*. 2019;23:1151-1158.

34. Roca GQ, Redline S, Claggett B, et al. Sex-specific association of sleep apnea severity with subclinical myocardial injury, ventricular hypertrophy, and heart failure risk in a community-dwelling cohort: the atherosclerosis risk in communities-sleep heart health study. *Circulation*. 2015;132:1329-1337.

35. Strausz S, Havulinna AS, Tuomi T, et al. Obstructive sleep apnoea and the risk for coronary heart disease and type 2 diabetes: a longitudinal population-based study in Finland. *BMJ Open*. 2018;8:e022752.

36. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394:861-867.

37. Ko W-Y, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75:722-733.

KEY WORDS artificial intelligence, convolutional neural network, electrocardiogram, machine learning, obstructive sleep apnea

APPENDIX For supplemental figures and tables, please see the online version of this paper.