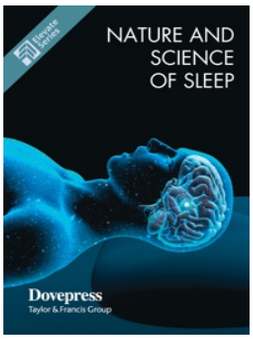

Taylor & Francis
Taylor & Francis Group

Nature and Science of Sleep

ISSN: 1179-1608 (Online) Journal homepage: www.tandfonline.com/journals/dnss20

Deep Learning for Obstructive Sleep Apnea Detection and Severity Assessment: A Multimodal Signals Fusion Multiscale Transformer Model

Yitong Zhang, Liang Zhou, Simin Zhu, Yanuo Zhou, Zitong Wang, Lina Ma, Yuqi Yuan, Yushan Xie, Xiaoxin Niu, Yonglong Su, Haiqin Liu, Xinhong Hei, Zhenghao Shi, Xiaoyong Ren & Yewen Shi

To cite this article: Yitong Zhang, Liang Zhou, Simin Zhu, Yanuo Zhou, Zitong Wang, Lina Ma, Yuqi Yuan, Yushan Xie, Xiaoxin Niu, Yonglong Su, Haiqin Liu, Xinhong Hei, Zhenghao Shi, Xiaoyong Ren & Yewen Shi (2025) Deep Learning for Obstructive Sleep Apnea Detection and Severity Assessment: A Multimodal Signals Fusion Multiscale Transformer Model, Nature and Science of Sleep, , 1-15, DOI: 10.2147/NSS.S492806

To link to this article: https://doi.org/10.2147/NSS.S492806

Published online: 07 Jan 2025.

Nature and Science of Sleep

Elevate Series

Open Access Full Text Article

ORIGINAL RESEARCH

Deep Learning for Obstructive Sleep Apnea Detection and Severity Assessment: A Multimodal Signals Fusion Multiscale Transformer Model

Yitong Zhang[1], Liang Zhou[2], Simin Zhu[1], Yanuo Zhou[1], Zitong Wang[1], Lina Ma[1], Yuqi Yuan[1], Yushan Xie[1], Xiaoxin Niu[1], Yonglong Su[1], Haiqin Liu[1], Xinhong Hei[2], Zhenghao Shi[2], Xiaoyong Ren[1], Yewen Shi[1]

[1]Department of Otorhinolaryngology Head and Neck Surgery, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi Province, People's Republic of China; [2]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi Province, People's Republic of China

Correspondence: Xiaoyong Ren; Yewen Shi, Department of Otorhinolaryngology Head and Neck Surgery, The Second Affiliated Hospital of Xi'an Jiaotong University, Address: NO. 157 Xi Wu Road, Xi'an, Shaanxi Province, Email cor_renxiaoyong@126.com; shiyewen59@outlook.com

Purpose: To develop a deep learning (DL) model for obstructive sleep apnea (OSA) detection and severity assessment and provide a new approach for convenient, economical, and accurate disease detection.

Methods: Considering medical reliability and acquisition simplicity, we used electrocardiogram (ECG) and oxygen saturation (SpO_2) signals to develop a multimodal signal fusion multiscale Transformer model for OSA detection and severity assessment. The proposed model comprises signal preprocessing, feature extraction, cross-modal interaction, and classification modules. A total of 510 patients who underwent polysomnography were included in the hospital dataset. The model was tested on hospital and public datasets. The hospital dataset was utilized to demonstrate the applicability and generalizability of the model. Two public datasets, Apnea-ECG dataset (consisting of 8 recordings) and UCD dataset (consisting of 21 recordings), were used to compare the results with those of previous studies.

Results: In the hospital dataset, the accuracy (Acc) values of per-segment and per-recording detection were 91.38 and 96.08%, respectively. The Acc values for mild, moderate, and severe OSA were 90.20, 88.24, and 92.16%, respectively. The Bland–Altman plots revealed the consistency of the true apnea–hypopnea index (AHI) and the predicted AHI. In the public datasets, the per-segment detection Acc values of the Apnea-ECG and UCD datasets were 95.04 and 90.56%, respectively.

Conclusion: The experiments on hospital and public datasets have demonstrated that the proposed model is more advanced, accurate, and applicable in OSA detection and severity assessment than previous models.

Keywords: obstructive sleep apnea, multimodal signals fusion, deep learning, detection model

Introduction

Obstructive sleep apnea (OSA) is defined as apnea and hypopnea caused by repeated collapse and upper airway obstruction during sleep.[1] Epidemiological studies have shown that OSA is highly prevalent in adults, with approximately one billion people worldwide experiencing this disease, 40–50% of whom have moderate-to-severe OSA; these populations are usually advised to receive prompt treatment.[2] In addition, severe OSA can be associated with dysfunctions in multiple organ systems, such as the cardiovascular, endocrine, and nervous system.[3] Given the high prevalence and severe dangers of OSA, timely detection and treatment are crucial. Polysomnography (PSG) is the gold standard for OSA diagnosis.[4] However, there are limitations such as the complexity and specialization of the technology, high economic costs, and long waiting times. OSA and its comorbidities will become an increasing health problem considering the global state of chronic health state transformation, such as the obesity epidemic and the aging population. Most OSA patients are not diagnosed and treated in time, resulting in a high economic and social burden. Therefore, developing simple and economical alternatives for OSA screening is essential.

Received: 23 August 2024
Accepted: 21 December 2024
Published: 6 January 2025

Automated detection by simplifying physiologic signals and developing artificial intelligence models is the most common alternative screening method for OSA. Various types of physiologic signals are used for OSA detection, such as electrocardiogram (ECG),[5–12] oxygen saturation (SpO_2)[13–16] and airflow signals.[17,18] Deep learning (DL) models, which are state-of-The-art artificial intelligence medical-assisted diagnostic models that extract deep features of signals and correlate them with respiratory events, have a wide range of potential for OSA detection.[5–13,15–18] Currently, single-lead signal detection is mainstream for OSA. However, relevant studies still have limitations, such as overlooking the potential correlation between multiple signals and having a limited detection capability.

Experts have suggested that effective integration of multimodal signals may improve model performance. For example, Taghizadegan et al fused electroencephalogram (EEG) and ECG signals and achieved a per-segment detection accuracy (Acc) of 91.74%.[19] Pathinarupothi et al reported that the fusion of ECG and SpO_2 signals resulted in a 3.10% increase in the detection Acc compared with that achieved with the single-lead ECG signal.[20] Although multimodal signal fusion methods have improved OSA detection performance, most of them only fuse the different modal information with fixed weights instead of considering the effect of potential correlations between different signals, thus failing to extract key target features and remove redundant information.

Therefore, we propose an innovative DL model for OSA detection and severity assessment to further improve OSA screening capabilities. Considering medical reliability and signal acquisition simplicity, we fuse ECG and SpO_2 signals. OSA contributes to cardiac electrophysiological remodeling, which appears as ECG signal changes and directly affects the amplitude of SpO_2 signal. The study of both signals alterations has a positive significance in the detection of respiratory events. Unlike previous studies, our contribution of the model design focuses on the novel multimodal signal fusion module, which effectively and efficiently achieves cross-modal information interaction. The results show that the model improves the detection capability of OSA and demonstrate the applicability in the hospital dataset.

Materials and Methods

Datasets

Hospital Dataset

All patients (n = 510) were enrolled from the Second Affiliated Hospital of Xi ‘an Jiaotong University, China, between 2022.04 and 2023.12. The inclusion criteria were as follows: (i) aged 18–65 years; (ii) symptoms suggestive of OSA, such as snoring or breathing cessation during sleep; and (iii) willingness to cooperate with data collection and accept PSG. Patients were excluded if they had (i) craniofacial abnormalities/disorders; (ii) severe neuromuscular or respiratory or cardiovascular disorders; (iii) other sleep disorders; (iv) previous sleep treatment; (v) long-term usage of medications known to affect sleep and heart rate; or (vi) signal damage and serious data loss. Figure 1 shows the enrollment process. This study was approved by the Ethics Committee of the Second Affiliated Hospital of Xi ‘an Jiaotong University (Ethics approval number: 2020–1122). All participants provided informed consent for the data collection and analysis. All the data were anonymized.

Clinical information, including demographic information, body measurements, and medical history, was collected from all patients. All patients underwent PSG (PHILIPS, Alice 6) supervised by night-shift staff. Two experienced sleep physicians scored the PSG findings according to the American Academy of Sleep Medicine (AASM) criteria. The apnea–hypopnea index (AHI) was defined as the total number of apnea and hypopnea events per hour of sleep. OSA was defined as AHI $\geq$ 5, mild OSA was defined as 5 $\leq$ AHI < 15, moderate OSA was defined as 15 $\leq$ AHI < 30, and severe OSA was defined as AHI $\geq$ 30. The demographics and characteristics of the patients in the hospital dataset is shown in Supplementary Table 1. We collected raw ECG and SpO_2 signals, respiratory event annotations, and PSG parameters. According to the AASM criteria recommendation, ECG signal was recorded from modified ECGII lead electrodes. The SpO_2 signal was continuously recorded by finger sensors. Notably, considering the diagnostic conditions for OSA, we classified both apnea and hypopnea events as respiratory events.

