Computer Methods and Programs in Biomedicine 260 (2025) 108558

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Time-hybrid OSAformer (THO): A hybrid temporal sequence transformer for accurate detection of obstructive sleep apnea via single-lead ECG signals

Lingxuan Hou [a], Yan Zhuang [a,*], Heng Zhang [b], Gang Yang [a], Zhan Hua [c], Ke Chen [a], Lin Han [a,d], Jiangli Lin [a,*]

[a] College of Biomedical Engineering, Sichuan University, Chengdu, 610065, Sichuan, China
[b] College of Electrical Engineering, Sichuan University, Chengdu, 610065, Sichuan, China
[c] China-Japan Friendship Hospital, Beijing, China
[d] Highong Intellimage Medical Technology (Tianjin) Co., Ltd, Tianjin, China

ARTICLE INFO

Keywords:
Multi-head Attention
Obstructive sleep apnea
Deep learning
Signal processing
Decision-making

ABSTRACT

Background and Objective: Obstructive Sleep Apnea (OSA) is among the most sleep-related breathing disorders, capable of causing severe neurological and cardiovascular complications if left untreated. The conventional diagnosis of OSA relies on polysomnography, which involves multiple electrodes and expert supervision. A promising alternative is single-channel Electrocardiogram (ECG) based diagnosis due to its simplicity and relevance. However, extracting respiratory-related features from ECG is challenging since ECG signals do not directly reflect respiratory patterns. Consequently, the accuracy of most deep learning models that predict OSA using ECG data remains to be improved.

Methods: In this study, we propose the Time-Hybrid OSA transformer (THO), a novel method that leverages single-lead ECG signals for accurate OSA detection. The THO enhances feature extraction using a hybrid architecture combining dilated convolution and Long Short-Term Memory (LSTM), along with a multi-scale feature fusion strategy. Additionally, THO integrates an embedded memory decay mechanism within a multi-head attention model to capture real-time characteristics of time series data. Finally, a voting mechanism is incorporated to enhance decision reliability.

Results: Evaluation of the THO model demonstrates superior performance with prediction accuracy (ACC) and area under the receiver operating characteristic curve (AUC) values of 95.03 % and 96.85 %, respectively, representing improvements of 11 % and 8 % over comparative models. Moreover, the ACC shows a 5 % enhancement relative to state-of-the-art models.

Conclusions: These results prove the THO model's efficacy in predicting OSA, offering a robust alternative to traditional diagnostic approaches.

1. Introduction

Respiratory activity during sleep often reflects an individual's health and sleep quality. Abnormal respiratory activities occurring in sleep, such as snoring, sleep apnea (SA), and sleep hypopnea (SH), can lead to sleep-related breathing disorders, significantly compromising sleep quality. Prolonged sleep breathing abnormalities could induce psychiatric illnesses and negatively impact health, resulting in numerous cardiocerebrovascular diseases [1–4], amongst which obstructive sleep apnea (OSA) is the most representative. American Academy of Sleep Medicine (AASM) [5] defined OSA as a sleep-breathing disorder in which apnea events occur during sleep. In China, the prevalence of sleep apnea syndrome has reached 176 million individuals, with projections estimating an increase to 210 million [6]. Notably, OSA often remains undetected for extended periods, contributing to a diagnosis rate of <1 % in China. Consequently, monitoring OSA is relevant not only for patients already diagnosed with sleep breathing disorders but also for essential real-time monitoring in the general healthy population.

The detection of OSA mainly relies on polysomnography (PSG) [7,8], unfortunately, the equipment supported by PSG technology is typically substantial and requires numerous electrodes connected to the body during sleep, significantly affecting sleep comfort. Hence, conducting

* Corresponding author.
E-mail addresses: zhuangy@scu.edu.cn (Y. Zhuang), linjlscu@163.com (J. Lin).

https://doi.org/10.1016/j.cmpb.2024.108558
Received 18 July 2024; Received in revised form 21 November 2024; Accepted 7 December 2024
Available online 7 December 2024

home-based OSA testing via PSG is unrealistic. Supposed OSA is not promptly identified and addressed. In that case, it can lead to various health issues such as fatigue, depression, weight gain, etc., affecting daily life and significantly increasing the risk of cardiovascular diseases, including hypertension, arrhythmia, myocardial infarction, and stroke. Therefore, timely and accurate detection of OSA is essential, and the development of home-based OSA detection has emerged. In recent years, numerous studies have employed various physiological signals for the detection of Obstructive Sleep Apnea (OSA), including Oxygen Saturation (SaO_2) [9,10] and Electroencephalogram (EEG) signals [11, 12]. These modalities, however, present certain limitations that hinder their efficacy and patient compliance. For instance, EEG-based OSA detection requires the attachment of multiple electrodes to the scalp, significantly disrupting sleep quality due to the discomfort associated with the device setup. This method thus fails to achieve a non-intrusive, wearable solution that could facilitate patient acceptance and continuous monitoring [13]. Moreover, although SaO_2 signals are less invasive, the extraction of relevant sleep respiratory data from these signals is challenging. As a result, the predictive accuracy of SaO_2-based methods for OSA diagnosis remains suboptimal, as indicated by several studies [10,14]."In 2002, T. Penzel [15] et al. introduced an approach to detecting OSA. They proposed the utilization of single-lead Electrocardiogram (ECG) signals for the identification of OSA. Due to ECG's noninvasive, convenient, and portable nature, using ECG for OSA detection has become a significant research direction. However, since ECG directly reflects heart rate and rhythm changes rather than sleep breathing conditions, sometimes OSA might not result in noticeable ECG changes. However, the early detection and treatment of OSA are of great significance, both in terms of personal health and cost savings of treatment. Consequently, enhancing the accuracy and precision of OSA detection via ECG remains a substantial challenge.

There are primarily two approaches to the analysis of OSA based on single-lead ECG signals. The first approach is the direct interpretation of raw ECG signals to determine OSA. The second method involves feature extraction from the raw ECG signals, such as the RR Interval (RRI), ECG-derived respiration (EDR), and heart rate variability (HRV), and then judging and detecting OSA based on these extracted features. Moreover, with the rapid advancement of artificial intelligence, numerous researchers have already conducted related studies in the area of ECG-based OSA detection. For instance, Chen et al. [16] introduced an end-to-end spatiotemporal learning method for OSA detection, composed of multiple spatiotemporal blocks. Each block shares a consistent architecture, including a convolutional neural network (CNN) layer, a maxpooling layer, and a Bidirectional Gated Recurrent Unit (BiGRU) layer. This structure effectively captures both the morphological spatial characteristics and temporal features of electrocardiogram signals. The model demonstrated promising predictive results when tested on the Apnea-ECG dataset. Yang et al. [17] proposed a one-dimensional Squeeze-and-Excitation (SE) residual group network to thoroughly extract the complementary information between Heart Rate Variability (HRV) and EDR. This approach enhances the feature extraction capability of OSA signals. The method was tested using the Apnea-ECG dataset, and the results indicated a significant improvement in the predictive accuracy of OSA. This provides a novel approach for the automatic prediction of OSA through ECG analysis. Cheng et al. [18] developed a methodology for OSA detection by decomposing electrocardiogram signals into 15 sub-band signals using a filter bank. They employed a 1D CNN model to independently work with each sub-band, extracting and classifying features of the given sub-band signals. This approach, with its fine-grained segmentation and meticulous feature extraction, effectively enhanced the precision of OSA detection. Zhou et al. [19] proposed a new method for classifying OSA using a single-lead ECG signal transformation and a composite deep convolutional neural network. This approach involves converting ECG signals into images reflecting HRV and temporal features, then analyzing them with a model combining finetuned AlexNet and ResNet, and a custom CNN with residual blocks. This technique achieved high accuracy and is suitable for real-time OSA detection on mobile devices. However, models currently proposed for OSA detection based on ECG analysis generally exhibit accuracy and precision rates of <90 % [20–23]. Moreover, many existing studies demonstrate a lack of balance between precision and recall metrics. For instance, Liu et al. [24] reported a precision of 85.0 % but a recall of only 76.1 %. Similarly, Yan et al. [21] observed a discrepancy of approximately 4 % between precision and recall. The absence of a proper tradeoff process between accuracy and recall in these studies results in reduced predictive accuracy. These fall short of fulfilling the requirement for precise detection of OSA. Hence, developing a model capable of accurately detecting OSA is of significant necessity and practical importance.

This study focuses on developing a more precise method for detecting OSA, addressing a critical challenge in current healthcare practice. Our proposed solution involves the construction of the Time-Hybrid OSAformer (THO) model, a novel hybrid that integrates CNN, (Long Short-Term Memory) LSTM, and Transformer models. The uniqueness of THO resides in integrating an innovative CNN-based block (Intra-Expanded Convolution) and a new Multihead-attention block (Time-Based Multihead-Attention, TBMA). In the THO model, each layer serves a specific purpose: the CNN convolution layer extracts initial features, laying the groundwork for classification; the LSTM layer recognizes long-term dependencies in time-series data; and the Transformer layer processes this data, managing long-distance dependencies and extracting rich features, ultimately leading to the prediction of OSA probability. To further enhance OSA detection, we selected three specific feature values (ECG signal, EDR, RRI, and Respiration Amplitude Modulation Pattern (RAMP)) as inputs to the model. Additionally, we introduced a deep learning model designed for feature extraction based on a sliding window and LSTM. The LSTM is applied to the window for constructing a prewarning mechanism to predict the OSA probability of the next time step from current time step. This approach excels in capturing subtle changes through the window and prewarning mechanism, substantially improving THO's performance. Our validation of this innovative model involved a rigorous process, encompassing five-fold cross-validation and external testing, which yielded exemplary results in OSA detection. Together, these advancements mark a significant step forward in the field and offer promising avenues for further research and practical application. The workflow of this study is shown in Fig. 1.

