Sensors and Materials, Vol. 37, No. 12 (2025) 5535–5540
MYU Tokyo

S & M 4261

Severity Prediction of Obstructive Sleep Apnea Using Transformed 2D Oxygen Saturation Signals

Yi-Cheng Wu, Cheng-Yu Yeh,* and Chun-Cheng Lin

Department of Electrical Engineering, National Chin-Yi University of Technology,
57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan

(Received June 2, 2025; accepted December 1, 2025)

Keywords: obstructive sleep apnea (OSA), apnea-hypopnea index (*AHI*), oxygen saturation (SpO_2), deep learning

In this paper, we present a deep-learning-based method for predicting obstructive sleep apnea (OSA) severity using 2D oxygen saturation (SpO_2) signal representations. Our previous works have demonstrated that unsegmented overnight 1D SpO_2 signals can achieve high accuracy in predicting the apnea-hypopnea index and classifying OSA severity. Extending this approach, we introduce a novel transformation of the original 1D SpO_2 signal into a 2D representation, enabling it to be processed similarly to an image signal and then trained by deep neural networks. This signal transformation is designed to facilitate subsequent studies into model interpretability. Experimental results show that the proposed approach reaches competitive performance compared with state-of-the-art baselines.

1. Introduction

Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by recurrent upper airway collapse during sleep, leading to intermittent airflow cessation, decreased blood oxygen saturation (SpO_2), and sleep fragmentation.[1] The severity of OSA is typically classified into four categories on the basis of the apnea-hypopnea index (*AHI*), which quantifies the number of apnea and hypopnea events per hour of sleep: normal ($AHI < 5$), mild ($5 \leq AHI < 15$), moderate ($15 \leq AHI < 30$), and severe ($AHI \geq 30$).[2]

The clinical diagnosis of OSA has traditionally depended on polysomnography (PSG) as the gold standard. Despite its well-established diagnostic reliability, PSG implementation faces three critical barriers: (1) prolonged wait times for hospital-based testing appointments, (2) substantial equipment costs, and (3) requirement for specialized technical operation. These limitations have significantly hindered its adoption in home-based settings.

In response to these practical constraints, contemporary research has increasingly focused on developing automated diagnostic systems utilizing single-channel physiological signals. Notably, SpO_2 signals have emerged as a particularly viable alternative due to their dual advantages: (1)

*Corresponding author: e-mail: cy.yeh@ncut.edu.tw
https://doi.org/10.18494/SAM5765

ISSN 0914-4935
https://myukk.org/

noninvasive acquisition through readily available pulse oximeters and (2) strong pathophysiological correlation with respiratory events.

Most existing deep learning studies for OSA prediction still rely on presegmented and annotated signal data, which imposes high manual labeling costs and limits practical applicability.[3–5] In our previous works,[6,7] we demonstrated that deep learning models using unsegmented overnight SpO_2 signals can achieve marked performance in predicting OSA severity. On the basis of the foundation in Ref. 7, in this work, we propose to split the signal into 60 s segments followed by 2D transformation, which is then fed into a convolutional neural network model to perform the *AHI* prediction and OSA severity classification.

Notably, our method maintains the use of unsegmented SpO_2 signals as model input, without requiring manually labeled event segments. The transformed 2D SpO_2 signals can potentially be integrated with visualization techniques[8] to provide clinicians with explainable cues regarding possible abnormal regions in the signal as auxiliary diagnostic information.

2. Methodology and Model

Figure 1 shows the proposed method for transforming a 1D SpO_2 signal into a 2D SpO_2 signal. First, an 8 h segment is extracted from the original SpO_2 signal sampled at 1 Hz, resulting in a total signal length of 28800 samples. This 1D SpO_2 signal is then segmented into 60 s segments, yielding 480 individual segments. Each segment is subsequently arranged as a row in the 2D representation, with the first segment forming the first row, the second segment forming the second row, and so on. This process results in a 2D SpO_2 signal with dimensions of 480 × 60, that is, $N = 480$ in Fig. 1.

