Neural Networks 162 (2023) 571–580

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

RAFNet: Restricted attention fusion network for sleep apnea detection

Ying Chen [b,1], Huijun Yue [c,1], Ruifeng Zou [b], Wenbin Lei [c], Wenjun Ma [b], Xiaomao Fan [a,*]

[a] College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China
[b] School of Computer Science, South China Normal University, Guangzhou, China
[c] Otorhinolaryngology Hospital, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

ARTICLE INFO

Article history:
Received 13 July 2022
Received in revised form 2 February 2023
Accepted 14 March 2023
Available online 21 March 2023

Keywords:
Sleep apnea detection
Restricted attention
ECG signals
Feature fusion

ABSTRACT

Sleep apnea (SA) is a common sleep-related breathing disorder, which would lead to damage of multiple systemic organs or even sudden death. In clinical practice, portable device is an important tool to monitor sleep conditions and detect SA events by using physiological signals. However, SA detection performance is still limited due to physiological signals with time-variability and complexity. In this paper, we focus on SA detection with single lead ECG signals, which can be easily collected by a portable device. Under this context, we propose a restricted attention fusion network called RAFNet for sleep apnea detection. Specifically, RR intervals (RRI) and R-peak amplitudes (Rpeak) are generated from ECG signals and divided into one-minute-long segments. To alleviate the problem of insufficient feature information of the target segment, we combine the target segment with two pre- and post-adjacent segments in sequence, (*i.e.* a five-minute-long segment), as the input. Meanwhile, by leveraging the target segment as the query vector, we propose a new restricted attention mechanism with cascaded morphological and temporal attentions, which can effectively learn the feature information and depress redundant feature information from the adjacent segments with adaptive assigning weight importance. To further improve the SA detection performance, the target and adjacent segment features are fused together with the channel-wise stacking scheme. Experiment results on the public Apnea-ECG dataset and the real clinical FAH-ECG dataset with sleep apnea annotations show that the RAFNet greatly improves SA detection performance and achieves competitive results, which are superior to those achieved by the state-of-the-art baselines.

1. Introduction

Sleep apnea (SA) is a sleep-related breathing disorder characterized by repeated complete or incomplete cessation of airflow during sleep. It would cause hypertension (Floras, 2015), coronary heart disease (Javaheri et al., 2017), stroke (Baillieul et al., 2022), heart failure (Wang et al., 2007), or even sudden death (Somboon, Grigg-Damberger, & Foldvary-Schaefer, 2019). According to the report in Benjafield et al. (2019), there are approximately 1 billion people with the SA problem worldwide. Especially in the adult population, more than 14% of men and 5% of women suffer from SA (Peppard et al., 2013). In clinical settings, polysomnography (PSG) is a gold standard for diagnosing SA. However, PSG is complex, low efficient, and heavy-load, as well as requiring a trained physician, which seriously blocks the application of SA detection in a large number of population. Therefore, how to effectively and accurately detect SA out of hospital is still a challenge.

As reported in Salari et al. (2022), breathing airflow has a high relation with heart rhythm. SA decreases the patient's oxygen level in the heart cells and reduces the heart rate. The brain then increases the respiration rate by using instant pulses to the respiratory system, leading to an increase of heart rate. ECG, as one kind of physiological signals, records the electrical activity of the heart, which can capture the changes of heart rhythm. Meanwhile, the amplitudes of a respiration signal also have a closed relationship with the ECG envelope (Wang, Lu, Shen, & Hong, 2019). With the merits of abundant physiological information and easy collection by a portable device, many researchers utilize ECG signals to design automatical SA detection methods, which can be mainly grouped into two categories: hand-crafted feature engineering and end-to-end feature engineering. It is noted that ECG signals are time-varying, complex, and sensitive to noise. How to accurately extract hand-crafted features is a big problem which would lead to greatly degrade the SA detection performance in a real clinical setting. Deep learning is powerful to automatically learn an end-to-end feature representation and achieves a big success in computer vision, natural language processing, and many other domains. Therefore, researchers attempt to leverage deep learning technologies to

* Corresponding author.
E-mail address: fanxiaomao@sztu.edu.cn (X. Fan).
1 Contributed equally.

https://doi.org/10.1016/j.neunet.2023.03.019
0893-6080/

address the SA detection problem, such as LeNet (Wang et al., 2019), LSTM (Faust, Barika, Shenfield, Ciaccio, & Acharya, 2021), CNN-LSTM (Almutairi, Hassan, & Datta, 2021), SEMSCNN (Chen, Chen, Ma, Fan, & Li, 2021), ConCAD (Huang & Ma, 2021), and MSDA-1DCNN (Shen, Qin, Wei, & Liu, 2021). As reported in Chen et al. (2021), Huang and Ma (2021), Wang et al. (2019), leveraging adjacent-ECG-segment information would greatly boost the SA detection performance.

Although SA detection models with adjacent-segment information can achieve comparable promising results, there still exists a major limitation: how to effectively learn the adjacent-segment feature representation and depress redundant noise information. Some studies directly utilized adjacent segments as input (Wang et al., 2019) or query vector (Chen et al., 2021; Huang & Ma, 2021) for an attention module. By this paradigm, SA models learn the adjacent-segment feature representation containing effective information as well as introducing redundant information (*i.e.* noise).

In this study, we follow the paradigm in Huang and Ma (2021), Wang et al. (2019) to utilize adjacent-segment information and firstly propose a new restricted attention fusion network (*i.e.* RAFNet) for SA detection. By using the target segment as the query vector, the restricted attention mechanism can effectively learn the adjacent-segment feature representation, as well as depressing the redundant feature information. Furthermore, we fuse the target-segment feature representation and adjacent-segment feature representation to boost the SA detection performance. In addition, to further verify generic performance of SA detection, the proposed RAFNet is validated on a real clinical FAH-ECG dataset collected from the First Affiliated Hospital, Sun Yat-sen University in China.

Experiment results on the public domain Apnea-ECG dataset and the real clinical FAH-ECG dataset show that RAFNet achieves competitive SA detection results, both of which are better than those achieved by the state-of-the-art SA detection baselines.

To sum up, our contributions can be summarized as follows:

- By leveraging target segment as the query vector, we propose a novel restricted attention mechanism with cascaded morphological and temporal attention modules, which can effectively learn the adjacent-segment feature information and depress redundant feature information from adjacent segments.
- To further boost the SA detection performance, we propose a new restricted attention fusion network (*i.e.* RAFNet) by fusing target segment feature representation and adjacent-segment feature representation.
- Our proposed RAFNet is validated on two ECG datasets: a public domain Apnea-ECG dataset and a real clinical FAH-ECG dataset. Experiment results show that RAFNet can achieve competitive results, which outperforms state-of-the-art baselines.