The patients in the hospital dataset were randomly divided into training, validation, and test sets at a ratio of 6:2:2. The recordings of 306, 102, and 102 patients were used to train, validate, and test the proposed model, respectively.

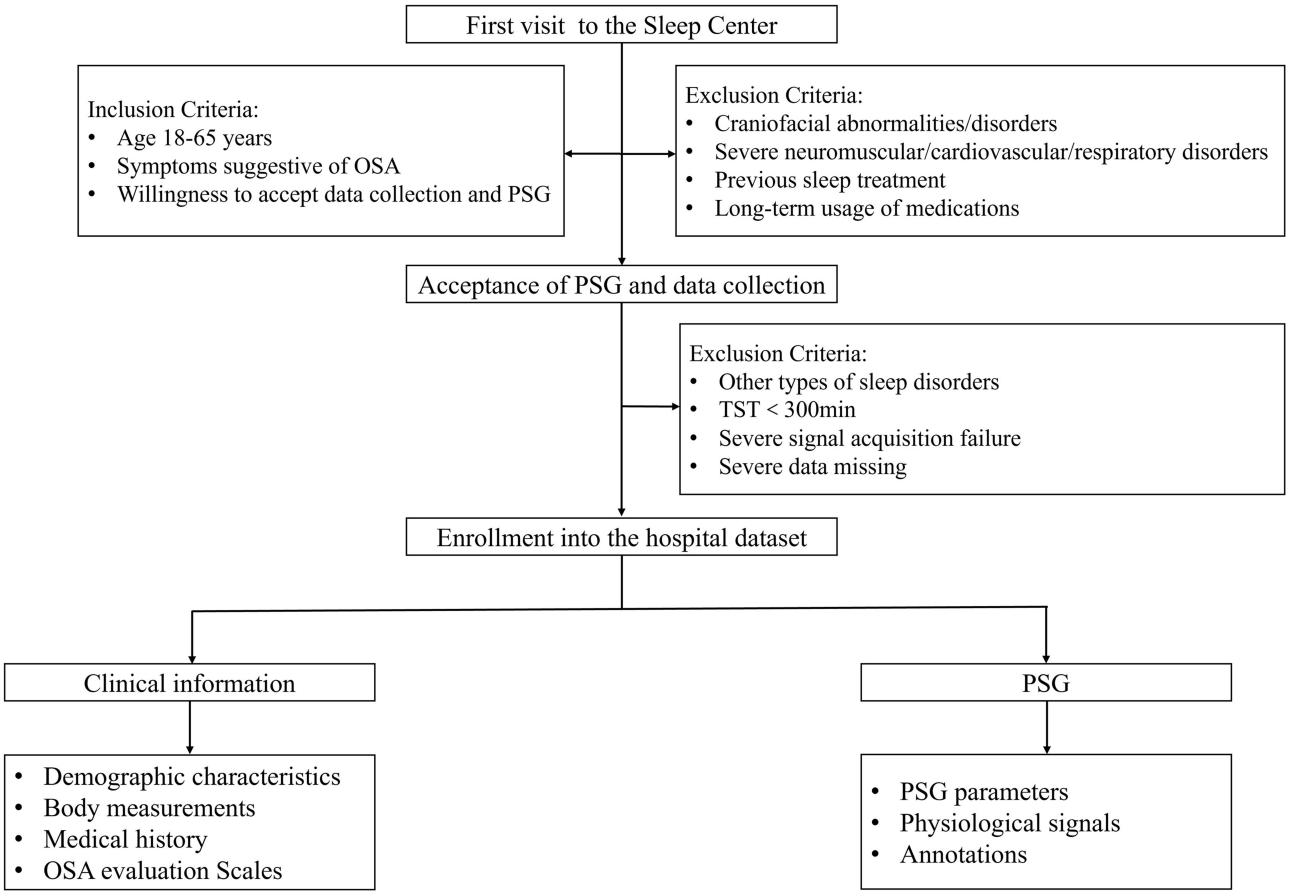

Figure 1 Enrollment process for the hospital dataset.
Abbreviations: OSA, obstructive sleep apnea; PSG, polysomnography; TST, total sleep time.

Public Datasets

In this study, the detection performance of the proposed model was compared with that of studies on public datasets. The demographics and characteristics of the patients in the public datasets are shown in Supplementary Table 2.

1) Apnea-ECG dataset:[21] This dataset is the most popular public dataset for OSA detection. The dataset comprises 70 recordings; however, only eight recordings (a01–a04, b01, and c01–c03) collected ECG and SpO_2 signals. The dataset is available at https://www.physionet.org/content/apnea-ecg/1.0.0/. We divided the eight recordings at a ratio of 6:2 for leave-out cross-validation. The model's performance was assessed by averaging the results over 5 repetitions to eliminate errors in randomly dividing the data.

2) St.Vincent's University Hospital/University College Dublin Sleep Apnea Dataset (UCD datasets):[22] This dataset comprises 25 recordings, which all collect ECG and SpO_2 signals and provide annotations on the occurrence and duration of respiratory events. Notably, we removed four subjects (ucddb08, ucddb015, ucddb018, and ucddb022) as they had few OSA segment and the data were incredibly unbalanced. The dataset is available at https://www.physionet.org/content/ucddb/1.0.0/. We divided the 21 recordings at a ratio of 4:1 for leave-out cross-validation. The model's performance was assessed by averaging the results over 5 repetitions to eliminate errors in randomly dividing the data.

Deep Learning Model

Model Structure

We propose a multimodal multiscale Transformer model for OSA detection and severity assessment. The model consists of four main modules: signals preprocessing, feature extraction, cross-modal interaction, and classification. Figure 2 shows the overall structure of the model.

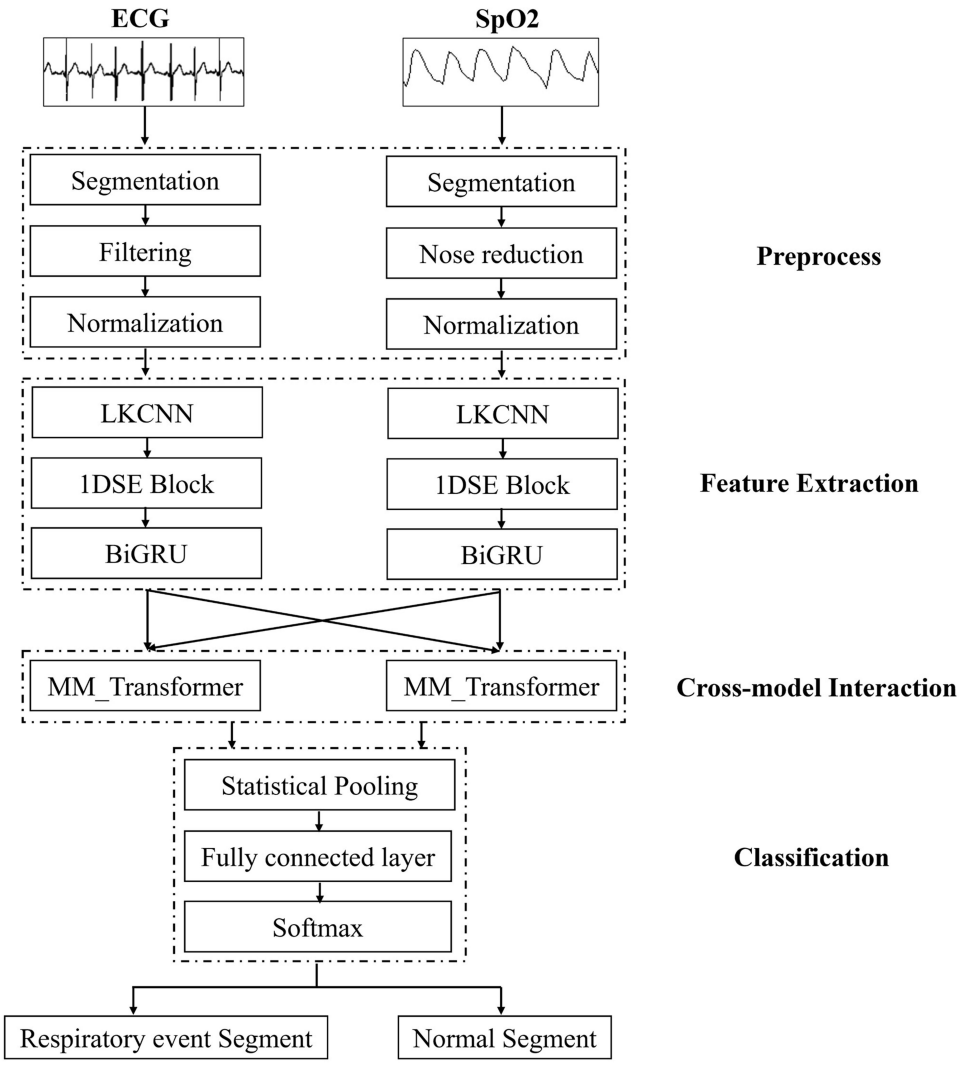

Figure 2 Overall structure of the proposed model.
Abbreviations: ECG, electrocardiogram; SpO_2, oxygen saturation; LKCNN, large convolutional kernels convolutional neural network; 1DSE, 1D squeeze-and-excitation; BiGRU, bi-directional gated recurrent unit; MM_Transformer, multimodal multiscale Transformer.