2. Method

2.1. Data collection

2.1.1. Apnea-ECG database

To ensure reliability of this study, we used the popular and widely-used Apnea-ECG database [25], supplied by PhysioNet [26] to detect OSA. The database encompasses ECG signals sampled at a frequency of 100 Hz with a resolution of 16 bits. These ECG signals are contributed by 70 males and females, with ages ranging from 27 to 60 years, and weights varying from 53 to 153 kgs (BMI between 20.3 and 42.1), all presenting with varying degrees of OSA, and have been recruited for full-night sleep monitoring in a medical setting. The duration of the collected ECG data spans from 401 to 578 min, segmented into per-minute ECG data, subsequently annotated by experts for sleep apnea by integrating other signals from PSG. Herein, the "A" label stands for apnea, and the "N" label signifies normal. Furthermore, individuals in the dataset were categorized into four severity levels of sleep apnea—Normal, Mild, Moderate, and Severe—based on the Apnea-Hypopnea Index (AHI) to facilitate individual-level diagnoses. All data is impartially partitioned into training data (35 recordings with the index of a01-a20, b01-b05, and c01-c10) and testing data (35 recordings with the index of $\times$ 01- $\times$ 35). The training and testing data comprise 17,125 (6514 apnea and 10,611 normal) and 17,303 (6552 apnea and 10,751 normal) per-minute data points, respectively. Thus,

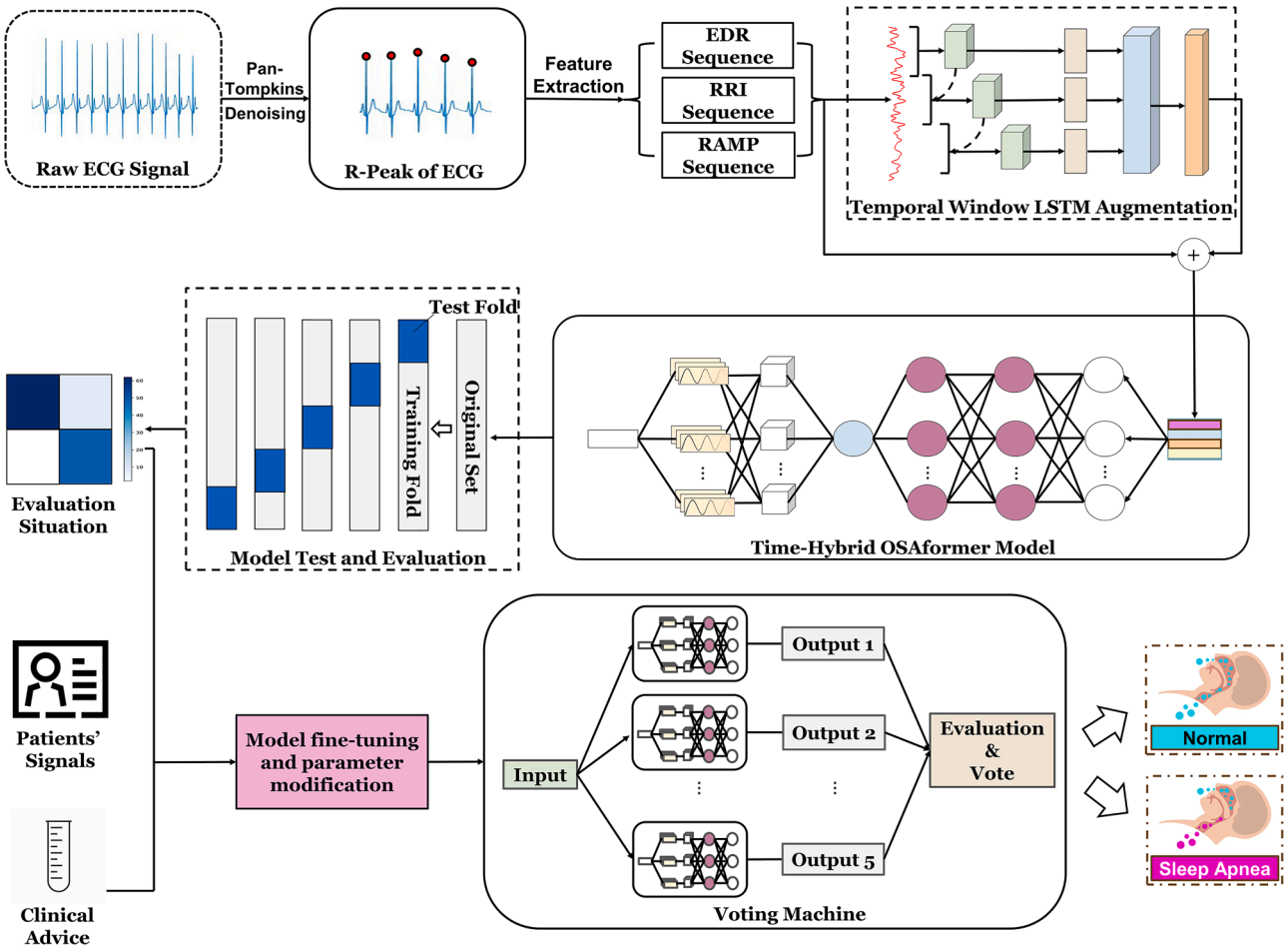

Fig. 1. The workflow of this study. EDR, RRI and RAMP are extracted from ECG signals as the input of THO model. Features are initially processed by the TWLA, then combined with EDR, RRI, and RAMP for input into subsequent model modules.

ample training and testing samples are provided for model training and validation. The specific participant information of the Apnea-ECG database is shown in Table 1.

2.1.2. *University college Dublin sleep Apnea database (UCDDB)*

To further validate the generalization capability of the THO model, this study employed an additional dataset, the University College Dublin Sleep Apnea Database (UCDDB) [26], for a series of tests and comparative analyses. This dataset encompasses 25 complete overnight PSG from adults suspected of having sleep-disordered breathing. The ECG signals in these PSG were collected using an improved V2 lead configuration, with a sampling rate of 128 Hz over durations ranging from 355 to 462 min. Each PSG includes precisely timed annotations of sleep apnea events. For our analysis, we segmented the ECG signals into one-minute intervals. The labeling method of Chen et al. [16] was applied in this study, if a sleep apnea event, lasting five seconds or longer, occurred within any given minute, that minute was labeled as 'Apnea'; otherwise, it was labeled as 'Normal'. Ultimately, 9193 instances were labeled as 'Normal' and 496 as 'Apnea'.

Table 1
Specific information of Apnea-ECG database.

	Train set	Test set	Average
Patient ratio	65.71 %	65.71 %	65.71 %
Male ratio in participant	85.71 %	77.14 %	81.43 %
Average age	46	44	45
Average BMI	28	28.2	28.1
Average record minutes	489.29	494.37	491.83

2.2. *Data preprocessing*

ECG signals are faint electrical signals detectable on the body's surface, typically characterized by amplitudes ranging from 10μV to 4 mV and frequencies spanning from 0.05 to 100Hz. The human body can be regarded as a complex system where multiple electrical signals are transmitted across its surface. Consequently, the randomness and nonstationary of ECG signals, coupled with various forms of noise interference, present significant challenges. These factors can easily degrade the signal quality when ECG devices capture these signals. Without noise elimination, it becomes challenging to recognize the acquired ECG signals and extract various features for tasks such as sleep monitoring. Therefore, it is crucial to conduct filtering and denoising procedures on ECG signals. Typically, the frequency of electromyography (EMG) signals lies between 20 and 5000 Hz, whereas the frequency of ECG signals is between 5 and 20Hz. Thus, this study employed a bilinear transformation method to design a Butterworth low-pass filter to eliminate EMG interference. Baseline drift is due to low-frequency interference brought about by the body's movement and the testing electrode, typically characterized by frequencies below 0.5 Hz. Hence, we designed an Infinite Impulse Response (IIR) zero-phase high pass filter with a cutoff frequency of 0.5 Hz. This approach effectively eliminates baseline drift while preserving the integrity of higher frequency components essential for accurate ECG analysis. After the denoising procedure, given that the Apnea-ECG database provides per-minute annotations, the denoised ECG signals were segmented into per-minute ECG data for subsequent feature extraction and corresponding to the labels, which can significantly improve the performance of OSA

detection. Furthermore, given the issue of data imbalance in the original training data from the Apnea-ECG and UCDDB database, we employed data augmentation techniques to balance the dataset and thereby enhance the performance of our model. Specifically, we utilized Time Warping and Noise Injection as augmentation methods. Time Warping was executed by distorting the time axis of the input signals through the addition of Gaussian noise, scaled relative to the signal length. This adjustment resulted in signals that were either stretched or compressed, thereby diversifying the temporal dynamics represented in the training data. Additionally, we introduced Gaussian noise to the original signals, setting a noise factor at 0.02. This technique generated noisy versions of the input data, which are instrumental in improving the model's robustness against anomalous inputs.

2.3. Feature extraction

We selected three features: the RRI, the EDR, and the RAMP [27,28] which are critical indicators of respiratory rhythm in ECG signals. The RRI reflects HRV, an essential measure of the autonomic nervous system's control over the heart. In patients with OSA, the autonomic nervous system may be affected by repetitive cycles of apnea and recovery, leading to changes in the RRI. EDR represents the influence of respiration on ECG signals, allowing us to assess respiratory conditions by analyzing ECG signals. In OSA patients, abnormal changes in respiratory patterns, such as apnea and shallow breathing, can cause abnormal variations in the EDR signal. The RAMP reflects the amplitude pattern of respiration, indicating variations in breathing depth and rhythm. In OSA patients, repeated cycles of apnea and recovery can lead to significant pattern changes in respiratory amplitude. For instance, patients may exhibit deep breathing for a period after recovery from apnea, leading to a temporary increase in respiratory amplitude. By analyzing the RAMP, we can assess the patient's respiratory pattern and the possible presence of OSA.

In this study, we first used the Pan-Tompkins algorithm [29] to detect R-waves in the pre-classified per-minute denoised ECG signals. We then calculated the distance between two adjacent R-waves to obtain the RRI in milliseconds. On account of the normal breathing frequency in humans generally falling between 0.15–0.4 Hz, we processed the extracted RRI through a low-pass filter to obtain the EDR signal. As physical activity and emotional changes during sleep negligibly affect RRI and EDR, no additional processing is required to exclude these factors' interference with OSA detection. After obtaining the EDR signal, we calculated the amplitude of each breath (i.e., a cycle of the respiratory signal) within a minute and analyzed the continuous change in respiratory amplitude to obtain the RAMP. Moreover, we have proposed a feature extraction method named Temporal Window LSTM Augmentation (TWLA). It operates by creating a shorter, customizable sliding window over the input signal and extracting features from each window to capture subtle changes in the input signal, thereby facilitating the detection of the faint impacts on ECG signals at the onset of sleep apnea. Simultaneously, we established a deep learning based early warning system on each window, where a miniature LSTM model is individually trained. The inherent temporal sequence prediction capability of LSTM is leveraged to anticipate potential OSA occurrences in the following window, thereby enhancing the model's perceptibility to subtle variations. Ultimately, a fully connected layer is employed to reshape the output feature dimension to match the model input. We used EDR, RAMP and RRI as inputs to the TWLA to extract corresponding feature values, eventually reshaping their output feature length to 240 and merging them with the original input for model prediction. The extracted features were rendered as four-dimensional inputs to our model, significantly enhancing its predictive accuracy.

2.4. Time-Hybrid OSAformer (THO)

2.4.1. Overview of THO model

Since the introduction of the Transformer model by Vaswani et al. in 2017 [30], artificial intelligence has embarked on a new era of research. Transformer model demonstrates substantial potential in enhancing the accuracy of ECG signal analysis and application, illustrating considerable advantages. In this study, we integrated two CNN layers, one LSTM layer, and one Transformer layer to extract features and perform subsequent classification from the input RRI, EDR, and RAMP feature sequences. The two CNN layers are composed of an IntraExpanded Convolution and a one-dimensional residual convolutional layer (ResidualConv1d). The IntraExpanded Convolution performs feature extraction in the input sequence data with the characteristic of extending the receptive field of the convolutional kernel, enabling the capture of long-range dependencies without increasing computational complexity. The ResidualConv1d adds a residual connection to the one-dimensional convolution, thereby effectively enhancing the expressivity of the model and alleviating the issue of gradient vanishing. Following this, the output from the CNN is reordered. It feeds it into the LSTM layer to handle the time-series sequence, further capturing the long-term dependencies in the input sequence. Subsequently, the sequence data enters the Transformer layer. Following the self-constructed TBMA in the Transformer which allows the model to further understand the time-correlated series. Subsequently, we combine the outputs from the CNN, LSTM, and TBMA layers through addition, integrating multi-scale features to enhance the model's grasp of detailed and macroscopic features within the ECG signals. Then, the possibility of having OSA is predicted through a decoder composed of a fully connected layer and a Rectified Linear Unit (ReLU) activation function. Subsequently, sequences with probabilities exceeding 50 % are classified as having OSA to derive the final results. Moreover, we implemented a voting mechanism utilizing five models trained during a five-fold cross-validation process. Prior to the deployment of our voting model, we performed parameter tuning on all of the constituent model in 5-fold cross validation to optimize their predictive performance. This voting system allows us to set the threshold for the number of votes required to classify a case as OSA-positive, thereby facilitating the achievement of a Precision-Recall Trade-off. In accordance with insights from clinical practitioners who emphasize the importance of reducing misdiagnosis rates, our approach prioritized high precision. Therefore, we set a higher threshold for determining OSA-positive cases, significantly enhancing precision. The structure of the THO model is shown in Fig. 2.

