Sensors and Materials, Vol. 36, No. 11 (2024) 4695–4712
MYU Tokyo

S & M 3825

Semisupervised Method with Adaptive Adjustment of Threshold for Detecting Obstructive Sleep Apnea Based on Oxygen Saturation Signal

Linqing Yang,[1] Na Ying,[1*] Hongyu Li,[1] Xinyu Lin,[1]
Yinfeng Fang,[1] Yong Zhou,[2] and Huahua Chen[1]

[1]School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Regional Medical Center for National Institute of Respiratory Disease, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou 310016, China

(Received August 7, 2024; accepted October 22, 2024)

Keywords: obstructive sleep apnea, oxygen saturation, deep learning, semisupervised learning

Obstructive sleep apnea (OSA) is a prevalent sleep disorder that seriously affects patients' quality of life and health status. Traditional diagnostic methods are time-consuming and labor-intensive, and apnea detection using deep learning algorithms also faces the problems of insufficient sample size and class imbalance. Therefore, in this paper, we propose a semisupervised apnea detection algorithm (Semi-DynaSeqNet) based on the oxygen saturation (SpO2) signal. In this study, we first extracted local features of the SpO2 signal using a one-dimensional convolutional neural network, and then combined the gate recurrent unit for time series modeling to capture the signal's long-term dynamic features. On this basis, a self-attention mechanism is introduced to further enhance the recognition of key features. Considering the small-sample classification task of the OSA detection, we further proposed the semisupervised learning method with the adaptive adjustment of threshold. By iteratively training a model to generate pseudo-labeled samples of unlabeled pulse oximetry signals and incorporating them into the training set, while adaptively adjusting the semisupervised threshold to fully utilize the unlabeled sample information, we thereby improved the generalization ability of Semi-DynaSeqNet. The experimental results showed that the algorithm proposed in this paper achieves an *F*1-*score* of 90.94% for the model on the St. Vincent's University Hospital/University College Dublin Sleep Apnea Database with one-second detection, whereas the algorithm achieves an *F*1-*score* of 94.65% on the PhysioNet Apnea-ECG Database with one-minute detection, indicating that the algorithm can perform well in both sleep apnea detection tasks with different time scales, demonstrating its flexibility and scalability.

*Corresponding author: e-mail: yingna@hdu.edu.cn
https://doi.org/10.18494/SAM5265

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

1. Introduction

With the fast pace of contemporary life, sleep disorders such as insomnia, difficulty initiating sleep, and snoring have become increasingly prominent, and the quality of sleep has drawn increasing attention. Among these, obstructive sleep apnea (OSA), as one of the common disorders affecting sleep, is characterized by recurrent involuntary episodes of partial or complete reduction of respiration in patients at night.[1] This disorder is not only accompanied by a decrease in blood oxygen saturation, but also triggers autonomic responses, which in turn interfere with normal sleep and usually lead to neurophysiological arousal.[2,3] As a result, patients with OSA often suffer from excessive daytime sleepiness, mood swings, depression, and inattention, which can seriously affect their social, work, and family life. However, it is worrying that owing to the patients' usual lack of knowledge about their symptoms, the diagnosis rate of OSA is relatively low.[4]

In clinical practice, the apnea–hypopnea index (AHI) is widely utilized for evaluating the severity of OSA in patients.[5] The AHI is calculated by assessing the frequency of respiratory pauses and hypopneas per hour during sleep, a process known as polysomnography (PSG).[6] However, PSG monitoring involves collecting multiple physiological signals, including electrocardiogram (ECG), electroencephalogram (EEG), respiratory signals, airflow signals, and oxygen saturation (SpO2), making it a time-consuming and labor-intensive task. For clinicians, analyzing an entire night's PSG record not only requires significant time and effort but also may lead to decreased analysis quality owing to the complexity of the task and fatigue resulting from prolonged analysis.[7]

Given the substantial resource demands in medical PSG analysis, the automated detection of sleep apnea using artificial intelligence algorithms holds significant importance. This approach not only mitigates medical burdens and enhances diagnostic efficiency but also extends its applicability from clinical settings to everyday homes. Specifically focusing on acquiring the SpO2 signal as an optimal measurement technique offers advantages such as portability, user-friendliness and cost-effectiveness.[8] Research has established a clear link between the SpO2 signal and OSA; hence, there is an urgent need for developing automated OSA identification technology based on the SpO2 signal. Further exploration could enable patients to self-identify sleep apnea syndrome through the readily available SpO2 signal readings—providing robust support for early intervention and treatment.

In recent years, the rapid advancement of deep learning technologies has led to their expanding applications in the medical field. In the domain of automated OSA detection systems, significant progress has also been made using these technologies. The SpO2 signal is readily obtainable, and leveraging it for OSA detection can more efficiently identify and diagnose this prevalent yet frequently overlooked sleep disorder. In the domain of OSA detection, researchers have been at the forefront of innovation and advancement by constructing and refining deep learning models. Pourbabaee *et al.*[9] devised a dense recurrent convolutional neural network (DRCNN) capable of accurately identifying various sleep disorders, such as arousals, apneas, and hypopneas. John *et al.*[10] proposed SomnNET, a one-dimensional convolutional neural network (1D-CNN) designed specifically for the real-time detection of OSA at a frequency of 1

event per second. Chaw *et al.*[11] proposed a deep CNN model utilizing the SpO2 signal from smart sensors to predict OSA with superior accuracy compared with traditional machine learning approaches. Wang *et al.*[12] proposed four methods based on convolutional and long short-term memory neural networks that use raw data from three respiratory signals (32 Hz sampling of nasal flow, abdomen, and chest) to predict OSA. Levy *et al.*[13] designed OxiNet, a deep learning model capable of estimating the AHI based on SpO2 measurements. The model has demonstrated an exceptionally low false-negative rate across patients of diverse races, ages, genders, and comorbidities. Yook *et al.*[14] employed a combined approach utilizing nasal airflow (RF), SpO2, and ECG signals obtained during PSG to enhance the accuracy of sleep apnea/hypopnea detection and OSA severity screening. Their study utilized the Xception network and integrated demographic data to improve model performance, achieving high accuracy in detection and screening tasks. Jimenez-Garcia *et al.*[15] proposed an interpretable architecture that combines convolutional and recurrent neural networks for pediatric OSA detection and severity assessment while leveraging gradient-weighted class activation mapping (Grad-CAM) to enhance model explainability. Chi *et al.*[16] introduced a model capable of directly predicting the AHI from the unsegmented overnight SpO2 signal, showcasing the potential of deep learning in processing the SpO2 signal. The relevant work is summarized in Table 1. These studies not only demonstrate the significant potential of deep learning in medical diagnosis, but also offer novel tools and methodologies for the early detection, treatment, and management of OSA. The deep learning techniques employed in these models primarily rely on CNNs and RNNs. However, there are still limitations in handling the dynamic time series data of the SpO2 signal, such as the inability to capture long-range dependences and identify crucial features. Therefore, in this paper, we propose the DynaSeqNet model, which harnesses the robust data processing capabilities of deep learning models to extract key physiological information from the SpO2

Table 1
Brief introduction of related work.