1) Signals preprocessing

First, the ECG and SpO_2 signals were segmented into 1-minute time intervals. The respiratory event annotations were used to label the segments, when the respiratory event occurred in the signal segment, it was labeled with "respiratory event", and otherwise labeled with "normal". The ECG signal was then denoised via a bandpass filter with a 0.5–48 Hz passband to avoid noise effects.[23] The SpO2 signal had some artifacts below 50% that were not considered physiological significant. These segments were removed. Finally, Z-Score normalization was used to eliminate the incomparability between the ECG and SpO_2 signals.

2) Feature extraction

The feature extraction module aims to extract the shallow features of the ECG and SpO_2 signals and perform adaptive feature optimization. First, a convolutional neural network module with large convolutional kernels (LKCNN) was used to extract shallow features.[24] Then, a 1D squeeze-and-excitation (1DSE) block was used to correct and filter the features extracted by the LKCNN.[25] This process focuses the model's attention on the most informative channel features and suppresses unimportant features. Finally, we added a bidirectional gated recurrent unit (Bi-GRU) after the 1DSE block to learn long-term dependencies between different features.[26]

3) Cross-modal interaction

This module aims to better accomplish multimodal fusion. By designing models that can process and correlate information from different modalities, more information can be provided for classification decisions. The module consists of two parallel multimodal multiscale Transformer (MM_Transformer), and three attention mechanisms are the focus of the module design.[27] A self-attention (SA) mechanism was used to filter out important information from a large amount of information and allow us to notice internal correlations in the input features. The multihead attention (MHA) mechanism was designed to enhance the model's feature extraction and representation learning ability to solve the problem of the limited global view of self-attention due to overfocusing on local information. The co-attention (CA) mechanism was designed to focus more attention on the region where the ECG and SpO_2 signals were correlated and to learn the dependencies between different features from multimodal signals.

4) Classification

The output features from the two cross-modal interactions were concatenated, and the mean and standard deviation were obtained through statistical pooling. Then, the mean and standard deviation were spliced into the fully connected layers. Finally, the predicted values for OSA classification were obtained via a Softmax operation.

Experimental Setup and Hyperparameter Settings

All the experiments were conducted on a workstation configured with one GeForce RTX3090 GPU with 24 GB of video memory, an Intel Core I9-10920X CPU with 3.50 GHz and 64 GB of RAM based on PyTorch 1.10.1, CUDA 11.1, and Python 3.8.

Considering the large number of hyperparameters, we used the stochastic search algorithm to find the optimal neural network hyperparameters to reduce the search time while ensuring that the model achieves a certain accuracy rate. Different combinations of learning rates, weight decays, batch sizes, and dropout rates were optimized via the stochastic search algorithm, and the set of hyperparameters with the best overall performance and effective suppression of overfitting was selected as the optimal hyperparameter. Table 1 shows the hyperparameter settings.

Evaluation Indicators

The value of the model in assisting in the diagnosis and severity grading of OSA was assessed. We evaluated per-segment and per-recording detection. Per-segment detection aimed to divide long-duration ECG signals into 1-min segments, each assigned a label of "normal" or "respiratory event". Then, we trained the model with the training set and used the model to make predictions for each 1-min segment of the test set. These data were used to calculate the predicted AHI (pred-AHI) as follows: pred-AHI = 60*(the number of segments with "respiratory event")/(the total number of segments). According to the pred-AHI, each patient's diagnosis and severity were assessed to obtain per-record detection results.

The performance of the model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), accuracy (Acc), sensitivity (Sen), specificity (Spec), and F1 score (F1).

Table 1 Hyperparameter Settings

Hyperparameters	Value
Learning rate	0.00005
Weight decay	0.00001
Dropout Rate	0.50
Epoch	30
Batch size	128/32
MM_Transformer Pathsize	(768,384,192,96)
MM_Transformer HeadNum	4
Optimizer	Adam

Notes: MM_Transformer HeadNum means the number of heads of MM_Transformer; MM_Transformer PatchSize represents the dimension of different patch feature vectors.

https://doi.org/10.2147/NSS.S492806

Statistical Analysis

Statistical analysis was performed with SPSS 22.0 and OriginPro 2021. Continuous variables are presented as medians with interquartile ranges. Categorical variables are presented as numbers with proportions. Differences between groups were compared via *t* tests or Wilcoxon rank-sum tests. $P < 0.05$ was set as the threshold for significance. A Bland–Altman plot was generated to evaluate the consistency of the true-AHI and pred-AHI.

Results

Hospital Dataset Results

The demographics and characteristics of the patients in the hospital dataset are shown in Table 2. The clinical information and PSG metrics were not significantly different between the training, validation, and test sets (all P > 0.05). The balanced distribution of these data reduced bias in the model experiments.

Figure 3 shows the loss and accuracy curves for the training and validation sets. With increasing epochs, the training loss and accuracy curves tended to stabilize, the fluctuations in the validation loss and accuracy curves decreased, and the model converged. This finding indicates that the model learns the practical features to improve the detection performance, and the model's generalizability in the validation dataset gradually increases with training. As shown in Figure 4,

Table 2 Demographics and Characteristics of Patients in the Training Set, Validation Set and Test Set from the Hospital Dataset

	Training set (N=306)	Validation set (N=102)	Test set (N=102)	*P* value
Demographic information				
Gender [n (%)]				
Female	85(27.78%)	26(25.49%)	18(17.65%)	0.126
Male	221(72.22%)	76(74.51%)	84(82.35%)	
Age (years) [$\bar{x} \pm s$]	37.51±12.78	37.51±11.46	36.90±12.68	0.578
Body mass index (kg/m^2) [M (Q1, Q3)]	25.70(23.00,28.25)	25.05(22.53,28.28)	24.90(23.53,27.40)	0.653
Waist circumference (cm) [$\bar{x} \pm s$]	93.62±11.85	92.48±11.20	91.66±14.04	0.491
Neck circumference (cm) [$\bar{x} \pm s$]	38.42±7.02	37.74±3.89	38.64±6.86	0.595
Sleep quality				
Snoring [n (%)]				
0	37(12.09%)	13(12.75%)	14(13.73%)	0.909
1	269(87.91%)	89(87.25%)	88(86.27%)	
Snoring history (years) [$\bar{x} \pm s$]	5.41±7.47	3.26±3.97	5.47±7.35	0.382
Breathing cessations during sleep [n (%)]				
0	189(61.76%)	65(63.73%)	60(58.82%)	0.767
1	117(38.24%)	37(36.27%)	42(41.18%)	
Breathing cessations history (years) [$\bar{x} \pm s$]	0.59±2.38	0.28±1.37	0.81±3.47	0.841
Polysomnography				
Total sleep time (h) [M (Q1, Q3)]	7.07(6.29,7.87)	7.13(6.08,8.09)	7.15(6.14,7.63)	0.948
AHI (time/h) [M (Q1, Q3)]	21.88(8.67,47.90)	22.56(7.83,46.14)	21.79(9.40,44.97)	0.917
OSA severity [n (%)]				
Normal (AHI < 5)	53(17.32%)	16(15.69%)	15(14.71%)	0.994
Mild (5 ≤ AHI < 15)	63(20.59%)	23(22.55%)	24(23.53%)	
Moderate (15 ≤ AHI < 30)	67(21.90%)	22(21.57%)	22(21.57%)	
Severe (AHI≥30)	123(40.19%)	41(40.19%)	41(40.19%)	
Mean oxygen saturation (%) [M (Q1, Q3)]	95.00(93.00,96.00)	94.00(93.00,96.00)	95.00(93.00,95.00)	0.649
Lowest oxygen saturation (%) [M (Q1, Q3)]	83.50(76.00,89.00)	84.00(75.00,88.00)	84.00(78.00,89.00)	0.795
Mean heart rate (time/min) [M (Q1, Q3)]	65.00(60.53,69.50)	64.50(59.95,69.48)	65.15(58.98,71.88)	0.807
Highest heart rate (time/min) [M (Q1, Q3)]	106.00(100.00,112.00)	105.00(98.25,113.00)	108.50(100.25,115.75)	0.218
Lowest heart rate (time/min) [M (Q1, Q3)]	51.00(46.00,56.00)	49.50(46.00,53.00)	51.00(47.00,56.00)	0.171