The remained content of this paper is organized as follows. Related work is briefly introduced in Section 2. In Section 3, the details of the proposed SA detection method are described, which mainly includes data preprocessing, the proposed SA detection method (*i.e.* RAFNet), and performance metrics. Results and discussion are illustrated in Section 4. Section 5 concludes this paper.

2. Related work

In clinical practice, PSG is a gold standard for SA detection, which is complex, low efficient, and heavy-load. Therefore, many researchers attempted to utilize a single mode of physiological signals like EEG, SpO2, and ECG, to build SA detection models. Based on EEG signals, Taran, Bajaj, Sinha, and Polat (2021) proposed automatic SA detection models with using single feature Lampel–Ziv complexity of EEG signals. Mahmud et al. (2020) divided EEG data into smaller sub-frames and proposed convolutional neural network (FCNN) that can take each sub-frame of a single frame individually to extract local features to detect SA. Vimala, Ramar, and Ettappan (2019) selected important features from the decomposed EEG signals and used machine learning classifiers for the SA classification. Bhattacharjee, Saha, Fattah, Zhu, and Ahmad (2018) developed a unique multiband subframe based feature extraction scheme to capture the feature variation pattern of EEG, and used K Nearest Neighbor (KNN) to classify SA. Although EEG signals can be utilized to reflect SA events, it always needs complex machine with numerous probes and self-adhesive electrodes to collect signals and had a bad effect on patient's sleeping. As for SpO2 signals, Deviaene et al. (2018) designed a methodology for the automatic detection of SA events based on SpO2 signals, and achieved the best performance with a random forest classifier based on six desaturation severity. John, Nundy, Cardiff, and John (2021) developed an apnea detection a 1-dimensional convolutional neural network SomnNET based on SpO2. However, SpO2 signals are often unstable as they are generated from PPG signals which are susceptible to outside interference.

As ECG signals are with the merits of low cost, non-invasive, and easy collection, many researchers attempted to use them to design an automatical SA detection method. Generally, traditional machine learning algorithms are often quite simple and easy to be implemented, which are utilized for designing SA detection models. Song, Liu, Zhang, Chen, and Xian (2015) designed a SA detection model based on Hidden Markov Model (HMM) to capture the temporal dependence within segmented ECG signals. Tripathy, Gajbhiye, and Acharya (2020) generated cardiopulmonary (CP) signals from ECG signals, and developed two SA detection models based on support vector machines (SVM) and Random Forest (RF). Based on discrete wavelet transform technology, Rajesh, Dhuli, and Kumar (2021) extracted features from one-minute-long segmented ECG signals and built SA detection models based on linear discriminant analysis (LDA), K-nearest neighbors (KNN), SVM and RF. Li, Pan, Li, Jiang, and Liu (2018a) proposed SlTr-fApEn as a nonlinear index for analyzing HRV in SA. Yoon et al. (2017) developed an automatic slow-wave sleep (SWS) detection algorithm based on autonomic activations derived from the HRV. Zarei and Asl (2018) employed a non-linear feature extraction using Wavelet Transform (WT) coefficients obtained by an ECG signal and used different classification methods. Machine learning based SA detection models achieved promising results, however, they highly depend on the hand-crafted features, which are quite sensitive to the quality of ECG signals. Their SA detection performance would degrade sharply when ECG signals are contaminated by noise.

Deep learning has a powerful ability of end-to-end feature learning without the requirement of domain knowledge. Therefore, researchers shift their focus from traditional machine learning to deep learning technology. Van Steenkiste, Groenendaal, Deschrijver, and Dhaene (2018) proposed a method based on deep learning with LSTM using raw physiological respiratory signals to automatically learn relevant features and detect SA. Faust et al. (2021) extracted beat-to-beat interval traces, medically known as RR intervals, from ECG signals, and designed a LSTM-based deep neural network that classifies apnea versus non-apnea. Shen et al. (2021) proposed a SA detection method based on a multiscale dilation attention-based 1-D convolutional neural network (MSDA-1DCNN). Almutairi et al. (2021) developed a hybrid model of CNN and LSTM to detect SA events. Ye et al. (2021) proposed a novel frequency extraction network for SA

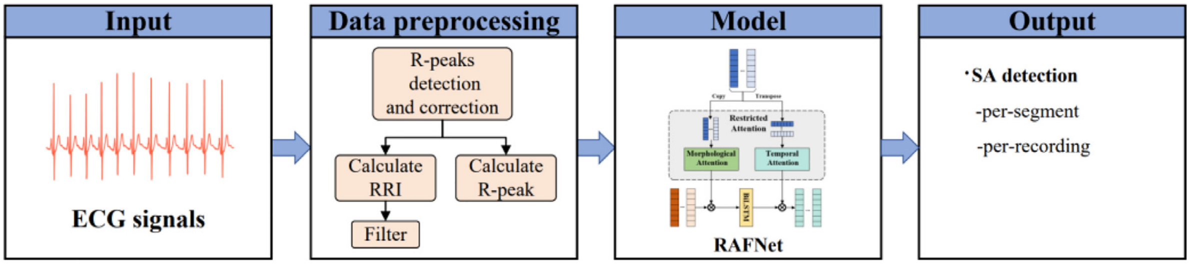


Fig. 1. The procedure of SA detection. It mainly includes data preprocessing and proposed SA detection model (i.e. RAFNet).

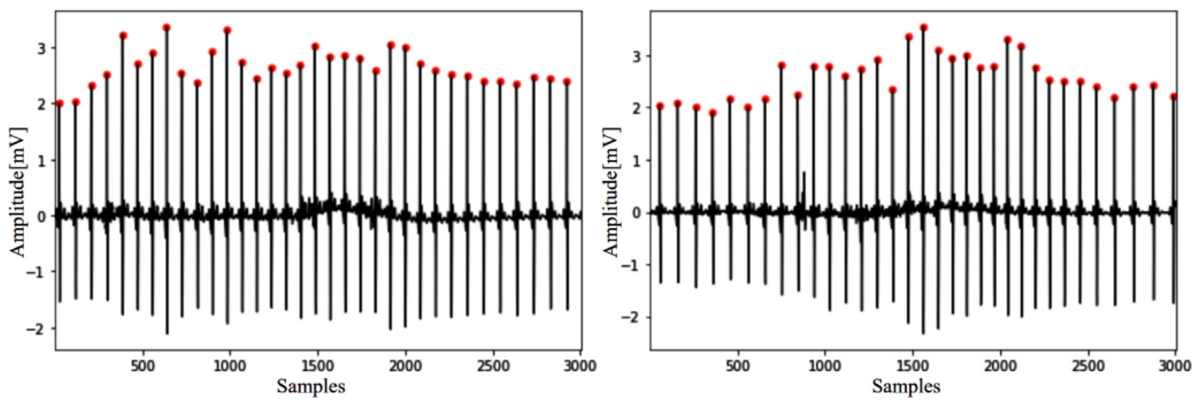


Fig. 2. Examples of filtered ECG and corrected positions of R peaks. The line in black represents the filtered ECG and the points in red are the corrected positions of R peaks.

detection. Bahrami and Forouzanfar (2021) extracted the R peak amplitudes and RR intervals, and implemented a multiple deep learning model with CNN, LSTM, BiLSTM, GRU, and deep mixed models for SA detection. Wang et al. (2019) proposed a modified LeNet-5 convolutional neural network with adjacent segments, demonstrating that adjacent segments information contributes to SA detection significantly. Chen et al. (2021) proposed a multi-scaled fusion network called SE-MSCNN for SA detection. Huang and Ma (2021) employed the contrastive learning mechanism for SA detection. Deep learning based SA detection methods obviously obtained much better results with adjacent segment information. However, they are still not aware of the affection by introducing the redundant feature information from the adjacent segments.