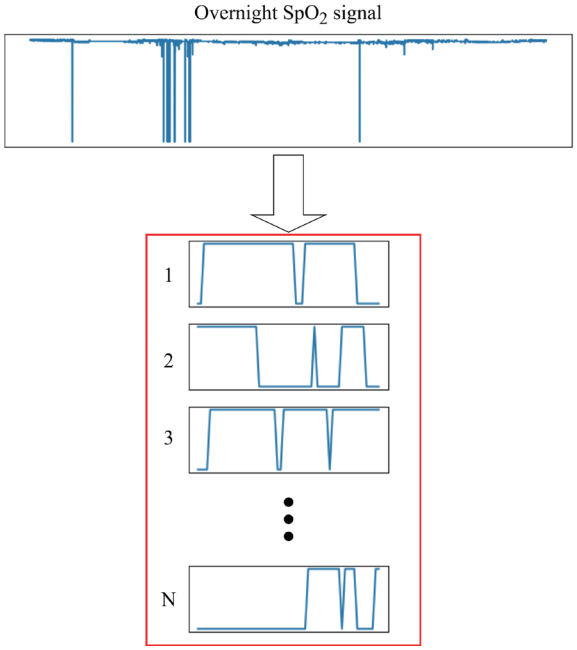

Fig. 1. (Color online) Schematic diagram for transforming 1D SpO_2 signal into 2D SpO_2 signal.

Figure 2 is the framework of the *AHI* prediction model proposed in this paper. The model input is 2D SpO_2 signals with a shape of 480 × 60. The block abbreviated as "Conv2D, (1, 9, 1, 1)" in Fig. 2 indicates that the parameters of kernel size and stride are set to 1 × 9 and 1 × 1, respectively. The dimension of each feature map appears at the lower right corner of the corresponding block in Fig. 2. Similarly to Ref. 7, the feature extraction with multiple receptive fields conducted on input SpO_2 signals is used in the model. This work builds three paths to extract respective feature maps from the input signal. The flowchart of the residual network (ResNet)[9] block, i.e., Res_Block, in Fig. 2 is presented in Fig. 3, where W, C_{in}, and C_o are known quantities.

This work is built by using Python and PyTorch programing with a GeForce RTX 3090 graphics card. A mean square error loss function and an Adam optimizer were used to train the presented model with a batch size of 64 and 500 epochs.

3. Experimental Results

For comparison, we trained and tested the model using the same datasets as those in Refs. 6 and 7, as listed in Table 1. These data comprise three publicly available datasets, namely, the MrOS, MESA, and Sleep Heart Health Study (SHHS) datasets, all of which were obtained from the National Sleep Research Resource.[10] The first two datasets were used as training data, whereas the last was applied as test data. A total of 5988 and 8444 recordings were employed as the training and test sets, respectively.

Figures 4(a) and 4(b) show the 4 × 4 confusion matrices tested on the SHHS1 and SHHS2 datasets, respectively. Additionally, the performance metrics including sensitivity, specificity, precision, $F1$-score, and overall accuracy are listed in Tables 2 and 3. The $F1$-scores of 85.25,