2.4.2. Temporal window LSTM augmentation

In the THO model, we proposed TWLA method for feature extraction, which takes advantage of applying LSTM within a sliding window. Initially, the sliding window is utilized on the sequence, inputting a window into the LSTM each time to extract a small portion of the features. Meantime, owing to the LSTM's inherent capability for time-step-based prediction, this LSTM layer also can predict the likelihood of sleep apnea occurrences in the subsequent time steps which serves as an pre-warning mechanism. Following that, the output from the LSTM is conveyed to a fully-connected layer for integration. By sliding the window, the entire sequence is processed and features are extracted. Finally, once the whole sequence has been loaded onto the fully-connected layer, the layer will merge and reshape the individual sequences into an appropriate form. In this process, the model input is assumed to be x[b, n, f], where 'b' represents the batch size, 'n' stands for time steps, and 'f' corresponds to the number of features. As the window slides across the input sequence, a portion of the samples is extracted each time as demonstrated by the following equation:

$$window_t = x[:, t : t + w, :] \tag{1}$$

where 'window' denotes the current window, 'w' is the window size,

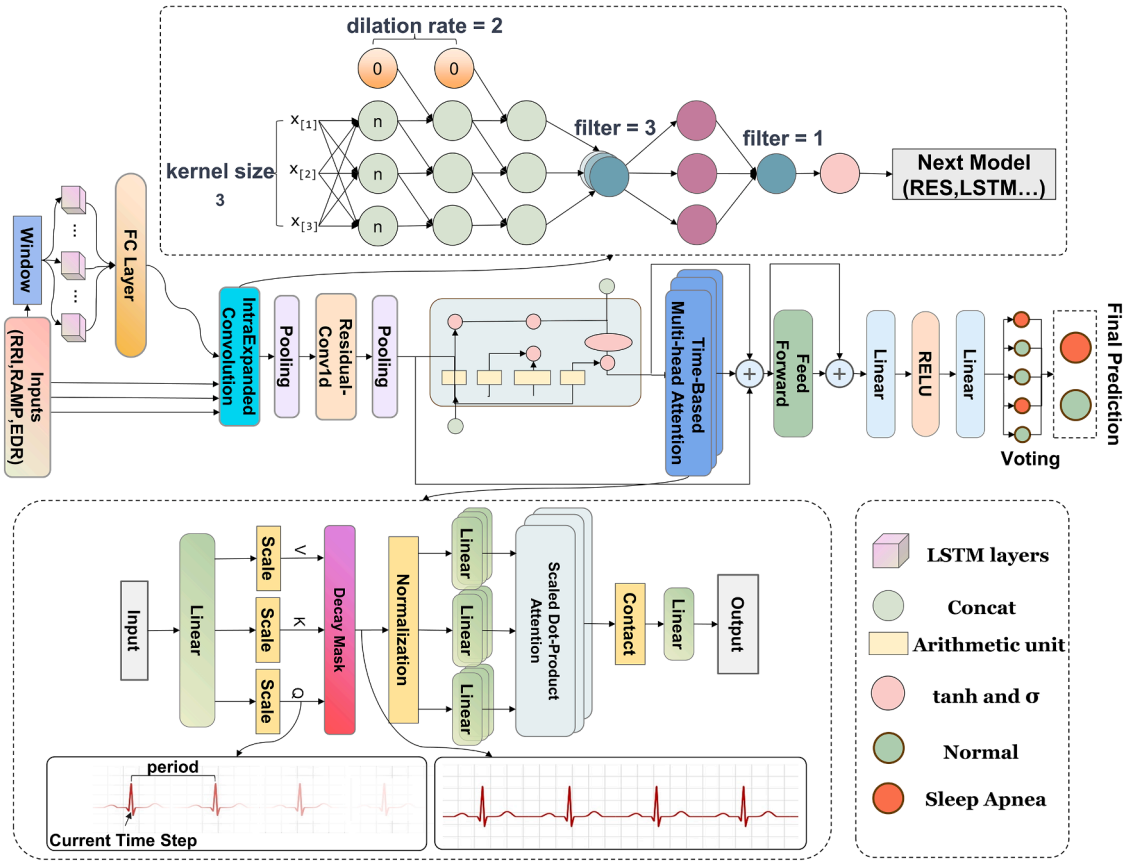

Fig. 2. The construction of the Time Hybrid OSAformer (THO) model. The core network comprises three optimized deep learning modules, namely Temporal Window LSTM Augmentation (TWLA), IntraExpanded Convolution and Time Based Multi-head Attention (TBMA).

which is set to 16 in this study, and 't' represents the current time step. The window is then inputted into the LSTM model. The LSTM model returns the final hidden layer state as the result of feature extraction, and the fully-connected layer is applied to this as the output 'o' for a single window. The formulation is depicted as follows:

$$o_t = fc(LSTM(window_t)) \tag{2}$$

In this LSTM block, we focus on its internal gating mechanisms (forget gate, input gate, output gate) and their interactions to implement an early warning mechanism. The forget gate determines how much of the old information is retained for attention over longer time steps, expressed as:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, window_t] + b_f\right) \tag{3}$$

Here, f_t is the output of the forget gate, with W_f and b_f as the weights and bias of the forget gate, respectively, and h_{t-1} as the previous time step's hidden state.

The input gate is responsible for incorporating new information, expressed as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, window_t] + b_i) \tag{4}$$

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1}, window_t] + b_c) \tag{5}$$

Where i_t is the output of the input gate, and $\widetilde{C}_t$ is the candidate value for new information. W_i, W_c, b_i, and b_c are the weights and biases for the input gate.

After obtaining new cell information $\widetilde{C}_t$ from the input gate, the current cell state C_t is updated:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

Upon merging new and old information, this is used for updating the hidden layer, with the previously hidden layer utilized in the output gate to extract predictive information:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * tanh(C_t) \tag{8}$$

In the LSTM network within TWLA, o_t represents the output of the output gate, while h_t is the hidden state at the current time step, with W_o and b_o being the weight and bias of the output gate, respectively. The hidden state is a crucial output of the LSTM in TWLA, updated at each time step to reflect the network's cumulative understanding of current and previous inputs. This accumulation encapsulates patterns and trends within the time series, enabling LSTM to effectively address long-term dependencies and predict signal information for subsequent time steps. In this study, the current hidden layer state h_t and output o_t are jointly fed into a fully connected layer for the integration of LSTM outputs and hidden layer predictions. Finally, a resize layer in the fully connected architecture reshapes this into the final predicted sequence size, aligning it with the model's input requirements. The pre-warning mechanism of the TWLA module detects subtle variations within model inputs and predicts changes of next window, thereby significantly enhancing feature extraction capabilities within and across windows, capturing detailed signal characteristics, and augmenting the model's predictive accuracy for OSA. The structure of TWLA is depicted in Fig. 3.

2.4.3. *IntraExpanded convolution*

Within the THO model, we introduced an adaptation of the Dilated Convolution-based Convolutional Neural Network model, coined as IntraExpanded Convolution. Dilated Convolution is an effective strategy to enhance the receptive field of the network without increasing

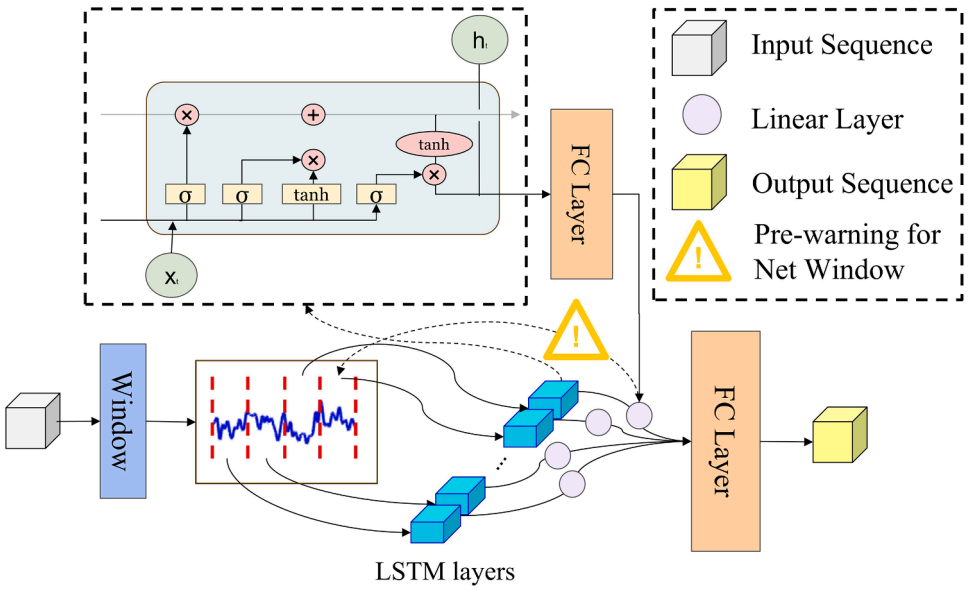

Fig. 3. The structure of TWLA, consisting of the LSTM and time window, forms the pre-warning mechanisms for predicting the likelihood of sleep apnea occurrences in subsequent time steps.