3. Methods

Fig. 1 shows the whole procedure of SA detection. It mainly includes data preprocessing and SA detection model, which are described in detail as follows.

3.1. Data preprocessing

According to breathing airflow signals, physicians can directly diagnose whether a subject suffered from SA. It is noted that R-peak amplitudes (Rpeak) are with a close relation to breathing airflow, and RR intervals (RRI) are indirect indicators of heart rhythm which are another close relation to breathing airflow. Both Rpeak and RRI can be extracted from raw ECG signals directly. Many studies (Huang & Ma, 2021; Sharan, Berkovsky, Xiong, & Coiera, 2020; Sharma & Sharma, 2016; Wang et al., 2019) also reported that RRI and Rpeak are effective for SA detection. As reported in Huang and Ma (2021), Wang et al. (2019), an overnight ECG signal is divided into segments with the length of one minute. Besides, they also manifested that adjacent ECG segments can greatly improve the SA detection performance. In this study, we follow this paradigm with a sequential ECG segments which are composed of target ECG segment and its surrounding two adjacent ECG segments. Specifically, we firstly segment overnight ECG signals into five-minute-long segments (i.e. one-minute-long target segment and its surrounding two one-minute-long segments) with an overlap of four minutes. As ECG signals are easy to be contaminated by noise, we use a Finite Impulse Response (FIR) filter to depress noise from raw ECG signals and obtain filtered ECG signals. Then we apply the Hamilton algorithm (Hamilton, 2002) to locate the positions of R peaks and correct positions to the maximum within the time interval of 0.1 s. As shown in Fig. 2, we randomly selected two examples to show filtered ECG and corrected positions of R peaks. It is observed that R peaks can be detected effectively. After that, we extract the amplitudes of the R peaks as Rpeak and calculate the distances between R peaks as RRI. As reported in Wang et al. (2019), there may be some physiologically uninterpretable points existed in RRI. Thus, the extracted RRI is further processed with a median filter (Chen, Zhang, & Song, 2014) to improve the signal quality. Finally, in order to satisfy the input of model, we employ the cubic interpolation technique to fix the length of RRI/Rpeak and make them equal in length of 900 points. Both RRI and Rpeak together are as input to be fed into the RAFNet described in the subsequent section.

3.2. Proposed RAFNet

3.2.1. Overview

In this study, our goal is to build an automatic SA detection framework using single-lead ECG signals. As RRI and Rpeak extracted from the target ECG segments as well as adjacent ECG segments (*i.e.* five-minute-long ECG segments) manifest the effectiveness in SA detection task (Huang & Ma, 2021; Wang et al., 2019), in this study we also follow this paradigm, extracting RRI and Rpeak from raw ECG signals (see in Section 3.1). To

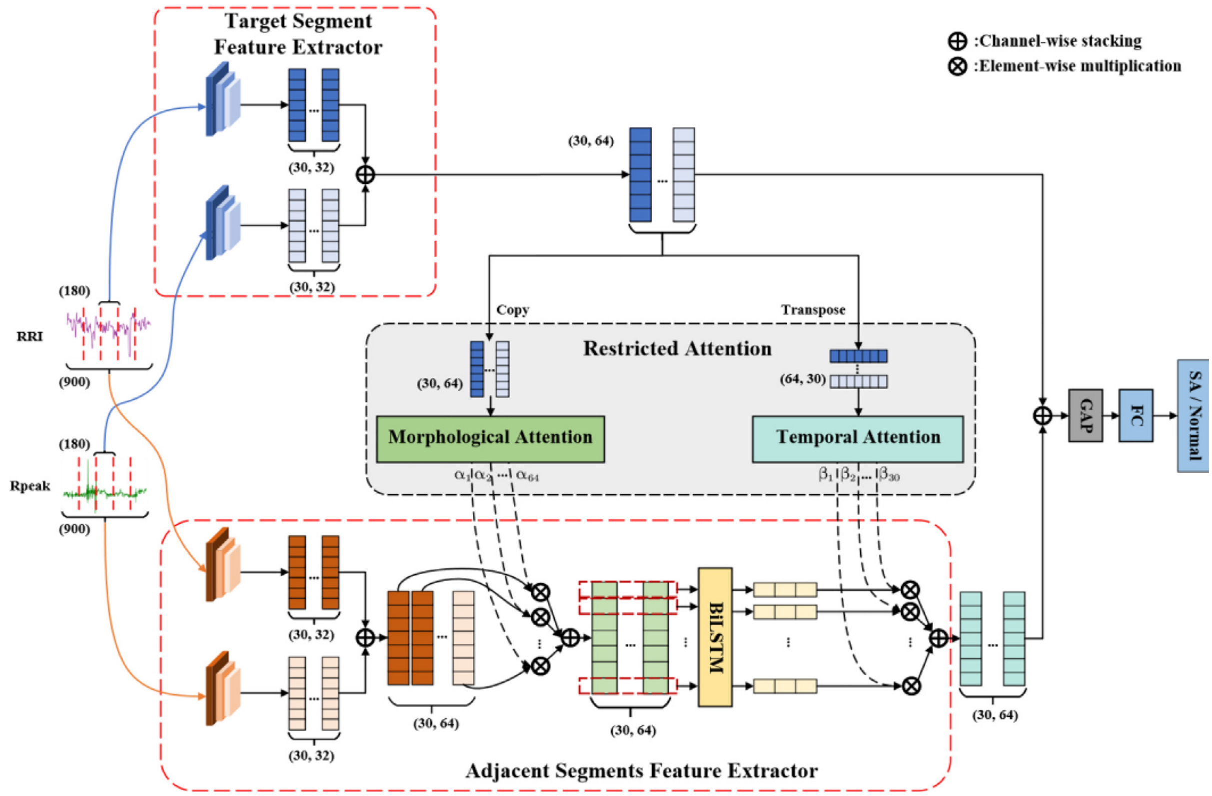


Fig. 3. The architecture of RAFNet. It mainly includes target segment feature extractor, adjacent segments feature extractor, restricted attention module, and a classifier of fully connected network.

meet the requirement of input format, the RRI/Rpeak is reorganized into target segment and adjacent segments shown in Fig. 4. Here, in order to describe concisely, we use the term of target segment and adjacent segments for target RRI/Rpeak and adjacent RRI/Rpeak segments in the subsequent sections. To fully utilize adjacent segments information to enhance the SA detection performance, we propose a new restricted attention fusion network called RAFNet as shown in Fig. 3. By leveraging the target segment as the query vector, RAFNet with a restricted attention can effectively extract adjacent-segment feature information and depress the redundant feature information.

