

Original Paper

Machine Learning–Based Multidimensional Oximetry for Obstructive Sleep Apnea Screening: Development and External Validation

Xuanyu Qian^{1,2*}, BS; Haitong Luo^{1*}, MBBS; Rong Ding^{3*}, MSc; Tianming Gao⁴, MBBS; Haoan Wang¹, MBBS; Pengliang Wu¹, MBBS; Ning Li^{1,5}, MD

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

²Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai, China

³Department of Respiratory and Critical Care Medicine, Taizhou Fourth People's Hospital, Taizhou, Jiangsu Province, China

⁴Department of Respiratory and Critical Care Medicine, Pingliang Municipal Hospital of Traditional Chinese Medicine, Pingliang, Gansu Province, China

⁵Institute of Respiratory Diseases, Shanghai Jiao Tong University School of Medicine, Shanghai, China

*these authors contributed equally

Corresponding Author:

Ning Li, MD

Department of Respiratory and Critical Care Medicine

Ruijin Hospital

Shanghai Jiao Tong University School of Medicine

No. 197 Rui Jin 2nd Road

Shanghai, 200025

China

Phone: 86 21 64370045

Email: drningbaby@163.com

Abstract

Background: Obstructive sleep apnea (OSA) affects nearly one billion people globally and poses a substantial public health threat. Effective and accessible methods for OSA risk identification are urgently needed.

Objective: This study aims to develop and externally validate a machine learning model derived from multi-parameter pulse oximetry (SpO₂) for OSA screening, and to evaluate its performance, interpretability, and robustness across sex and age subgroups.

Methods: Of 4156 screened participants, 2195 underwent polysomnography (internal cohort) and 446 received home sleep apnea testing (external cohort). Eight SpO₂-derived parameters, including oxygen desaturation index (ODI), hypoxic burden (HB), and ST90 (percentage of sleep time with SpO₂ < 90%), were used to construct models. Six machine learning algorithms were trained, with F_1 -score as the primary metric and area under the curve as the secondary metric. Model interpretability was assessed using Shapley additive explanations and intrinsic feature importance scores.

Results: Nonlinear parameter-risk relationships were observed between oximetry indices and OSA probability. The 4-parameter ODI-HB-MinSpO₂-ST90 model achieved optimal performance (F_1 -score = 0.9516, area under the curve = 0.9879), surpassing all single-parameter models. Shapley additive explanations analysis identified ODI, HB, and MinSpO₂ as key predictors. The ODI-HB-MinSpO₂-MeanSpO₂ configuration demonstrated superior performance in female and younger subgroups, whereas the ODI-HB-MinSpO₂-ST90 model remained optimal for male and older participants. Categorical boosting outperformed other algorithms across multiple metrics and remained robust in both subgroup and external validation analyses.

Conclusions: The multi-parameter oximetry model based on the categorical boosting algorithm provides a simple and accurate tool for OSA screening. Sex- and age-stratified strategies can further enhance its clinical applicability.

(*JMIR Med Inform 2026;14:e80384*) doi: [10.2196/80384](https://doi.org/10.2196/80384)

KEYWORDS

obstructive sleep apnea; pulse oximetry; machine learning; multi-parameter oximetry; screening; CatBoost; categorical boosting

Introduction

Obstructive sleep apnea (OSA) affects nearly one billion individuals globally [1], and untreated OSA significantly increases the comorbidity burden and the risk of motor vehicle crashes [2]. Although polysomnography (PSG) remains the diagnostic gold standard, its high cost and operational complexity limit widespread accessibility [3,4]. Current screening tools, such as the STOP-BANG questionnaire or single physiological parameters, demonstrate limited diagnostic accuracy, with reported area under the receiver operating characteristic curve (AUC) values ranging from 0.55 to 0.83 [5,6]. Therefore, developing robust OSA screening tools using readily available physiological parameters remains imperative.

The pathophysiology of OSA is characterized by recurrent upper airway collapse, resulting in intermittent nocturnal hypoxia. Pulse oximetry-derived metrics, including the oxygen desaturation index (ODI), percentage of sleep time with $SpO_2 < 90\%$ (ST90), and minimum oxygen saturation (Min SpO_2), offer accessible alternatives to PSG [7], yet they reflect only a single dimension of nocturnal desaturation, thereby limiting their clinical utility [8]. ODI quantifies the frequency of desaturation events and correlates with PSG-derived apnea-hypopnea index (AHI), but does not capture hypoxic duration or desaturation depth [9]. ST90 reflects cumulative hypoxic burden (HB) but cannot distinguish between distinct hypoxic patterns, such as single prolonged versus multiple brief desaturations [9]. Min SpO_2 identifies the instantaneous nadir but does not characterize cumulative hypoxic exposure [6]. The novel integrated HB metric, which combines desaturation depth, duration, and frequency, demonstrates superior predictive performance for OSA-related comorbidities compared with AHI and ODI [8], though direct comparisons with conventional metrics within the same datasets remain scarce [10]. Entropy and frequency-domain analyses of SpO_2 complexity can capture dynamic nocturnal fluctuations overlooked by traditional metrics [11,12]. However, most existing studies evaluate parameters in isolation or focus on linear associations, leaving multidimensional feature integration and nonlinear relationships among parameters largely unexplored [4,13,14]. Thereby, multi-parameter models leveraging complementary oximetric indices may yield improved robustness for OSA screening [4,15,16].

Machine learning (ML) holds great potential for OSA diagnosis. Although deep learning models (eg, OxiNet) enable

high-precision AHI estimation [17], their “black box” nature compromises clinical interpretability and raises clinical skepticism, thus limiting real-world utility [18,19]. Traditional ML algorithms show inconsistent performance across cohorts, with support vector machines (SVM) and random forests (RF) demonstrating variable performance [20,21]. Extreme gradient boosting (XGBoost) models for moderate-to-severe OSA exhibit limited accuracy (sensitivity 72.5%, specificity 62.8%) [22]. Least squares boosting for AHI estimation illustrates the benefits of ensemble approaches but does not resolve generalization issues between community and clinical cohorts [23]. Recent evidence suggests that categorical boosting (CatBoost) is superior for OSA classification, outperforming XGBoost, light gradient boosting machine (LightGBM), and RF in several studies [24,25], yet its application to oximetry-based OSA screening has not been evaluated.

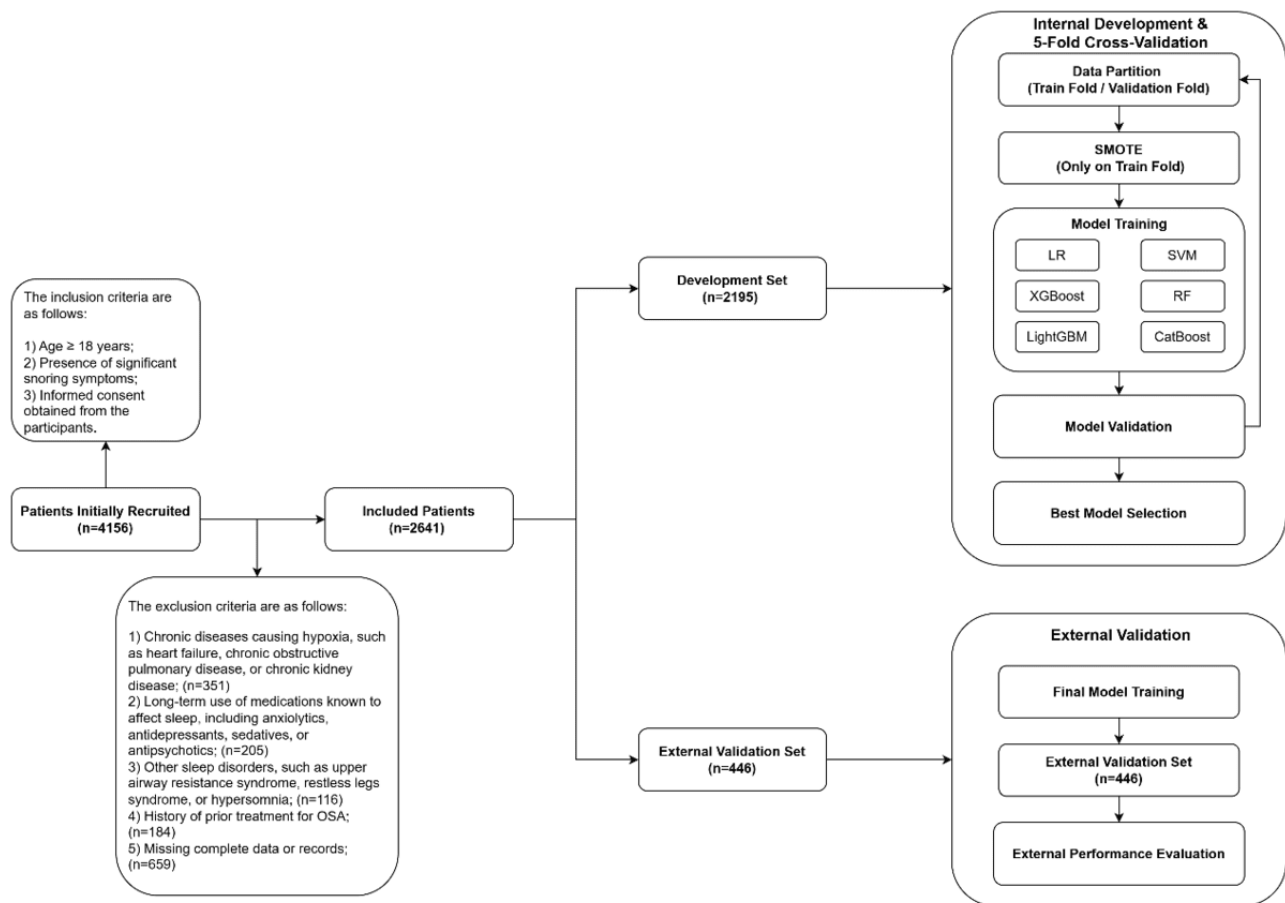
Our study has three main aims: (1) to develop a parsimonious and robust OSA screening tool by evaluating multidimensional oximetric parameters using ML; (2) to validate model generalizability across community and clinical populations in an independent external cohort undergoing home sleep apnea test (HSAT); and (3) to assess performance heterogeneity across sex and age subgroups to inform personalized screening strategies.

Methods

Study Design and Population

We consecutively enrolled adults with suspected OSA who underwent in-laboratory PSG at the Sleep Center of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, between June 2022 and July 2024. During the same period, we additionally recruited community-based participants who underwent HSAT. Inclusion criteria were age ≥ 18 years, prominent snoring, and provision of informed consent. Exclusion criteria included: (1) chronic diseases that may contribute to hypoxemia, such as heart failure, chronic obstructive pulmonary disease, chronic kidney disease; (2) chronic use of medications affecting sleep, including sedative-hypnotics, anxiolytics, antidepressants, and antipsychotics; (3) other concurrent sleep disorders, such as upper airway resistance syndrome, restless legs syndrome, or hypersomnia; (4) prior treatment for OSA; and (5) incomplete data. The participant flowchart is shown in Figure 1.

Figure 1. Flowchart of the overall study. CatBoost: categorical boosting; LightGBM: light gradient boosting machine; LR: logistic regression; OSA: obstructive sleep apnea; PSG: polysomnography; RF: random forest; SMOTE: synthetic minority over-sampling technique; SVM: support vector machine; XGBoost: extreme gradient boosting.



Ethical Considerations

The study protocol complied with the Declaration of Helsinki and was approved by the Ruijin Hospital Ethics Committee (approval number: 2018-107). Due to its retrospective design, all data were fully deidentified, and the ethics committee waived the requirement for informed consent. All data were securely stored in accordance with institutional research data management standards, and no compensation was provided to participants.

PSG and HSAT

Participants abstained from sedatives, alcohol, and caffeinated beverages for at least 24 hours prior to the study. In-laboratory PSG was performed using the Alice 6 system (Philips Respironics, Murrysville, PA, USA) with standard monitoring including electroencephalography, submental electromyography, bilateral electrooculography, electrocardiography, pulse oximetry, oronasal airflow via thermistor and pressure transducer, thoracoabdominal effort, snoring, and body position. HSAT used the Alice NightOne device (Philips Respironics, Murrysville, PA, USA) to record nasal airflow, respiratory effort, and fingertip SpO₂. Recordings with more than 4 hours of analyzable data following manual review were considered valid. Two certified sleep specialists independently scored PSG and HSAT data according to the AASM scoring manual [3]: apnea was defined as ≥ 90% airflow reduction for ≥ 10 seconds, and hypopnea as ≥ 30% airflow reduction for ≥ 10 seconds

accompanied by ≥ 4% SpO₂ desaturation. AHI was calculated as the total number of apneas and hypopneas per hour of sleep, and OSA was defined as AHI ≥ 5 events/hour.

Definition and Calculation of Pulse Oximetry Parameters

Summary of Signal Processing

During PSG and HSAT, SpO₂ signals were collected at a sampling rate of 500 Hz and down-sampled to 1 Hz for computational efficiency. Eight parameters were extracted to quantify different aspects of nocturnal hypoxemia:

Mean SpO₂ (MeanSpO₂) and Minimum SpO₂ (MinSpO₂)

The average and the lowest SpO₂ values during sleep:

$$\text{MeanSpO}_2 = \frac{1}{n} \sum_{i=1}^n \text{SpO}_{2i}$$

$$\text{MinSpO}_2 = \min(\text{SpO}_{2_1}, \text{SpO}_{2_2}, \dots, \text{SpO}_{2_n})$$

These reflect overall oxygenation status and the most severe desaturation [26].

Oxygen Desaturation Index (ODI)

Number of desaturation events ($\geq 4\%$ drop from baseline) per hour of sleep:

$$ODI = \frac{N_{\Delta \geq 4\%}}{TST}$$

Where TST is the total sleep time (hours). ODI is a key indicator of the frequency of respiratory disturbances and serves as a surrogate for the AHI [27].