parameter count or computational complexity, thereby allowing the network to encompass a broader scope of contextual information. The Dilated Convolution can be expressed as:

$$Y[n] = \sum_{k=0}^{k-1} X[n - dk]H[k] \tag{9}$$

where X[n] represents the input sequence, H[k] is the convolution kernel, k is the length of the convolution kernel, Y[n] is the network output, and d denotes the dilation factor. In this context, when d equals 1, it devolves into a standard convolution. To ensure that each channel independently learns features without mutual interference, each channel is equipped with an exclusive convolution filter post input. This can be represented for each channel as:

$$Y_i[n] = \sum_{k=0}^{k-1} X_i[n - k]H_i[k] \tag{10}$$

where i stands for the individual index allocated for each channel. Following this, a 1 × 1 convolution kernel is utilized to convolve the previous output, altering the channel count and delivering the output to the subsequent model. This can be mathematically expressed as:

$$Y[n] = \sum_{i=0}^{C-1} Y_i[n]H[i] \tag{11}$$

where C denotes the input channel count. Consequently, the Intra-Expanded Convolution can be expressed as:

$$Y[n] = \sum_{i=0}^{C-1} \sum_{k=1}^{k-1} X_i[n - dk]H_i[k] \tag{12}$$

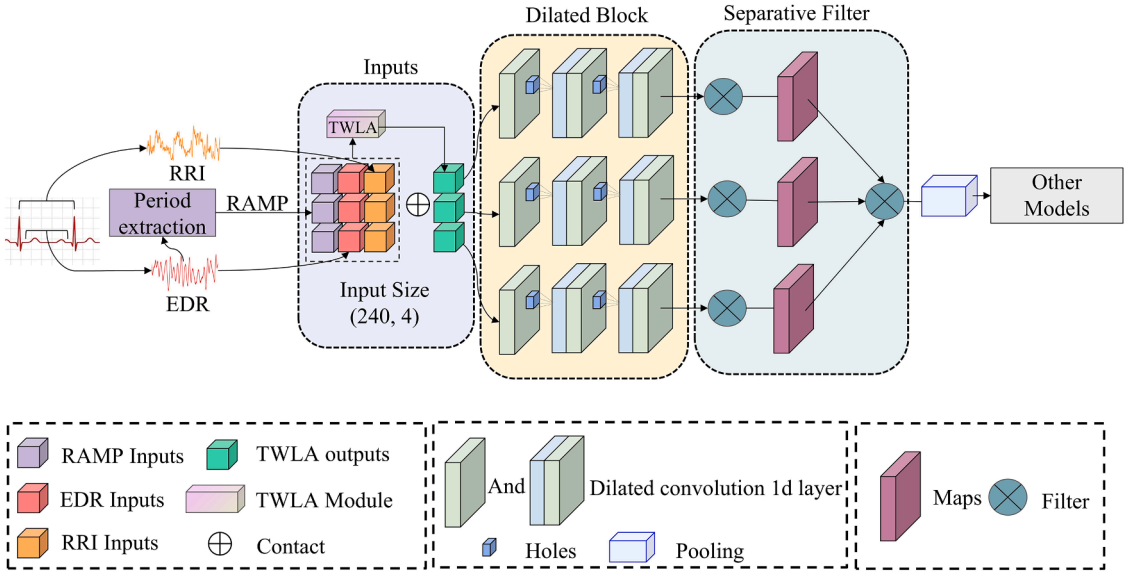

Fig. 4. The structure of IntraExpanded Convolution, which is constructed with dilated block and each channel is allocated an individual filter for enhanced contextual information extraction.

Following empirical testing, it has been found that the modified CNN structure of IntraExpanded Convolution significantly reduces parameters and computational complexity, while simultaneously expanding the receptive field to capture a broader range of contextual information, and augmenting the diversity of features learned by the model. This has yielded remarkable results in the current study, as depicted in the network structure in Fig. 4.

2.4.4. Time-Based Multihead-Attention

Due to the temporal correlation inherent in physiological signals, we've improved the Multihead-Attention in the Transformer [30] by proposing a time decay function, thus concentrating attention more on the current time point or recent inputs. This modification, called TBMA, helps minimize disturbances during classification while boosting detection accuracy. In the application of the time decay function, we assume an attention matrix A of size B x H x S, where B is the batch size, H is the number of heads, and S is the sequence length. The element A[i] represents the degree to which the model focuses on the i th batch. This attention matrix is derived from the dot product of Q and K, followed by a softmax operation, which can be expressed as follows:

$$A[i] = softmax(Q_i \cdot K_i) \tag{13}$$

We developed a time decay mask by multiplying the attention matrix with a decay mask. The function of this time decay mask is described as follows:

$$A[i] = A[i] \times decay\_mask \tag{14}$$

In this section, decay_mask represents a temporal mask, the selection and construction of which play a pivotal role in optimizing the model's performance. Considering the temporal correlation and periodicity of ECG signals involved in this study, we designed a periodic decay function to design the decay_mask. This approach enhances computational efficiency by reducing the weight of early data points and focusing on its periodic change, enabling the model to focus more on recent data and thereby diminishing the need for complex computations across the entire sequence. Moreover, by directing the model's attention towards recent inputs, it adapts to the temporal correlations of ECG signals, significantly improving its performance in processing time-correlated sequences. The mathematical representation of decay_mask is as follows:

$$decay = sigmoid(decay\_rate)^{arange(num\_steps)} \tag{15}$$

$$periodic = \sin\left(\frac{arange(num\_steps)}{sigmoid(raw_{periodic})} \times \omega\right) \tag{16}$$

$$decay\_mask = decay \times (1.0 + periodic) \tag{17}$$

In this mathematical representation, decay_rate denotes the rate of decay, which is a significant constant between 0 and 1, controlling the speed of the decay. In this case, the initial value of the parameters decay_rate and periodic were set to 0.9 and 10.0, respectively. These are learnable parameters mapped to a specific range through the sigmoid function, providing the mechanism to adjust the impact of time steps based on their distance from the current step. The num_steps represent the total number of time steps. It can be seen that a lower decay rate results in a faster decay of the curve. The parameter ω controls the initial period of decay_mask. We set it to 20 to adapt to our task after numerous experiments. Moreover, with the increase in time steps, the weights decrease periodically.

2.5. Experimental set up

In this research, we utilized the pre-classified Apnea-ECG database, wherein the training set comprises 35 recordings with the title of a01-a20, b01-b05, and c01-c10, and the testing set includes 35 recordings with the title of × 01- × 35. For UCDDB database to ensure that segments from the same individual were not present in both training and testing datasets, we initially divided the participants randomly into training and testing sets in a 7:3 ratio. Following this, the recordings were segmented into one-minute intervals, which then underwent subsequent feature extraction steps. This methodology ensures robustness in the handling and analysis of data, thereby preventing potential data leakage between training and testing phases. To evaluate and compare different models, we adopted a 5-fold cross-validation approach objectively and comprehensively. Specifically, we initially divided the 17,125 training data randomly into five equal parts, each part referred to as a "fold". In each iteration, a fold was selected as the test set, and the rest were used as the validation set, thus training and testing each model five times (i.e., each fold served as the test dataset once). Concurrently, we calculated the corresponding evaluation metrics for each training and their averages over the five iterations. Moreover, we have constructed an ensemble classifier based on the five models trained via five-fold cross-validation for external validation. The ensemble operates on a voting mechanism, where all 17,303 instances from the external test set were used for validation. For balancing the precision and recall of the THO model, a voting mechanism is configured such that if three or more models predict the occurrence of OSA, the final prediction is determined to be an OSA event. Before external testing, numerous experiments were conducted for optimal parameter tuning depending on the results of 5-fold cross validation. Ultimately, the Adam optimizer was selected for all models due to its efficient handling of sparse gradients and its adaptability in large datasets. For the THO model, the learning rate was set to 0.0001 to moderate the convergence rate and prevent the model from settling into local minima. Additionally, L2 regularization was employed with a weight decay of 0.0003 to mitigate overfitting. The training epoch was set to 150 to ensure thorough learning. Moreover, considering the comprehensive utilization of global information by the ensemble model and the limitations of hardware performance, the batch size was set at 128. Through rigorous experimentation and parameter adjustment, all models achieved optimal performance, fully converging without any signs of overfitting.

All experiments in this study were performed under the Windows 11 Professional Edition system, utilizing the Python language within the Python 3.10.9 framework, and also leveraged packages such as Pytorch 2.0.1+cu117, Scikit-learn, Sklearn 0.0.post1, scipy 1.10.0, and matplotlib to support model architecture and results validation. In terms of hardware configuration, our CPU was an Intel Core i7 10750H (base frequency 2.6 GHz, maximum turbo frequency 5 GHz, core/thread count six cores/twelve threads), and the GPU was an NVIDIA GeForce GTX 1650Ti (memory size 4GB, memory bus width 128bit).

2.6. Model evaluation

In this experiment, we selected Accuracy (ACC), Precision (PRE), Recall (REC), and the Receiver Operating Characteristic curve (ROC) as performance metrics.

The accuracy, precision and recall are calculated as:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{18}$$

$$precision = \frac{TP}{TP + FP} \tag{19}$$

$$recall = \frac{TP}{TP + FN} \tag{20}$$

The ROC curve, reflecting both sensitivity and specificity of a classifier, is a crucial means of evaluating classification performance. The X-axis represents the True Positive Rate (TPR), and the Y-axis represents the False Positive Rate (FPR). Generally, the closer the ROC curve is to the upper left corner, the better the classification performance.

Furthermore, the Area Under the Curve (AUC) of the ROC is calculated to measure the overall performance of the model, with a value range of 0–1.

For more objective evaluation of THO, we also built baseline models for comparison. For the deep learning model, we opted for Bi-directional Long Short-Term Memory (biLSTM) and Gate Recurrent Unit (GRU) as deep learning models for comparison with THO. Prior to the proposal of the Transformer model, LSTM had gained significant attention in the field of artificial intelligence, with numerous researchers utilizing LSTM for OSA detection or other processing of ECG signals [31,32]. Additionally, biLSTM, as an enhanced model of LSTM, has also emerged as a significant research focus in the fields of ECG and OSA detection [33]. GRU [34] was proposed after LSTM [35], and despite not having undergone as extensive validation as LSTM, it has attracted the attention of many researchers due to its similar experimental performance to LSTM and its computational efficiency [36]. We selected these mature and popular models as their viability has been substantiated through extensive learning and validation by many researchers. Our proposed model in this paper, being a new model, still needs to be learned, compared, and further validated with traditional models.

Similarly, we selected three machine learning models — k-nearest neighbor classification (KNN), Random Forest (RF), and extreme learning machine (ELM) for comparison. As traditional models, these models have been the subject of much attention and research. Pang et al. [37] compared the results of OSA detection using RF and Support Vector Machine (SVM) models, and found that the RF model performed similarly to the SVM model, and either model could be used as a rapid OSA screening tool. Hamidi et al. [38]. introduced a KNN model-based approach for identifying artifacts in ECG signals using 3-axis accelerometer data. This method achieved a positive predictive accuracy of 94.7 %. It is applicable to ECG monitoring wearable devices, enabling the capture of patients' ECG history and marking artifact samples for noise filtering or signal reconstruction systems. Kuila et al. [39] enveloped a hybrid model combining ELM and CNN for detecting arrhythmias in ECG signals. In external testing, this method achieved a classification accuracy of 98.82 %. It holds potential for implementation in medical devices for early detection of cardiovascular diseases. A large number of previous studies and models have provided reference for the establishment and improvement of our model. Through comparison, we also can see the improvements and enhancements of our model relative to traditional machine learning models.

In this study, all models were constructed using the Pytorch 2.0.1+cu117 environment. For the deep learning models, a classification module identical to that employed in the THO model was integrated following feature extraction to perform the classification tasks outlined in this research. For the machine learning models, we utilized the classification modules directly available in the Pytorch package. Parameter tuning for all models was conducted using a gradient thresholding method to optimize settings. These optimized parameters are detailed in Additional File 2. Simultaneously, in the interest of objectivity, we employed identical five-fold cross-validation and external validation methods, using the same dataset to validate all the models.