Specifically, as shown in Fig. 3, RAFNet is mainly composed of two streams of neural networks, named target segment feature extractor and adjacent segments feature extractor, respectively. The one-minute-long target segment is as the input of target segment feature extractor and the five-minute-long adjacent segments are as the input of adjacent segments feature extractor. In regard to the first-stream neural network $F^1(\cdot)$ as shown in the top part of 3, we call it target segment feature extractor, which learns feature representations from one-minute-long target RRI/Rpeak segment. We denote the input as $X^1 = \left[x^1_{RRI}, x^1_{Rpeak}\right]$, where $x^1_{RRI} \in R^{d^1}$ and $x^1_{Rpeak} \in R^{d^1}$ refer to target RRI and Rpeak segments, respectively. d^1 is the dimension size of target RRI/Rpeak segment. As for the second-stream neural network $F^2(\cdot)$ shown in the bottom part of Fig. 3, we call it adjacent segments feature extractor, which learns the feature representations from five-minute-long adjacent RRI/Rpeak segments (*i.e.* one-minute-long target segment and surrounding two adjacent segments). We denote the input as $X^2 = \left[x^2_{RRI}, x^2_{Rpeak}\right]$, where $x^2_{RRI} \in R^{d^2}$ and $x^2_{Rpeak} \in R^{d^2}$ refer to adjacent RRI/Rpeak segments, respectively. d^2 is the dimension size of five-minute-long RRI/Rpeak segment. To effectively extract adjacent segments information and depress redundant feature information, the restricted attention $Att_{res}(\cdot)$ is employed, which is shown in the middle part of Fig. 3. Finally, the learned target- and adjacent-segment feature representations fused by channel-wise stacking scheme are fed into the classifier $F_{clf}(\cdot)$ and output the probability $\hat{y}$ of SA events in the right part of Fig. 3. More formally, it can be defined to be

$$\hat{y} = F_{clf}\left(F^1\left(X^1; \mathcal{W}^1\right), F^2\left(X^2, Att_{res}\left(F^1\left(X^1; \mathcal{W}^1\right); \alpha, \beta\right); \mathcal{W}^2\right); \mathcal{W}^3\right) \quad (1)$$

where $\mathcal{W}^1$, $\mathcal{W}^2$ and $\mathcal{W}^3$ refer to the weight parameters of neural networks in terms of target segment feature extractor, adjacent segments feature extractor, and classifier of fully connected network, respectively.

3.2.2. Target segment feature extractor

As shown in the top part of Fig. 3, the target segment feature extractor is composed of two streams of 1-dimensional convolutional neural networks for learning target RRI/Rpeak segment feature representations. As reported in literature (Chen et al., 2021), it employed 1-dimensional convolutional neural network for SA detection and achieved a promising result. Therefore, the architecture of convolutional neural networks in this study follows its paradigm but with different kernel size and stride in the convolutional layers. Since the input size in this study is much less than that in Chen et al. (2021), we set much smaller kernel size and stride in convolutional and max-pooling layers. The detailed configuration of target segment feature extractor is shown in Table 1.

3.2.3. Adjacent segments feature extractor

As shown in the bottom of Fig. 3, the adjacent segments feature extractor is composed of two-stream fusion network followed by a BiLSTM network. In the two-stream fusion network,

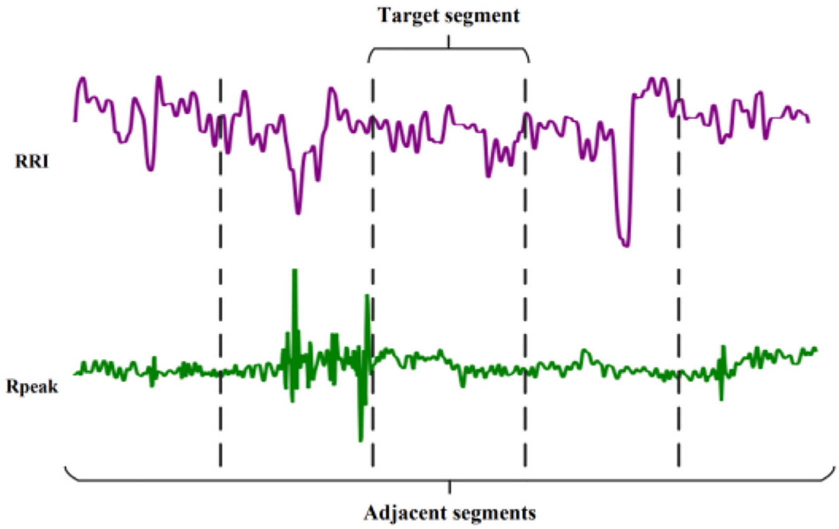


Fig. 4. Example illustrations of target segment and adjacent segments. RRI and R peak segmented signals are extracted from an ECG signal. The line in purple is the RRI segmented signal and the line in green is the Rpeak segmented signal. Both of RRI and Rpeak segments are composed of five one-minute-long segments, including one target segment and its two pre- and post-segments.

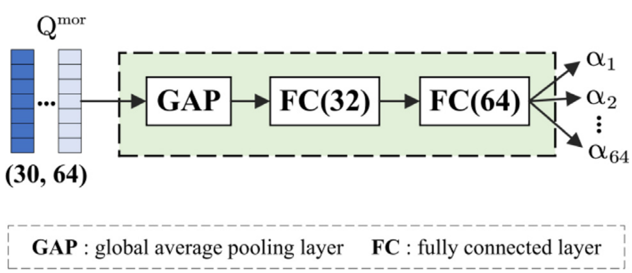


Fig. 5. The architecture of the morphological attention.

Table 1
The configuration of target segment feature extractor.

Layer	Type	Parameters
1	conv1d	filters = 16, kernel = 7, stride = 1
2	conv1d	filters = 24, kernel = 7, stride = 1
3	max-pooling1d	kernel = 2, stride = 2
4	conv1d	filters = 32, kernel = 7, stride = 1
5	max-pooling1d	kernel = 3, stride = 3

like the target segment feature extractor, it also employ convolutional neural networks to learn the morphological feature representation from five-minute-long adjacent segments. The configuration of the convolutional neural network follows the scheme in Chen et al. (2021) which has obtained a quite well performance in SA detection. To further learn the temporal feature representation among morphological features, BiLSTM with merits of well capturing temporal information of forward and backward directions is presented. In addition, the adjacent segments feature extractor with a restricted attention mechanism can effectively extract the morphological and temporal information and depress redundant noise information from adjacent segments. Here, we must point out that the input of BiLSTM is the transpose of morphological feature representation learned by two-stream fusion network. More concretely, each row of the transposed matrix is as the time-step input to BiLSTM. It means each time-step input has cross-channel information from morphological representation, which would greatly enhance the SA detection performance.