T90 and ST90

Total time and percentage of sleep spent with SpO₂ below 90%:

$$T90 = \sum \Delta t_{SpO_2 < 90\%}$$

$$ST90 = \frac{T90}{TST} \times 100\%$$

These quantify cumulative exposure to clinically significant hypoxemia [27].

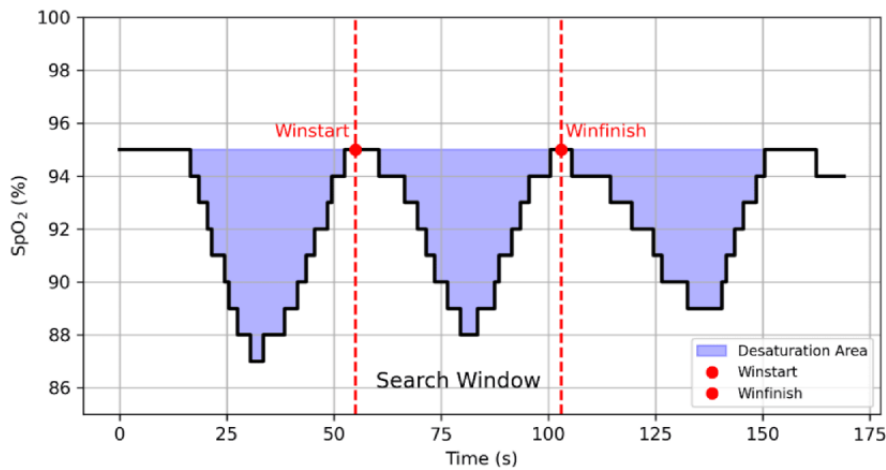
Hypoxic Burden (HB)

The normalized total area under the SpO₂ desaturation curve associated with respiratory events.

$$HB = \frac{1}{TRT} \sum_{i=1}^m AUC_i$$

where AUC_{*i*} is the area of the *i*-th desaturation event identified by the “Trapping Rain Water” algorithm (Figure 2), and TRT is the total recording time. HB integrates frequency, depth, and duration of desaturations, representing total oxygen debt [28].

Figure 2. Hypoxia burden calculation using the rainwater collection algorithm applied to pulse oximetry signals.



Attention Entropy (AttnEn)

AttnEn is a complexity measure of the SpO₂ signal waveform variability [29].

$$AttnEn = - \sum_{i=1}^k P_i \ln P_i$$

Where *P_i* is the distribution of intervals between adjacent local extrema. Higher entropy reflects fragmented, unstable desaturation patterns typical of severe OSA.

Total Spectral Power (TotalPower)

Integrated Lomb–Scargle periodogram power within the ultradian band (0.014-0.035 Hz), corresponding to respiratory cycles of 30-70 seconds.

$$Total Power = 10 \log_{10} \left(\sum_{f=0.014}^{0.035} P_{LS}(f) \right)$$

Elevated power in this band indicates the repetitive oscillatory desaturation dynamics characteristic of OSA [30,31].

The algorithm proceeds as follows: (1) Event Identification: The pulse oximetry (SpO₂) signal is analyzed to detect all local minima (valleys), thereby identifying the nadir (lowest saturation) of each desaturation event. (2) Window Initialization: From each nadir, a bidirectional search is performed to delineate the event window (Win_{start} and Win_{finish}). Boundaries are established at the nearest peaks that recover to $\geq 75\%$ of the preceding peak-to-nadir amplitude. (3) Boundary Refinement: The search window is further adjusted based on the mean event duration to ensure temporal consistency. (4) Baseline Determination: The baseline for each event is defined as the maximum SpO₂ value within the 100-second window preceding the event onset. (5) Area Integration: The under the curve (AUC) for each event is computed by integrating the deficit between the baseline and the SpO₂ signal within the defined window. (6) Hypoxia burden (HB) Calculation: All individual AUCs are summed to obtain the total desaturation area, which is then divided by the total recording time to derive the HB.

Establishment and Validation of ML Models

Data Preprocessing

To mitigate bias from varying feature magnitudes, the data were first standardized via Z-score normalization, transforming each feature to a mean of 0 and SD of 1 using:

$$Z = \frac{x - \mu}{\sigma}$$

where μ and σ represent the mean and SD of the feature. This step ensures stable distance-based computations and gradient optimization. Subsequently, class imbalance was addressed using the synthetic minority over-sampling technique (SMOTE) [32,33]. SMOTE synthesizes minority-class samples by interpolating between an instance x_i and a randomly chosen neighbor $x_{\hat{i}}$ from its k -nearest neighbors.

$$x_{\text{new}} = x_i + \text{rand}(0,1) \times (\hat{x}_i - x_i)$$

Algorithm Introduction

This study evaluates multiple ML models, grouped into three categories: (1) linear and kernel-based models, (2) ensemble learning methods, and (3) gradient boosting decision trees, to balance interpretability with predictive performance. For linear and kernel-based models, logistic regression (LR) is a foundational model for clinical binary classification. It extends linear regression by applying the Sigmoid function to map linear outputs to a probability range between 0 and 1:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i x_i)}}$$

Where P is the predicted probability, β_0 is the bias, β_i are coefficients, and x_i represent input features. Its transparency and low computational cost make it a standard benchmark in medical research [34,35].

SVM constructs an optimal separating hyperplane by maximizing the margin between classes. Its decision function is:

$$w \cdot x + b = 0$$

Where w is the normal vector, x is the input feature, and b is the bias. The model is trained by minimizing $\frac{1}{2} \|w\|^2$ subject to the constraint $y_i (w \cdot x_i + b) \geq 1$, ensuring correct classification with a margin of at least one. SVMs excel in high-dimensional spaces and can capture nonlinear patterns through kernel functions, making them a widely adopted method [36].

For ensemble learning methods, RF is a bagging ensemble method that reduces overfitting by aggregating predictions from multiple decision trees. Each tree is trained on a bootstrap sample of the data and a random subset of features. The final prediction is obtained through majority voting:

$$\hat{y} = \text{Majority Vote}(h_1(x), h_2(x), \dots, h_T(x))$$

where \hat{y} is the final predicted result, $h_t(x)$ denotes the prediction of the t -th tree, and T is the total number of trees. By averaging across trees, RF improves stability and accuracy, making it effective for high-dimensional data and widely used in practice [37].

For gradient boosting decision trees, this kind of method iteratively combines weak learners, typically decision trees, to minimize a regularized objective function:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{j=1}^m \Omega(f_j)$$

Where $l(y_i, \hat{y}_i)$ represents the loss function, $\Omega(f_j)$ controls model complexity; and θ denotes the parameters.

Three prominent variants, including XGBoost, LightGBM, and CatBoost, share this framework but differ in optimization and implementation: XGBoost uses second-order gradient approximation and explicit regularization, offering high precision and efficiency, especially with structured or sparse data [38,39]. LightGBM uses a leaf-wise growth strategy with gradient-based sampling and feature bundling, enabling faster training on large-scale datasets [40]. CatBoost is optimized for categorical features, using ordered target statistics and symmetric trees to prevent prediction shift and effectively handle high-dimensional categorical variables [41].

Modeling Process

The modeling pipeline followed a 2-stage design: internal development with cross-validation followed by independent external validation (Figure 1). In the internal phase, a cohort of 2195 subjects was preprocessed and evaluated using 5-fold cross-validation. To prevent data leakage, SMOTE was applied exclusively to the training folds, with validation sets retaining the original class distribution. Six ML algorithms were trained under fixed random seeds to ensure reproducibility, and hyperparameters are detailed in Table 1. Model selection was based on the average performance across validation folds. The best-performing model was subsequently retrained on the full internal dataset (n=2195) without SMOTE to preserve the original data distribution. The selected model's generalization ability was then assessed on an independent external cohort (n=446). Performance on this external set reflects the model's robustness for real-world OSA screening.

Table 1. Hyperparameters of the 6 machine learning models for obstructive sleep apnea screening.

Model	Hyperparameters
SVM ^a	'C': 1.0, 'gamma': 'scale', 'kernel': 'rbf'
RF ^b	'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100
LR ^c	'C': 1.0, 'penalty': 'l2', 'tol': 1e ⁻⁴
XGBoost ^d	'learning_rate': 0.3, 'reg_lambda': 1, 'n_estimators': 100, 'booster': 'gbtree'
LightGBM ^e	'learning_rate': 0.1, 'n_estimators': 100, 'boosting_type': 'gbdt'
CatBoost ^f	'learning_rate': 0.03, 'n_estimators': 100, 'loss_function': 'Logloss', 'l2_leaf_reg': 3

^aSVM: support vector machine.

^bRF: random forest.

^cLR: logistic regression.

^dXGBoost: extreme gradient boosting.

^eLightGBM: light gradient boosting machine.

^fCatBoost: categorical boosting.

Model Evaluation Metrics

Predictive performance was quantified using accuracy, sensitivity, specificity, F_1 -score, AUC, positive predictive value (PPV), and negative predictive value. The specific formulas are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{PPV} = \frac{TP}{TP+FP}$$

$$\text{Negative predictive value (NPV)} = \frac{TN}{TN+FN}$$

$$F_1\text{-score} = 2 \cdot \frac{\text{PPV} \cdot \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}}$$

Where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. Given the class imbalance in the clinical cohort, the F_1 -score was selected as the primary evaluation metric because it balances PPV and sensitivity (recall). In imbalanced clinical settings, AUC may overestimate performance by reflecting overall discriminability while masking poor sensitivity to the minority class. Unlike the threshold-independent AUC, the F_1 -score directly captures misclassification costs for minority samples, thereby ensuring robust diagnostic accuracy across classes. AUC is reported as a complementary measure of overall discriminative ability [42].

Statistical Analysis

All analyses were conducted using Python (version 3.11; Python Software Foundation). Continuous variables are presented as median and IQR, and categorical variables as frequency and percentage. The Anderson–Darling test was used to assess normality. Group differences were evaluated with the Kruskal–Wallis H test, followed by Dunn's post-hoc test (significance threshold $P < .05$). To examine linear and nonlinear associations between continuous predictors and the binary outcome, restricted cubic spline (RCS) regression was fitted within an LR framework. Likelihood-ratio tests compared RCS models against linear specifications, and spline curves were used to visualize dose-response relationships. To further interpret model predictions, Shapley additive explanations (SHAP) were used to quantify the contribution of each feature. Finally, stratified analyses by sex and age were conducted for these oximetry parameters.

Results

Characteristics of Study Participants

Among 4156 screened participants, 2641 were included in the final analysis: 2195 undergoing PSG and comprised the internal development cohort, and 446 undergoing HSAT and formed the external validation cohort (Figure 1). The internal cohort consisted of 943 non-OSA and 1252 OSA participants. Compared with the non-OSA group, the OSA group was significantly older, a higher male proportion, experienced more frequent hypoxic episodes, and had longer hypoxic durations. The external cohort comprised 76 non-OSA and 370 OSA participants. These OSA patients displayed higher AHI, ODI, ST90, T90, and HB values alongside lower oxygen saturation, yet they were younger than non-OSA participants, with no significant between-group difference in sex distribution. Demographic and clinical characteristics are summarized in Table 2. Violin plots (Figure 3) revealed a higher median age (60.0 vs 45.0 years) and more severe nocturnal hypoxemia in the external validation cohort, underscoring distinct disease severity and physiological profiles between the 2 cohorts. These

differences provide a robust foundation for validating the diverse clinical scenarios. generalizability of the multi-parameter oximetry model across

Table 2. Baseline characteristics of non-obstructive sleep apnea and obstructive sleep apnea patients in the internal development and external validation cohorts.

Characteristics	Internal development cohort				External cohort			
	All (n=2195)	Non-OSA ^a (n=953)	OSA (n=1242)	P value	All (n=446)	Non-OSA (n=76)	OSA (n=370)	P value
Age (years), median (IQR)	45.00 (36.00-57.00)	45.00 (35.00-57.00)	46.00 (37.00-57.00)	.001	60.0 (45.00-69.00)	63.00 (49.75-71.25)	58.00 (44.00-68.00)	<.001
Male, n (%)	1651 (75.22)	684 (71.77)	967 (77.86)	<.001	351 (78.70)	55 (72.37)	296 (80.00)	.14
AHI ^b (events/h), median (IQR)	8.20 (2.10-34.10)	1.80 (0.90-3.00)	29.15 (13.83-54.08)	<.001	19.40 (9.03-37.48)	2.25 (0.90-3.58)	24.80 (13.65-42.45)	<.001
ODI ^c (events/h), median (IQR)	12.80 (3.80-37.90)	3.30 (1.70-5.40)	33.45 (16.70-58.55)	<.001	23.85 (10.40-45.28)	3.40 (1.18-5.13)	29.00 (17.35-49.53)	<.001
MinSpO ₂ ^d (%), median (IQR)	86.00 (78.00-90.00)	90.00 (88.00-92.00)	79.00 (71.00-85.00)	<.001	82.00 (75.00-86.00)	89.00 (88.00-91.00)	80.00 (72.00-85.00)	<.001
MeanSpO ₂ (%), median (IQR)	95.00 (93.00-96.00)	96.00 (95.00-96.00)	94.00 (92.00-95.00)	<.001	94.00 (92.00-95.00)	95.00 (94.00-97.00)	93.00 (92.00-95.00)	<.001
ST90 ^e (%), median (IQR)	0.44 (0.02-5.12)	0.02 (0-0.11)	3.38 (0.72-14.42)	<.001	2.91 (0.32-13.80)	0.0 (0-0.20)	4.51 (1.12-16.48)	<.001
T90 ^f (minute), median (IQR)	2.00 (0.10-23.70)	0.10 (0.00-0.50)	16.20 (3.5-67.10)	<.001	14.50 (1.53-65.52)	0.00 (0.00-0.10)	22.20 (5.55-78.9)	<.001
HB ^g (%·min/h), median (IQR)	3.90 (0.90-16.40)	0.70 (0.20-1.70)	13.20 (5.50-36.80)	<.001	51.83 (20.22-112.84)	6.29 (2.32-9.70)	65.97 (35.07-138.08)	<.001
AttnEn ^h , median (IQR)	2.19 (1.78-2.84)	1.74 (1.54-1.97)	2.70 (2.25-3.37)	<.001	5.87 (5.62-6.10)	6.10 (5.99-6.25)	5.82 (5.57-6.01)	<.001
TotalPower ⁱ (dB), median (IQR)	38.17 (35.83-40.81)	37.61 (35.62-40.64)	38.52 (36.10-41.10)	<.001	46.95 (45.56-49.16)	45.06 (44.81-45.40)	47.33 (46.11-50.28)	<.001

^aOSA: obstructive sleep apnea.