2.7. Individual-level classification

In addition to classifying per-minute segments, this study also employs the AHI for diagnosing and grading OSA at the individual level. The formula for AHI is defined as follows:

$$AHI_{Real} = \frac{Num\ of\ OSA\ Events}{Total\ Sleep\ Time\ (in\ hours)} \times 60 \tag{21}$$

Based on this formula, an AHI value below 5 is considered normal; an AHI of 5 to <15 is categorized as Mild OSA; 15 to <30 as Moderate OSA; and 30 or above as Severe OSA [40]. In this study, the occurrence of OSA events is calculated using the minute-segmented data. A segment is counted as one OSA event when classified as OSA by the THO model. Utilizing the model's diagnostic outcomes for segments, we have categorized the severity of OSA at the individual level.

3. Results

3.1. Result of feature extraction

In this experiment, we successfully extracted three significant features: the EDR, RRI, and RAMP from the ECG signals. Owing to the effects of heart rate variability in practical physiological conditions, the feature dimensions extracted per minute for each feature are not uniform. Hence, an interpolation process was executed to standardize the data across each feature value. For the RRI and RAMP, smoothing and spline interpolation were employed, while for the EDR, a down sampling process was adopted. Following this processing series, we obtained uniform feature dimensions of 240 for each of the attributes. The visualization of these features, EDR and RRI is presented in Fig. 5.

3.2. Result of 5-fold cross-validation

Based on Apnea-ECG dataset, we conducted 5-fold cross-validation for the THO model and six comparison models as shown in Table 2. For the THO model, we conducted numerous trials and ultimately determined to set the training iterations to 200 rounds. We plotted the average ROC curves from the five validation processes of each model as a representation of the 5-fold cross-validation results. Simultaneously, we also plotted boxplots for each performance metric from the cross-validation to make the results more intuitive. The THO model stands out with exemplary stability and performance, characterized by its high accuracy and narrow 95 % confidence interval (CI) of 92.38 % ± 0.56 %. On the other hand, the biLSTM model, while ranking second in terms of performance, registered an ACC and its accompanying 95 % CI of 85.95 % ± 1.52 %, revealing a noticeable discrepancy in both stability and efficiency when juxtaposed with the THO model. Among the baseline models, the KNN model took the lead in stability with a 95 % CI for ACC of merely 1.04 %. Nonetheless, its ACC was the nadir, standing at a mere 70.99 %. Hence, in comparative terms, the THO model unfailingly showcased a substantial edge in both model efficacy and stability.

Meantime the box plot for THO and ROC curves are presented in Figs. 6 and 7 respectively, and the box plot for other models is shown in the appendix, which intuitively indicates the model's central tendency and stability. The results of the five-fold cross-validation demonstrate that the average AUC of the THO model reached 96.72 %, which is superior to the comparison models. In the boxplots, performance metrics such as accuracy, recall, and precision were all superior to those of the comparison models. In comparison with traditional models, the THO model demonstrated significant advantages, further reinforcing the superiority and feasibility of the THO model.

3.3. Result of external validation

In the context of maintaining consistent parameter settings, we implemented external validation of all models using the test set of Apnea-ECG database. For the THO model, we utilized a voting mechanism for external validation by the five models trained in the 5-fold cross-validation. Compared to single model derived from the 5-fold cross-validation, the voting mechanism enabled us to enhance the PRE from approximately 90 % to 94.4 %. By numerous experiments, the model has fully converged without any signs of overfitting, as demonstrated in Fig. 8. We also carried out similar verification and tuning tasks on other models. However, as they are not the main focus of this paper, they are not elaborated further here. Finally, we conducted external tests on all models, calculating the ACC, PRE, REC, and AUC, in addition to plotting the ROC curves for each model. Remarkably, the THO model achieved an AUC of 96.85 %, vastly surpassing the best performer among the baseline models, the biLSTM, which reached only 88.05 %.

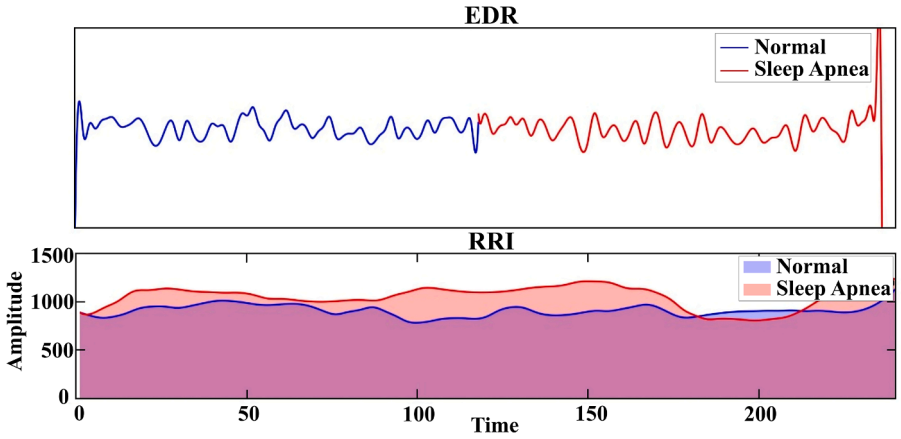

Fig. 5. The visualization of EDR, RRI. In EDR, the frequency of the signal during episodes of Sleep Apnea (red) is significantly higher compared to Normal (blue) conditions. For RRI, the amplitude of the signal notably increases during Sleep Apnea (red) as opposed to Normal (blue) conditions.

Table 2
Model performance on Apnea-ECG database.

Model	5-fold cross validation				External validation		
	ACC ± 95 %CI	PRE ± 95 %CI	REC ± 95 %CI	AUC ± 95 %CI	ACC	PRE	REC
biLSTM	85.96 % ± 1.52 %	83.30 % ± 4.48 %	79.90 % ± 9.37 %	89.71 % ± 2.45 %	83.8 %	78.2 %	79.8 %
ELM	74.95 % ± 4.16 %	69.23 % ± 9.03 %	63.84 % ± 8.32 %	77.17 % ± 8.06 %	75.7 %	65.7 %	76.7 %
GRU	81.02 % ± 1.43 %	75.68 % ± 6.56 %	76.22 % ± 16.37 %	83.86 % ± 6.24 %	79.3 %	77.6 %	64.4 %
RF	75.64 % ± 13.00 %	70.84 % ± 18.66 %	66.93 % ± 30.17 %	84.53 % ± 18.79 %	82.0 %	74.9 %	79.7 %
KNN	70.99 % ± 1.04 %	80.51 % ± 1.92 %	32.31 % ± 7.24 %	75.28 % ± 3.28 %	76.0 %	80.2 %	70.6 %
OURS	**92.38 % ± 0.56 %**	**91.47 % ± 0.56 %**	**89.64 % ± 2.19 %**	**96.72 % ± 0.44 %**	**95.0 %**	**94.4 %**	**93.3 %**

Following closely were the GRU, Random Forest, KNN, and ELM models with respective AUCs of 81.79 %, 81.61 %, 79.74 %, and 75.79 %. Evidenced by the ROC curves, the THO model stood out as a superior classifier in comparison to the baseline models, offering heightened clinical significance. The ROC curves of constructed models are shown in Fig. 9.

Compared with all baseline models, the THO model performed exceptionally well, with a peak ACC of 95.0 %, PRE of 94.4 %, and REC of 93.3 %. The top-performing model among the baseline set, the biLSTM, secured ACC, PRE, and REC scores of 83.8 %, 78.2 %, and 79.8 % respectively. These figures are all >10 % below the metrics achieved by the THO model. Moreover, the only other baseline model surpassing an accuracy of 80 % was the Random Forest, recording an ACC of 82.0 %. The remaining models from the baseline, namely the GRU, KNN, and ELM, failed to reach the 80 % mark in ACC, posting results of 79.3 %, 76.0 %, and 75.7 % respectively. Their performances substantially lagged behind that of the THO model. In consequence, when compared with the baseline models, the THO model unequivocally demonstrated marked superiority across all evaluation metrics. Moreover, to demonstrate the efficacy of the voting mechanism employed in this study, the performance of a standalone THO model was also evaluated. This model was trained using the entire training dataset and subsequently tested using the test dataset. It achieved ACC, PRE, REC and AUC of 91.4 %, 89.0 %, 88.7 % and 96.3 % respectively. These results significantly

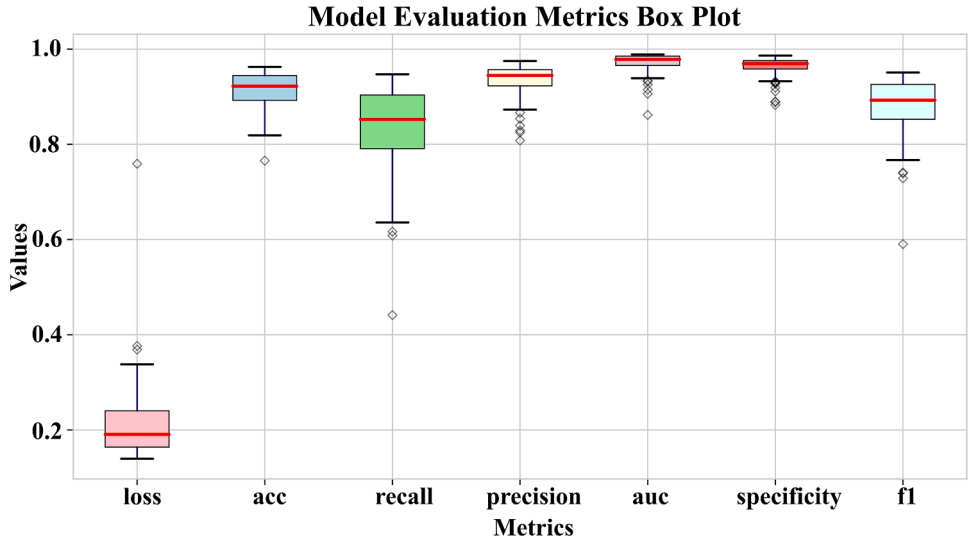

Fig. 6. The Box Plot of THO model (the Box Plot of other models in the Additional File 1).