Author	Model	Description
Pourbabaee *et al.*[9]	DRCNN	Devised a DRCNN capable of accurately identifying various sleep disorders such as arousals, apneas, and hypopneas
John *et al.*[10]	SomnNET	Proposed SomnNET, a 1D-CNN designed specifically for the real-time detection of OSA
Chaw *et al.*[11]	CNN	Utilized intelligent sensors for the acquisition of the SpO2 signal and designed a deep CNN model for the prediction of OSA
Wang *et al.*[12]	CNN, LSTM	Proposed four methods based on convolutional and long short-term memory neural networks that use raw data from three respiratory signals (32 Hz sampling of nasal flow, abdomen, and chest) to predict OSA
Levy *et al.*[13]	OxiNet	Designed OxiNet, a deep learning model capable of estimating the AHI based on SpO2 measurements
Yook *et al.*[14]	Xception	Proposed the Xception network and used RF, SpO2, and ECG signals obtained during PSG recordings to improve the accuracy of sleep apnea/hypopnea detection and OSA severity screening
Jiménez-García *et al.*[15]	CNN, RNN	Proposed an interpretable architecture that integrates convolutional and recurrent neural networks to detect pediatric OSA and assess its severity, with the utilization of Grad-CAM to enhance model interpretability

signal, thereby enabling the precise monitoring of OSA. The main contributions of this study are as follows.

(1) We propose a semisupervised algorithm for OSA detection using the SpO2 signal to address the challenge of sparse samples in the small-sample classification task of OSA detection by integrating unlabeled data, effectively mitigating overfitting caused by insufficient samples and significantly enhancing the model's generalization capability.
(2) The proposed semisupervised algorithm incorporates an adaptive threshold adjustment mechanism, which can dynamically adjust the classification boundary according to the distribution of the SpO2 signal categories, optimize the model's recognition accuracy for abnormal samples belonging to a minority class, and improve the fairness and accuracy of the model.
(3) We propose the Semi-DynaSeqNet model, which integrates a 1D-CNN and the gate recurrent unit (GRU) to concurrently capture local features and long-term dynamic trends in the SpO2 signal. This integration enhances sensitivity to subtle changes in the SpO2 signal. Moreover, Semi-DynaSeqNet incorporates the self-attention mechanism, thereby enhancing the model's capacity to capture long-term dependences in time series and improving its sensitivity to critical events within the sequence.

The sections of this paper are organized as follows. In Sect. 1, we describe the related work of OSA detection and the main contributions of this paper. In Sect. 2, we introduce the architecture of the Semi-DynaSeqNet model and its specific implementation steps. In Sect. 3, we show the experimental results of the algorithm on different databases. In Sect. 4, we summarize the whole paper and present our outlook.

2. Methods

Figure 1 depicts the Semi-DynaSeqNet architecture, designed to boost minority class recognition accuracy in sleep apnea detection through semisupervised learning with adaptive thresholds and optimized time series processing. Initially, a 1D-CNN extracts and represents local SpO2 signal features, facilitating sequence analysis. A Gated Recurrent Unit then models time series, capturing long-term signal dependences and enhancing temporal sensitivity. A self-attention mechanism focuses on critical features, refining feature capture. Additionally, Semi-DynaSeqNet integrates semisupervised learning with adaptive thresholds, iteratively generating pseudo-labels to leverage unlabeled data. This enhances model generalization and classification performance, particularly with small samples. The following text will provide a detailed introduction of the main method.

2.1 Extraction and representation of local features

In signal processing, the effective information contained in the original sequence is often difficult to completely extract when the data dimensionality is low. To solve this problem, we employed the CNN to increase dimensionality on the raw SpO2 signal, which is conducive to extracting rich information from different dimensions. The operation of dimensionality

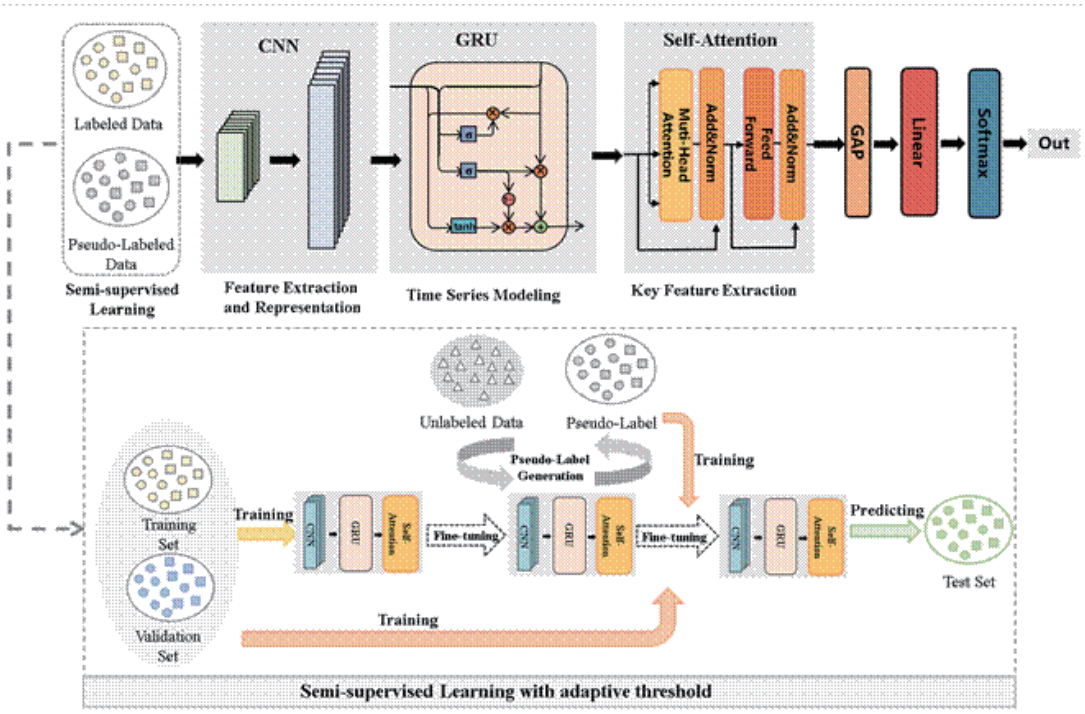


Fig. 1. (Color online) Architecture of Semi-DynaSeqNet.