^bAHI: apnea-hypopnea index.

^cODI: oxygen desaturation index.

^dMinSpO₂: minimal SpO₂.

^eST90: percentage of sleep time with SpO₂ < 90%.

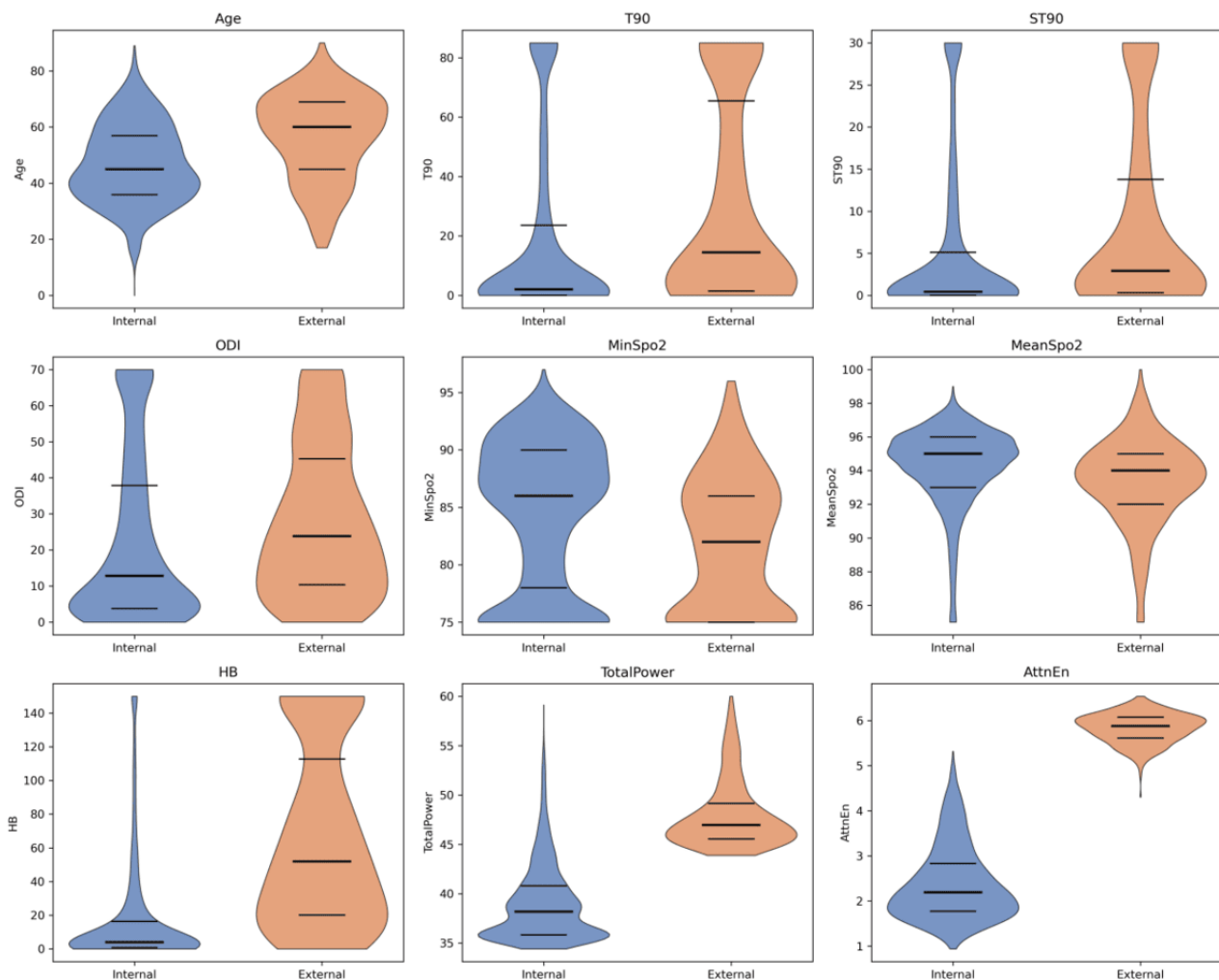
^fT90: total sleep time spent with SpO₂ < 90%.

^gHB: hypoxia burden.

^hAttnEn: attention entropy.

ⁱTotalPower: integrated power from power spectral density estimates in the 14-35 mHz frequency band.

Figure 3. Comparison of baseline characteristics between the internal development and external validation cohorts. Violin plots comparing baseline characteristics between the internal development cohort (blue) and external validation cohort (orange). Each plot depicts the kernel density estimate, with bold horizontal lines representing medians and thin lines indicating IQRs. AttnEn: attention entropy; HB: hypoxia burden; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; ST90: percentage of sleep time with SpO₂ < 90%; TotalPower: integrated power from power spectral density estimates in the 14-35 mHz frequency band.



Performance of Single-Parameter Oximetry Models

We evaluated the predictive performance of 8 OSA-related oximetry parameters using 6 ML algorithms. Given the class imbalance between OSA and non-OSA groups in the internal cohort, the F_1 -score was selected as the primary metric to balance precision and recall, with AUC used to assess overall discriminative performance [43]. Substantial heterogeneity was observed in model performance, with F_1 -scores ranging from 0.5332 to 0.9269 and AUC values from 0.5660 to 0.9808. Notably, ODI and HB exhibited the strongest discriminative ability. Figure 4A summarizes the top 4 single-parameter oximetry models ranked by F_1 -score. The SVM model achieved optimal performance for ODI (F_1 -score = 0.9269, AUC = 0.9712), and the LightGBM model performed best for HB (F_1 -score = 0.9043, AUC = 0.9590). By contrast, MeanSpO₂ and TotalPower showed comparatively weaker discriminative capacity, with F_1 -score of 0.7073 (LR model) and 0.6713 (CatBoost model), respectively.

Beyond the linear association of MinSpO₂, all other oximetry parameters exhibited nonlinear relationships with OSA risk ($P < .001$), accounting for the heterogeneous predictive performance across indicators. Strong predictors, including ODI, HB, T90, and ST90, exhibited steep dose-response curves with pronounced threshold effects (Figure 5). For instance, ODI showed a rapid risk escalation at lower values followed by a plateau, thereby providing distinct decision boundaries that enhanced the model's discriminative ability and optimized F_1 -scores. Conversely, weaker predictors exhibited contrasting profiles: MeanSpO₂ showed shallow gradients within the clinically critical 88%-92% range, resulting in classification ambiguity, whereas TotalPower displayed marked variability with widened 95% CIs at higher values, indicating substantial noise that limited predictive utility (Figure 5). Given the complex nonlinear patterns of most key predictors, traditional linear regression models fail to capture these critical features.

Figure 4. Heatmap of F1-scores for multi-parameter oximetry models across 6 machine learning algorithms. (A) single-parameter; (B-D) combinations of 2, 3, and 4 parameters, respectively. The top 4 F1-scores are shown for each model configuration, with darker colors representing higher classification performance. CatBoost: categorical boosting; HB: hypoxia burden; LightGBM: light gradient boosting machine; LR: logistic regression; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; OSA: obstructive sleep apnea; RF: random forest; ST90: percentage of sleep time with SpO₂ <90%; SVM: support vector machine; T90: total sleep time spent with SpO₂ < 90%; XGBoost: extreme gradient boosting.

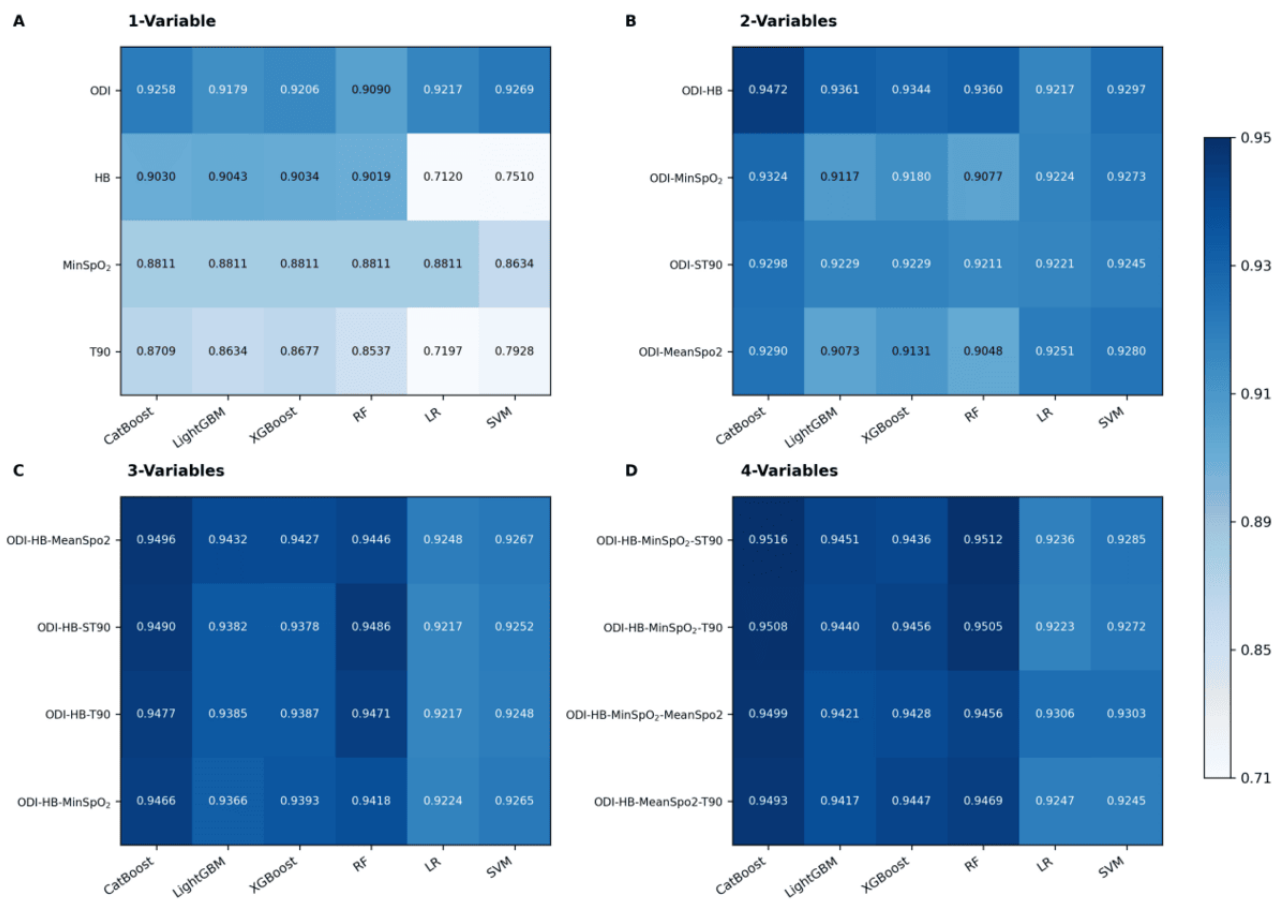
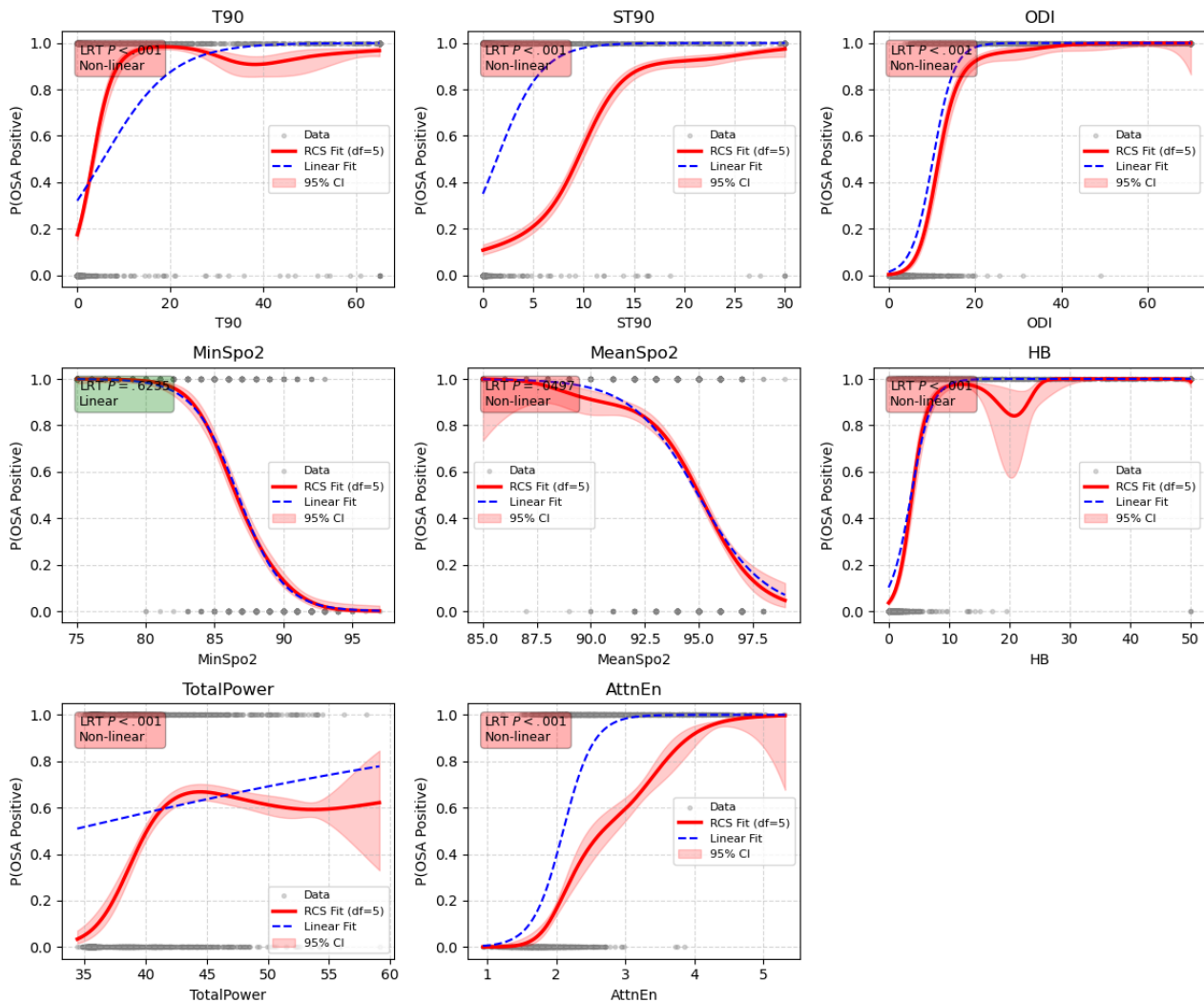


Figure 5. RCS curves showing associations between oximetry parameters and OSA risk. The analysis was performed on the internal development cohort (n=2195). The solid red lines indicate RCS fits with 5 degrees of freedom, and the red shaded areas indicate the 95% CIs. The blue dashed lines represent the linear fit for comparison. The y-axis represents the predicted probability of OSA. The P values were derived from Likelihood Ratio Tests to evaluate nonlinearity (nonlinearity: $P < .05$, red boxes; linear: $P \geq .05$, green boxes). The gray dots (top and bottom) represent individual data distributions for OSA-positive and OSA-negative participants, respectively. AttnEn: attention entropy; HB: hypoxia burden; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; OSA: obstructive sleep apnea; RCS: restricted cubic spline; ST90: percentage of sleep time with SpO₂ <90%; T90: total sleep time spent with SpO₂ < 90%; TotalPower: integrated power from power spectral density estimates in the 14-35 mHz frequency band.