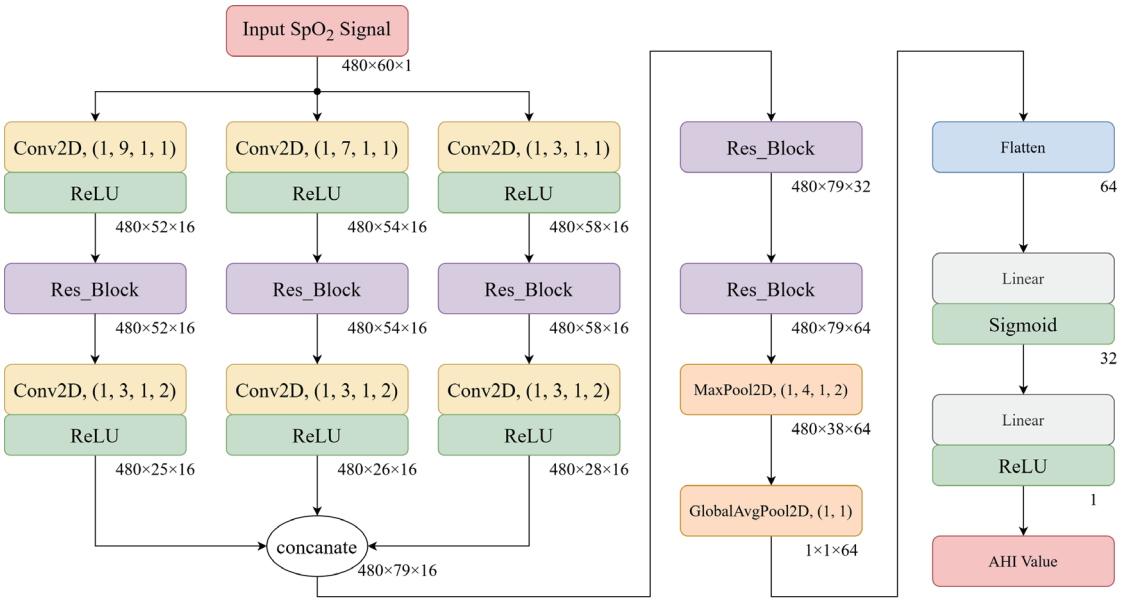

Fig. 2. (Color online) Architecture of the presented AHI prediction model for OSA detection.

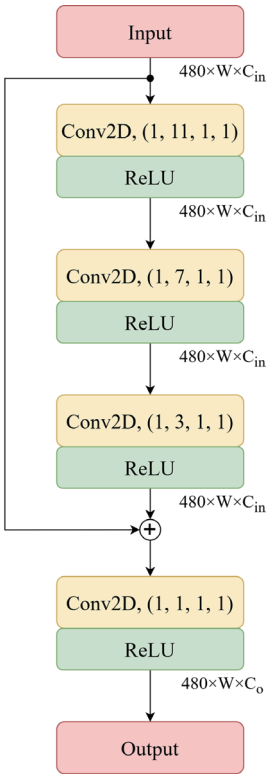

Fig. 3. (Color online) Framework of ResNet block.

Table 1
Training and test datasets used in this work.

Dataset	OSA severity category				Amount
	Normal	Mild	Moderate	Severe	
Training set	1070	2032	1593	1293	5988
Test–SHHS1	1766	2031	1237	759	5793
Test–SHHS2	629	960	637	425	2651

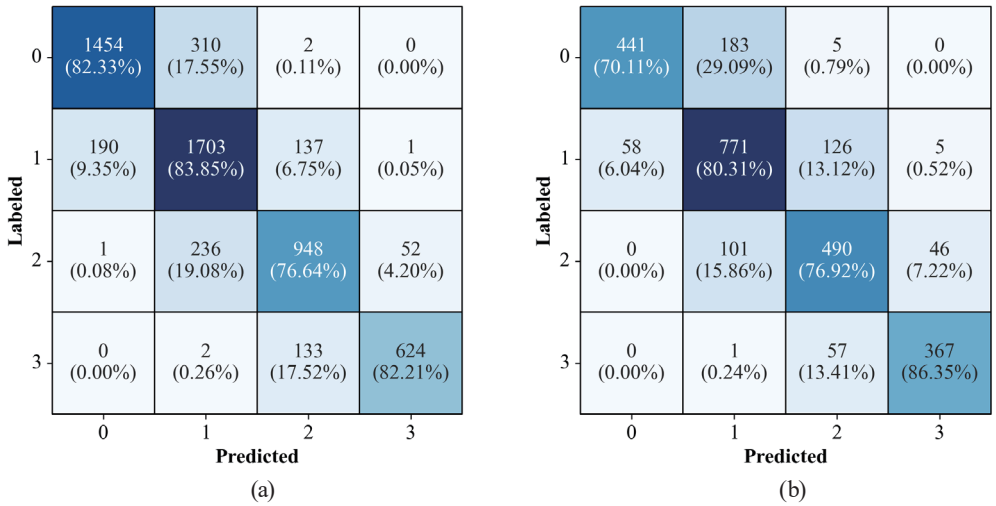

Fig. 4. (Color online) Confusion matrices for performance analysis tested on (a) SHHS1 and (b) SHHS2 datasets.

Table 2
Performance metrics of the presented model tested on SHHS1 dataset.