Abbreviations: OSA, obstructive sleep apnea; AHI, apnea–hypopnea index.

https://doi.org/10.2147/NSS.S492806

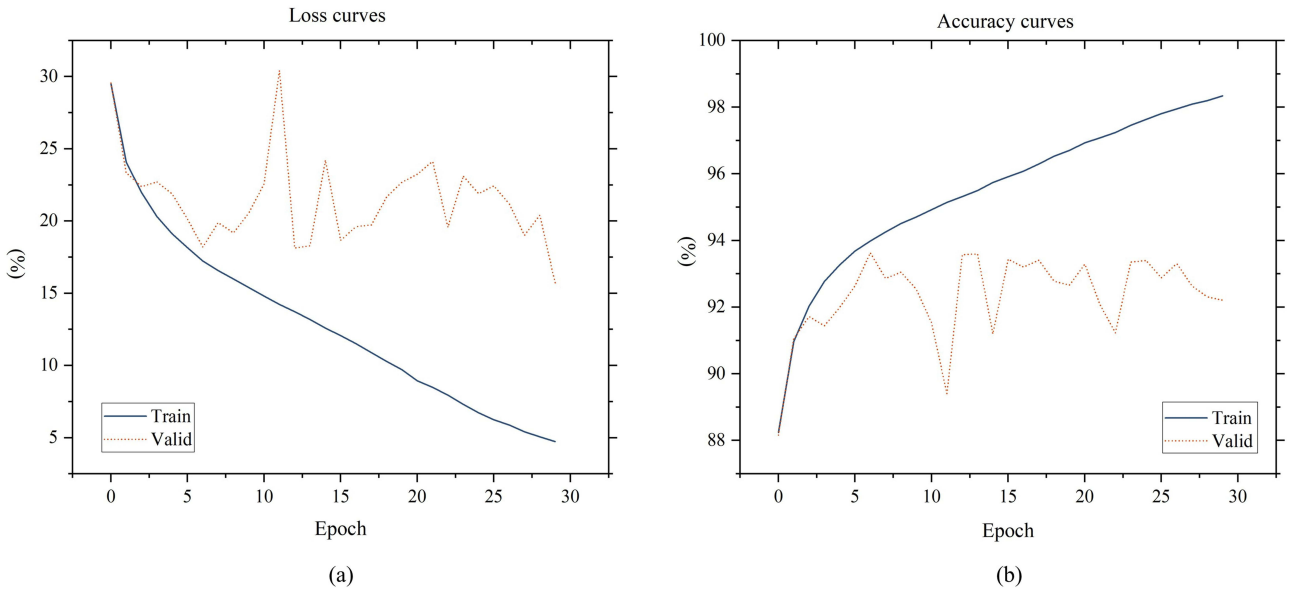

Figure 3 Training and validation loss and accuracy curves.
Notes: (**a**) Loss curves of training and validation sets; (**b**) Accuracy curves of training and validation sets.

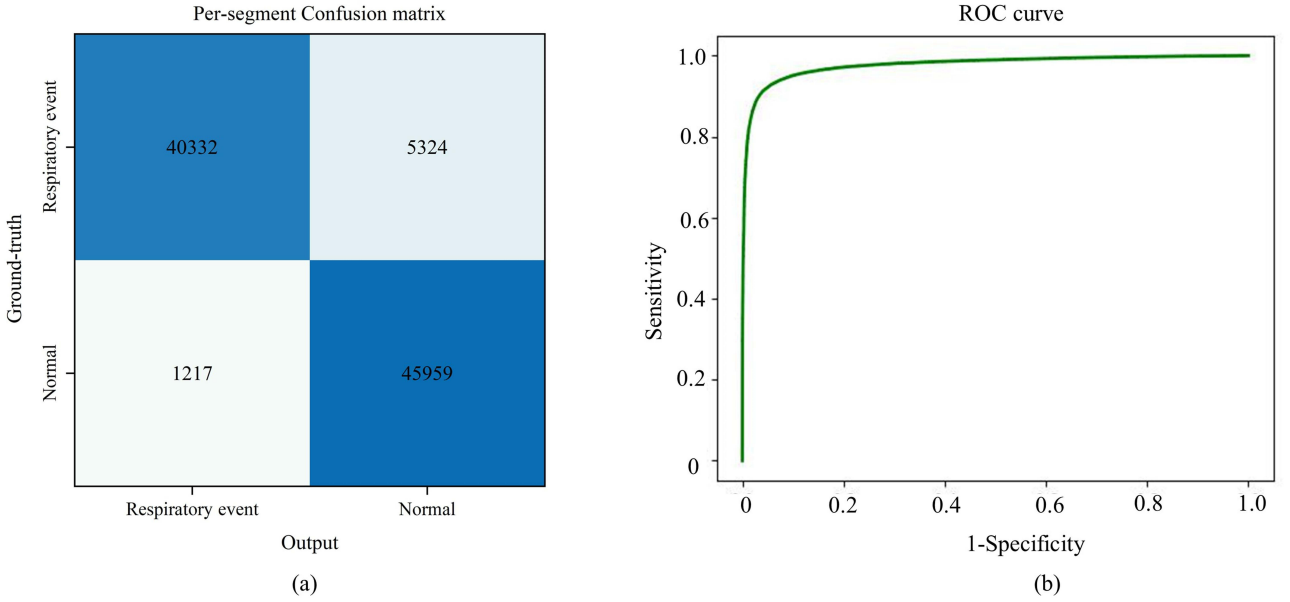

Figure 4 Per-segment detection confusion matrix and ROC curve of the validation set.
Notes: (**a**) Per-segment confusion matrix of the validation set; (**b**) Per-segment ROC curve of the validation set.

the per-segment detection Acc, Sen, Spec, and F1 values and the AUC of the validation set were 92.95%, 88.34%, 97.42%, 92.50%, and 0.978, respectively.

We evaluated the per-segment detection performance of the test set and calculated the pred-AHI for per-recording. The confusion matrix and ROC curve for per-segment detection of the test set are shown in Figure 5. As shown in Table 3, the Acc, Sen, Spec, F1 value, and AUC were 91.38%, 85.55%, 96.69%, 90.44%, and 0.968, respectively.

https://doi.org/10.2147/NSS.S492806

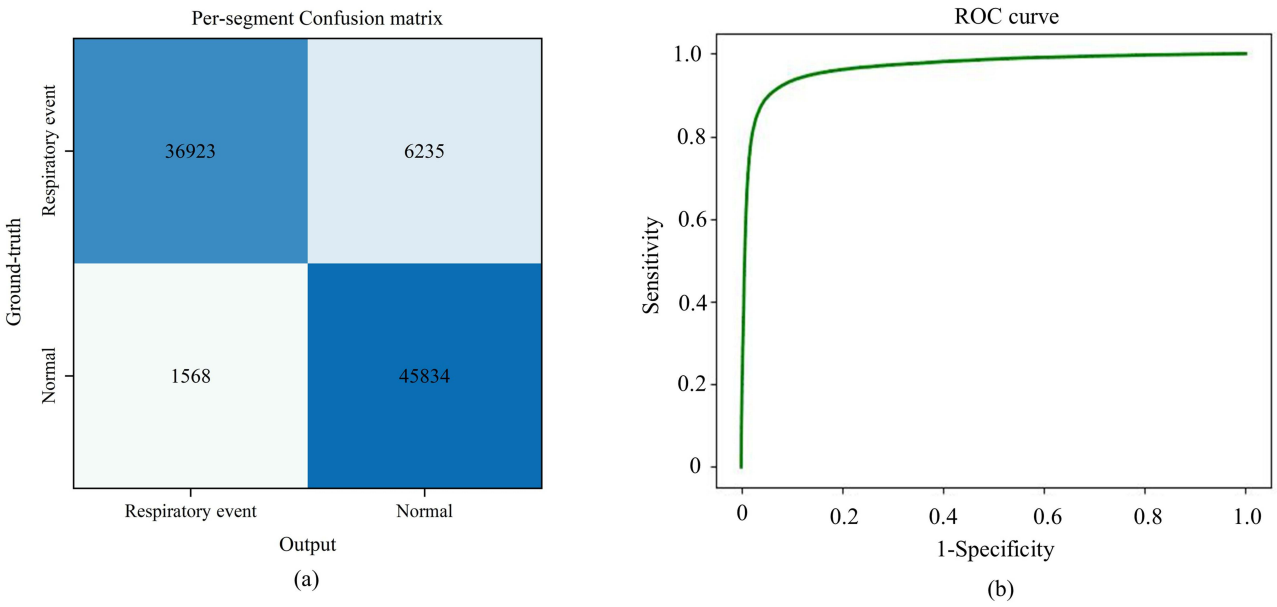

Figure 5 Per-segment detection confusion matrix and ROC curve of the test set.
Notes: (**a**) Per-segment confusion matrix of the test set; (**b**) Per-segment ROC curve of the test set.