Predictive Performance of Multi-Parameter Oximetry Models

We constructed and evaluated multi-parameter oximetry models, including 28 dual-, 56 triple-, and 70 quadruple-parameter combinations, with top-performing models illustrated in Figure 4B-D. Among the dual-parameter models, the CatBoost-trained ODI-HB model achieved optimal performance (F_1 -score = 0.9472, AUC = 0.9865; Table 3, Figure 4B). The ODI-HB-MinSpO₂ model performed best in the triple-parameter

category (F_1 -score = 0.9496, AUC = 0.9869; Table 3, Figure 4C), whereas the quadruple-parameter ODI-HB-MinSpO₂-ST90 model attained the highest overall discriminative ability (F_1 -score = 0.9516, AUC = 0.9879), significantly outperforming single-parameter oximetry models (Table 3, Figure 6). CatBoost demonstrated consistent superiority across all evaluation metrics (Table 3). Notably, adding 5 or more oximetry parameters yielded only marginal gains, underscoring the importance of selecting informative and complementary features rather than an indiscriminate increase in input dimensionality.

Table 3. Comparison of machine learning algorithms for obstructive sleep apnea screening using multi-parameter oximetry.

Feature sets and machine learning model	AUC ^a	F ₁ -score	Accuracy	Sensitivity	Specificity	PPV ^b	NPV ^c
ODI-HB^{d,e}							
CatBoost ^f	0.9865	0.9472	0.9408	0.9412	0.9402	0.9537	0.9253
LightGBM ^g	0.9280	0.9361	0.9280	0.9332	0.9213	0.9396	0.9143
XGBoost ^h	0.9812	0.9344	0.9262	0.9300	0.9213	0.9392	0.9104
RF ⁱ	0.9794	0.9360	0.9280	0.9316	0.9234	0.9409	0.9129
LR ^j	0.9809	0.9217	0.9134	0.8881	0.9496	0.9586	0.8655
SVM ^k	0.9774	0.9297	0.9226	0.9066	0.9434	0.9545	0.8863
ODI-HB-MeanSpO₂^l							
CatBoost	0.9869	0.9496	0.9435	0.9420	0.9454	0.9575	0.9265
LightGBM	0.9848	0.9432	0.9367	0.9332	0.9413	0.9540	0.9165
XGBoost	0.9831	0.9427	0.9358	0.9372	0.9339	0.9487	0.9205
RF	0.9803	0.9446	0.9380	0.9348	0.9423	0.9550	0.9179
LR	0.9816	0.9248	0.9180	0.8921	0.9517	0.9604	0.8720
SVM	0.9796	0.9267	0.9194	0.9018	0.9423	0.9536	0.8813
ODI-HB-MinSpO₂-ST90^m							
CatBoost	0.9879	0.9516	0.9458	0.9444	0.9475	0.9592	0.9296
LightGBM	0.9862	0.9451	0.9385	0.9388	0.9381	0.9520	0.9227
XGBoost	0.9842	0.9436	0.9367	0.9380	0.9349	0.9496	0.9212
RF	0.9856	0.9512	0.9453	0.9444	0.9465	0.9585	0.9299
LR	0.9811	0.9236	0.9162	0.8970	0.9412	0.9525	0.8760
SVM	0.9815	0.9285	0.9212	0.9050	0.9423	0.9537	0.8844

^aAUC: area under the receiver operating characteristic curve.

^bPPV: positive predictive value.

^cNPV: negative predictive value.

^dODI: oxygen desaturation index.

^eHB: hypoxia burden.

^fCatBoost: categorical boosting.

^gLightGBM: light gradient boosting machine.

^hXGBoost: extreme gradient boosting.

ⁱRF: random forest.

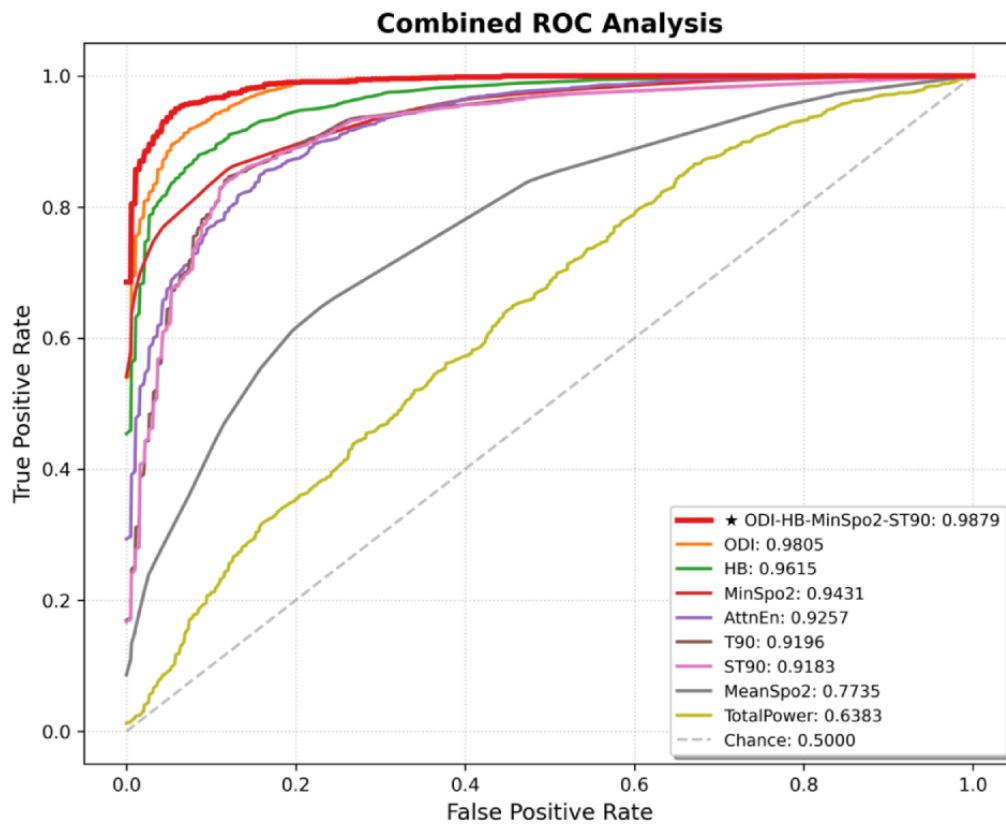
^jLR: logistic regression.

^kSVM: support vector machine.

^lMinSpO₂: minimal SpO₂.

^mST90: percentage of sleep time with SpO₂ < 90%.

Figure 6. ROC curves of the optimal 4-parameter oximetry model versus single-parameter oximetry models for OSA screening. AUC was used to quantify model discrimination, with values closer to 1 indicating better predictive ability. AttnEn: attention entropy; AUC: area under the receiver operating characteristic curve; HB: hypoxia burden; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; OSA: obstructive sleep apnea; ROC: receiver operating characteristic; ST90: percentage of sleep time with SpO₂ < 90%; T90: total sleep time spent with SpO₂ < 90%; TotalPower: integrated power from power spectral density estimates in the 14-35 mHz frequency band.



Stratified Analysis by Sex and Age

Subgroup analyses revealed significant performance variations across demographics. In the male subgroup, the optimal model (ODI-HB-MinSpO₂-ST90) achieved an F_1 -score of 0.9460 and an AUC of 0.9853, with CatBoost outperforming other algorithms (Table 4, Figure 7A). In the female subgroup, the best-performing combination was ODI-HB-MinSpO₂-MeanSpO₂ (F_1 -score = 0.9543, AUC = 0.9919; Table 4, Figure 7B), suggesting sex-specific differences in OSA-related hypoxic

patterns. In the age-stratified analysis, the older subgroup demonstrated superior overall performance (F_1 -score = 0.9398-0.9701, AUC = 0.9913-0.9933) with ODI-HB-MinSpO₂-ST90 as the optimal model, whereas the younger subgroup exhibited stable but slightly lower performance (F_1 -score = 0.9163-0.9467, AUC = 0.9774-0.9863), favoring ODI-HB-MinSpO₂-MeanSpO₂ (Figure 7C-D). Across all subgroups, CatBoost maintained consistently superior classification performance (Table 4).

Table 4. Performance of optimal predictive models for obstructive sleep apnea screening across sex and age subgroups in the internal development cohort.

Feature Sets	Subgroup ^a	AUC ^b	F ₁ -score	Accuracy	Sensitivity	Specificity	PPV ^c	NPV ^d
ODI-HB-Min-SpO ₂ -ST90 ^{e,f,g,h}	Male	0.9853	0.9460	0.9376	0.9338	0.9438	0.9587	0.9102
ODI-HB-Min-SpO ₂ -MeanSpO ₂	Female	0.9919	0.9543	0.9541	0.9527	0.9554	0.9572	0.9532
ODI-HB-Min-SpO ₂ -ST90	Older (≥ 60 years)	0.9942	0.9701	0.9657	0.9664	0.9647	0.9741	0.9552
ODI-HB-Min-SpO ₂ -MeanSpO ₂	Younger (< 60 years)	0.9863	0.9467	0.9404	0.9384	0.9429	0.9552	0.9224

^aAll subgroup models used CatBoost as the optimal classifier.

^bAUC: area under the receiver operating characteristic curve.

^cPPV: positive predictive value.

^dNPV: negative predictive value.

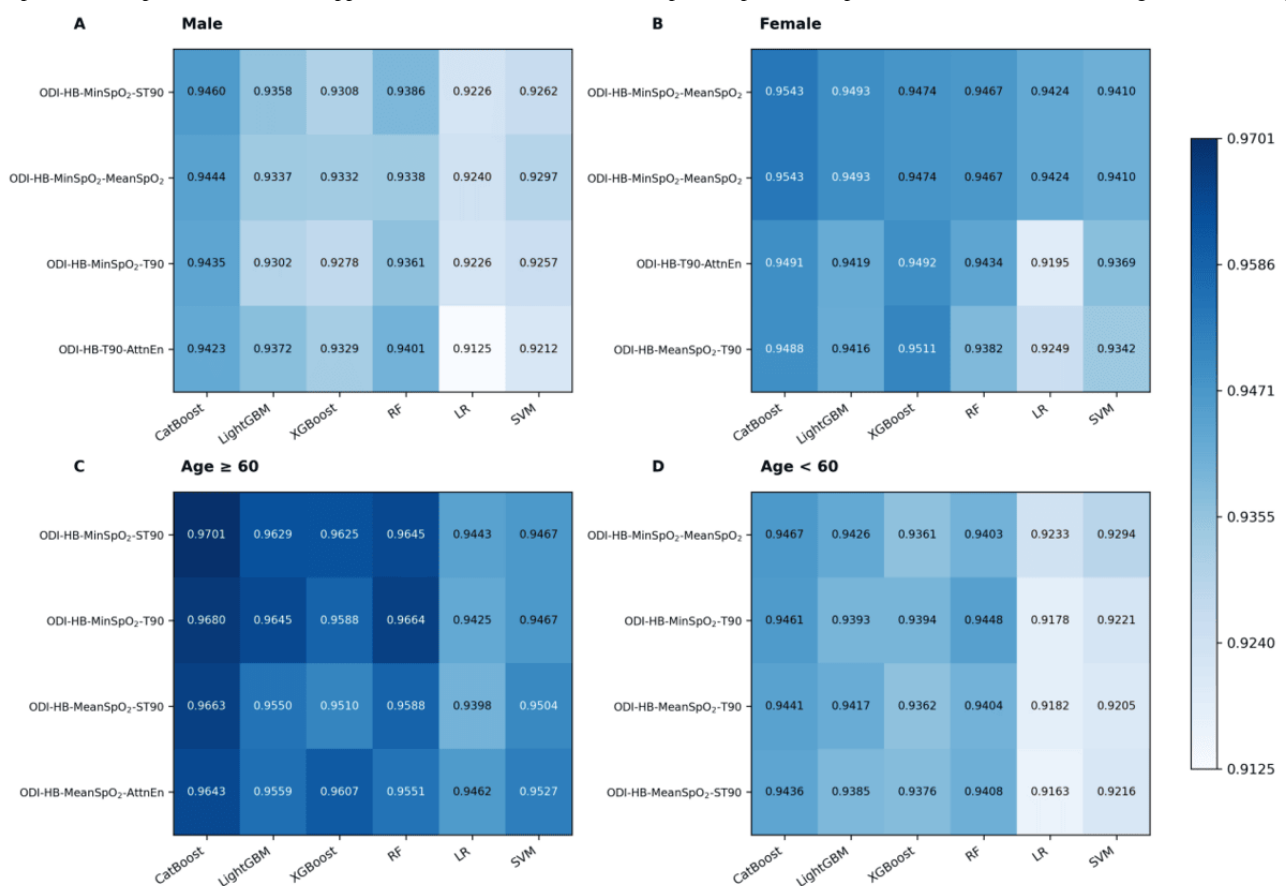
^eODI: oxygen desaturation index.