Category	Sensitivity (%)	Specificity (%)	Precision (%)	*F*1-score (%)
Normal	82.33	95.26	88.39	85.25
Mild	83.85	85.43	75.66	79.54
Moderate	76.64	94.03	77.70	77.17
Severe	82.21	98.95	92.17	86.91
Accuracy (%)		81.63		

Table 3
Performance metrics of the presented model tested on SHHS2 dataset.

Category	Sensitivity (%)	Specificity (%)	Precision (%)	*F*1-score (%)
Normal	70.11	97.13	88.38	78.19
Mild	80.31	83.15	73.01	76.49
Moderate	76.92	90.67	72.27	74.52
Severe	86.35	97.71	87.80	87.07
Accuracy (%)		78.05		

79.54, 77.17, and 86.91% in the categories of normal, mild, moderate, and severe were obtained in the SHHS1 case, whereas in the SHHS2 case, the *F*1-scores of 78.19, 76.49, 74.52, and 87.07% were obtained, respectively. In this work, we obtained overall accuracies of 81.63% in the SHHS1 case and 78.05% in the SHHS2 case for four-level OSA severity classification.

In terms of the overall accuracy, in our original work,[6] we obtained 80.51 and 77.63% in the SHHS1 and SHHS2 cases, meaning that this work gives improvements of 1.12 and 0.42% in the SHHS1 and SHHS2 cases, respectively. We also provided competitive performance compared with that reported in Ref. 7. Note that this study significantly outperforms a counterpart Ref. 5 regarding the accuracy, that is, 81.63 vs 67.11% in the SHHS1 case and 78.05 vs 72.46% in the SHHS2 case.

We aimed to employ 2D representations of SpO_2 signals, in conjunction with visualization techniques, to provide clinicians with interpretable cues regarding potential abnormal regions in the signals as auxiliary diagnostic information. A prerequisite for achieving this goal is to ensure that the 2D SpO_2 representation can maintain high accuracy in predicting OSA severity. Visualization analysis carries meaningful clinical value only when it is supported by a reliable and accurate OSA prediction model.

The experimental results demonstrate that the proposed method achieves overall accuracy comparable to state-of-the-art baselines, thereby confirming its effectiveness. These findings not only validate the feasibility of the approach but also establish a solid foundation for future research on interpretability analysis.

4. Conclusions

In this paper, we presented a deep-learning-based method for predicting *AHI* and OSA severity using 2D representations of SpO_2 signals. Experimental results showed that this work achieves relatively high overall accuracy compared with state-of-the-art approaches. Notably, the method operates directly on unsegmented SpO_2 signals, without the need for manually

labeled event segments. Furthermore, the transformed 2D SpO_2 representations can be integrated with visualization techniques to provide clinicians with explainable cues, highlighting potential abnormal regions in the signal as auxiliary diagnostic information.

References

1 T. Matsumoto, T. Hirai, and K. Chin: Curr. Sleep Med. Rep. **7** (2021) 186.
2 D. W. Hudgel: Sleep **39** (2016) 1165.
3 K. Abu, M. L. Khraiche, and J. Amatoury: Sleep Med. **113** (2024) 260.
4 G. C. Gutiérrez-Tobal, D. Álvarez, F. Vaquerizo-Villar, A. Crespo, L. Kheirandish-Gozal, D. Gozal, F. del Campo, and R. Hornero: Appl. Soft Comput. **111** (2021) 107827.
5 R. Liu, C. Li, H. Xu, K. Wu, X. Li, Y. Liu, J. Yuan, L. Meng, J. Zou, W. Huang, H. Yi, B. Sheng, J. Guan, and S. Yin: Nat. Sci. Sleep **14** (2022) 927.
6 J. W. Chen, C. M. Liu, C. Y. Wang, C. C. Lin, K. Y. Qiu, C. Y. Yeh, and S. H. Hwang: Eng. Appl. Artif. Intell. **122** (2023) 106161.
7 C. Y. Yeh, H. Y. Chi, and C. C. Lin: IEEJ Trans. Electr. Electron. Eng. **20** (2025) 823.
8 R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra: Int. J. Comput. Vis. **128** (2020) 336.
9 K. He, X. Zhang, S. Ren, and J. Sun: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 770. https://doi.org/10.1109/CVPR.2016.90
10 G. Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline: J. Am. Med. Inf. Assoc. **25** (2018) 1351.