3.2.4. Restricted attention

It is noted that adjacent segments include useful information for boosting SA detection performance as well as redundant feature information. Towards this end, we propose a restricted attention mechanism with cascaded morphological and temporal attention modules, which is shown in the middle part of Fig. 3. By leveraging target segment as the query vector, the restricted attention can effectively capture morphological and temporal information from adjacent segments. As for the cascaded morphological attention module and temporal module, they are described as follows:

Morphological Attention. It is noted that the channel-wise attention can greatly enhance the performance in computer vision task by adaptively assigning weight importance for feature maps (Hu, Shen, & Sun, 2018). In this study, we also design a morphological attention to learn the weight importance for the learned adjacent-segment morphological feature maps, whose network architecture follows the scheme in Hu et al. (2018). Differed in Hu et al. (2018), we utilize the target-segment feature representation as the query vector, which would help to focus on the useful morphological information and ignore the redundant feature information. Here, we denote the query vector $Q^{mor} \in R^{d^3 \times c}$, where c is the number of the adjacent-segment morphological feature maps and d^3 is the dimension size of the target-segment morphological feature maps. Concretely, input the query vector Q^{mor} into the morphological attention $Att_{mor}(\cdot)$, output the weight importance $\alpha \in R^c$, which can be defined to be

$$\alpha = \mathrm{Att}_{\mathrm{mor}}\left(\mathrm{Q}^{\mathrm{mor}}; \mathcal{W}^{\mathrm{mor}}\right) \tag{2}$$

where $\mathcal{W}^{mor}$ is the weight parameters of morphological attention. The detailed configuration of morphological attention is shown in Fig. 5.

Temporal Attention. As shown in the bottom part of Fig. 3, the temporal attention can capture the adjacent-segment temporal feature information with the weight importance. Owing to generating many outputs from BiLSTM, the temporal attention generates the corresponding weight importance with the transpose of target-segment feature representation. Specifically, each column of the transpose matrix is as input to a fully connection layer with sharing weight parameters scheme, which can greatly

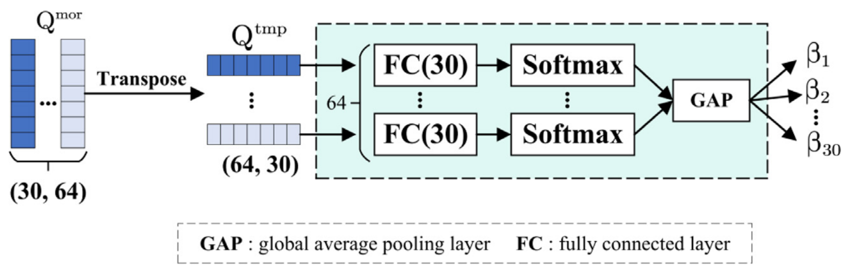


Fig. 6. The structure of temporal attention.

Table 2
Description of five-minute-long ECG segments on Apnea-ECG dataset and FAH-ECG dataset.

	Apnea-ECG			FAH-ECG		
	Normal	SA	Total	Normal	SA	Total
Training	8189	5178	13367	11976	7721	19197
Validation	2047	1295	3342	2994	1806	4800
Test	10455	6490	16945	15064	9327	24391

reduce the computational complexity and remain the high performance. Followed by a Softmax layer and global average pooling layer, the temporal attention Att_{tmp} can adaptively generate the weight importance $\beta \in R^{d^3}$ for outputs by BiLSTM, which can be defined to be

$$\beta = Att_{tmp}\left(Q^{tmp}; \mathcal{W}^{tmp}\right) \tag{3}$$

where $\mathcal{W}^{tmp}$ is the weight parameters of the temporal attention and $Q^{tmp} = (Q^{mor})^T$ is the query vector. The detailed configuration of temporal attention is shown in Fig. 6.

3.2.5. Classifier

As shown in the right part of Fig. 3, the classifier includes a global average pooling layer (GAP) and two fully connected layers (FC), followed by a Softmax layer to output the probability of SA events. Here, the input of the classifier is the flatten vector of the channel-wise stacking feature maps from the learned target segment and adjacent segments feature maps.

3.3. Performance metrics

Evaluating the performance of SA detection mainly consists of two stages: per-segment and per-recording (*i.e.* overnight ECG signal). For per-segment performance, it means that RAFNet must detect whether each one-minute-long ECG segment is with SA event. In regard to per-recording, RAFNet gives an overall evaluation on overnight ECG signals to diagnose a subject whether he/she suffered from SA.

3.3.1. Per-segment evaluation

In this study, we utilize evaluation metrics of accuracy (Acc), sensitivity (Sn), specificity (Sp), F_1 score and area under the receiver operating characteristic curve (AUC) for SA detection on per-segment.

3.3.2. Per-recording evaluation

The performance metrics of per-recording include Acc, Sn, Sp, Pearson correlation coefficient ($Corr$). $Corr$ is the value between the ground truth and the predicted apnea-hypopnea index (AHI), which is a widely used SA detection evaluation metric. If AHI is greater than or equal to 5, the individual suffers from SA; otherwise, the individual is Normal.

$$AHI = \frac{60}{T} \times N \tag{4}$$

where N denotes the number of SA segments, T represents the number of one-minute long segments, so $\frac{T}{60}$ is the number of recording hours for the entire recording process.

4. Results and discussion

4.1. Data sources

4.1.1. Apnea-ECG dataset

The Apnea-ECG (Penzel, Moody, Mark, Goldberger, & Peter, 2000) is a public domain ECG dataset provided by PhysioNet for SA detection, where ECG signals are sampled at 100 Hz with a resolution of 16 bits. 70 subjects aged from 27 to 63 years old were recruited to monitor their sleep conditions overnight in the hospital. All collected ECG signals ranging from 401 to 578 min were divided into one-minute-long segments, which were annotated by a trained expert. Each segment was annotated by SA or Normal. A total of 70 overnight ECG signals were grouped into training set (35 records) and test set (35 records). In order to obtain the optimized hyper-parameters of RAFNet, 20% of ECG records in training set are randomly selected as the validation set. After data preprocessing (see in Section 3.1), one-minute-long ECG segments are reorganized into sequential five-minute-long ECG segments. As shown in Table 2, there are a total of 13,367, 3342, and 16,945 five-minute-long ECG segments in training set, validation set, and test set, respectively.