^fHB: hypoxia burden.

^gMinSpO₂: minimal SpO₂.

^hST90: percentage of sleep time with SpO₂ < 90%.

Figure 7. Heatmap of F1-scores for 4-parameter oximetry models across age and sex subgroups. (A) Male subgroup; (B) female subgroup; (C) older subgroup (≥ 60 years); (D) younger subgroup (< 60 years). The top 4 F1-scores are displayed for each subgroup, with darker colors indicating superior classification performance. AttnEn: attention entropy; CatBoost: categorical boosting; HB: hypoxia burden; LightGBM: light gradient boosting machine; LR: logistic regression; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; OSA: obstructive sleep apnea; RF: random forest; ST90: percentage of sleep time with SpO₂ < 90%; SVM: support vector machine; T90: total sleep time spent with SpO₂ < 90%; XGBoost: extreme gradient boosting.

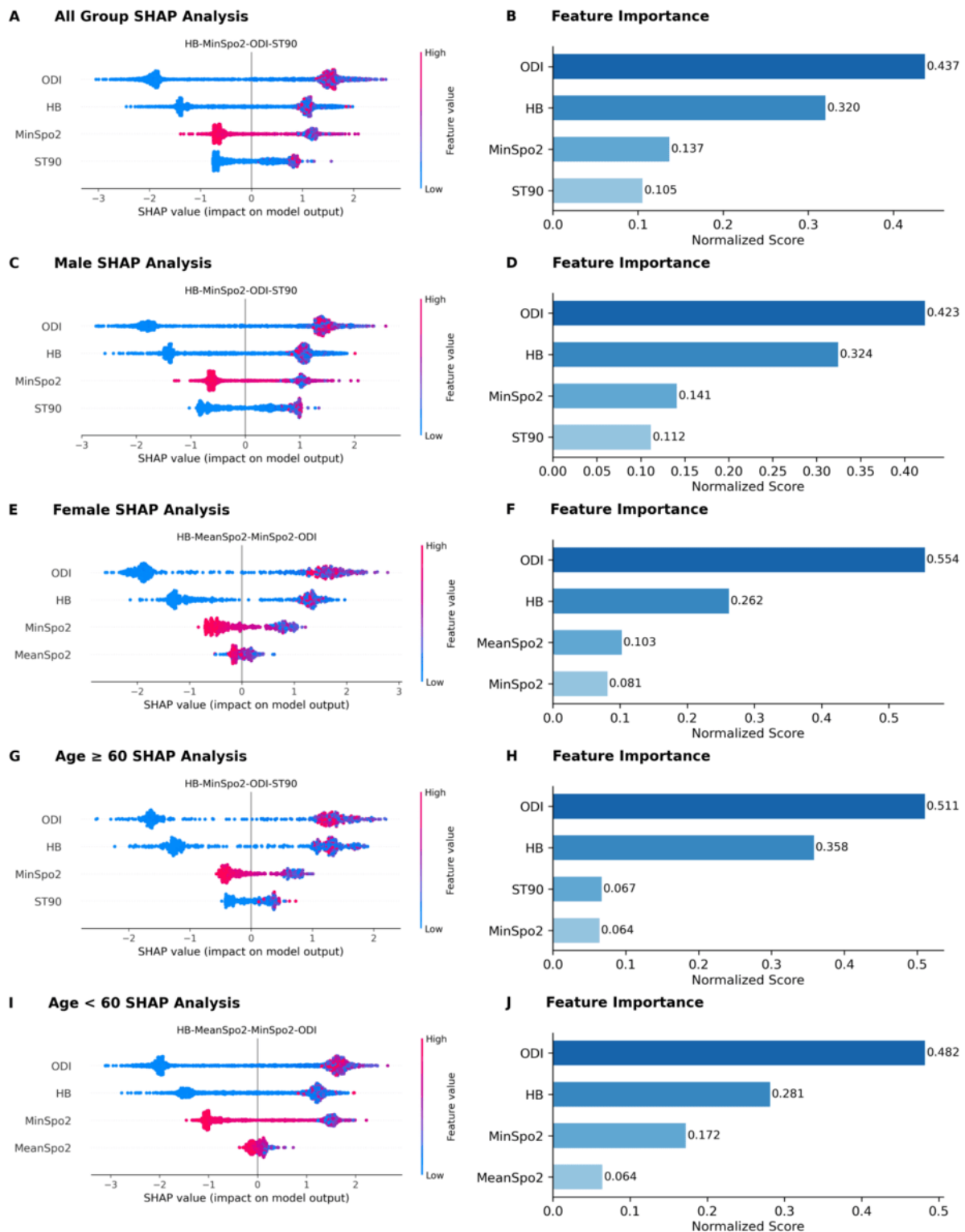


Model Interpretability

To elucidate the predictive mechanisms of the optimal 4-parameter model, we integrated SHAP analysis with normalized feature importance scores. SHAP values quantified each feature's marginal contribution and revealed nonlinear relationships between oximetry parameters and OSA risk (Figure 8A), while normalized scores reflected relative contribution weights (Figure 8B). In the internal cohort, ODI, HB, and MinSpO₂ emerged as the top 3 predictors, with importance scores of 0.437, 0.320, and 0.137, respectively (Figure 8B).

Subgroup analyses revealed heterogeneous contribution patterns across sex and age strata (Figure 8C-J). Notably, male and older subgroups showed consistent dominance of ODI, HB, MinSpO₂, and ST90 (Figure 8C, D, G, H), with ODI exhibiting the highest contribution in the older subgroup (importance score: 0.511). Conversely, younger and female subgroups were characterized by ODI-HB-MinSpO₂-MeanSpO₂, where MeanSpO₂ replaced ST90 as a stronger predictor (Figure 8E,F,I,J). Particularly in females, MeanSpO₂ surpassed MinSpO₂ in contribution strength (Figure 8F).

Figure 8. Interpretability analysis of oximetry parameters across sex and age subgroups for OSA screening. (A, C, E, G, I) SHAP summary plots illustrating feature contributions; dot color denotes feature magnitude (red: high, blue: low) and horizontal position indicates the SHAP value. (B, D, F, H, J) Normalized feature importance scores. Results are presented for all participants (A,B), male subgroup (C,D), female subgroup (E,F), older subgroup (≥ 60 years) (G,H), and younger subgroup (< 60 years) (I,J). HB: hypoxia burden; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; OSA: obstructive sleep apnea; SHAP: Shapley additive explanations; ST90: percentage of sleep time with SpO₂ $< 90\%$.



External Validation

To assess generalizability, we tested model performance on an independent external cohort. The CatBoost algorithm demonstrated robust generalizability, achieving an F_1 -score of 0.9667 with single-parameter configurations and maintaining

high performance as oximetry parameter complexity increased (Table 5). Specifically, the optimal 4-parameter oximetry model (ODI-HB-MinSpO₂-ST90) achieved an F_1 -score of 0.9838 and an AUC of 0.9881 (Table 5, Figure 9D), suggesting that the model captures shared OSA pathophysiological features rather than overfitting the internal cohort. Subgroup analysis further

confirmed the robustness of sex- and age-stratified models in the external cohort (Table 6). Sex-optimized models achieved F_1 -scores of 0.9848 (male subgroup: ODI-HB-MinSpO₂-ST90) and 0.9799 (female subgroup: ODI-HB-MinSpO₂-MeanSpO₂),

with AUCs exceeding 0.98 (Figure 9E, F). Age-stratified models similarly achieved F_1 -scores exceeding 0.98 across subgroups (Figure 9G, H). These results validate the excellent generalizability of the CatBoost-based oximetry model for diverse OSA screening applications (Table 6).

Table 5. Performance of multi-parameter oximetry models in external validation. All subgroup models used categorical boosting (CatBoost) as the optimal classifier.

Feature sets	AUC ^a	F_1 -score	Accuracy	Sensitivity	Specificity	PPV ^b	NPV ^c
ODI ^d	0.9877	0.9667	0.9462	0.9405	0.9737	0.9943	0.7708
ODI-HB ^e	0.9861	0.9727	0.9552	0.9622	0.9211	0.9834	0.8333
ODI-HB-MeanSpO ₂ ^f	0.9863	0.9810	0.9686	0.9784	0.9211	0.9837	0.8974
ODI-HB-MinSpO ₂ -ST90 ^g	0.9881	0.9838	0.9731	0.9838	0.9211	0.9838	0.9211

^aAUC: area under the receiver operating characteristic curve.

^bPPV: positive predictive value.

^cNPV: negative predictive value.

^dODI: oxygen desaturation index.

^eHB: hypoxia burden.

^fMinSpO₂: minimal SpO₂.

^gST90: percentage of sleep time with SpO₂ < 90%.

Figure 9. Heatmap of F1-scores in external validation across parameter combinations and demographic subgroups. (A-D) Performance in all participants by parameter complexity: (A) single-parameter, (B) 2-parameter, (C) 3-parameter, and (D) 4-parameter models. (E-H) Performance of 4-parameter oximetry models across subgroups: (E) male, (F) female, (G) older (≥ 60 years), and (H) younger (< 60 years). The heatmap displays the top 4 F1-scores for each parameter combination across 6 machine learning algorithms. Darker colors indicate higher F1-scores, reflecting superior classification performance. AttnEn: attention entropy; CatBoost: categorical boosting; HB: hypoxia burden; LightGBM: light gradient boosting machine; LR: logistic regression; MinSpO₂: minimal SpO₂; ODI: oxygen desaturation index; RF: random forest; ST90: percentage of sleep time with SpO₂ < 90%; SVM: support vector machine; T90: total sleep time spent with SpO₂ < 90%; XGBoost: extreme gradient boosting.

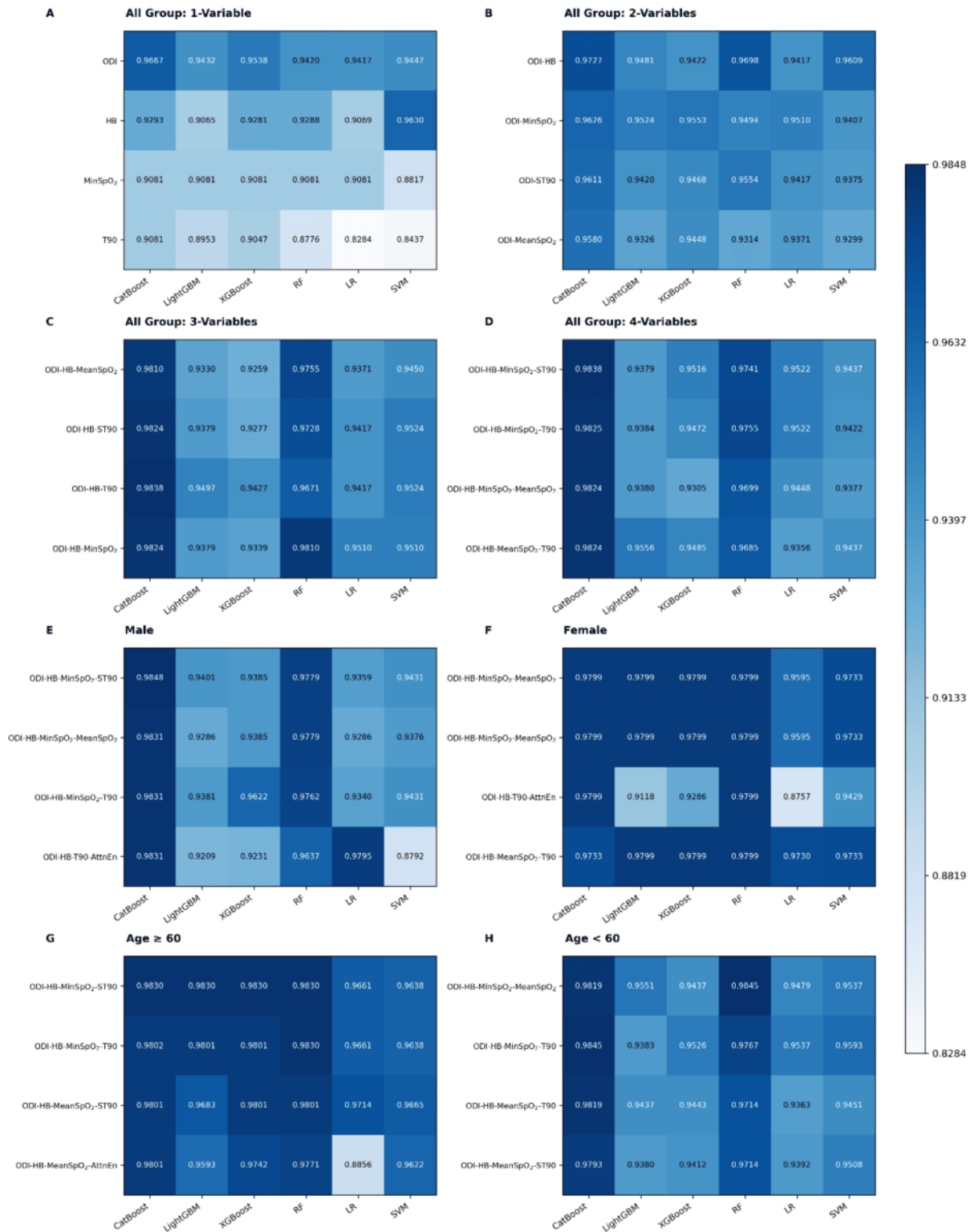


Table 6. Performance of optimal predictive models in external validation across sex and age subgroups. All subgroup models used categorical boosting (CatBoost) as the optimal classifier.