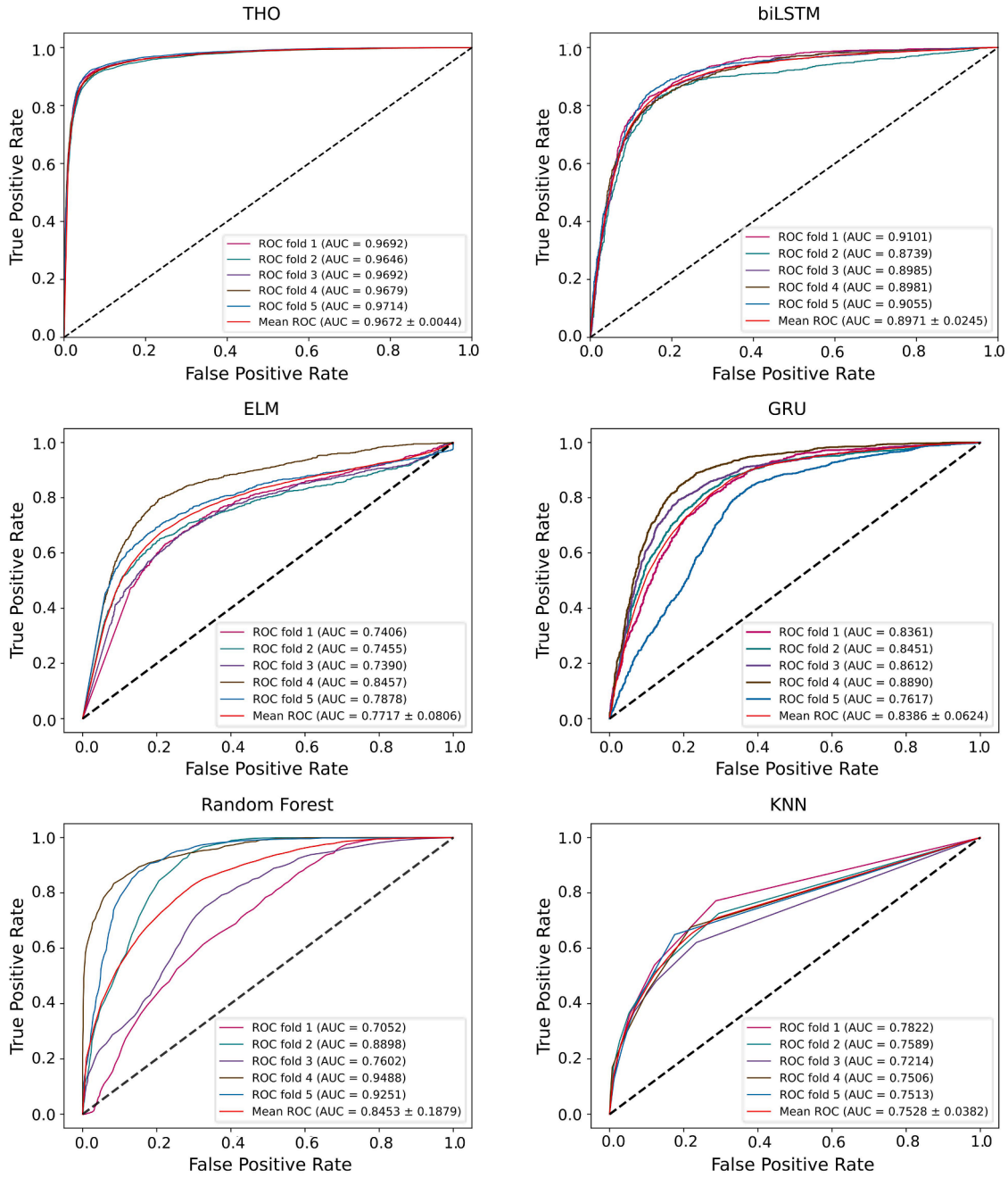

Fig. 7. The ROC curve of models (the first row from left to right is THO and biLSTM, the second row from left to right is ELM and GRU, the third row from left to right is Random Forest and KNN).

surpassed those of the baseline models, highlighting the superior predictive capability of the THO model. However, there was a nearly 4 % decrease in accuracy compared to the model utilizing the voting mechanism, underscoring the effectiveness of the voting approach proposed in this study. Comprehensive data regarding this can be consulted in Table 2.

3.4. Result of ablation study

To validate the effectiveness of our proposed method and its individual components, we conducted an ablation study on the Apnea ECG dataset. The study involved decomposing the TWLA, TBMA, and IntraExpanded Convolution modules and testing their feature extraction capabilities both individually and in various combinations. The results of the ablation experiments are presented in Table 3. It is evident that when operating independently, each module demonstrated substantial predictive capabilities, with models based solely on the TWLA, TBMA, and IntraExpanded Convolution achieving accuracy rates of 86.22 %, 85.04 %, and 81.13 %, respectively.

Notably, the IntraExpanded Convolution module, when used in isolation, showed some limitations in extracting features from time-related sequences, achieving an accuracy of only 81.13 %. However, the integration of this module with others significantly enhanced its ability to capture detailed features, thereby improving the overall feature extraction capability of the model. Specifically, when IntraExpanded Convolution was combined with TWLA and TBMA, the prediction accuracy increased by 1.71 % and 5.78 %, respectively.

In summary, the ablation study confirmed the effectiveness of each proposed module, demonstrating that their synergistic operation significantly enhances the model's ability to extract and predict features

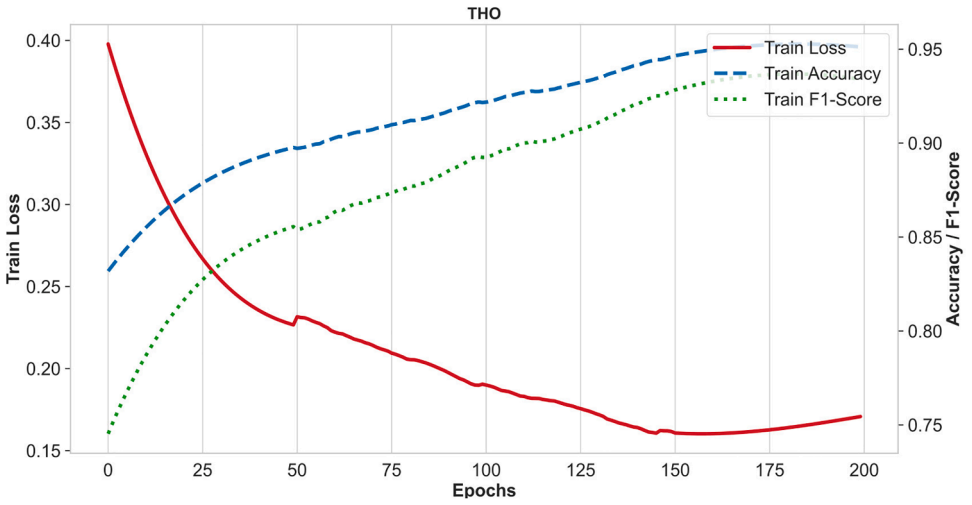

Fig. 8. The visualization of the training process.

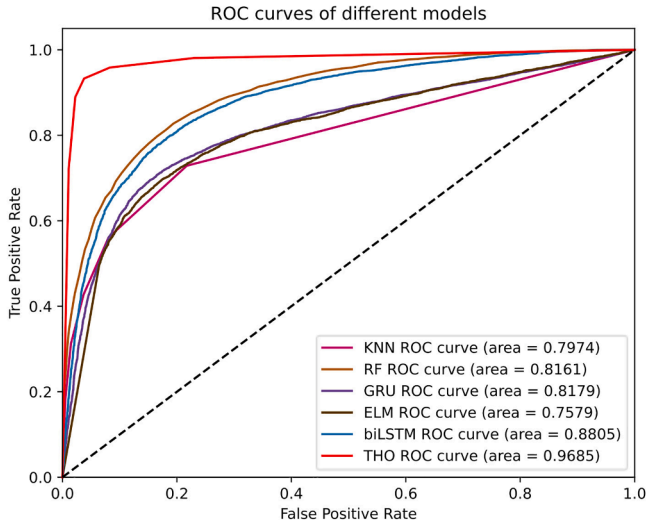

Fig. 9. The ROC curve of external validation.

Table 3
Ablation study of THO model. The THO model is decomposed into three parts: TWLA, IntraExpanded Convolution, and TBMA, to test the performance of the model under various combinations in ablation experiments.

Methods			Evaluation Metrics			
TWLA	IntraExpanded Convolution	TBMA	ACC	PRE	REC	AUC
√			86.22 %	82.31 %	81.49 %	85.32 %
	√		81.13 %	75.36 %	75.31 %	80.02 %
		√	85.04 %	80.25 %	80.79 %	84.23 %
√	√		87.93 %	84.68 %	83.61 %	87.11 %
√		√	83.39 %	78.84 %	77.34 %	82.24 %
	√	√	90.82 %	87.60 %	88.56 %	90.39 %
√	√	√	95.03 %	94.43 %	93.28 %	96.85 %

relevant to OSA.

3.5. Result of individual-level classification

Due to the pivotal role of individual-level diagnosis results in selecting treatment methods for OSA, we calculated the AHI from the segment-level diagnostic outputs of the THO model to classify the severity of OSA at the individual level. Fig. 10 presents the confusion matrix of the individual-level diagnostic results as calculated by the THO model. The model achieved a classification accuracy of 100 % at this level, thereby providing precise individual-level diagnosis and classification of OSA, which demonstrates significant potential for clinical application.

3.7. Model performance on UCDDB database

To further validate the generalization capability of the THO model, we segmented the ECG signals from the UCDDB database into one-minute intervals, performing five-fold cross-validation and external testing. The results, compared with SOTA models, are detailed in Table 4. In the cross-validation, the THO model exhibited an average

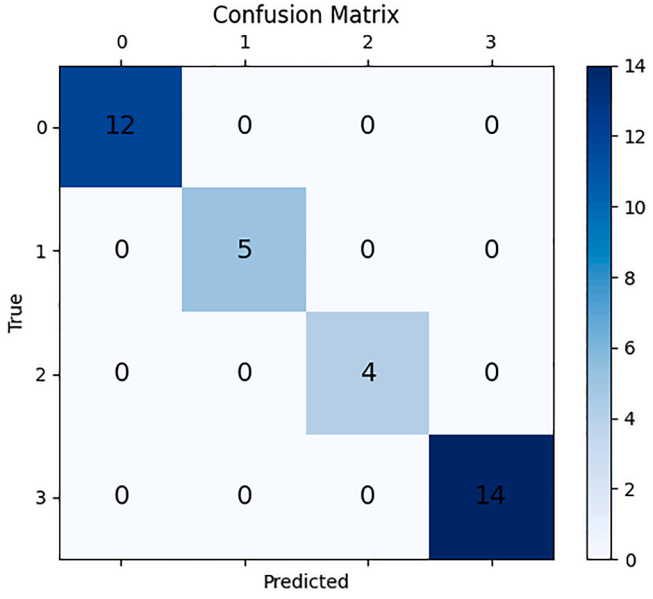

Fig. 10. The confusion matrix of individual-level classification.

Table 4
Model performance on UCDDB database.

model	ACC	PRE	REC	AUC
Yang et al. [17].	75.10 %	–	61.10 %	–
Fatimah et al. [41].	80.40 %	–	68.90 %	–
Srivastava et al. [42].	81.86 %	–	71.62 %	–
Nguyen et al. [43].	81.30 %	–	40.30 %	–
OURS	**85.35 %**	**24.57 %**	**81.13 %**	**90.04 %**

accuracy with a 95 % confidence interval of 86.20 % ±6.15 %. During external testing, the model achieved an accuracy of 85.35 % and an AUC of 90.04 %, marking a significant improvement of >3.49 % in accuracy relative to SOTA models.

Fig. 11 illustrates the ROC curve from the five-fold cross-validation, where the THO model consistently achieved an AUC of over 86 % across each fold, with an average AUC and a 95 % confidence interval of 87.57 %±1.22 %. These results underscore the model's robust predictive ability and its considerable potential for clinical application.