4.1.2. FAH-ECG dataset

FAH-ECG is a real clinical ECG dataset from the First Affiliated Hospital, Sun Yat-sen University, China. 140 subjects aged from 18 to 67 years old were recruited to conduct sleep conditions monitoring in hospital with a signed consent form (approval IRB No. 2,019,119). And a total of 140 overnight ECG signals ranging from 311 to 658 min were recorded, which are randomly divided into training set and test set with 50% to 50%. It means that there are 70 ECG records in the training set and 70 ECG records in the test set. Similar to the process on Apnea-ECG dataset, 20% of ECG records in training set are randomly selected as the validation set. After data preprocessing, as shown in Table 2, there are a total of 19,197, 4800, and 24,391 five-minute-long ECG segments in training set, validation set, and test set, respectively.

4.2. Classification performance

In this study, the maximum number of iterated training epochs is set to 50 as well as batch size is set to 64. The initial learning rate is set to 0.001, reduced by a factor of 0.5 for every 10 iterated epochs. Meanwhile, the weight parameters of RAFNet are initialized by the HeNormal initializer (He, Zhang, Ren, & Sun, 2015). In order to overcome overfitting problem to some extent, L_2 regulation is employed with the penalty coefficient of 0.001. Here, we use the binary cross-entropy loss and optimize it with Adam optimizer (Kingma & Ba, 2014). Other hyperparameters

Table 3
Per-segment performance comparison on Apnea-ECG dataset.

Method		*Acc* (%)	*Sn* (%)	*Sp* (%)
Feature engineering	HMM-SVM (Song et al., 2015)	86.2	82.6	88.4
	SVM (Surrel, Aminifar, Rincón, Murali, & Atienza, 2018)	85.7	81.4	88.4
	LS-SVM (Sharma & Sharma, 2016)	83.4	79.5	88.4
Deep learning	Auto-encoder (Li, Pan, Li, Jiang, & Liu, 2018b)	83.8	88.9	88.4
	LeNet-5 (Wang et al., 2019)	87.6	83.1	90.3
	1-D CNN (Sharan et al., 2020)	88.2	82.7	91.6
	LSTM (Faust et al., 2021)	81.3	59.9	91.8
	CNN and LSTM (Almutairi et al., 2021)	90.9	91.2	90.4
	SE-MSCNN (Chen et al., 2021)	90.6	86.0	93.5
	ConCAD (Huang & Ma, 2021)	91.2	–	–
	MSDA-1DCNN (Shen et al., 2021)	89.4	89.8	89.1
This work	RAFNet	91.4	88.75	93.05

Table 4
Per-recording performance comparison on Apnea-ECG dataset.

Method		*Acc* (%)	*Sn* (%)	*Sp* (%)	*Corr*
Feature engineering	HMM-SVM (Song et al., 2015)	97.1	95.8	100	0.860
	LS-SVM (Sharma & Sharma, 2016)	97.1	95.8	100	0.841
Deep learning	Auto-encoder (Li et al., 2018b)	100	100	100	–
	LeNet-5 (Wang et al., 2019)	97.1	100	91.7	0.943
	SE-MSCNN (Chen et al., 2021)	100	100	100	0.979
	MSDA-1DCNN (Shen et al., 2021)	100	100	100	–
This work	RAFNet	100	100	100	0.985

are kept with default settings. All experiments are implemented based on Keras framework at version 2.3.1.

4.2.1. Per-segment classification

In this section, we conduct extensive experiments with RAFNet on the public domain Apnea dataset and real clinical FAH-ECG dataset. Regarding the Apnea-ECG dataset, the RAFNet can achieve competitive results on per-segment classification task, with the SA detection performance of 91.40%, 88.75%, 93.05%, 88.77% and 96.80% in terms of accuracy, sensitivity, specificity, F1 score, and AUC value, respectively. Even on the real clinical FAH-ECG dataset, the RAFNet still can obtain promising results, with the SA detection performance of 84.70%, 72.91%, 92.01%, 78.47% and 89.71% in terms of accuracy, sensitivity, specificity, F1 score, and AUC value, respectively. The SA detection performance on the public domain ApneaECG dataset is obviously better than that on the real clinical FAH-ECG dataset. The reason may be that the quality of ECG signals on FAH-ECG dataset is much lower than the public Apnea-ECG dataset as ECG signals in FAH-ECG dataset were collected in hospital and directly fed into RAFNet without any special treatment during data collecting phase.

4.2.2. Per-recording classification

According to the per-segment results, physician can diagnose whether a subject suffered from SA based on the AHI metric. On the public domain Apena-ECG dataset, the RAFNet achieves 100% on per-recording metrics of accuracy, sensitivity, specificity, as well as *Corr* of 0.985. As for the real clinical FAH-ECG dataset, the RAFNet can still be up to 85.0% accuracy, 90.0% sensitivity and 92.0% specificity, as well as 0.922 *Corr*. It means that the proposed RAFNet can be potentially deployed into a medical system to provide sleep monitoring service with a portable device out of hospital.

Table 5
Per-segment and per-recording performance comparison on FAH-ECG dataset.

Method	Per-segment			Per-recording			
	Acc (%)	*Sn* (%)	*Sp* (%)	*Acc* (%)	*Sn* (%)	*Sp* (%)	*Corr*
LeNet-5 (Wang et al., 2019)	81.86	71.22	88.44	80.14	86.60	64.00	0.866
CNN and LSTM (Almutairi et al., 2021)	75.79	49.57	92.03	66.94	58.57	87.86	0.740
SE-MSCNN (Chen et al., 2021)	83.97	71.26	91.84	85.43	86.40	83.00	0.898
RAFNet	84.41	73.32	91.27	85.56	91.80	71.00	0.915

4.3. Performance comparison

4.3.1. Apnea-ECG dataset

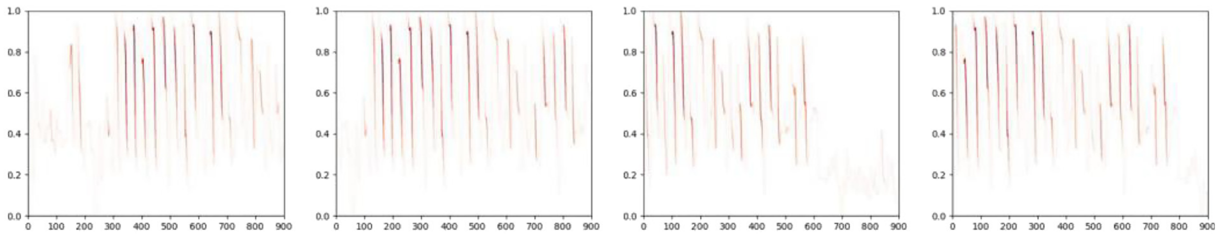
As shown in Table 3, it is noted that deep learning based SA detection methods are obviously superior to machine learning based ones on persegment performance metrics. The reason would be that deep learning based methods are powerful to learn feature representation from ECG signals with an end-to-end paradigm. While, machine learning based methods are highly dependent on the feature engineering technologies, which are sensitive to the quality of ECG signals. If the ECG signal is contaminated with noise, it is very difficult to precisely extract hand-crafted features, resulting in degrading the SA detection performance greatly. Among the deep learning based methods, it can be observed that the proposed RAFNet with a restricted attention outperforms the state-of-the-art baselines.