Feature sets	Subgroup	AUC ^a	F_1 -score	Accuracy	Sensitivity	Specificity	PPV ^b	NPV ^c
ODI-HB-Min-SpO ₂ -ST90 ^{d,e,f,g}	Male	0.9854	0.9848	0.9744	0.9831	0.9273	0.9864	0.9107
ODI-HB-Min-SpO ₂ -MeanSpO ₂	Female	0.9916	0.9799	0.9684	0.9865	0.9048	0.9733	0.9500
ODI-HB-Min-SpO ₂ -ST90	Older (≥ 60 years)	0.9855	0.9830	0.9733	0.9719	0.9787	0.9943	0.9020
ODI-HB-Min-SpO ₂ -MeanSpO ₂	Younger (< 60 years)	0.9914	0.9819	0.9683	0.9896	0.8276	0.9774	0.9231

^aAUC: area under the receiver operating characteristic curve.

^bPPV: positive predictive value.

^cNPV: negative predictive value.

^dODI: oxygen desaturation index.

^eHB: hypoxia burden.

^fMinSpO₂: minimal SpO₂.

^gST90: percentage of sleep time with SpO₂ $< 90\%$.

Discussion

Principal Findings and Model Development

This study presents the first comprehensive evaluation of multidimensional oximetric parameters for OSA screening. Using 6 ML algorithms, we developed and rigorously validated an integrated multi-parameter model that overcomes the inherent limitations of conventional single- or dual-parameter approaches [44,45]. Through a novel algorithm-parameter matching framework, we established a CatBoost model combining ODI, HB, MinSpO₂, and ST90. This model showed robust performance in external validation, providing a streamlined and high-precision tool for OSA screening. Furthermore, our findings elucidate the heterogeneous contributions of oximetric parameters across sex- and age-specific subgroups, addressing a critical gap in population-specific research [46] and providing the foundation for personalized risk stratification.

Model Generalizability and Oximetric Parameter Performance

We used a large internal development set derived from PSG and an independent external validation set derived from HSAT, which included older patients with severe nocturnal hypoxemia, representing distinct clinical phenotypes. This integration of community and clinical data improves model generalizability and may streamline OSA diagnosis [4,23]. Consistent with previous reports, HSAT showed high diagnostic accuracy and strong correlation with PSG in older patients with severe OSA [47]. As expected, the OSA group was predominantly male and exhibited worse oximetry profiles [1,22,23]. Single-parameter oximetry models demonstrated marked variability in predictive performance, with ODI and HB emerging as the strongest predictors. ODI quantifies hourly desaturation frequency and correlates strongly with AHI, acting as an independent predictor of OSA severity regardless of sleep stage or body position [48,49]. Although ODI >20 events/h showed high sensitivity

(96.6%) for severe OSA, its AUC for mild disease was only 0.62, suggesting limited standalone utility [50,51]. In contrast, HB integrates desaturation depth and duration, better capturing the cumulative physiological burden of intermittent hypoxemia [44,46,52]. Notably, RCS analysis revealed that the MeanSpO₂ risk curves plateaued within the 88%-92% range, indicating that mean values fail to capture transient desaturation events and lack diagnostic sensitivity without complementary parameters [53]. TotalPower reflects global signal fluctuations without a specific mechanistic link to respiratory events [54]. These nonlinear parameter-risk relationships explain the suboptimal performance of linear models such as LR [10,23,44,55]. The steep threshold effects for ODI and HB suggest that even mild OSA can trigger substantial risk escalation, supporting the hypothesis of a critical threshold for hypoxic exposure [56,57].

Multidimensional Synergy of Oximetric Indices

We further quantified interactions among multiple oximetric indices. While ODI tracks event frequency, it neglects desaturation depth and duration, failing to capture physiological nuances of OSA heterogeneity [10,57]. Our 2-parameter ODI-HB CatBoost model achieved an F_1 -score of 0.9472. By characterizing both temporal and intensity dimensions, this synergy explains why ODI, as the strongest predictor, requires HB integration to improve performance [57,58]. The superior performance of our primary model, ODI-HB-MinSpO₂-ST90 (F_1 -score = 0.9516, AUC = 0.9879), stems from the inherent complementarity of these parameters. MinSpO₂ captures severe hypoxic nadirs [16], while ST90 quantifies nocturnal hypoxemia duration [59]. Together, this 4-parameter ensemble facilitates comprehensive multidimensional phenotyping of OSA, encompassing the frequency, depth, duration, and cumulative burden of hypoxic events [57,60]. In terms of benchmarking, our model outperformed the multidimensional oximetry approach proposed by Kong et al [8] (AUC=0.939) and the least

squares boosting model (AUC=0.889-0.924) developed by Gutiérrez-Tobal et al [23]. Moreover, integrated models incorporating demographics, questionnaires, and facial photography achieve AUCs ranging from 0.88 to 0.89 [1,61], whereas our oximetry-only approach achieved superior accuracy without auxiliary clinical data. This suggests multidimensional oximetry indices serve as effective PSG surrogates, encapsulating more direct pathophysiological information than traditional clinical markers [4].

Benchmarking of Prediction Models and ML Algorithms

ML algorithms demonstrate variable performance in OSA screening [62]. LR yielded an AUC of approximately 0.77 [63], whereas SVM exhibited the highest accuracy exclusively for mild OSA [62]. RF reached 84.4% accuracy in predicting severe OSA but remained insensitive to complex feature interactions [62,64]. In contrast, gradient boosting frameworks excelled at handling heterogeneous interactions, class imbalances, and missing data [4,23]. By implementing ordered boosting, CatBoost effectively mitigates gradient bias, thereby enhancing model robustness and consistently outperforming established benchmarks such as XGBoost and LightGBM [4,65,66]. Our study represents the first application of CatBoost to multidimensional oximetry-based OSA screening, underscoring its capacity to resolve complex nonlinear relationships [24,25]. Notably, performance did not improve with 5 or more oximetry parameters. Indiscriminate feature addition introduces multicollinearity and overfitting without incremental gains [4]. By prioritizing core parameter selection over feature stacking, our model ensures both robustness and clinical feasibility, providing a basis for developing portable screening devices [7,23].

Sex- and Age-Specific Performance Heterogeneity

Given the substantial phenotypic heterogeneity of OSA, sex-specific differences have been inadequately addressed in existing models, often leading to underdiagnosis among women [67-69]. To address this, we conducted stratified analyses by sex and age. In males and older subgroups, the ODI-HB-MinSpO₂-ST90 model proved optimal, consistent with our overall findings. This result likely attributable to our predominantly male sample (median age 45 years). Conversely, the ODI-HB-MinSpO₂-MeanSpO₂ configuration demonstrated superior performance in females. Patiño et al [70] reported that despite lower AHI values, women exhibit mean SpO₂ reductions comparable to those in men, while Poka-Mayap et al [71] confirmed significantly lower mean SpO₂ levels in women with OSA, suggesting that the female OSA phenotype may be more closely linked to sustained hypoxemia. This may reflect heightened hypoxic sensitivity in women, where SpO₂ fluctuations at subclinical AHI thresholds are sufficient to induce end-organ damage [56]. Notably, in older adults, the ODI-HB-MinSpO₂-ST90 model achieved an exceptional AUC of 0.9950, surpassing that in younger participants. This enhanced accuracy likely stems from reduced hypoxic tolerance and pronounced SpO₂ variability characteristic of aging. Indeed,

older patients, particularly older women, consistently exhibit lower SpO₂ levels independent of AHI burden [72].

Model Interpretability and External Validation Robustness

Previous ML-based OSA diagnostic studies have focused predominantly on predictive performance, frequently neglecting to quantify parameter contributions, thereby limiting clinical trust and adoption of oximetry-based models [17,19,23]. We addressed this “black-box” limitation through SHAP values and feature importance analysis, which corroborated the central role of ODI and HB while revealing significant feature hierarchy shifts across subgroups. Notably, the importance of MeanSpO₂ increased substantially in the female subgroup, reinforcing hypotheses regarding sex-specific physiological signatures in which sustained hypoxemia may characterize the female phenotype more prominently [56,69,71]. Furthermore, our CatBoost-based model maintained exceptional performance across independent external validation sets, with sex- and age-specific models demonstrating sustained stability, attributable to CatBoost’s ability to mitigate gradient bias [73]. Despite the relatively small non-OSA control group in the external validation cohort (n = 76), the high precision maintained across populations indicates strong potential for cross-cohort generalizability [7,23,74]. This robustness suggests that ML-integrated multidimensional oximetry can streamline OSA screening protocols and reduce reliance on resource-intensive PSG [75].

Limitations

This study has several limitations. First, despite the large sample size, participants were recruited from a single center and comprised individuals referred for suspected OSA. This may have resulted in a higher OSA prevalence than in the general population, thereby increasing the pretest probability and potentially overestimating diagnostic performance. Second, the lack of longitudinal follow-up precludes assessment of the model’s temporal consistency or long-term predictive capacity. Third, our cohort lacked racial and ethnic diversity, and as skin pigmentation can introduce systematic biases in SpO₂ measurements, this may limit generalizability to individuals with darker skin tones [76]. Finally, the single-center design and absence of prospective community-based validation restricts broader external validity.

Conclusions

In conclusion, we developed and validated a robust CatBoost-based multidimensional oximetry model that enables accurate OSA screening. Although the ODI-HB-MinSpO₂-ST90 combination demonstrated optimal performance in the general population, substituting ST90 with MeanSpO₂ proved superior for female and younger subgroups. By integrating nocturnal hypoxemia indices spanning frequency, depth, and duration, our approach overcomes the limitations of single-parameter screening and offers multidimensional physiological assessment beyond conventional AHI-centric methods. These models can be readily integrated into portable monitoring devices or wearable technologies to facilitate early OSA diagnosis. Future research should prioritize multicenter prospective trials and

multi-ethnic validation studies to establish standardized protocols for personalized OSA risk stratification.

Acknowledgments

This work was supported by the Open Project of the State Key Laboratory of Pollution Control and Resource Reuse, Tongji University (No. PCRRF21013).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy concerns, ethical restrictions, and protected health information regulations but are available from the corresponding author on reasonable request. The underlying code for this study (and training/validation datasets) is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Funding

This study was supported by the grants from the National Natural Science Foundation of China (82100103), the National Science and Technology Major Project of China (2025ZD01902405), and the Shanghai Shengkang Hospital Management Center, Shanghai Clinical Cohort (SHDC2025CCS044).

Authors' Contributions

XQ developed software, visualized results, contributed to the original draft writing, validation, and funding acquisition. HL conceptualized the project, designed the methodology, validated results, and contributed to manuscript writing. RD curated data, validated results, visualized data, and conducted formal analysis. TG, HW, and PW contributed to the methodology and visualization. NL designed the methodology, curated data, validated results, provided supervision, project administration, resources, and contributed to writing and funding acquisition. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698. [FREE Full text] [doi: [10.1016/S2213-2600\(19\)30198-5](https://doi.org/10.1016/S2213-2600(19)30198-5)] [Medline: [31300334](https://pubmed.ncbi.nlm.nih.gov/31300334/)]
2. Lv R, Liu X, Zhang Y, Dong N, Wang X, He Y, et al. Pathophysiological mechanisms and therapeutic approaches in obstructive sleep apnea syndrome. *Signal Transduct Target Ther*. 2023;8(1):218. [FREE Full text] [doi: [10.1038/s41392-023-01496-3](https://doi.org/10.1038/s41392-023-01496-3)] [Medline: [37230968](https://pubmed.ncbi.nlm.nih.gov/37230968/)]
3. Troester MM, Quan SF, Berry RB. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL. American Academy of Sleep Medicine; 2023.
4. Kuo NY, Tsai HJ, Tsai SJ, Yang AC. Efficient screening in obstructive sleep apnea using sequential machine learning models, questionnaires, and pulse oximetry signals: mixed methods study. *J Med Internet Res*. 2024;26:e51615. [FREE Full text] [doi: [10.2196/51615](https://doi.org/10.2196/51615)] [Medline: [39699950](https://pubmed.ncbi.nlm.nih.gov/39699950/)]
5. Holfinger SJ, Lyons MM, Keenan BT, Mazzotti DR, Mindel J, Maislin G, et al. Diagnostic performance of machine learning-derived OSA prediction tools in large clinical and community-based samples. *Chest*. 2022;161(3):807-817. [FREE Full text] [doi: [10.1016/j.chest.2021.10.023](https://doi.org/10.1016/j.chest.2021.10.023)] [Medline: [34717928](https://pubmed.ncbi.nlm.nih.gov/34717928/)]
6. Yang H, Lu S, Yang L. Clinical prediction models for the early diagnosis of obstructive sleep apnea in stroke patients: a systematic review. *Syst Rev*. 2024;13(1):38. [FREE Full text] [doi: [10.1186/s13643-024-02449-9](https://doi.org/10.1186/s13643-024-02449-9)] [Medline: [38268059](https://pubmed.ncbi.nlm.nih.gov/38268059/)]
7. da Silva Dantas EL, Stelzer FG, Bernardo WM, Eckeli AL. Oximetry-based devices in diagnosis of obstructive sleep apnea: a systematic review and meta-analysis. *Sleep Med Rev*. 2025;83:102139. [doi: [10.1016/j.smrv.2025.102139](https://doi.org/10.1016/j.smrv.2025.102139)] [Medline: [40749523](https://pubmed.ncbi.nlm.nih.gov/40749523/)]
8. Kong D, Hu C, Zhu H. Oxygen desaturation index, lowest arterial oxygen saturation and time spent below 90% oxygen saturation as diagnostic markers for obstructive sleep apnea. *Am J Transl Res*. 2023;15(5):3597-3606. [FREE Full text] [Medline: [37303658](https://pubmed.ncbi.nlm.nih.gov/37303658/)]
9. Chen S, Redline S, Eden UT, Prerau MJ. Dynamic models of obstructive sleep apnea provide robust prediction of respiratory event timing and a statistical framework for phenotype exploration. *Sleep*. 2022;45(12):zsac189. [FREE Full text] [doi: [10.1093/sleep/zsac189](https://doi.org/10.1093/sleep/zsac189)] [Medline: [35932480](https://pubmed.ncbi.nlm.nih.gov/35932480/)]
10. Hui X, Cao W, Xu Z, Guo J, Luo J, Xiao Y. Hypoxic indices for obstructive sleep apnoea severity and cardiovascular disease risk prediction: a comparison and application in a community population. *Respirology*. 2024;29(9):825-834. [doi: [10.1111/resp.14754](https://doi.org/10.1111/resp.14754)] [Medline: [38773880](https://pubmed.ncbi.nlm.nih.gov/38773880/)]