We performed OSA diagnosis and severity assessment of patients in the test according to the pred-AHI. Figure 6 shows the confusion matrix and ROC curve for per-recording detection of the test set. As shown in Table 3, the Acc, Sen, Spec, F1 value, and AUC were 96.08%, 97.70%, 86.67%, 97.70%, and 0.922, respectively.

Figure 7 and Table 4 show the confusion matrix, ROC curves and performance for severity assessment. The Acc values of patients with mild OSA, moderate OSA, and severe OSA were 90.20%, 88.24%, and 92.16%, respectively, with an overall Acc of 83.33%. The AUC with mild OSA, moderate OSA, and severe OSA were 0.891, 0.835, and 0.918, respectively, with an overall AUC of 0.881.

In addition, we plotted Bland–Altman plots to assess the consistency of the true-AHI and pred-AHI. As shown in Figure 8, 93 (91.18%) points fell within the 95% concordance interval overall, indicating that the pred-AHI had a favorable concordance profile. When grouped by OSA severity, 12 (80.00%), 73 (83.91%), 20 (83.33%), 19 (86.36%), and 34 (82.93%) points fell within their 95% concordance intervals for non-OSA, OSA, mild OSA, moderate OSA, and severe OSA, respectively.

Table 3 Per-Segment and Per-Recording Results and Performance of the Test Set

	Per-segment	Per-recording
Total segments/recordings	90560	102
Normal segments/recordings	47402	15
Respiratory event segments/OSA recordings	43158	87
Correct predicted segments/recordings	82757	98
Wrong predicted segments/recordings	7803	4
Acc (%)	91.38	96.08
Sen (%)	85.55	97.70
Spec (%)	96.69	86.67
F1 (%)	90.44	97.70
AUC	0.968	0.922

Abbreviations: OSA, obstructive sleep apnea; Acc, accuracy; Sen, sensitivity; Spec, specificity; F1, F1 score; AUC, area under the receiver operating characteristic curve.

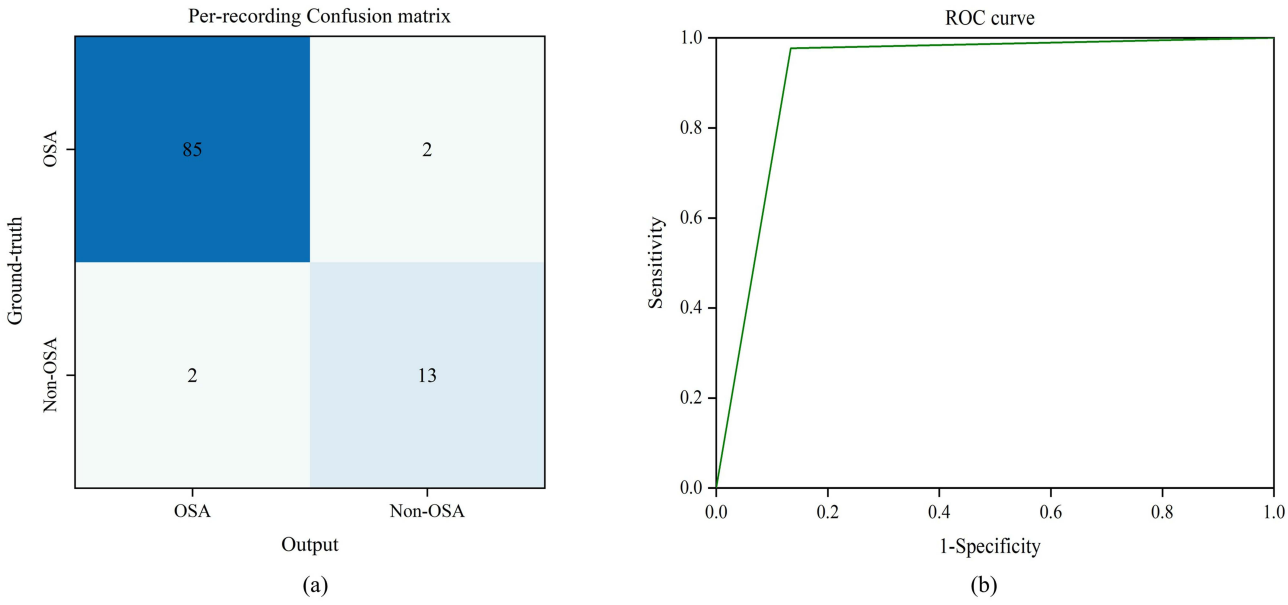

Figure 6 Per-recording detection confusion matrix and ROC curve of the test set.
Notes: (**a**) Per-recording confusion matrix of the test set; (**b**) Per-recording ROC curve of the test set.

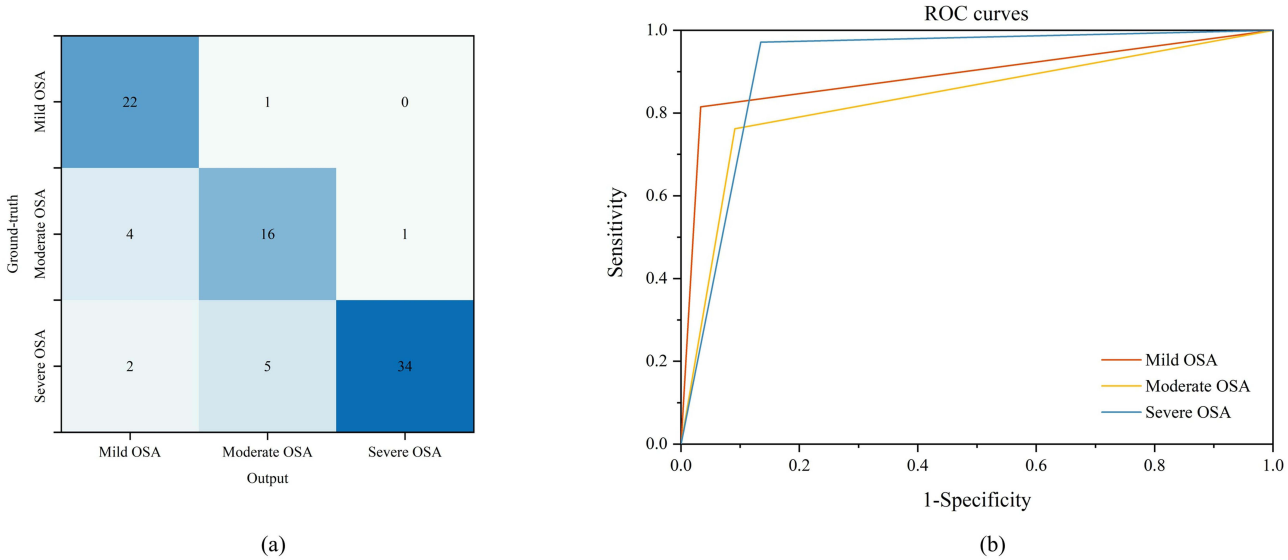

Figure 7 Confusion matrix and ROC curves of OSA severity in the test set.
Notes: (**a**) Confusion matrix of OSA severity; (**b**) ROC curves of OSA severity.
Abbreviations: OSA, obstructive sleep apnea.

Public Datasets Results

The model was validated on the Apnea-ECG and UCD datasets to make the model performance comparable.

The Apnea-ECG dataset included 8 records with 13694 segments, including 10469 "normal" and 3225 "respiratory event" segments. We performed leave-out cross-validation at a ratio of 6:2 and averaged the results 5 times. As shown in Figure 9, the average Acc, Sen, and Spec of pre-segment detection were 95.04%, 88.05%, and 96.21%, respectively.

The UCD dataset included 21 records with 3153 segments, including 2369 "normal" and 784 "respiratory event" segments. We performed leave-out cross-validation at a ratio of 4:1 and averaged the results 5 times. As shown in Figure 10, the average Acc, Sen, and Spec of pre-segment detection were 90.56%, 69.22%, and 96.21%, respectively.

https://doi.org/10.2147/NSS.S492806

Table 4 OSA Severity Prediction Results of the Test Set

	Normal	Mild OSA	Moderate OSA	Severe OSA
Acc (%)	96.08	90.20	88.24	92.16
Sen (%)	86.67	91.67	72.73	97.14
Spec (%)	97.70	89.74	92.50	89.55
F1 (%)	86.67	81.48	72.73	89.47

Abbreviations: OSA, obstructive sleep apnea; Acc, accuracy; Sen, sensitivity; Spec, specificity; F1, F1 score.

Discussion

The persistence and progression of OSA cause enormous health and economic burdens.[2] Symptom atypicality and diagnostic capability deficiencies contribute to the low diagnostic rate of OSA. Given the demand for new methods for OSA screening and assisting in diagnosis, this study proposes an innovative model based on multimodal signal fusion for OSA detection and severity assessment. The model performs initial OSA screening through simple signals acquisition and detection, and can be widely used in hospitals at all levels, especially in economically underdeveloped areas. It provides a new scheme to further improve the OSA detection capability.