augmentation enhances the network's feature learning ability and improves its performance in classification tasks. CNNs were initially widely used in the field of image recognition, and owing to the 2D matrix nature of images, the internal structures of the networks, such as the convolution kernel and the feature mapping system, were also primarily designed using 2D designs. However, when faced with 1D data such as the SpO2 signal, the traditional 2D-CNN performs poorly because it cannot effectively capture the features contained in 1D signals. The 1D-CNN is characterized by parameter sharing and sparse connectivity, which enables the network to reduce the number of parameters, lower computational complexity, and improve the model's generalization ability when extracting features. Therefore, we chose to use the 1D-CNN to extract features from the SpO2 signal.

First, the SpO2 signal after windowed segmentation is layer-normalized to obtain the input $\boldsymbol{I}_{Signal}$ for the convolutional layer, ensuring the uniformity of the input data distribution and facilitating subsequent feature extraction. Subsequently, the 1D-CNN layer is used to deeply explore the features of $\boldsymbol{I}_{Signal}$, and no pooling layer is added during the design of this layer to preserve its adjacent position information. Finally, the local feature embedding vector $\boldsymbol{F}_{Local}$ generated on the basis of the changes in local trends is obtained, providing a rich feature representation for the subsequent analysis of the SpO2 signal. The working principle of the 1D-CNN is shown in Eq. (1).

$$\boldsymbol{F}_{Local} = \boldsymbol{W}_{CNN} * \boldsymbol{I}_{Singal} \tag{1}$$

Here, $\boldsymbol{F}_{Local}$ is the output, $\boldsymbol{W}_{CNN}$ is the convolutional layer weight matrix, $\boldsymbol{I}_{Signal}$ is the SpO2 signal that serves as the input, and $*$ is the convolution operation.

Specifically, the convolutional layer convolves the input $\boldsymbol{I}_{Signal}$ by sliding a fixed size window and extracts local features $\boldsymbol{F}_{Local}$ on the basis of the amplitude trend changes of the SpO2 signal within the window. These local features $\boldsymbol{F}_{Local}$ are then mapped to the next layer, forming a more abstract feature representation. The layer-by-layer extraction and mapping approach enables the 1D-CNN to effectively capture short-term fluctuations and periodic patterns from the signal.

The 1D-CNN has significant advantages in processing 1D data, as it can effectively extract feature information from signals through dimensionality increase and layer-by-layer convolution operations, thereby improving the performance of classification tasks. Therefore, the 1D-CNN is used in the feature extraction module to process the SpO2 signal data to achieve better results in the subsequent classification task.

2.2 Sequence modeling and dynamic feature extraction

A CNN has limitations in processing long sequences, which cannot fully perceive global information and is prone to lose positional information during processing. When data is fed to the attention mechanism in parallel, the lock of positional information can lead to the inaccurate judgment of trends based on adjacent times. On the basis of this issue, we introduce GRU [17] into the algorithm for the time series modeling of the SpO2 signal. GRU dynamically adjusts the retention and transmission of information through its internal gating mechanism, which can adapt to changes in time series data and better capture dynamic characteristics.

GRU exhibits powerful capabilities in processing dynamically changing sequence data owing to its unique gating mechanism and memory units. It can learn and retain key dynamic features in time series, providing deep insights into complex data patterns for models. By integrating GRU, embedding vectors containing implicit relative position information can be generated, which not only enrich the representation of data, but also enable the model to capture long-range dependences between sequences. The improvement significantly enhances the model's ability to process long sequence data, providing more accurate and comprehensive information for the subsequent attention mechanism, thereby improving the model's ability to discriminate trends in time series data. The module uses GRU to capture the dynamic features of the SpO2 signal. The specific working principle is as follows.

$$\boldsymbol{r}^t = \sigma\left(\boldsymbol{W}_r \boldsymbol{F}_{Local}^t + \boldsymbol{U}_r \boldsymbol{F}_{Dynamic}^{t-1} + \boldsymbol{b}_r\right), \tag{2}$$

$$\boldsymbol{z}^t = \sigma\left(\boldsymbol{W}_z \boldsymbol{F}_{Local}^t + \boldsymbol{U}_z \boldsymbol{F}_{Dynamic}^{t-1} + \boldsymbol{b}_z\right), \tag{3}$$

$$\widetilde{\boldsymbol{h}^t} = \tanh\left(\boldsymbol{W}_c \boldsymbol{F}_{local}^t + \boldsymbol{U}\left(\boldsymbol{r}^t \bullet \boldsymbol{F}_{Dynamic}^{t-1}\right)\right), \tag{4}$$

$$\boldsymbol{F}_{Dynamic}^{t} = \boldsymbol{z}^{t} \bullet \boldsymbol{F}_{Dynamic}^{t-1} + \left(1 - \boldsymbol{z}^{t}\right) \bullet \widetilde{\boldsymbol{h}^{t}} \ , \tag{5}$$

$$\boldsymbol{F}_{Dynamic} = \left[\boldsymbol{F}_{Dynamic}^{1}, \boldsymbol{F}_{Dynamic}^{2}, \ldots, \boldsymbol{F}_{Dynamic}^{t} \right], \tag{6}$$

where $\boldsymbol{r}^t$ is the reset gate, $\boldsymbol{z}^t$ is the update gate, $\widetilde{\boldsymbol{h}^t}$ is the candidate hidden state, $\boldsymbol{F}_{Dynamic}^{t}$ is the dynamic feature at the current moment, $\boldsymbol{F}_{Dynamic}$ is the final output, $\boldsymbol{W}$ and $\boldsymbol{U}$ are the weight matrices of the GRU layer, $\boldsymbol{b}$ is the bias term, $\boldsymbol{F}_{Local}^{t}$ is the local feature at the current moment, and $\boldsymbol{F}_{Dynamic}^{t-1}$ is the dynamic feature at the previous moment. σ is the sigmoid activation function, σ is the hyperbolic tangent function, and the calculations are as follows.

$$\sigma\left(z\right) = \frac{1}{1 + e^{-z}} \tag{7}$$

$$\tanh\left(z\right) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{8}$$

GRU receives the local features of the current input $\boldsymbol{F}_{Local}$ and the hidden state from the last time, and uses reset and update gates to control the degrees of discarding and retaining $\boldsymbol{F}_{Local}^{t}$ at each time. The reset gate determines the impact of previous information on the current input, while the update gate determines the weight of the current input and previous information in the current hidden state. Subsequently, GRU calculates candidate hidden states and combines them with $\boldsymbol{F}_{Dynamic}^{t-1}$ through the update gate to obtain $\boldsymbol{F}_{Dynamic}^{t}$, which can be used as both current output and input next time. Through continuous iteration, GRU can better capture the long-term dependences relationship and dynamic features in the SpO2 signal and obtain $\boldsymbol{F}_{Dynamic}$ as the final output in the module.