As for per-recording performance comparison on Apnea-ECG dataset, the RAFNet also achieves the best SA detection performance which is shown in Table 4. RAFNet can obtain 100% on SA detection performance metrics of accuracy, sensitivity, and specificity. It is noted that both SE-MSCNN (Chen et al., 2021) and MSDA-1DCNN (Shen et al., 2021) also achieved 100% on all SA detection performance metrics. However, the proposed RAFNet achieves Corr of 0.985, which is obviously superior to SE-MSCNN and MSDA-1DCNN.

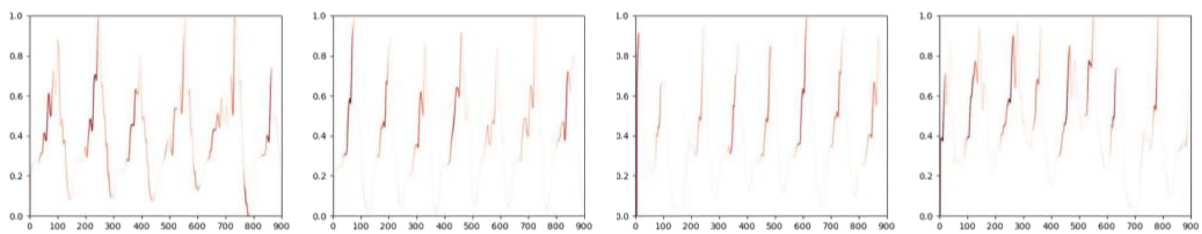
To sum up, the proposed RAFNet achieves the state-of-the-art on persegment and per-recording SA detection performance metrics.

4.3.2. FAH-ECG dataset

We also validate the SA detection performance on the real clinical FAHECG dataset. Here, we implement the deep learning based methods with promising SA detection results where the codes are available. As shown in Table 5, the RAFNet achieves 84.41% per-segment accuracy and 85.56% per-recording accuracy, which are improved a lot to the SE-MSCNN, which achieved



(a)



(b)

Fig. 7. Illustration examples of class activation maps of RRI: (a) Normal (b) SA.

Table 6
Per-segment performance for ablation studies.

Model	*Acc* (%)	*Sn* (%)	*Sp* (%)	F_1 (%)	*AUC* (%)
A	78.89 ± 0.33	70.66 ± 2.43	84.01 ± 1.35	71.93 ± 0.89	87.29 ± 0.27
B	88.95 ± 1.19	86.97 ± 1.60	92.09 ± 1.31	87.09 ± 0.21	96.05 ± 0.24
C	90.24 ± 0.22	87.84 ± 1.66	91.73 ± 1.03	87.33 ± 0.35	96.43 ± 0.20
D	90.42 ± 0.27	86.24 ± 0.95	92.91 ± 0.47	87.33 ± 0.38	96.44 ± 0.13
E	90.44 ± 0.37	**88.33 ± 0.68**	91.74 ± 0.79	87.62 ± 0.42	96.44 ± 0.18
F	**91.04 ± 0.21**	87.89 ± 0.91	**93.00 ± 0.61**	**88.26 ± 0.28**	**96.65 ± 0.12**

the top-2 per-segment and per-recording SA detection performance. It means that RAFNet with a restricted attention can effectively learn the morphological and temporal feature representation from adjacent segments.

4.4. Ablation study

In order to verify the effectiveness of adjacent segments and the restricted attention, extensive ablation experiments on Apnea-ECG dataset are conducted. Here, each experiment is performed 10 times to obtain general SA detection performance. Detailed experimental setups are as follows:

(A) **Target-segment**: Target segment as input to build the SA detection model with target segment feature extractor.
(B) **Adjacent-segments**: Adjacent segments as input to build the SA detection model with adjacent segments feature extractor.
(C) **Target-segment** + **Adjacent-segments**: Fusing Model (A) and (B).
(D) **Target-segment** + **Adjacent-segments** + **Morphological-attention**: Model (C) with morphological attention.
(E) **Target-segment** + **Adjacent-segments** + **Temporal- attention**:Model (C) with temporal attention.
(F) **RAFNet**: Model (C) with morphological and temporal attention (*i.e.* restricted attention).

As shown in Table 6, it is observed that Model (A) based on one-minute-long target segments achieves accuracy of 78.89 ± 0.33%, as well as Model (B) based on five-minute-long adjacent segments achieves accuracy of 88.95 ± 1.19%. It further validates that adjacent segments can greatly improve the SA detection performance. With the fusion of Model (A) and (B), Model (C) obtains 90.24 ± 0.22%, which outperforms Model (A) and (B) with a big margin. Based on Model (C), merely introducing the morphological attention, Model (D) achieves better accuracy. Or by merely introducing the temporal attention, Model (E) obtains much better performance than Model (D). The reason may be that temporal information has a much higher relationship than the morphological information among adjacent segments. Finally, by introducing both morphological and temporal attentions (*i.e. restricted attention*), Model (F) (*i.e.* RAFNet) further improves the SA detection performance on most metrics. The accuracy of Model (F) can be up to 91.04 ± 0.21%. To sum up, RAFNet with a restricted attention can effectively learn the morphological and temporal information from adjacent segments and achieve promising results.

4.5. Classification signatures

As reported in Zhou, Khosla, Lapedriza, Oliva, and Torralba (2016), the class activation mapping (CAM) is an effective way to enable convolutional neural network to perform object localization and highlight the discriminative object parts from original image. In this work, we apply the CAM to the RRI and Rpeak segments and get their corresponding activation maps, which highlight the important parts that contribute to the task of SA detection.