Currently, the use of single-lead signal is still mainstream in DL-based OSA detection models,[5–18] and despite some progress, several shortcomings remain. First, multiple physiological signals change with the respiratory event, and the single-lead signal ignores the potential relationship between different signals. Second, the single-lead signal has limited information and restricts detection performance. Finally, the single-lead signal has poor stability during acquisition and is easily disturbed and distorted by interference. With the rapid development of DL technology and its wide application in biosignal processing, experts have proposed that the effective integration of multimodal information can significantly

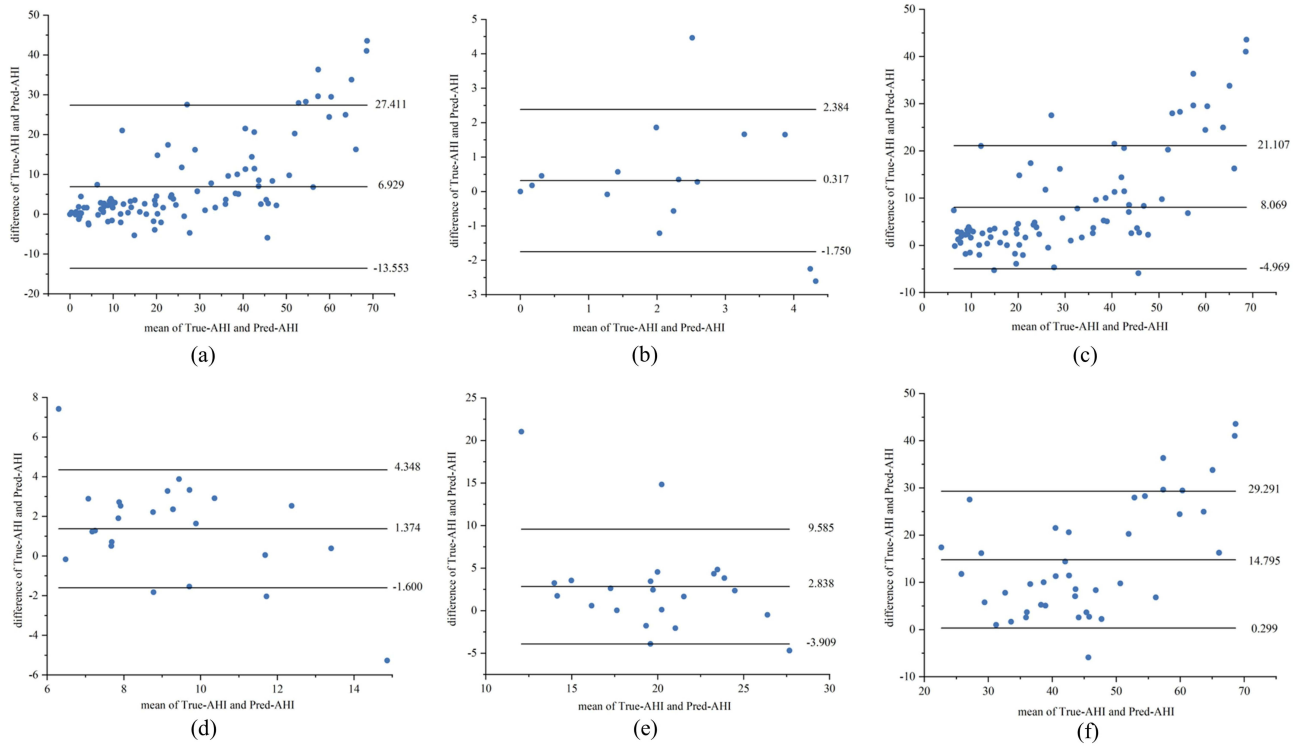

Figure 8 Bland–Altman plots of true-AHI and pred-AHI of the test set.
Notes: (**a**) All recordings; (**b**) non-OSA group; (**c**) OSA recordings; (**d**) mild OSA recording; (**e**) moderate OSA recordings; (**f**) severe OSA recordings.
Abbreviations: AHI, apnea–hypopnea index.

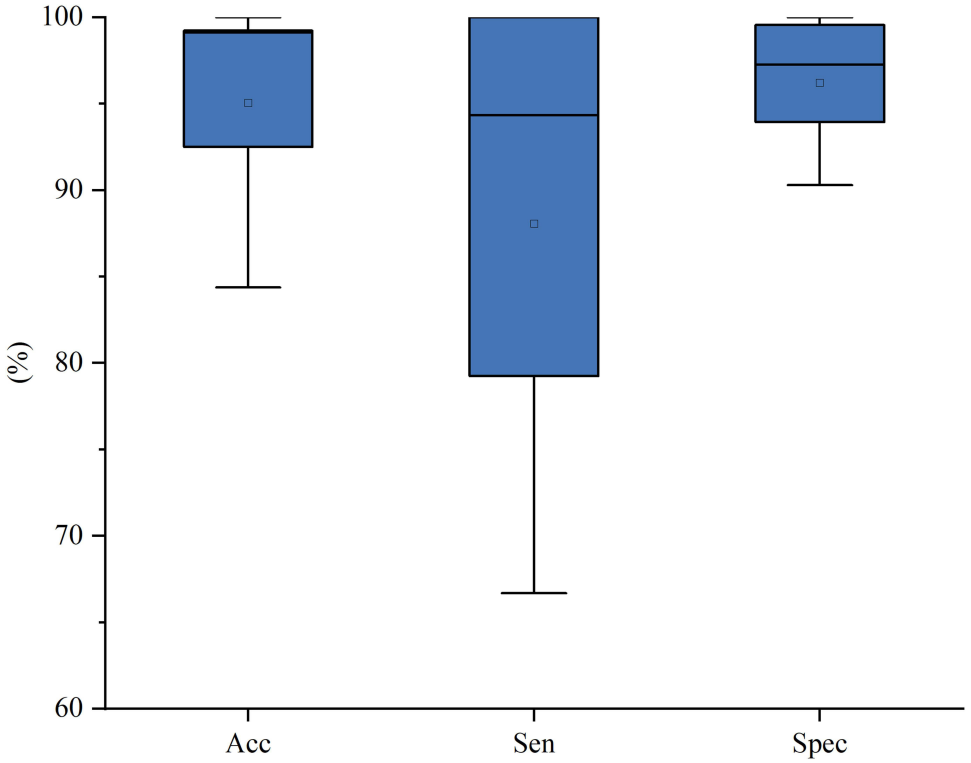

Figure 9 Performance on the Apnea-ECG dataset.
Abbreviations: Acc, accuracy; Sen, sensitivity; Spec, specificity.

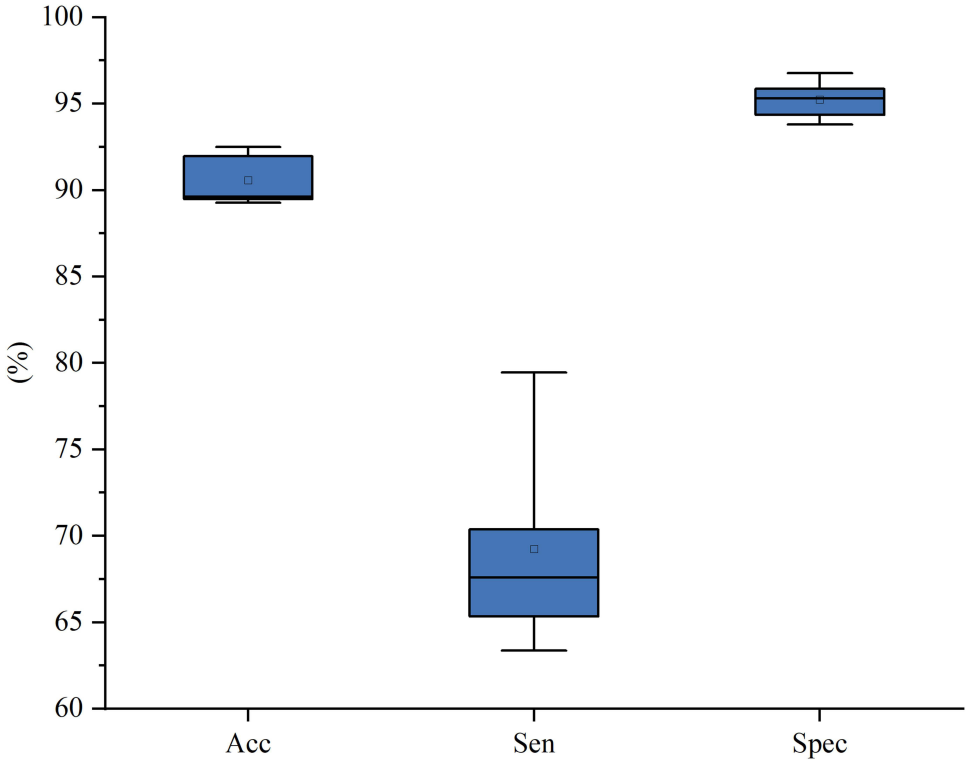

Figure 10 Performance on the UCD dataset.
Abbreviations: Acc, accuracy; Sen, sensitivity; Spec, specificity.

https://doi.org/10.2147/NSS.S492806

improve the ability to detect a target.[19,20] Critical information related to the detection target is made available by taking advantage of different modal signals. This method may be an effective approach to solving the problem of low detection accuracy of a single lead signal.