2.3 Enhancement of key features

In the GRU model, inputs for all moments are considered equally weighted. To further enhance the model's attention of key features in the SpO2 signal, we introduce the self-attention layer. As the core part of the Transformer[18] model, the self-attention mechanism adaptively selects and weighs key features in the input sequence by calculating self-attention weights, thereby enhancing the model's ability to recognize and process important information. The specific steps for using self-attention to extract key features in this module are as follows.

(1) The inputs $\boldsymbol{F}_{Dynamic}$ multiply each of the three weight matrices $\boldsymbol{W}^{Q}$, $\boldsymbol{W}^{K}$, and $\boldsymbol{W}^{V}$ separately to obtain three matrix queries $\boldsymbol{Q}$, keys $\boldsymbol{K}$, and values $\boldsymbol{V}$.

$$\boldsymbol{Q} = \boldsymbol{W}^{Q} \boldsymbol{F}_{Dynamic}$$

$$\boldsymbol{K} = \boldsymbol{W}^{K} \boldsymbol{F}_{Dynamic} \tag{9}$$

$$\boldsymbol{V} = \boldsymbol{W}^{V} \boldsymbol{F}_{Dynamic}$$

(2) The correlation between the dynamic features of each pair of input vectors $\boldsymbol{f}_{Dynamic}$ is calculated using the obtained $\boldsymbol{Q}$ and $\boldsymbol{K}$ to obtain the attention score $\boldsymbol{A}$.

$$\boldsymbol{A} = \boldsymbol{Q} \cdot \boldsymbol{K}^{T} \tag{10}$$

(3) Softmax calculation is performed on matrix $\boldsymbol{A}$ to obtain $\boldsymbol{A'}$.
(4) The key feature $\boldsymbol{F}_{Key}$, which is the output of the self-attention layer determined using the obtained $\boldsymbol{A'}$ and $\boldsymbol{V}$, is calculated as

$$\boldsymbol{F}_{\boldsymbol{Key}} = \boldsymbol{V} \cdot \boldsymbol{A'}. \tag{11}$$

In detail, the parameters such as queries, keys, and values in the self-attention mechanism enable the model to calculate the similarity and dependence between the values of the SpO2 signal at different times when processing the signal. Through this mechanism, Semi-DynaSeqNet can sensitively capture subtle signal changes, especially the abnormal patterns that are closely related to OSA. For example, when there is a sharp decrease or sustained low level of blood oxygen saturation, which are potential signs of OSA, the self-attention mechanism is key to helping models accurately identify these signs. Under the effect of the self-attention mechanism, Semi-DynaSeqNet can adaptively select and weigh key features in the SpO2 signal on the basis of attention weights. It means that Semi-DynaSeqNet can focus on information that is crucial for task determination while ignoring irrelevant details. The introduction of the self-attention mechanism enables the model to process signals more accurately and efficiently, bringing higher accuracy and reliability to the detection of OSA.

Finally, the obtained output $\boldsymbol{F}_{Key}$ is put through average pooling computation and fully connected layers to obtain the result $\boldsymbol{Z}$. The calculation is as follows.

$$\boldsymbol{Z} = \boldsymbol{W}_{Linear} \cdot GAP(\boldsymbol{F}_{Key}) + \boldsymbol{B}_{Linear}, \tag{12}$$

where $\boldsymbol{W}_{Linear}$ denotes the weight matrix of the final fully connected output layer, $\boldsymbol{B}_{Linear}$ is the bias term, and GAP is the average pooling computation.

2.4 Semisupervised method with adaptive adjustment of threshold

By combining the above modules, the DynaSeqNet model is obtained. In this section, we incorporate the semisupervised method to obtain the final Semi-DynaSeqNet model proposed in this paper.

The number of samples in the published dataset that contains the SpO2 signal is relatively small, and the sample categories are unbalanced in the actual tasks of OSA detection. Therefore, we propose the semisupervised method with the adaptive adjustment of threshold. By using unlabeled samples to generate pseudo-labels and jointly training with labeled samples, the generalization ability of Semi-DynaSeqNet has been improved. The method enables Semi-DynaSeqNet to demonstrate superior performance in complex scenarios.

The semisupervised method with the adaptive adjustment of threshold is shown in Fig. 1. First, select the labeled sample training set to train the DynaSeqNet model. Through periodic evaluations of the performance of the DynaSeqNet model on the validation set, the best performing Semi-DynaSeqNet model is selected for storage. Then, put the unlabeled data into the Semi-DynaSeqNet model and obtain the output z by inferring and predicting the input samples. The SoftMax function is used to transform the probability distribution and obtain the category confidence $\boldsymbol{p}$ of the OSA classification task. The calculation is as follows.

$$\boldsymbol{p}_0 = SoftMax\left(\boldsymbol{z}_0\right) = \frac{e^{\boldsymbol{z}_0}}{e^{\boldsymbol{z}_0} + e^{\boldsymbol{z}_1}}, \tag{13}$$

$$\boldsymbol{p}_1 = SoftMax\left(\boldsymbol{z}_1\right) = \frac{e^{\boldsymbol{z}_1}}{e^{\boldsymbol{z}_0} + e^{\boldsymbol{z}_1}}, \tag{14}$$

where $\mathbf{z}_0$ is the output of the sample with the category of 0, $\mathbf{z}_0$ is the output of the sample with the category of 1, $\boldsymbol{p}_0$ is the confidence level of the sample with the category of 0, and $\boldsymbol{p}_1$ is the confidence level of the sample with the category of 1.