11. Cabanas AM, Fuentes-Guajardo M, Sáez N, Catalán DD, Collao-Caiconte PO, Martín-Escudero P. Exploring the hidden complexity: entropy analysis in pulse oximetry of female athletes. *Biosensors (Basel)*. 2024;14(1). [[FREE Full text](#)] [doi: [10.3390/bios14010052](https://doi.org/10.3390/bios14010052)] [Medline: [38275305](#)]
12. Al Ashry HS, Hilmisson H, Ni Y, Thomas RJ, APPLES Investigators. Automated apnea-hypopnea index from oximetry and spectral analysis of cardiopulmonary coupling. *Ann Am Thorac Soc*. 2021;18(5):876-883. [doi: [10.1513/AnnalsATS.202005-510OC](https://doi.org/10.1513/AnnalsATS.202005-510OC)] [Medline: [33472017](#)]
13. Nigro CA, Castaño G, Bledel I, Colombi A, Zicari MC. A novel, simple, and accurate pulse oximetry indicator for screening adult obstructive sleep apnea. *Sleep Breath*. 2022;26(3):1125-1134. [doi: [10.1007/s11325-021-02439-4](https://doi.org/10.1007/s11325-021-02439-4)] [Medline: [34554375](#)]
14. Wan C, Nnamdi MC, Shi W, Smith B, Purnell C, Wang MD. Advancing sleep disorder diagnostics: a transformer-based EEG model for sleep stage classification and OSA prediction. *IEEE J Biomed Health Inform*. 2025;29(2):878-886. [doi: [10.1109/JBHI.2024.3512616](https://doi.org/10.1109/JBHI.2024.3512616)] [Medline: [40030422](#)]
15. Yook S, Kim D, Gupte C, Joo EY, Kim H. Deep learning of sleep apnea-hypopnea events for accurate classification of obstructive sleep apnea and determination of clinical severity. *Sleep Med*. 2024;114:211-219. [doi: [10.1016/j.sleep.2024.01.015](https://doi.org/10.1016/j.sleep.2024.01.015)] [Medline: [38232604](#)]
16. Liu Y, Huang W, Zou J, Xu H, Qian Y, Zhu H, et al. Sea level nocturnal minimal oxygen saturation can accurately detect the presence of obstructive sleep apnea in a population with high pretest probability. *Sleep Breath*. 2021;25(1):171-179. [doi: [10.1007/s11325-020-02014-3](https://doi.org/10.1007/s11325-020-02014-3)] [Medline: [32306175](#)]
17. Levy J, Álvarez D, Del Campo F, Behar JA. Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry. *Nat Commun*. 2023;14(1):4881. [[FREE Full text](#)] [doi: [10.1038/s41467-023-40604-3](https://doi.org/10.1038/s41467-023-40604-3)] [Medline: [37573327](#)]
18. Garcia-Vicente C, Gutierrez-Tobal GC, Vaquerizo-Villar F, Martin-Montero A, Gozal D, Hornero R. SleepECG-Net: Explainable deep learning approach with ECG for pediatric sleep apnea diagnosis. *IEEE J Biomed Health Inform*. 2025;29(2):1021-1034. [doi: [10.1109/JBHI.2024.3495975](https://doi.org/10.1109/JBHI.2024.3495975)] [Medline: [39527413](#)]
19. Pare G, Raymond L, Etindele Sosso FA. Nurses' intention to integrate AI into their practice: survey study in Canada. *JMIR Nurs*. 2025;8:e76795. [[FREE Full text](#)] [doi: [10.2196/76795](https://doi.org/10.2196/76795)] [Medline: [40911813](#)]
20. Liu K, Geng S, Shen P, Zhao L, Zhou P, Liu W. Development and application of a machine learning-based predictive model for obstructive sleep apnea screening. *Front Big Data*. 2024;7:1353469. [[FREE Full text](#)] [doi: [10.3389/fdata.2024.1353469](https://doi.org/10.3389/fdata.2024.1353469)] [Medline: [38817683](#)]
21. Kim YJ, Jeon JS, Cho S, Kim KG, Kang S. Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques. *Diagnostics (Basel)*. 2021;11(4):612. [[FREE Full text](#)] [doi: [10.3390/diagnostics11040612](https://doi.org/10.3390/diagnostics11040612)] [Medline: [33808100](#)]
22. Talukder A, Li Y, Yeung D, Shi M, Umbach DM, Fan Z, et al. OSApredictor: A tool for prediction of moderate to severe obstructive sleep apnea-hypopnea using readily available patient characteristics. *Comput Biol Med*. 2024;178:108777. [[FREE Full text](#)] [doi: [10.1016/j.combiomed.2024.108777](https://doi.org/10.1016/j.combiomed.2024.108777)] [Medline: [38901189](#)]
23. Gutiérrez-Tobal GC, Álvarez D, Vaquerizo-Villar F, Crespo A, Kheirandish-Gozal L, Gozal D, et al. Ensemble-learning regression to estimate sleep apnea severity using at-home oximetry in adults. *Appl Soft Comput*. 2021;111:107827. [doi: [10.1016/j.asoc.2021.107827](https://doi.org/10.1016/j.asoc.2021.107827)] [Medline: [39544517](#)]
24. Chen Q, Liang Z, Wang Q, Ma C, Lei Y, Sanderson JE, et al. Self-helped detection of obstructive sleep apnea based on automated facial recognition and machine learning. *Sleep Breath*. 2023;27(6):2379-2388. [doi: [10.1007/s11325-023-02846-9](https://doi.org/10.1007/s11325-023-02846-9)] [Medline: [37278870](#)]
25. Han H, Oh J. Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity. *Sci Rep*. 2023;13(1):6379. [[FREE Full text](#)] [doi: [10.1038/s41598-023-33170-7](https://doi.org/10.1038/s41598-023-33170-7)] [Medline: [37076549](#)]
26. Alotaibi N, Cheung M, Shah A, Hurst JR, Mani AR, Mandal S. Changes in physiological signal entropy in patients with obstructive sleep apnoea: a systematic review. *Physiol Meas*. 2024;45(9). [doi: [10.1088/1361-6579/ad79b4](https://doi.org/10.1088/1361-6579/ad79b4)] [Medline: [39260403](#)]
27. Suen C, Ryan CM, Mubashir T, Ayas NT, Abrahamyan L, Wong J, et al. Sleep study and oximetry parameters for predicting postoperative complications in patients with OSA. *Chest*. 2019;155(4):855-867. [[FREE Full text](#)] [doi: [10.1016/j.chest.2018.09.030](https://doi.org/10.1016/j.chest.2018.09.030)] [Medline: [30359618](#)]
28. Bokov P, Dudoignon B, Delclaux C. Hypoxic burden in children with sleep-disordered breathing: determinants and correlates. *J Sleep Res*. 2026;35(2):e70211. [doi: [10.1111/jsr.70211](https://doi.org/10.1111/jsr.70211)] [Medline: [40974193](#)]
29. Yang J, Choudhary GI, Rahardja S, Fränti P. Classification of interbeat interval time-series using attention entropy. *IEEE Trans. Affective Comput*. 2023;14(1):321-330. [doi: [10.1109/TAFFC.2020.3031004](https://doi.org/10.1109/TAFFC.2020.3031004)]
30. Terrill PI. A review of approaches for analysing obstructive sleep apnoea-related patterns in pulse oximetry data. *Respirology*. 2020;25(5):475-485. [doi: [10.1111/resp.13635](https://doi.org/10.1111/resp.13635)] [Medline: [31246376](#)]
31. Kainulainen S, Töyräs J, Oksenberg A, Korkalainen H, Afara IO, Leino A, et al. Power spectral densities of nocturnal pulse oximetry signals differ in OSA patients with and without daytime sleepiness. *Sleep Med*. 2020;73:231-237. [doi: [10.1016/j.sleep.2020.07.015](https://doi.org/10.1016/j.sleep.2020.07.015)] [Medline: [32861188](#)]
32. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]

33. Shi H, Wu C, Bai T, Chen J, Li Y, Wu H. Identify essential genes based on clustering based synthetic minority oversampling technique. *Comput Biol Med.* 2023;153:106523. [doi: [10.1016/j.combiomed.2022.106523](https://doi.org/10.1016/j.combiomed.2022.106523)] [Medline: [36652869](https://pubmed.ncbi.nlm.nih.gov/36652869/)]
34. Schober P, Vetter TR. Logistic regression in medical research. *Anesth Analg.* 2021;132(2):365-366. [FREE Full text] [doi: [10.1213/ANE.0000000000005247](https://doi.org/10.1213/ANE.0000000000005247)] [Medline: [33449558](https://pubmed.ncbi.nlm.nih.gov/33449558/)]
35. Boateng EY, Abaye DA. A review of the logistic regression model with emphasis on medical research. *J Data Anal Inf Process.* 2019;07(04):190-207. [doi: [10.4236/jdaip.2019.74012](https://doi.org/10.4236/jdaip.2019.74012)]
36. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing.* 2020;408:189-215. [doi: [10.1016/j.neucom.2019.10.118](https://doi.org/10.1016/j.neucom.2019.10.118)]
37. Geetha R, Sivasubramanian S, Kaliappan M, Vimal S, Annamalai S. Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier. *J Med Syst.* 2019;43(9):286. [doi: [10.1007/s10916-019-1402-6](https://doi.org/10.1007/s10916-019-1402-6)] [Medline: [31312985](https://pubmed.ncbi.nlm.nih.gov/31312985/)]
38. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016. Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17; San Francisco California USA. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
39. Li Y, Li M, Li C, Liu Z. Forest aboveground biomass estimation using landsat 8 and sentinel-1A data with machine learning algorithms. *Sci Rep.* 2020;10(1):9952. [FREE Full text] [doi: [10.1038/s41598-020-67024-3](https://doi.org/10.1038/s41598-020-67024-3)] [Medline: [32561836](https://pubmed.ncbi.nlm.nih.gov/32561836/)]
40. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. 2017. Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9:3149-3157; Long Beach California USA.
41. Prokhorenkova L, Gusev G, Vorobev A, Dorogush A, Gulin A. CatBoost: unbiased boosting with categorical features. 2018. Presented at: NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems; December 3-8:6639-6649; Montréal, Canada.
42. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep.* 2022;12(1):5979. [FREE Full text] [doi: [10.1038/s41598-022-09954-8](https://doi.org/10.1038/s41598-022-09954-8)] [Medline: [35395867](https://pubmed.ncbi.nlm.nih.gov/35395867/)]
43. DeVries Z, Locke E, Hoda M, Moravek D, Phan K, Stratton A, et al. Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability. *Spine J.* 2021;21(7):1135-1142. [doi: [10.1016/j.spinee.2021.02.007](https://doi.org/10.1016/j.spinee.2021.02.007)] [Medline: [33601012](https://pubmed.ncbi.nlm.nih.gov/33601012/)]
44. Martinez-Garcia MA, Sánchez-de-la-Torre M, White DP, Azarbarzin A. Hypoxic burden in obstructive sleep apnea: present and future. *Arch Bronconeumol.* 2023;59(1):36-43. [doi: [10.1016/j.arbres.2022.08.005](https://doi.org/10.1016/j.arbres.2022.08.005)] [Medline: [36115739](https://pubmed.ncbi.nlm.nih.gov/36115739/)]
45. Polytarchou A, Ohler A, Moudaki A, Koltzida G, Kanaka-Gantenbein C, Kheirandish-Gozal L, et al. Nocturnal oximetry parameters as predictors of sleep apnea severity in resource-limited settings. *J Sleep Res.* 2023;32(1):e13638. [doi: [10.1111/jsr.13638](https://doi.org/10.1111/jsr.13638)] [Medline: [35624085](https://pubmed.ncbi.nlm.nih.gov/35624085/)]
46. Li C, Gao Y, Huang W, Xu H, Liu Y, Zou J, et al. The use of the sleep apnea-specific hypoxic burden to predict obstructive sleep apnea hypopnea syndrome: evidence from a large cross-sectional study. *Sleep Med.* 2023;111:94-100. [doi: [10.1016/j.sleep.2023.09.007](https://doi.org/10.1016/j.sleep.2023.09.007)] [Medline: [37742592](https://pubmed.ncbi.nlm.nih.gov/37742592/)]
47. Corral-Peñañiel J, Pepin JL, Barbe F. Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome. *Eur Respir Rev.* 2013;22(129):312-324. [FREE Full text] [doi: [10.1183/09059180.00004213](https://doi.org/10.1183/09059180.00004213)] [Medline: [23997059](https://pubmed.ncbi.nlm.nih.gov/23997059/)]
48. Duarte RLM, Magalhães-da-Silveira FJ, Gozal D. Nocturnal oximetry in bariatric surgery patients referred to overnight in-lab polysomnography. *Obesity (Silver Spring).* 2021;29(9):1469-1476. [doi: [10.1002/oby.23231](https://doi.org/10.1002/oby.23231)] [Medline: [34328276](https://pubmed.ncbi.nlm.nih.gov/34328276/)]
49. Sforza E, Roche F, Chapelle C, Pichot V. Internight variability of apnea-hypopnea index in obstructive sleep apnea using ambulatory polysomnography. *Front Physiol.* 2019;10:849. [FREE Full text] [doi: [10.3389/fphys.2019.00849](https://doi.org/10.3389/fphys.2019.00849)] [Medline: [31354515](https://pubmed.ncbi.nlm.nih.gov/31354515/)]
50. Mashaqi S, Staebler D, Mehra R. Combined nocturnal pulse oximetry and questionnaire-based obstructive sleep apnea screening - a cohort study. *Sleep Med.* 2020;72:157-163. [doi: [10.1016/j.sleep.2020.03.027](https://doi.org/10.1016/j.sleep.2020.03.027)] [Medline: [32653766](https://pubmed.ncbi.nlm.nih.gov/32653766/)]
51. Varghese L, Rebekah G, Oliver A, Kurien R. Oxygen desaturation index as alternative parameter in screening patients with severe obstructive sleep apnea. *Sleep Sci.* 2022;15(Spec 1):224-228. [FREE Full text] [doi: [10.5935/1984-0063.20200119](https://doi.org/10.5935/1984-0063.20200119)] [Medline: [35273770](https://pubmed.ncbi.nlm.nih.gov/35273770/)]
52. Coso C, Solano-Pérez E, Romero-Peralta S, Castillo-García M, Silgado-Martínez L, López-Monzoni S, et al. The hypoxic burden, clinical implication of a new biomarker in the cardiovascular management of sleep apnea patients: a systematic review. *Rev Cardiovasc Med.* 2024;25(5):172. [FREE Full text] [doi: [10.31083/j.rcm2505172](https://doi.org/10.31083/j.rcm2505172)] [Medline: [39076480](https://pubmed.ncbi.nlm.nih.gov/39076480/)]
53. Feng T, Shan G, He H, Pei G, Tan J, Lu B, et al. The association of nocturnal hypoxemia with dyslipidemia in sleep-disordered breathing population of Chinese community: a cross-sectional study. *Lipids Health Dis.* 2023;22(1):159. [FREE Full text] [doi: [10.1186/s12944-023-01919-8](https://doi.org/10.1186/s12944-023-01919-8)] [Medline: [37752495](https://pubmed.ncbi.nlm.nih.gov/37752495/)]
54. Stewart GM, Tong BK, de Chazal P, Flood J, Bin YS, Kairaitis K, et al. Sydney Sleep Biobank Investigators. Relationship between framingham 10-year cardiovascular disease risk score and pulse wave amplitude drop characteristics in a sleep clinic cohort. *Chest.* 2026;169(2):514-526. [doi: [10.1016/j.chest.2025.09.136](https://doi.org/10.1016/j.chest.2025.09.136)] [Medline: [41106590](https://pubmed.ncbi.nlm.nih.gov/41106590/)]