4. Discussion

In this study, we presented a deep learning model based on the Hybrid Neural Network to accurately detect OSA. The model extracts three features from the original ECG signal - EDR, RRI, and RAMP, moreover, these features are put into the TWLA for more precise feature exaction. The output of TWLA is combined with EDR, RRI and RAMP as model input. The input is first processed through an LSTM layer and two CNN layers for feature extraction, enhancing the model's understanding of the data. Subsequently, the processed data is fed into the Transformer layer and processed with TBMA to reduce the interference of early input on classification and improve detection accuracy. Then, the data is classified and output through the ReLU activation function and a fully connected layer. Moreover, we developed a voting mechanism to balance the predictive outcomes of five models trained through a five-fold cross-validation process, thereby achieving a Precision-Recall Trade-off. In the external validation conducted on the Apnea-ECG database, THO model's overall ACC reached 95.03 %, and the AUC reached 96.85 %, achieving the goal of accurate OSA detection. Furthermore, additional training and testing on the UCDDB dataset enhanced the THO model's performance in external evaluations, where it reached an accuracy of 85.35 % and a recall of 90.04 %. This represents an improvement of over 3 % compared to SOTA models [17,41–43], demonstrating the model's exceptional predictive power and generalization ability.

Conventional machine learning models often overly rely on input features, failing to recognize the features of input signals automatically. Moreover, they often fall short when dealing with complex or highly non-linear problems. Regular deep learning models also have issues with poor generalization ability, demonstrating suboptimal performance in OSA prediction. Therefore, the THO model proposed in this study effectively addresses these issues. Firstly, we extract three features - EDR, RRI, and RAMP, reducing the model's feature extraction workload, making it easier for the model to understand the correlation between the input and sleep apnea. The application of IntraExpanded Convolution and ResidualConv1d layers enables the model to effectively learn the local and global pattern information of respiratory signals in ECG. Meanwhile, the enhancement of transformer's multi-head attention mechanism enables the model to capture long-term time series dependencies without being disturbed by overly early sequences. This strategy also can capture dependencies between any two points in the sequence, greatly enhancing the model's understanding of time-dependent sequences. Lastly, the fully connected layer integrates the extracted features and makes non-linear combinations, enabling the model to learn higher-level feature representations, further improving the model's predictive capabilities. All these factors together contribute to the THO model's accurate predictive ability for OSA.

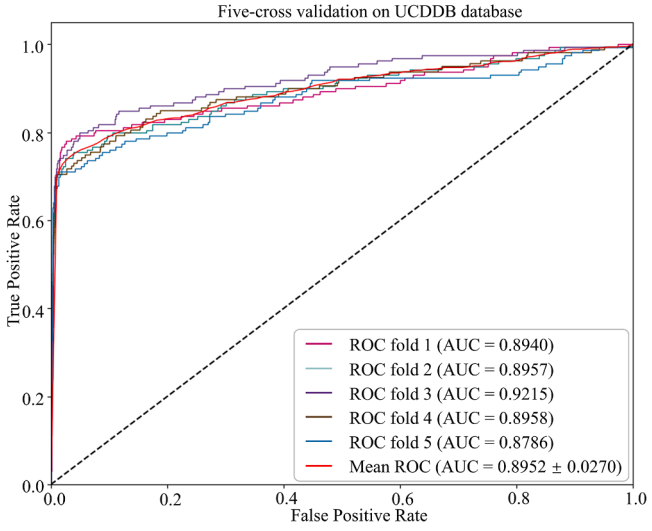

Fig. 11. The ROC curve of five-fold cross validation on the UCDDB dataset.

The outcomes of this study have been compared with other germane research utilizing the identical database. These research predominantly employ similar feature extraction methodologies or directly process raw ECG signals. In the realm of Transformer model application, numerous studies exhibit commendable performance. For instance, Liu et al. [44] developed a hybrid model integrating CNN with transformer and adopted global average pooling layer instead of the traditional fully connected layer. This model discerns OSA by processing raw ECG signals, achieving an ACC, PRE, and REC of 88.2 %, 89.0 %, and 78.5 %, respectively. Meanwhile, Hu et al. [44] designed a model with a Multiperspective Channel Attention (MPCA) block, which concurrently extracts and analyzes features from raw ECG, RRI, and more, followed by classification and OSA detection via Transformer layer, achieving an accuracy of 90.5 %, precision of 88.5 %, and recall of 86.5 %, indicative of exemplary performance. It is noteworthy that traditional models have also rendered significant results in OSA detection. For example, Chen et al. [50] developed a novel attention mechanism termed 'restricted attention,' which enhances the analysis of target segments by sequentially combining them with their preceding and following segments (each of five-minute duration) as input. Concurrently, the target segment is utilized as a query vector to synthesize information from adjacent segments to determine the presence of sleep apnea. Applied to the Apnea-ECG dataset, this method achieved an accuracy of 91.4 %. Additionally, Tyagi et al. [47] introduced an innovative approach in Deep Belief Networks (DBN), where they cascaded two distinct types of Restricted Boltzmann Machines (RBMs) to detect OSA through HRV and EDR signals extracted from ECG. Upon rigorous testing, this method achieved an accuracy rate of 89.11 %. Moreover, we conducted comparisons with OSA detection algorithms constructed based on HMM, LSTM, KNN, and other models, as detailed in Table 5.

Consequently, through meticulous comparison with other pertinent studies, this research has significantly enhanced the detection accuracy of OSA, and effectively addressed the issue of insufficient accuracy when utilizing ECG for OSA detection. Looking ahead, this method holds promise for widespread application in scenarios that demand precise OSA detection. Upon optimization through model lightweight, it possesses the potential to be integrated into wearable devices, facilitating the early detection and treatment of OSA, and delivering superior service in clinical management and OSA prevention.

Nonetheless, our study is not without limitations. Firstly, the THO model we developed incorporates the CNN, LSTM, and Transformer, which results in a substantial model size that is unsuitable for deployment on portable devices with limited computational capabilities, hence model lightweight is a crucial direction for development. Secondly, as this study solely employs data from the Apnea-ECG database, the limited

Table 5
The comparison of performance between proposed model and other works. The best results for each evaluation metric are displayed using boldface numbers.

Work	Method	Input type	ACC (%)	PRE (%)	REC/SEN (%)	AUC (%)
Liu et al. [24]	CNN+Transformer	Raw ECG	88.2	89	78.5	94.7
Hu et al. [44]	CNN+Transformer	Raw ECG, RA, RRI, RRID	90.5	88.5	86.5	96.1
Deng et al. [45]	CNN+Self-Attention	RRI, R-peak	91.4	–	88.75	–
Fan et al. [46]	CNN+Self-Attention	RRI, R-peak	91.3	92.6	89.1	**96.9**
Chen et al. [16]	CNN+GRU	RRI, R-peak	91.2	86.5	94.2	–
Feng et al. [47]	HMM	RRI	85.1	77.2	86.2	–
Yang et al. [17]	CNN	RRI, RA, and QA	90.3	–	87.6	–
Cheng et al. [18]	CNN	Raw ECG	87.0	–	81.6	–
Zhou et al. [19]	CNN	HRV and GAF	90.9	83.9	95.3	89.0
Hassan et al. [48]	CNN	RRI	88.0	–	94.0	–
Bahrami et al. [49]	LSTM	RRI, RA	80.2	–	75.1	–
Tyagi et al. [50]	DBN	EDR, HRV	86.2	–	82.6	94.0
Wang et al. [51]	Multiscale Neural Network	RRI	90.4	–	83.3	–
Bahrami et al. [52]	CNN+BiLSTM	RRI, R-peak amplitude	88.1	–	81.5	–
Nasifoglu et al. [53]	CNN	Raw ECG	82.3	–	83.2	90.0
Ours	**Time-Hybrid OSAformer**	**EDR, RRI, RAMP**	**95.0**	**94.4**	**93.3**	**96.9**

data volume does not fully take advantage of proposed model, coupled with a lack of validation using real-world clinical data; therefore, there is an urgent need to amass more authentic clinical data for validation. Concerning the UCDDB database, the model exhibits a decline in performance relative to the results obtained on the Apnea-ECG dataset. This discrepancy can be attributed to two primary factors. One reason is that the labels in the UCDDB database were derived based on the reported occurrence and duration of apnea events as provided by the dataset, which we converted into minute-by-minute labels. In contrast, the Apnea-ECG database features expert-annotated labels for each minute. Consequently, the potential inaccuracies in our derived labels for the UCDDB might have contributed to lower model performance compared to the expert-annotated Apnea-ECG database. Another contributing factor is the significant class imbalance in the UCDDB database, with 6445 normal signals compared to only 337 apnea signals in the training set. This imbalance may not adequately reflect the complexity of real apnea events, likely impacting the model's learning efficacy and its subsequent predictive performance. This observation of a performance discrepancy underscores potential avenues for enhancing our proposed model, particularly in addressing sample imbalances and improving generalization across diverse datasets. Lastly, the three features we extracted, namely EDR, RRI, and RAMP, have a large scale of 240 dimensions per minute for one feature, resulting in a total of 720 dimensions per minute for input data. This significantly prolongs the model's training time and complicates the interpretation of the input, making the simplification of input features, while maintaining the same results, a worthwhile avenue to explore.

5. Conclusion

In this study, a model named THO is proposed for precisely detecting OSA through single-lead ECG signals. Initially, EDR, RRI, and RAMP are extracted from the ECG signals, concurrently, features derived from the TWLA were integrated with them to serve as input features for the model. Subsequently, the model employs two CNN layers and one Transformer layer to accomplish the classification task of OSA. Notably, within the Transformer layer, we enhanced the Multihead-Attention by incorporating a temporal decay mechanism, thus refining it into a Time-Based Multihead-Attention, further improving the model's interpretability. Upon validation through testing, the proposed methodology achieved accuracy of 95.03 %, substantially augmenting the precision of OSA detection. With the continual advancements in research, the next steps entail the lightweight of the model and the development of a system capable of real-time processing and detection of OSA. This method holds promise for application in clinical settings, facilitating the precise detection and timely prevention of OSA.

Funding

This study was supported by the National Natural Science Foundation of China (No. 62406211) and the Natural Science Foundation of Sichuan Province (No. 2024NSFSC0654 and No. 2023NSFSC0636).

Availability of data and materials

The dataset for model training and testing is available at https://physionet.org/content/apnea-ecg/1.0.0/ (accessed on 15 Sep 2023). The codes of this study are available on request from the corresponding author.

CRediT authorship contribution statement

Lingxuan Hou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Yan Zhuang:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Heng Zhang:** Writing – review & editing, Investigation. **Gang Yang:** Methodology, Formal analysis. **Zhan Hua:** Software, Data curation. **Ke Chen:** Software. **Lin Han:** Validation. **Jiangli Lin:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that we have no conflict of interest. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

Acknowledgments

The authors gratefully acknowledge technical and financial support provided by the Sichuan University. The authors would like to thank Dr. Thomas Penzel of Phillips University for providing the Apnea-ECG dataset, which was used in this research and is available in the public domain. We also acknowledge the computing services provided by the Biomedical Engineering Experimental Teaching Center of Sichuan University (Chengdu, China).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2024.108558.

References

[1] A. Abbasi, S.S. Gupta, N. Sabharwal, V. Meghrajani, S. Sharma, S. Kamholz, Y. Kupfer, A comprehensive review of obstructive sleep apnea, Sleep. Sci. 14 (2021) 142.