Then, the adaptive threshold is calculated to infer the distribution of unlabeled data categories, and the adaptive threshold τ is recursively calculated on the basis of the distribution information. On the basis of the threshold τ, high-confidence samples are selected to generate pseudo-labels, thereby improving the accuracy and reliability of the pseudo-labels. Afterwards, the pseudo-labeled and labeled data will be included in the Semi-DynaSeqNet training to further fine-tune the model. Finally, the adaptive adjustment of threshold is repeated multiple times. During iteration, the adaptive threshold is dynamically adjusted to obtain high-quality pseudo-labels and achieve the continuous optimization of semisupervised learning. The calculations of the adaptive threshold are as follows.

$$R_t = \frac{\underset{s}{max}\left(N_{s(t)}\right)}{\sum_{s=0}^{S-1} N_{s(t)}}, \tag{15}$$

$$\tau_t = \frac{1}{S} + \left(1 - \frac{1}{S}\right) \times \frac{t}{T} \times R_{t-1}, \tag{16}$$

where S is the number of categories in the OSA classification task, T is the total number of iterations of the reprogrammed semisupervised learning, t is the number of iterations currently in progress in the semisupervised task, $N_{s(t)}$ is the number of pseudo-labels generated for the s-th class sample in the t-th training, $\max_s(N_s)$ is the number of sample categories with the most generated pseudo-labels, and R_t is the ratio of majority class samples among all samples generated in the t-th iteration.

At the beginning, the threshold τ is set lower to facilitate the addition of more unlabeled samples to the training. With the increase in the number of iterations (t), the threshold τ gradually increases to incorporate more accurate pseudo-labeled data into training, which improve the classification performance of OSA and promote the continuous optimization of semisupervised learning. Moreover, for categories with fewer samples, a lower threshold is set to increase the number of pseudo-labels and balance the sample distribution between categories. For categories with a large sample size, a higher threshold is set to ensure the accuracy and reliability of pseudo-labels. The semisupervised method with the adaptive adjustment of threshold is helpful in effectively utilizing the unlabeled SpO2 signal and improving the overall performance of the Semi-DynaSeqNet model. The pseudocode of the Semi-DynaSeqNet model is shown in Algorithm 1.

Algorithm 1
Semi-DynaSeqNet.

Input: the Spo2 singal set I_{Singal}, the number of iterations of the semisupervised learning T, the number of categories in the task of OSA classification S.
Output: pseudo-labeled data set Data.

Initialize Data = set()
Initialize $R_{t-1} = 0.5$
for t = 1 to T do
  /* calculate features */
  $F_{Local} = W_{CNN} * I_{Singal}$
  $F_{Dynamic} = GRU(F_{Local})$
  $F_{Key} = Attention(F_{Dynamic})$
  $Z = W_{Linear} \cdot GAP(F_{Key}) + B_{Linear}$
  $P = softmax(Z)$
  /* calculate the adaptive threshold */
  $\tau_t = \frac{1}{S} + \left(1 - \frac{1}{S}\right) \times \frac{t}{T} \times R_t$.
  /* obtain pseudo-labeled data set*/
  for s in 1 to S do
    $N_{s(t)} = 0$
  done
  for i in 1 to $|I_{Singal}|$ do
    $s = argmax(P_i)$
    $input = I^i_{Singal}$
    $N_{s(t)} = 1 + N_{s(t)}$
    Data.add([$input$, s])
  done
  $R_t = \frac{max_s\left(N_{s(t)}\right)}{\sum_{s=0}^{S-1} N_{s(t)}}$
done
return Data

3. Experiments and Results

3.1 Datasets

(1) PhysioNet Apnea-ECG Database (PAEDB)[19]

PAEDB contains 70 single-lead ECG recordings, all of which were collected at a sampling rate of 100 Hz using modified lead V2. Among them, eight recordings (numbered a01 to a04, b01, and c01 to c03) were additionally equipped with four important signals, namely, Resp C and Resp A (breathing effort signals obtained through the inductive volume recording of the chest and abdomen), Resp N (nasal thermistor-measured oral and nasal airflow signals), and SpO2 (oxygen saturation signals). These additional signals provide valuable information on the breathing and oxygenation status of the subjects. PAEDB covers male and female subjects aged 27 to 63, with a weight range of 53 to 135 kg. The sleep events of each participant were marked by sleep experts in units of minutes. The annotations include "A" labels for apnea and "N" labels for normal sleep. In this study, only the eight individual recordings with SpO2 signals were selected from the database.

(2) St. Vincent's University Hospital/University College Dublin Sleep Apnea Database (UCDDB) [20]

UCDDB includes polysomnographic data from 25 participants, with sleep durations ranging from 5.9 to 7.7 h. The participants (21 males and 4 females) in the database range in age from 28 to 68 years, and their weights range from 59.8 to 128.6 kg. The physiological signals recorded in the database include electroencephalography, electrocardiography, nasal and oral airflow, thoracic movement, respiratory sounds, and blood oxygen saturation. Sleep events for each participant were precisely marked by sleep experts in units of seconds. Additionally, the peripheral blood oxygen saturation signal used in this study was a detailed data recorded at a sampling rate of 8 Hz.

3.2 Experimental settings and environment

The experiments were conducted using Python 3.6 and PyTorch 1.6.0 versions on a machine equipped with AMD Ryzen 5 5600H series CPU and NVIDIA GeForce RTX 3080 GPU. The sampling rate of the SpO2 signal from the two databases was uniformly processed to 8 Hz, the batch size of the experiment was set as 32, and the initial learning rate was set as 0.0001.

3.3 Evaluation metrics

In practice, classification tasks often face the challenge of data imbalance. In response to this challenge, the experiments adopt a variety of evaluation metrics to comprehensively measure the performance of the model, including *accuracy*, *precision*, *recall* and *F*1-*score*. Among them, accuracy reflects the ratio of the number of samples correctly classified by the model to the total number of samples, which is an important measure of the overall performance of the model. Precision rate mainly measures the proportion of samples predicted by the model to be positive

cases that are actually positive cases, which reflects the reliability or precision of the model to find positive cases. Recall rate focuses on the proportion of all samples that are actually positive cases that are correctly predicted by the model, which reflects the model's ability or sensitivity to find positive cases. *F*1-*score* is the reconciled average of precision and recall rates, especially in the case of unbalanced data, which integrates the precision and recall rates of the classifier and provides a comprehensive evaluation perspective. They are defined in Eqs. (17)–(20).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{20}$$

Here, *TP* is the sample correctly identified as OSA, *TN* is the sample correctly identified as normal, *FP* is the normal sample incorrectly identified as OSA, and *FN* is the OSA sample incorrectly identified as normal.