55. Haghghat S, Joghatayi M, Issa J, Azimian S, Brinz J, Ashkan A, et al. Diagnostic accuracy of artificial intelligence for obstructive sleep apnea detection: a systematic review. *BMC Med Inform Decis Mak.* 2025;25(1):278. [FREE Full text] [doi: [10.1186/s12911-025-03129-x](https://doi.org/10.1186/s12911-025-03129-x)] [Medline: [40722158](https://pubmed.ncbi.nlm.nih.gov/40722158/)]
56. Sutherland K, Sadr N, Bin YS, Cook K, Dissanayake HU, Cistulli PA, et al. Comparative associations of oximetry patterns in obstructive sleep apnea with incident cardiovascular disease. *Sleep.* 2022;45(12):zsac179. [FREE Full text] [doi: [10.1093/sleep/zsac179](https://doi.org/10.1093/sleep/zsac179)] [Medline: [35896039](https://pubmed.ncbi.nlm.nih.gov/35896039/)]
57. Martinez-Garcia MA, Sánchez-de-la-Torre M, White DP, Azarbarzin A. Hypoxic burden in obstructive sleep apnea: present and future. *Arch Bronconeumol.* 2023;59(1):36-43. [doi: [10.1016/j.arbres.2022.08.005](https://doi.org/10.1016/j.arbres.2022.08.005)] [Medline: [36115739](https://pubmed.ncbi.nlm.nih.gov/36115739/)]
58. Sanchez-Azofra A, Orr JE, Sanz-Rubio D, Marin-Oto M, Alarcon-Sisamon S, Vicente E, et al. Apnea-specific pulse-rate response is associated with early subclinical atherosclerosis in obstructive sleep apnea. *Arch Bronconeumol.* 2025;61(2):82-89. [doi: [10.1016/j.arbres.2024.07.003](https://doi.org/10.1016/j.arbres.2024.07.003)] [Medline: [39084963](https://pubmed.ncbi.nlm.nih.gov/39084963/)]
59. Labarca G, Dreyse J, Salas C, Letelier F, Schmidt A, Rivera F, et al. Clinical utility of oximetric parameters to identify a high-risk phenotype of moderate-severe obstructive sleep apnea (OSA). *Clin Respir J.* 2020;14(12):1166-1175. [doi: [10.1111/crj.13256](https://doi.org/10.1111/crj.13256)] [Medline: [32780496](https://pubmed.ncbi.nlm.nih.gov/32780496/)]
60. Yook S, Park HR, Seo D, Joo EY, Kim H. Obstructive sleep apnea subtyping based on apnea and hypopnea specific hypoxic burden is associated with brain aging and cardiometabolic syndrome. *Comput Biol Med.* 2025;185:109604. [FREE Full text] [doi: [10.1016/j.combiomed.2024.109604](https://doi.org/10.1016/j.combiomed.2024.109604)] [Medline: [39721413](https://pubmed.ncbi.nlm.nih.gov/39721413/)]
61. Chapman JL, Hoyos CM, Killick R, Sutherland K, Cistulli PA, Zwar N, et al. Sydney OSA-GP study Investigators†. Development and validation of a model for diagnosis of obstructive sleep apnoea in primary care. *Respirology.* 2021;26(10):989-996. [doi: [10.1111/resp.14122](https://doi.org/10.1111/resp.14122)] [Medline: [34342088](https://pubmed.ncbi.nlm.nih.gov/34342088/)]
62. Russo S, Martini A, Luzzi V, Garbarino S, Pietrafesa E, Polimeni A. Exploring the complexity of obstructive sleep apnea: findings from machine learning on diagnosis and predictive capacity of individual factors. *Sleep Breath.* 2024;29(1):49. [doi: [10.1007/s11325-024-03191-1](https://doi.org/10.1007/s11325-024-03191-1)] [Medline: [39636493](https://pubmed.ncbi.nlm.nih.gov/39636493/)]
63. Liu X, Shu Y, Yu P, Li H, Duan W, Wei Z, et al. Classification of severe obstructive sleep apnea with cognitive impairment using degree centrality: a machine learning analysis. *Front Neurol.* 2022;13:1005650. [FREE Full text] [doi: [10.3389/fneur.2022.1005650](https://doi.org/10.3389/fneur.2022.1005650)] [Medline: [36090863](https://pubmed.ncbi.nlm.nih.gov/36090863/)]
64. Tsai CY, Liu WT, Hsu WH, Majumdar A, Stettler M, Lee KY, et al. Screening the risk of obstructive sleep apnea by utilizing supervised learning techniques based on anthropometric features and snoring events. *Digit Health.* 2023;9:20552076231152751. [FREE Full text] [doi: [10.1177/20552076231152751](https://doi.org/10.1177/20552076231152751)] [Medline: [36896329](https://pubmed.ncbi.nlm.nih.gov/36896329/)]
65. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J Big Data.* 2020;7(1):94. [FREE Full text] [doi: [10.1186/s40537-020-00369-8](https://doi.org/10.1186/s40537-020-00369-8)] [Medline: [33169094](https://pubmed.ncbi.nlm.nih.gov/33169094/)]
66. Moon J, Kim JH, Hong SJ, Yu CW, Kim YH, Kim EJ, et al. Deep learning model for identifying acute heart failure patients using electrocardiography in the emergency room. *Eur Heart J Acute Cardiovasc Care.* 2025;14(2):74-82. [doi: [10.1093/ehjacc/zuaf001](https://doi.org/10.1093/ehjacc/zuaf001)] [Medline: [39787045](https://pubmed.ncbi.nlm.nih.gov/39787045/)]
67. Sands SA, Alex RM, Mann D, Vena D, Terrill PI, Gell LK, et al. Pathophysiology underlying demographic and obesity determinants of sleep apnea severity. *Ann Am Thorac Soc.* 2023;20(3):440-449. [FREE Full text] [doi: [10.1513/AnnalsATS.202203-271OC](https://doi.org/10.1513/AnnalsATS.202203-271OC)] [Medline: [36287615](https://pubmed.ncbi.nlm.nih.gov/36287615/)]
68. Romero-Peralta S, García-Rio F, Resano Barrio P, Viejo-Ayuso E, Izquierdo JL, Sabroso R, et al. Defining the heterogeneity of sleep apnea syndrome: a cluster analysis with implications for patient management. *Arch Bronconeumol.* 2022;58(2):125-134. [doi: [10.1016/j.arbres.2021.02.022](https://doi.org/10.1016/j.arbres.2021.02.022)] [Medline: [33820676](https://pubmed.ncbi.nlm.nih.gov/33820676/)]
69. Chinese Thoracic Society. [Expert consensus on the diagnosis and treatment of obstructive sleep apnea in women]. *Zhonghua Jie He He Hu Xi Za Zhi.* 2024;47(6):509-528. [doi: [10.3760/cma.j.cn112147-20240206-00072](https://doi.org/10.3760/cma.j.cn112147-20240206-00072)] [Medline: [38858201](https://pubmed.ncbi.nlm.nih.gov/38858201/)]
70. Patiño MC, Bueno Florez SJ, Gallo L, Ortiz PA, Payán-Gómez C, Molano-Gonzalez N, et al. Gender and polysomnographic profiles findings in obstructive sleep apnea syndrome patients living in high altitude. *Nat Sci Sleep.* 2021;13:547-556. [FREE Full text] [doi: [10.2147/NSS.S287165](https://doi.org/10.2147/NSS.S287165)] [Medline: [33994817](https://pubmed.ncbi.nlm.nih.gov/33994817/)]
71. Poka-Mayap V, Balkissou Adamou D, Massongo M, Voufouo Sonwa S, Alime J, Moutlen BPM, et al. Obstructive sleep apnea and hypopnea syndrome in patients admitted in a tertiary hospital in Cameroon: prevalence and associated factors. *PLoS One.* 2020;15(1):e0227778. [FREE Full text] [doi: [10.1371/journal.pone.0227778](https://doi.org/10.1371/journal.pone.0227778)] [Medline: [31945127](https://pubmed.ncbi.nlm.nih.gov/31945127/)]
72. Soria Robles AI, Aguado Blanco C, Juárez España M, Andrés Pretel F, Massó Núñez ML, Vizcaíno García MS, et al. Obstructive sleep apnea and oxygenation in very old adults: a propensity-score match study. *J Am Med Dir Assoc.* 2024;25(7):105023. [doi: [10.1016/j.jamda.2024.105023](https://doi.org/10.1016/j.jamda.2024.105023)] [Medline: [38763163](https://pubmed.ncbi.nlm.nih.gov/38763163/)]
73. Li X, Zhou Z, Fan J, Zhao L, Xu R, Li D, et al. A machine learning algorithm to predict treatment effectiveness for Kawasaki disease in China: a retrospective model development and validation study. *Front Immunol.* 2025;16:1629600. [FREE Full text] [doi: [10.3389/fimmu.2025.1629600](https://doi.org/10.3389/fimmu.2025.1629600)] [Medline: [41383607](https://pubmed.ncbi.nlm.nih.gov/41383607/)]
74. Abd-Alrazaq A, Aslam H, AlSaad R, Alsahli M, Ahmed A, Damseh R, et al. Detection of sleep apnea using wearable AI: systematic review and meta-analysis. *J Med Internet Res.* 2024;26:e58187. [FREE Full text] [doi: [10.2196/58187](https://doi.org/10.2196/58187)] [Medline: [39255014](https://pubmed.ncbi.nlm.nih.gov/39255014/)]

75. Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P. Enabling early obstructive sleep apnea diagnosis with machine learning: systematic review. *J Med Internet Res*. 2022;24(9):e39452. [FREE Full text] [doi: [10.2196/39452](https://doi.org/10.2196/39452)] [Medline: [36178720](https://pubmed.ncbi.nlm.nih.gov/36178720/)]
76. Acevedo-Fontanez AI, Patel SR. Racial disparities in obstructive sleep apnea care in the United States. *Sleep*. 2025;48(6):zsaf078. [doi: [10.1093/sleep/zsaf078](https://doi.org/10.1093/sleep/zsaf078)] [Medline: [40130545](https://pubmed.ncbi.nlm.nih.gov/40130545/)]

Abbreviations

AHI: apnea-hypopnea index
AUC: area under the receiver operating characteristic curve
CatBoost: categorical boosting
HB: hypoxic burden
HSAT: home sleep apnea test
LightGBM: light gradient boosting machine
LR: logistic regression
ML: machine learning
ODI: oxygen desaturation index
OSA: obstructive sleep apnea
PPV: positive predictive value
PSG: polysomnography
RCS: restricted cubic spline
RF: random forest
SHAP: Shapley additive explanations
SMOTE: synthetic minority over-sampling technique
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by A Benis; submitted 09.Jul.2025; peer-reviewed by FAE Sosso, B Xiang; comments to author 16.Oct.2025; revised version received 25.Feb.2026; accepted 06.Mar.2026; published 08.May.2026

Please cite as:

Qian X, Luo H, Ding R, Gao T, Wang H, Wu P, Li N

Machine Learning–Based Multidimensional Oximetry for Obstructive Sleep Apnea Screening: Development and External Validation
JMIR Med Inform 2026;14:e80384

URL: <https://medinform.jmir.org/2026/1/e80384>

doi: [10.2196/80384](https://doi.org/10.2196/80384)

PMID:

©Xuanyu Qian, Haitong Luo, Rong Ding, Tianming Gao, Haoan Wang, Pengliang Wu, Ning Li. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 08.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.