[2] L.A. Salman, R. Shulman, J.B. Cohen, Obstructive sleep apnea, hypertension, and cardiovascular risk: epidemiology, pathophysiology, and management, Curr. Cardiol. Rep. 22 (2020) 1–9.

[3] J. Vanek, J. Prasko, S. Genzor, M. Ociskova, K. Kantor, M. Holubova, M. Slepecky, V. Nesnidal, A. Kolek, M. Sova, Obstructive sleep apnea, depression and cognitive impairment, Sleep Med. 72 (2020) 50–58.

[4] J.L. Chang, A.N. Goldberg, J.A. Alt, A. Mohammed, L. Ashbrook, D. Auckley, I. Ayappa, H. Bakhtiar, J.E. Barrera, B.L. Bartley, International consensus statement on obstructive sleep apnea. International Forum of Allergy & Rhinology, Wiley Online Library, 2023, pp. 1061–1482.

[5] R.B. Berry, R. Brooks, C. Gamaldo, S.M. Harding, R.M. Lloyd, S.F. Quan, M. T. Troester, B.V. Vaughn, AASM Scoring Manual Updates for 2017 (version 2.4), American Academy of Sleep Medicine, 2017, pp. 665–666.

[6] Y. Wei, Y. Liu, N. Ayas, I. Laher, A narrative review on obstructive sleep apnea in China: a sleeping giant in disease pathology, Heart Mind 6 (2022) 232–241.

[7] M. Hajipour, B. Baumann, A. Azarbarzin, A.H. Allen, Y. Liu, S. Fels, S. Goodfellow, A. Singh, R. Jen, N.T. Ayas, Association of alternative polysomnographic features with patient outcomes in obstructive sleep apnea: a systematic review, J. Clin. Sleep Med. 19 (2023) 225–242.

[8] M. Duarte, P. Pereira-Rodrigues, D. Ferreira-Santos, The role of novel digital clinical tools in the screening or diagnosis of obstructive sleep apnea: systematic review, J. Med. Internet. Res. 25 (2023) e47735.

[9] J. Levy, D. Álvarez, F. Del Campo, J.A. Behar, Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry, Nat. Commun. 14 (2023) 4881.

[10] J. Jiménez-García, M. García, G.C. Gutiérrez-Tobal, L. Kheirandish-Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo, D. Gozal, R. Hornero, An explainable deep-learning architecture for pediatric sleep apnea identification from overnight airflow and oximetry signals, Biomed. Signal Process. Control 87 (2024) 105490.

[11] X. Zhao, X. Wang, T. Yang, S. Ji, H. Wang, J. Wang, Y. Wang, Q. Wu, Classification of sleep apnea based on EEG sub-band signal characteristics, Sci. Rep. 11 (2021) 5824.

[12] L. Cheng, S. Luo, X. Yu, H. Ghayvat, H. Zhang, Y. Zhang, EEG-CLNet: collaborative learning for simultaneous measurement of sleep stages and OSA events based on single EEG signal, IEEE Trans. Instrum. Meas. 72 (2023) 1–10.

[13] A.F. Jackson, D.J. Bolger, The neurophysiological bases of EEG and EEG measurement: a review for the rest of us, Psychophysiology. 51 (2014) 1061–1071.

[14] M.A. Espinosa, P. Ponce, A. Molina, V. Borja, M.G. Torres, M. Rojas, Advancements in home-based devices for detecting obstructive sleep apnea: a comprehensive study, Sensors 23 (2023) 9512.

[15] T. Penzel, J. McNames, P. De Chazal, B. Raymond, A. Murray, G. Moody, Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings, Med. Biol. Eng. Comput. 40 (2002) 402–407.

[16] J. Chen, M. Shen, W. Ma, W. Zheng, A spatio-temporal learning-based model for sleep apnea detection using single-lead ECG signals, Front. Neurosci. 16 (2022) 972581.

[17] Q. Yang, L. Zou, K. Wei, G. Liu, Obstructive sleep apnea detection from single-lead electrocardiogram signals using one-dimensional squeeze-and-excitation residual group network, Comput. Biol. Med. 140 (2022) 105124.

[18] C.Y. Yeh, H.Y. Chang, J.Y. Hu, C.C. Lin, Contribution of different Subbands of ECG in sleep apnea detection evaluated using filter bank decomposition and a convolutional neural network, Sensors (Basel) (2022) 22.

[19] Y. Zhou, Y. He, K. Kang, OSA-CCNN: obstructive sleep apnea detection based on a composite deep convolution neural network model using single-lead ECG signal, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2022, pp. 1840–1845.

[20] D. Romero, R. Jané, Dynamic bayesian model for detecting obstructive respiratory events by using an experimental model, Sensors 23 (2023) 3371.

[21] X. Yan, L. Wang, J. Zhu, S. Wang, Q. Zhang, Y. Xin, Automatic obstructive sleep apnea detection based on respiratory parameters in physiological signals, in: 2022 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2022, pp. 461–466.

[22] N. Salari, A. Hosseinian-Far, M. Mohammadi, H. Ghasemi, H. Khazaie, A. Daneshkhah, A. Ahmadi, Detection of sleep apnea using Machine learning algorithms based on ECG Signals: a comprehensive systematic review, Expert Syst. Appl. 187 (2022) 115950.

[23] F. Xia, H. Li, Y. Li, X. Liu, Y. Xu, C. Fang, Q. Hou, S. Lin, Z. Zhang, J. Yang, Minimally invasive hypoglossal nerve stimulator enabled by ECG sensor and WPT to manage obstructive sleep apnea, Sensors 23 (2023) 8882.

[24] H. Liu, S. Cui, X. Zhao, F. Cong, Detection of obstructive sleep apnea from single-channel ECG signals using a CNN-transformer architecture, Biomed. Signal Process. Control 82 (2023) 104581.

[25] T. Penzel, G.B. Moody, R.G. Mark, A.L. Goldberger, J.H. Peter, The apnea-ECG database, in: Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163), IEEE, 2000, pp. 255–258.

[26] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J. E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.

[27] J.J. Goldberger, N.P. Johnson, H. Subacius, J. Ng, P. Greenland, Comparison of the physiologic and prognostic implications of the heart rate versus the RR interval, Heart. Rhythm. 11 (2014) 1925–1933.

[28] P. Przystup, A. Poliński, J. Wtorek, QRS morphology-based EDR Signal—Factors determining its properties, IEEe Access. 10 (2022) 34665–34676.

[29] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. (1985) 230–236.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv Neural Inf Process Syst (2017) 30.

[31] A. Iwasaki, C. Nakayama, K. Fujiwara, Y. Sumi, M. Matsuo, M. Kano, H. Kadotani, Screening of sleep apnea based on heart rate variability and long short-term memory, Sleep Breath. 25 (2021) 1821–1829.

[32] A. Iwasaki, K. Fujiwara, C. Nakayama, Y. Sumi, M. Kano, T. Nagamoto, H. Kadotani, RR interval-based sleep apnea screening by a recurrent neural network in a large clinical polysomnography dataset, Clin. Neurophysiol. 139 (2022) 80–89.

[33] Y. Wang, Z. Xiao, S. Fang, W. Li, J. Wang, X. Zhao, BI-Directional long short-term memory for automatic detection of sleep apnea events based on single channel EEG signal, Comput. Biol. Med. 142 (2022) 105211.

[34] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555, (2014).

[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.

[36] N.L.C. Group, R-NET: machine reading comprehension with self-matching networks, (2017).

[37] B. Pang, S. Doshi, B. Roy, M. Lai, L. Ehlert, R.S. Aysola, D.W. Kang, A. Anderson, S. H. Joshi, D. Tward, Machine learning approach for obstructive sleep apnea screening using brain diffusion tensor imaging, J. Sleep Res. 32 (2023) e13729.

[38] A.A. Hamidi, B. Robertson, J. Ilow, A new approach for ECG artifact detection using fine-KNN classification and wavelet scattering features in vital health applications, Procedia Comput. Sci. 224 (2023) 60–67.

[39] S. Kuila, N. Dhanda, S. Joardar, ECG signal classification to detect heart arrhythmia using ELM and CNN, Multimed. Tools Appl. 82 (2023) 29857–29881.

[40] L. Parrino, R. Ferri, M. Zucconi, F. Fanfulla, Commentary from the Italian Association of Sleep Medicine on the AASM manual for the scoring of sleep and associated events: for debate and discussion, Sleep Med. 10 (2009) 799–808.

[41] B. Fatimah, P. Singh, A. Singhal, R.B. Pachori, Detection of apnea events from ECG segments using Fourier decomposition method, Biomed. Signal Process. Control 61 (2020) 102005.

[42] G. Srivastava, A. Chauhan, N. Kargeti, N. Pradhan, V.S. Dhaka, ApneaNet: a hybrid 1DCNN-LSTM architecture for detection of obstructive sleep apnea using digitized ECG signals, Biomed. Signal Process. Control 84 (2023) 104754.

[43] H.X. Nguyen, D.V. Nguyen, H.H. Pham, C.D. Do, MPCNN: a novel matrix profile approach for CNN-based single lead sleep apnea in classification problem, IEEE J. Biomed. Health Inform. (2024).

[44] S. Hu, W. Cai, T. Gao, M. Wang, A hybrid transformer model for obstructive sleep apnea detection based on self-attention mechanism using single-lead ECG, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.

[45] J. Deng, Y. Jiang, Z.B. Chen, J.W. Rhee, Y. Deng, Z.V. Wang, Mitochondrial dysfunction in cardiac arrhythmias, Cells 12 (2023).

[46] X. Fan, X. Chen, W. Ma, W. Gao, BAFNet: bottleneck attention based fusion network for sleep apnea detection, IEEE J. Biomed. Health Inform. 28 (2024) 2473–2484.

[47] K. Feng, H. Qin, S. Wu, W. Pan, G. Liu, A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram, IEEE Trans. Instrum. Meas. 70 (2020) 1–12.

[48] O. Hassan, T. Paul, N. Amin, T. Titirsha, R. Thakker, D. Parvin, A.S.M. Mosa, S. K. Islam, An optimized hardware inference of SABiNN: shift-accumulate Binarized neural network for sleep apnea detection, IEEE Trans. Instrum. Meas. (2023).

[49] M. Bahrami, M. Forouzanfar, Detection of sleep apnea from single-lead ECG: comparison of deep learning algorithms, in: 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2021, pp. 1–5.

[50] P.K. Tyagi, D. Agrawal, Automatic detection of sleep apnea from single-lead ECG signal using enhanced-deep belief network model, Biomed. Signal Process. Control 80 (2023) 104401.

[51] Z. Wang, C. Peng, B. Li, T. Penzel, R. Liu, Y. Zhang, X. Yu, Single-lead ECG based multiscale neural network for obstructive sleep apnea detection, IoT 20 (2022) 100613.

[52] M. Bahrami, M. Forouzanfar, Sleep apnea detection from single-lead ECG: a comprehensive analysis of machine learning and deep learning algorithms, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.

[53] H. Nasifoglu, O. Erogul, Obstructive sleep apnea prediction from electrocardiogram scalograms and spectrograms using convolutional neural networks, Physiol. Meas. (2021) 42.