3.4 Experimental comparison of different baseline networks

To validate the effectiveness of Semi-DynaSeqNet in this paper, experiments were conducted on PAEDB and UCDDB datasets, with one-minute- and one-second-based time segmentations, respectively, and the experimental results were compared with those obtained by other methods. (1) Short time series apnea detection based on one second

In this experiment, the sleep data of the whole night is first segmented in one-minute window lengths, with the window movement step set to one second. Each window is taken as a sample, where the label of the 2nd second is taken as the window label. The experiments use the UCDDB dataset as the labeled training and testing set, while the PAEDB dataset is added as the unlabeled samples to introduce semisupervised learning, improve the sample size, and increase the robustness. The experimental results are shown in Table 2 and Fig. 2.

In Table 2, DynaSeqNet has the best scores among all comparison methods. The scores of DynaSeqNet under the four metrics are 0.8475, 0.9518, 0.7639 and 0.8626, respectively. Among them, *F*1-*score* has improved by 1.32% compared to the model combining LSTM and CNN[12]. Compared to other models, DynaSeqNet enhances the key feature extraction of the SpO2 signal and improves the detection rate of the algorithm by introducing the self-attention mechanism. The *F*1-*score*, *precision*, *recall*, and *accuracy* of the Semi-DynaSeqNet method with

Table 2
Comparison of model performance based on one-second detection.

Model	*F*1-*score*	*Precision*	*Recall*	*Accuracy*
CNN[10]	0.8143	0.9933	0.6900	0.8427
LSTM[21]	0.8324	0.9423	0.7454	0.8470
LSTM+CNN[12]	0.8343	0.9100	0.7702	0.8499
DynaSeqNet	0.8475	0.9518	0.7639	0.8626
Semi-DynaSeqNet	0.9094	0.9177	0.9013	0.9102

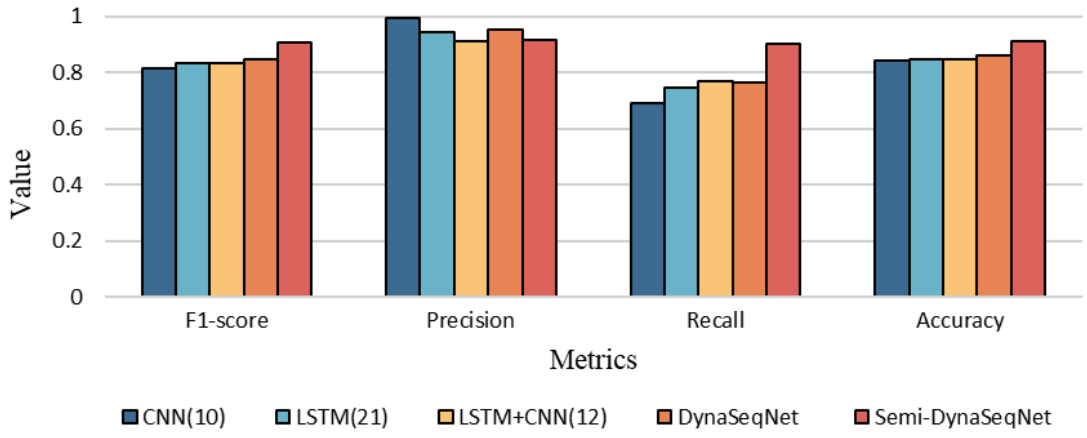


Fig. 2. (Color online) Model performance based on one second.

semisupervised learning are 0.9094, 0.9177, 0.9013 and 0.9102, respectively. The *F*1-*score* has improved by 6.19% compared to DynaSeqNet. In the original dataset, the sample distribution is unbalanced, and there is a phenomenon where the number of normal samples far exceeds the number of OSA samples. The introduction of the semisupervised method compensates for the lack of OSA samples in the original dataset and substantially increases the score of *recall*.

(2) Long time series apnea detection based on one minute

In this experiment, the SpO2 signal of the whole night is partitioned in one-minute window lengths with no overlap between neighboring windows, and each minute data corresponds to a label as a sample. The experiment uses the PAEDB dataset as the labeled training and testing set, while the UCDDB dataset is added as the unlabeled samples to introduce semisupervised learning, increase the sample size, and enhance the generalization ability. The experimental results are shown in Table 3 and Fig. 3.

As shown in Table 3, DynaSeqNet has the best scores in *F*1-*score*, *precision*, *recall*, and *accuracy* among all the comparison methods. The scores of DynaSeqNet under the four metrics are 0.9170, 0.9364, 0.8984, and 0.9229, respectively. Among them, *F*1-*score* has improved by 3.95% compared to the model combining LSTM and CNN[12]. The Semi-DynaSeqNet method, which introduces semisupervised learning, achieved an *F*1-*score* improvement of 2.95% compared to DynaSeqNet. The scores of Semi-DynaSeqNet under the four metrics are 0.9465, 0.9583, 0.9350, and 0.9499, respectively. By integrating unlabeled data, Semi-DynaSeqNet

Table 3
Comparison of model performance based on one-minute detection.

Model	*F*1-*score*	*Precision*	*Recall*	*Accuracy*
CNN[10]	0.8651	0.8862	0.8450	0.8690
LSTM[21]	0.8731	0.9228	0.8285	0.8728
LSTM+CNN[12]	0.8775	0.9024	0.8538	0.8805
DynaSeqNet	0.9170	0.9364	0.8984	0.9229
Semi-DynaSeqNet	0.9465	0.9583	0.9350	0.9499

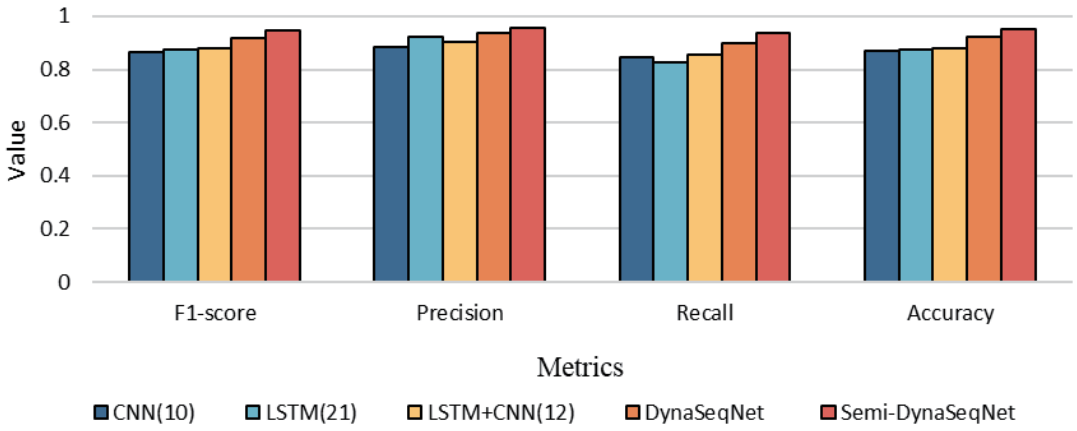


Fig. 3. (Color online) Model performance based on one-minute detection.

effectively alleviates overfitting caused by insufficient samples and significantly improves the performance of the model. The results verify the necessity and validity of incorporating the semisupervised method with the adaptive adjustment of threshold in the design of the proposed algorithm.