The selection of fused signals is one of the key points in model design. To a certain extent, multimodal signals increase the cost of acquisition and the complexity of the model. Therefore, balancing the simplicity of signal acquisition and the comprehensiveness of information to improve detection performance while reducing economic and computational costs is the key issue of the multimodal signal fusion model. We selected the ECG and SpO_2 signals for fusion. Medical reliability and acquisition simplicity were the main reasons for this choice. Pathophysiological changes in OSA, such as autonomic imbalance, chronic intermittent hypoxia, and changes in negative intrathoracic pressure, promote cardiac electrophysiological remodeling through direct and indirect pathways.[28–31] The corresponding changes in the ECG signal have established a reliable medical background for OSA detection.[32–35] The SpO_2 signal is directly affected by respiratory events and plays an essential role in determining OSA and its severity.[1] Moreover, as the earliest studied physiological signal, ECG and SpO_2 signals can be acquired by various devices. The preprocessing and feature extraction methods for ECG and SpO_2 signals are mature, providing a high-quality source of physiological signals. Other signals, such as snoring signals, are susceptible to environmental noise, EEG signal acquisition is very inconvenient, and the AF signal requires two transducers and, therefore, is a secondary choice.

Previous studies have also been conducted on performing OSA detection according to ECG and SpO_2 signals. Pathinarupothi et al and Paul et al used ECG signal, SpO2 signal, and a combination of the two as inputs, and the results demonstrated that the multimodal signals improved the accuracy of OSA detection compared with single-lead signal.[20,36] However, existing methods fuse the information from different modalities with fixed weights without considering the effect of potential correlations between different signals, thus limiting the detection performance. Therefore, the innovative design of the proposed model aims to accomplish cross-modal information interaction and improve detection performance.

We introduce Transformer structure and propose a cross-modal interaction module consisting of two parallel MM_Transformer. The attention mechanisms were the focus and innovation. SA allows the model to capture dependencies between input sequences at different locations and to consider relationships between multiple subsequences simultaneously; thus, more detailed modeling of the features and structure of long sequences is possible. MHA uses multiple attention heads to learn different information simultaneously, meaning the model can process it in parallel. MHA has significant advantages in terms of capturing richer features, improving generalization, and speeding up model computation with parallel processing. In addition, the CA focused more on the region of the ECG signal correlated with the SpO_2 signal. By simulating the process of manual diagnosis, the dependency between different features was learned, and feature fusion was performed on a parallel multiscale_co-attention module. Cross-modal information interaction was finally achieved by integrating the above features.

We compared the proposed model on public datasets with previous models. The models obtained the best detection accuracy on both public datasets, as shown in Supplementary Table 3. Compared with the previous best-performing model, the proposed model improved the accuracy, with accuracies of 1.11% and 6.16% for the Apnea-ECG and UCD datasets, respectively. In the Apnea-ECG dataset, Pathinarupothi et al used ECG signal to calculate the instantaneous heart rate and input with SpO_2 signal to the long-short time memory model; the Acc increased to 92.10% after the two-modal signal was fused.[20] Li et al proposed the fusion of ECG, SpO_2, airflow, and chest and abdomen signals for OSA detection by first extracting the time domain, frequency domain, and nonlinear features of the above signals, followed by evaluating and classifying the significance of the features; finally, different categories of features were input to a support vector machine classifier, and the model achieved an optimal Sen of 93.22%.[37] Paul et al obtained RR intervals from ECG signal, and combined RR intervals and SpO_2 signals with a feedforward neural network, and the detection accuracy was 92.00%.[36] In the UCD dataset, Xie et al extracted and selected the time and frequency domain features of ECG and SpO_2 signals and then conducted experiments on ten machine learning classifiers, which suggested that the combination of two signals resulted in an Acc of 84.40%.[38] By utilizing an existing mathematical model of the cardiopulmonary system, Gutta et al proposed the use of the likelihood ratio of the ECG signal to the SpO2 signal to detect OSA, obtaining a detection Acc of 82.33%.[39]

Two main features characterize previous multimodal signals fusion methods: one is that the signals need to be preprocessed, such as extracted RR intervals, heart rate and time-domain features, and the other is that the cross-modal

signals fusion is completed by simple classifiers and neural networks. In contrast, the proposed model directly uses the original signals. Shallow features extraction and selection were accomplished by the LKCNN, 1DSE, and BiGRU blocks. The MM_Transformer architecture was developed for cross-modal information interaction. Thus, this model maximizes feature extraction, selection and fusion capability, and efficiency to obtain the best detection accuracy. However, the Sen performance on the UCD dataset was poor, which was considered to be the result of a severe category imbalance between respiratory event and normal segments.

The construction and application of a hospital dataset with a large sample size was another highlight of our study. Most current OSA detection models are trained and validated on public datasets; however, practical testing on real hospital data is lacking. Therefore, to improve the model's applicability, we constructed a hospital dataset and trained, validated, and tested the model on this dataset. Moreover, increasing the dataset size gives the model a stronger learning ability, which is crucial for improving detection performance and generalizability.

The results were summarized and interpreted in terms of both segments and recordings. In the test set, the Acc and AUC of per-segment detection were 91.38% and 0.968, respectively, indicating that the proposed model can distinguish well between respiratory event and standard segments. The ability of individual detection was evaluated both quantitatively and qualitatively. From the quantitative analysis perspective, the Bland–Altman plots indicated strong agreement between the true and predicted AHI. Overall, 91.18% of the points fell within the 95% consistency interval. When the patients were divided into subgroups according to severity, 83.33%, 86.36%, and 82.93% of the points fell within the 95% consistency intervals for mild, moderate, and severe OSA, respectively. From the qualitative analysis perspective, the Acc and AUC of per-recording detection were 96.08% and 0.922, respectively. Notably, the model had favorable accuracy in both OSA classification and severity assessment (mild, 90.20%; moderate, 88.24%; and severe, 92.16%). The above results indicate that the proposed model has excellent generalizability, which is conducive to extended application in clinical practice. We also analyzed the possible reasons for the incorrectly predicted segments and records. First, the signals were segmented in 1-min time units, leading to underestimation in very severe OSA patients with True-AHI >60. Second, when respiratory events occurred across segments or lasted longer than 1-min, a single respiratory event was labeled as more than one "respiratory event" segment, resulting in an overestimation of Pred-AHI. Finally, there was a time delay between the onset of a respiratory event and the signal features change, leading to underdiagnosis of "respiratory event" segments.

Overall, the main contributions and innovations of our study are as follows: 1) the development of a multiscale Transformer model using ECG and SpO_2 signals, which is innovative in the design of the multimodal information fusion module; 2) the model shows optimal detection performance compared with previous studies; and 3) the construction of a larger sample of the hospital dataset and the experimental results of the hospital dataset demonstrate the applicability and generalizability of the proposed model.

This study has several limitations. First, this was a single-center study. Patients were from the same region and ethnicity, and the racial differences need to be considered, for example, in SpO_2 signal acquisition. And the model must be verified in various acquisition environments and on data obtained via different devices. Second, patients with suspected OSA were enrolled, which may affect the proportion of OSA and its severity levels. Meanwhile, patients received only one night sleep test, and the first-night effect is difficult to avoid. Third, compared to the typical PSG, OSA detection by ECG and SpO_2 signals cannot determine the type of respiratory events and classify sleep stages. Fourth, the computational cost of the model is greater than that of simple structural models. Fifty, the accuracy and stability of the model need to be evaluated and improved.

In the future, we will improve and apply the model in the following directions. First, a multi-center study will be conducted to increase the sample size and the region of the patients. In this process, we need to consider the ethics of medical data inclusion and the security of data collection and storage. Second, the model algorithm will be optimized in terms of adding more relevant features and reducing computational cost to improve the detection performance. Finally, we will design the OSA detection system equipped with the model, focusing on the realistic multi-tasking problem to enhance the clinical application value of the model.

Conclusion

The present study proposed a DL model for OSA detection and severity evaluation according to ECG and SpO_2 signals. The model achieved excellent detection results on both the hospital and public datasets. This approach provided new options for assisting in the diagnosis and grading of OSA.