3.5 Ablation experiments with semisupervised methods

To validate the effectiveness of the semisupervised method with the adaptive adjustment of threshold, semisupervised ablation experiments were designed. The results based on one-second detection are summarized in Table 4 and Fig. 4, and the results based on one-minute detection are summarized in Table 5 and Fig. 5.

From Table 4, it can be seen that the *F*1-*score* of the DynaSeqNet model is 0.8475 before incorporating the semisupervised method. After incorporating the semisupervised method with the adaptive adjustment of threshold, the *F*1-*score* of the Semi-DynaSeqNet model gradually improved with the increase in the number of iterations. Finally, the *F*1-*score* of the Semi-DynaSeqNet model is 0.9094. After seven iterations, the *F*1-*score* of the Semi-DynaSeqNet model has improved by 5.67% compared to the first iteration.

In Table 5, the *F*1-*score* of the DynaSeqNet model is 0.9190 before incorporating the semisupervised method. After incorporating the semisupervised method with the adaptive adjustment of threshold, the *F*1-*score* of the Semi-DynaSeqNet model gradually improved with

Table 4
Results of semisupervised experiments based on one-second detection.

No. of iterations	*F*1-*score*	*Precision*	*Recall*	*Accuracy*
0	0.8475	0.9518	0.7639	0.8626
1	0.8527	0.9304	0.7870	0.8640
2	0.8633	0.9225	0.8112	0.8715
3	0.8677	0.8862	0.8499	0.8704
4	0.8800	0.8555	0.9059	0.8764
5	0.8960	0.8824	0.9099	0.8943
6	0.9036	0.8849	0.9232	0.9016
7	0.9094	0.9177	0.9013	0.9102

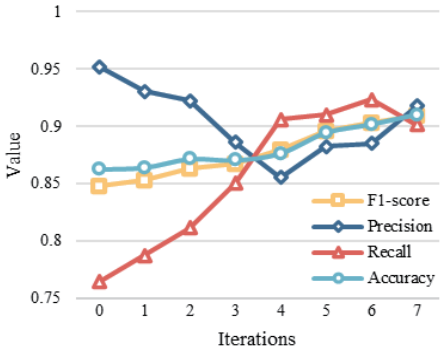


Fig. 4. (Color online) Results of semisupervised experiments based on one-second detection.

Table 5
Results of semisupervised experiments based on one-minute detection.

No. of iterations	*F*1-*score*	*Precision*	*Recall*	*Accuracy*
0	0.9170	0.9364	0.8984	0.9229
1	0.9281	0.9378	0.9187	0.9326
2	0.9400	0.9578	0.9228	0.9441
3	0.9419	0.9619	0.9228	0.9461
4	0.9446	0.9540	0.9267	0.9468
5	0.9465	0.9583	0.9350	0.9499

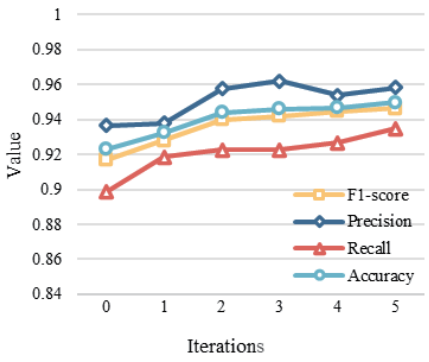


Fig. 5. (Color online) Results of semisupervised experiments based on one-minute detection.

the increase in the number of iterations. Finally, the *F*1-*score* of the Semi-DynaSeqNet model is 0.9465. After five iterations, the *F*1-*score* of the Semi-DynaSeqNet model has improved by 1.84% compared to the first iteration.

The experimental results showed that with increasing number of iterations, the model performance gradually improves. With the semisupervised threshold gradually increasing, only the samples with high confidence can be used as pseudo-labels, which have higher prediction accuracy in the training process. This method effectively improves the quality of the pseudo-labeled data. The semisupervised method with the adaptive adjustment of threshold not only improves the reliability of the pseudo-labeled data, but also optimizes the training effect of the model, which leads to a steady improvement in both learning and generalization abilities.

4. Conclusions

The Semi-DynaSeqNet model was proposed to improve the accuracy of sleep apnea detection using the SpO2 signal. This model integrates 1D-CNN, GRU, and the self-attention mechanism to extract local features, capture long-distance dependences of dynamic characteristics, and adaptively weigh critical information for efficient and stable classification. Additionally, the semisupervised learning method with the adaptive adjustment of threshold is employed to address the issues related to the insufficient training samples and class imbalance, thereby enhancing the model's generalization ability. The experimental results demonstrate that the Semi-DynaSeqNet model accurately captures dynamic changes, trends, and features of the SpO2 signal while improving decision precision.

In this study, we developed a composite deep learning architecture that integrates 1D-CNN, GRU, and self-attention mechanisms. This architecture achieves multilevel feature extraction by concatenating various modules. Although this theoretically improves the model's ability to capture data features, it also leads to the complexity of the model structure and a significant increase in the number of parameters. The expansion of this parameter quantity may lead to an increase in the demand for storage and computing resources of the model, affecting its deployment efficiency in resource constrained environments. For application scenarios that require a rapid response, such as the real-time monitoring of OSA, the current model's response speed may not be ideal. With further research, we will focus on exploring model optimization strategies aimed at reducing the computational complexity of the model and improving its processing speed. For example, considering integrating attention mechanisms more deeply into CNN or GRU modules to achieve adaptive focusing on key features, thereby improving computational efficiency without sacrificing model performance, while significantly enhancing the practicality and applicability of the model in real-time application scenarios.

Acknowledgments

This work was supported by Zhejiang Provincial Natural Science Foundation of China under Grant no. LTGS23F010001 and the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant no. GK239909299001-406.

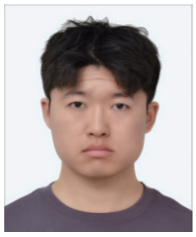
References

1 C. Guilleminault, F. L. Eldridge, and W. C. Dement: Science **181** (1973) 856. https://doi.org/10.1126/science.181.4102.856
2 D. Pitson and J. Stradling: J. Sleep Res. **7** (1998) 53. https://doi.org/10.1046/j.1365-2869.1998.00092.x
3 R. J. Thomas: Sleep **26** (2003) 1042. https://doi.org/10.1093/sleep/26.8.1042
4 M. J. Sateia: Chest **146** (2014) 1387. https://doi.org/10.1378/chest.14-0970
5 P. J. Strollo, Jr. and R. M. Rogers: New England J. Med. **334** (1996) 99. https://doi.org/10.1056/NEJM199601113340207
6 C. P. Langlotz and R. G. Carol: JAMA **273** (1995) 425. https://doi.org/10.1001/jama.1995.03520290079039
7 E. Urtnasan, J.-U. Park, E.-Y. Joo, and K.-J. Lee: J. Med. Syst. **42** (2018) 1. https://doi.org/10.1007/s10916-018-0963-0
8 S. O. Wali, B. Abaalkhail, I. AlQassas, F. Alhejaili, D. W. Spence, and S. R. Pandi-Perumal: Annals Thoracic Med. **15** (2020) 70. https://doi.org/10.4103/atm.ATM_215_19
9 B. Pourbabaee, M. H. Patterson, M. R. Patterson, and F. Benardet: Physiol. Meas. **40** (2019) 084005. https://doi.org/10.1088/1361-6579/ab3632
10 A. John, K. K. Nundy, B. Cardiff, and D. John: 2021 43rd Annu. Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE, 2021). https://doi.org/10.1109/EMBC46164.2021.9631037
11 H. T. Chaw, T. Kamolphiwong, S. Kamolphiwong, K. Tawaranurak, and R. Wongtanawijit: Appl. Comput. Intell. Soft Comput. **2023** (2023). https://doi.org/10.1155/2023/8888004
12 E. Wang, I. Koprinska, and B. Jeffries: IEEE J. Biomed. Health Inf. **27** (2023) 5644. https://doi.org/10.1109/JBHI.2023.3305980
13 J. Levy, D. Álvarez, F. Del Campo, and J. A. Behar: Nature Commun. **14** (2023) 4881. https://doi.org/10.1038/s41467-023-40604-3
14 S. Yook, D. Kim, C. Gupte, E. Y. Joo, and H. Kim: Sleep Med. **114** (2024) 211. https://doi.org/10.1016/j.sleep.2024.01.015
15 J. Jiménez-García, M. García, G. C. Gutiérrez-Tobal, L. Kheirandish-Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo D. Gozal, and R. Hornero: Biomed. Signal Process. Control **87** (2024) 105490. https://doi.org/10.1016/j.bspc.2023.105490
16 H.-Y. Chi, C.-Y. Yeh, J.-W. Chen, C.-Y. Wang, and S.-H. Hwang: IEEJ Trans. Electr. Electro. Eng. **19** (2024) 448. https://doi.org/10.1002/tee.23974
17 J. Chung, C. Gulcehre, K. Cho, and Y. Bengio: arXiv (2014). https://doi.org/10.48550/arXiv.1412.3555
18 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin: Adv. Neural Inf. Process. Syst. **31** (2017) 6000. https://doi.org/10.48550/arXiv.1706.03762
19 T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter: Computers in Cardiology 2000 Vol. 27 (Cat. 00CH37163) (IEEE, 2000).
20 C. Heneghan: Vincent's University Hospital/University College Dublin Sleep Apnea Database (2011).
21 A. Bernardini, A. Brunello, G. L. Gigli, A. Montanari, and N. Saccomanno: Artif. Intell. Med. **118** (2021) 102133. https://doi.org/10.1016/j.artmed.2021.102133

About the Authors

Linqing Yang received her bachelor's degree in communication engineering from Shandong University, Weihai, China. She is currently pursuing a master's degree in electronic information engineering at Hangzhou Dianzi University, Hangzhou, China. Her research interests include signal processing using deep learning techniques and their applications. (2220800071@hdu.edu.cn)

Na Ying received her B.S., M.S., and Ph.D. degrees in 2000, 2003, and 2006, respectively, all from the School of Communication Engineering of Jilin University, China. She is an associate professor of the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China, where she develops and applies models and algorithms with signal processing in the fields of artificial intelligence, Internet of Things, and deep learning. (yingna@hdu.edu.cn)



Hongyu Li received his bachelor's degree in communication engineering from Hangzhou Dianzi University, Hangzhou, China, in 2021. He is currently pursuing a master's degree in electronic information engineering at Hangzhou Dianzi University, Hangzhou, China. His research interests include signal processing using deep learning techniques and their applications. (222080250@hdu.edu.cn)

Xinyu Lin is currently pursuing her bachelor's degree in communication engineering at Hangzhou Dianzi University, Hangzhou, China. Her research interests include signal processing and its application. (22080404@hdu.edu.cn)

Yinfeng Fang received his Ph.D. degree in intelligent robotics from the University of Portsmouth, UK, in 2015, and his B.S. degree in electrical and information engineering and M.S. degree in pattern recognition and intelligent system from Zhejiang University of Technology, China, in 2009 and 2012, respectively. He joined Hangzhou Dianzi University in October 2018, where he is currently an associate professor in the College of Communication Engineering. (yinfeng.fang@hdu.edu.cn)

Yong Zhou received his Ph.D. degree from the School of Medicine of Zhejiang University, Hangzhou, China, in 2001. He joined the Pulmonary and Critical Care Medicine Department of Sir Run Run Shaw Hospital, Zhejiang University, in August 2001, where he is currently an attending doctor. His research interests include chronic respiratory disease, sleep medicine, lung cancer, and tobacco control. (211080039@hdu.edu.cn)

Huahua Chen received his B.S. degree in electronic engineering, M.S. degree in electronic science and technology, and Ph.D. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 1999, 2002, and 2005, respectively. He joined Hangzhou Dianzi University, in May 2005, where he is currently an associate professor in the College of Communication Engineering. His research interests include image processing, computer vision, video anomaly detection, and machine learning. (iseealv@hdu.edu.cn)