Data Sharing Statement

Research data are not publicly available but can be obtained from the corresponding author on request after approval from the institutional review boards of all participating institutions.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the Second Affiliated Hospital of Xi'an Jiaotong University (No. 2020-1122). Our study adheres to the principles of the Declaration of Helsinki. All data collection methods were performed per the relevant guidelines and regulations. All participants provided informed consent for data collection and analysis. All data were anonymized.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the National Natural Science Foundation of China (62076198), National Natural Science Foundation of China (82371129) and the Free Exploration and Innovation Project of the Basic Scientific Research Fund of Xi 'an Jiaotong University (xzy012023119). The funding bodies played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

Disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Gottlieb D, Punjabi N. Diagnosis and Management of Obstructive Sleep Apnea: a Review. *JAMA*. 2020;323(14):1389–1400. doi:10.1001/jama.2020.3514
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(8):687–698. doi:10.1016/S2213-2600(19)30198-5
3. Trzepizur W, Blanchard M, Ganem T, et al. Sleep apnea–specific hypoxic burden, symptom subtypes, and risk of cardiovascular events and all-cause mortality. *Am J Respir Crit Care Med*. 2022;205(1):108–117. doi:10.1164/rccm.202105-1274OC
4. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: an American Academy of Sleep Medicine Clinical Practice Guideline. *J Clin Sleep Med*. 2017;13(3):479–504. doi:10.5664/jcsm.6506
5. Bahrami M, Forouzanfar M. Sleep Apnea Detection From Single-Lead ECG: a Comprehensive Analysis of Machine Learning and Deep Learning Algorithms. *IEEE T Instrum Meas*. 2022;71:1–11.
6. Hu S, Cai W, Gao T, Wang M. A Hybrid Transformer Model for Obstructive Sleep Apnea Detection Based on Self-Attention Mechanism Using Single-Lead ECG. *IEEE T Instrum Meas*. 2022;71:1–11.
7. Liu H, Cui S, Zhao X, Cong F. Detection of obstructive sleep apnea from single-channel ECG signals using a CNN-transformer architecture. *Biomed Signal Proc Cont*. 2023;82.
8. Qin H, Liu G. A dual-model deep learning method for sleep apnea detection based on representation learning and temporal dependence. *Neurocomputing*. 2022;473:24–36. doi:10.1016/j.neucom.2021.12.001
9. Shao S, Han G, Wang T, Song C, Yao C, Hou J. Obstructive Sleep Apnea Detection Scheme Based on Manually Generated Features andParallel Heterogeneous Deep Learning Model under IoMT. *IEEE J Biomed Health Inform*. 2022;26(12):5841–5850. doi:10.1109/JBHI.2022.3166859
10. Wang Z, Pan X, Mei Z, et al. ECGAN-Assisted ResT-Net Based on Fuzziness for OSA Detection. *IEEE Trans Biomed Eng*. 2024.
11. Lin Y, Zhang H, Wu W, Gao X, Chao F, Lin J. Wavelet transform and deep learning-based obstructive sleep apnea detection from single-lead ECG signals. *Phys Eng Sci Med*. 2024;47(1):119–133. doi:10.1007/s13246-023-01346-0

12. Li C, Shi Z, Zhou L, et al. Tfformer: a time frequency information fusion based cnn-transformer model for osa detection with single-lead ecg. *IEEE Transactions on Instrumentation Measurement*. 2023.
13. Mostafa SS, Mendonça F, Morgado-Dias F, Ravelo-García A. SpO2 based sleep apnea detection using deep learning. Paper presented at: 2017 IEEE 21st international conference on intelligent engineering systems (INES). 2017.
14. Gutiérrez-Tobal GC, Álvarez D, Crespo A, Del Campo F, Hornero R. Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings. *IEEE j Biomedical Health Inform*. 2018;23(2):882–892. doi:10.1109/JBHI.2018.2823384
15. Vaquerizo-Villar F, Álvarez D, Kheirandish-Gozal L, et al. Convolutional neural networks to detect pediatric apnea-hypopnea events from oximetry. Paper presented at: 2019 41st annual international conference of the IEEE engineering in medicine and biology society 2019.
16. Leino A, Nikkonen S, Kainulainen S, et al. Neural network analysis of nocturnal SpO2 signal enables easy screening of sleep apnea in patients with acute cerebrovascular disease. *Sleep Medi*. 2021;79:71–78. doi:10.1016/j.sleep.2020.12.032
17. Barroso-García V, Gutiérrez-Tobal GC, Gozal D, et al. Wavelet analysis of overnight airflow to detect obstructive sleep apnea in children. *Sensors*. 2021;21(4):1491. doi:10.3390/s21041491
18. Yue H, Lin Y, Wu Y, et al. Deep learning for diagnosis and classification of obstructive sleep apnea: a nasal airflow-based multi-resolution residual network. *Nat Science of Sleep*. 2021;13:361–373. doi:10.2147/NSS.S297856
19. Taghizadegan Y, Jafarnia Dabanloo N, Maghooli K, Sheikhani A. Prediction of obstructive sleep apnea using ensemble of recurrence plot convolutional neural networks (RPCNNs) from polysomnography signals. *Med Hypotheses*. 2021;154:110659. doi:10.1016/j.mehy.2021.110659
20. Pathinarupothi RK, Prathap D, Rangan ES, Gopalakrishnan EA. Single Sensor Techniques for Sleep Apnea Diagnosis Using Deep Learning. Paper presented at: 2017 IEEE International Conference on Healthcare Informatics (ICHI). 2017.
21. Penzel T, Moody GB, Mark RG, Goldberger AL, Peter JH. The apnea-ECG database. *Compu Card*. 2000;27.
22. Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):E215–220. doi:10.1161/01.CIR.101.23.e215
23. Christiano LJ, Fitzgerald TJ. The band pass filter. *Inter economic revi*. 2003;44(2):435–465.
24. Eldele E, Chen Z, Liu C, et al. An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:809–818. doi:10.1109/TNSRE.2021.3076234
25. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. *IEEE T PATTERN ANAL*. 2020;42(8):2011–2023. doi:10.1109/TPAMI.2019.2913372
26. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *Arxiv*. 2014.
27. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Info Proc Syste*. 2017;30.
28. Abboud F, Kumar R. Obstructive sleep apnea and insight into mechanisms of sympathetic overactivity. *J Clin Invest*. 2014;124(4):1454–1457. doi:10.1172/JCI70420
29. May AM, Van Wagoner DR, Mehra R. OSA and Cardiac Arrhythmogenesis: mechanistic Insights. *Chest*. 2017;151(1):225–241. doi:10.1016/j.chest.2016.09.014
30. Orrù G, Storari M, Scano A, et al. Obstructive Sleep Apnea, oxidative stress, inflammation and endothelial dysfunction-An overview of predictive laboratory biomarkers. *Eur Rev Med Pharmacol Sci*. 2020;24(12):6939–6948. doi:10.26355/eurrev_202006_21685
31. Unnikrishnan D, Jun J, Polotsky V. Inflammation in sleep apnea: an update. *Rev Endocr Metab Disord*. 2015;16(1):25–34. doi:10.1007/s11154-014-9304-x
32. Alonso-Fernández A, García-Río F, Racionero M, et al. Cardiac rhythm disturbances and ST-segment depression episodes in patients with obstructive sleep apnea-hypopnea syndrome and its mechanisms. *Chest*. 2005;127(1):15–22. doi:10.1378/chest.127.1.15
33. Corotto PS, Kang H, Massaro B, et al. Obstructive sleep apnea and electrocardiographic P-wave morphology. *Ann Noninvasive Electrocardiol*. 2019;24(4):e12639. doi:10.1111/anec.12639
34. Pallas-Areny R, Colominas-Balague J, Rosell FJ. The effect of respiration-induced heart movements on the ECG. *IEEE Trans Biomed Eng*. 1989;36(6):585–590. doi:10.1109/10.29452
35. Penzel T, Kantelhardt JW, Bartsch RP, et al. Modulations of Heart Rate, ECG, and Cardio-Respiratory Coupling Observed in Polysomnography. *Front Physiol*. 2016;7:460. doi:10.3389/fphys.2016.00460
36. Paul T, Hassan O, Alaboud K, et al. ECG and SpO2 Signal-Based Real-Time Sleep Apnea Detection Using Feed-Forward Artificial Neural Network. Paper presented at: AMIA Jt Summits Transl Sci Proc2022.
37. Li X, Ling SH, Su S. A Hybrid Feature Selection and Extraction Methods for Sleep Apnea Detection Using Bio-Signals. *Sensors*. 2020;20(15).
38. Xie B, Minn H. Real-time sleep apnea detection by classifier combination. *IEEE Trans Inf Technol Biomed*. 2012;16(3):469–477. doi:10.1109/TITB.2012.2188299
39. Gutta S, Cheng Q, Nguyen HD, Benjamin BA. Cardiorespiratory Model-Based Data-Driven Approach for Sleep Apnea Detection. *IEEE J Biomed Health Inform*. 2018;22(4):1036–1045. doi:10.1109/JBHI.2017.2740120