Figs. 7 and 8 show the illustration examples of class activation maps for the RRI and Rpeak segments, where the importance is

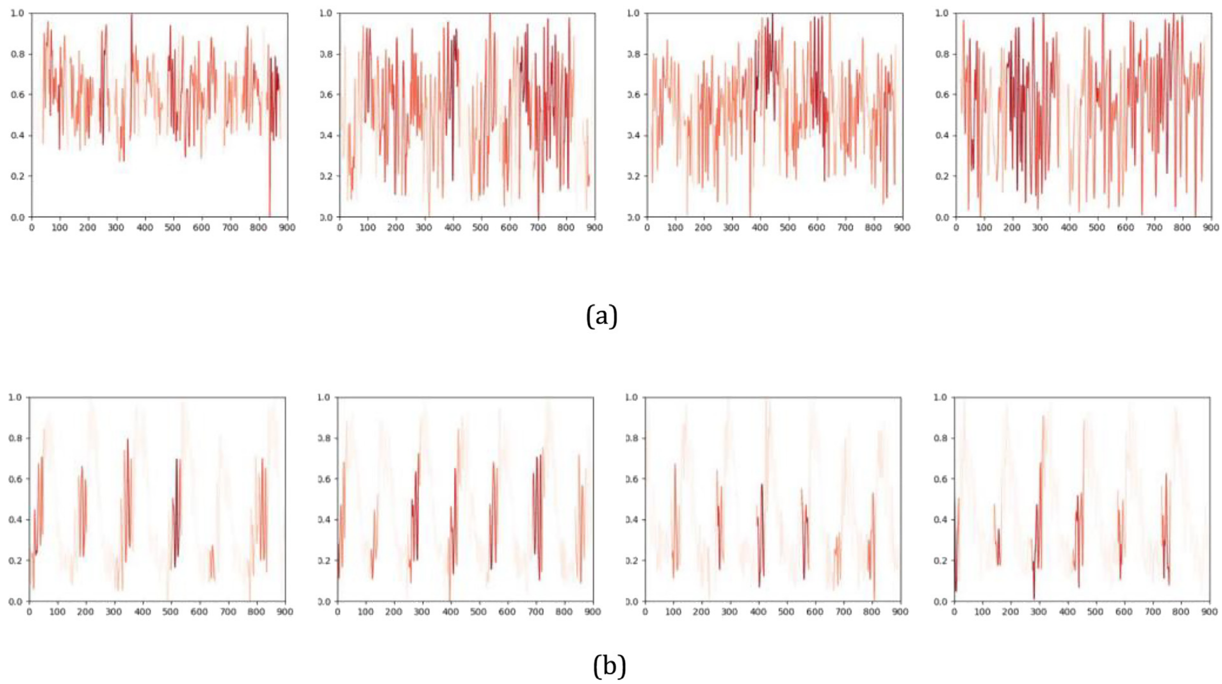


Fig. 8. Illustration examples of class activation maps of Rpeak: (a) Normal (b) SA.

Table 7
Per-segment and per-recording performance comparison of training on Apnea-ECG dataset and testing on FAH-ECG dataset as an external dataset.

Method	Per-segment *Acc* (%)	Per-recording *Acc* (%)	Per-recording *Corr*
LeNet-5 (Wang et al., 2019)	70.86	60.00	0.699
CNN and LSTM (Almutairi et al., 2021)	75.00	67.14	0.777
SE-MSCNN (Chen et al., 2021)	72.07	60.00	0.696
RAFNet	**76.73**	**70.00**	**0.802**

marked with red color from light to dark. It is observed that there exists significant difference in the morphological waveforms of Normal and SA in the RRI/Rpeak segments. In contrast to the SA segments, the waveform variations of Normal segments are much more stable as a subject without SA conditions has a stable heart rate and respiratory amplitude during his/her sleep. On the other hand, it can be noted that the class activation maps for Normal and SA in the RRI/Rpeak segments. For Normal segment, the RAFNet pays the attention on almost of RRI/Rpeak segments. Whereas, for SA segments, the RAFNet merely pays attention on one part of RRI/Rpeak segments. It means that the RAFNet can well learn the feature patterns from RRI/Rpeak segments to detect SA events.

4.6. Model limitations

In this study, RAFNet is trained and tested on the training and test datasets which are generated from the identical dataset. Since training and test datasets approximately belong to the identical data distribution, RAFNet can achieve the higher generic performance to that trained and tested on different individual datasets, which is a general problem machine learning models encounter. To verify this, we further conduct experiments in this manner, RAFNet model is trained on the public Apnea-ECG dataset and validated on the real clinical FAH-ECG data as an external test set. Experiment results show that the proposed RAFNet achieves per-segment accuracy of 76.73% and per-recording accuracy of 70.00% on FAH-ECG dataset. Compared with the performance of RAFNet in both training and validation on FAH-ECG dataset, its performances decrease a lot, which are 7.68% and 15.56% in terms of per-segment performance and per-recording performance, respectively. The cause of declining RAFNet performance may be the difference of data distribution between Apnea-ECG and FAH-ECG datasets, which are collected from American population conducted by Phillips University and Chinese population conducted by the Sun Yat-sen University, respectively. What is more, we also conduct the experiments with deep learning based advanced methods with hyper-parameters tuning in the aforementioned manner. As shown in Table 7, we can see that our proposed RAFNet still can achieve the competitive results, which are superior to the deep learning based methods. To alleviate this issue caused by the population difference, we will try to collect more data for RAFNet training from all kinds of populations in the future.

5. Conclusions

In this study, we propose a new restricted attention fusion network named RAFNet for SA detection. By utilizing the target segment as the query vector, the restricted attention with cascaded morphological and temporal attention modules can well extract adjacent-segment feature information and depress redundant feature information. What is more, by fusing target-segment feature information and adjacent-segment feature information, it can further boost the RAFNet performance for SA detection. In addition, the proposed RAFNet is validated on the public domain ECG dataset of Apnea-ECG and the real clinical ECG dataset of FAH-ECG. Extensive experiments show that RAFNet can achieve 91.40% accuracy on Apnea-ECG and 84.70% accuracy on FAH-ECG, which are superior to those state-of-the-art baselines achieved. It means that RAFNet can be potentially deployed into a medical system to provide sleep conditions monitoring service on a large scale of population.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the Basic and Applied Basic Research Project of Guangdong Province under the grant No. 2022B1515130009.

References

Almutairi, H., Hassan, G. M., & Datta, A. (2021). Classification of obstructive sleep apnoea from single-lead ECG signals using convolutional neural and long short term memory networks. *Biomedical Signal Processing and Control*, http://dx.doi.org/10.1016/j.bspc.2021.102906.

Bahrami, M., & Forouzanfar, M. (2021). Detection of sleep apnea from singlelead ECG: Comparison of deep learning algorithms. In *2021 IEEE international symposium on medical measurements and applications* (pp. 1–5). IEEE.

Baillieul, S., Dekkers, M., Brill, A.-K., Schmidt, M. H., Detante, O., Pépin, J.-L., et al. (2022). Sleep apnoea and ischaemic stroke: current knowledge and future directions. *The Lancet Neurology*, *21*(1), 78–88.

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., et al. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *The Lancet Respiratory Medicine*, *7*(8), 687–698.

Bhattacharjee, A., Saha, S., Fattah, S. A., Zhu, W.-P., & Ahmad, M. O. (2018). Sleep apnea detection based on rician modeling of feature variation in multiband EEG signal. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 1066–1074.

Chen, X., Chen, Y., Ma, W., Fan, X., & Li, Y. (2021). Se-mscnn: A lightweight multi-scaled fusion network for sleep apnea detection using single-lead ECG signals. In *2021 IEEE international conference on bioinformatics and biomedicine* (pp. 1276–1280). IEEE.

Chen, L., Zhang, X., & Song, C. (2014). An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Transactions on Automation Science and Engineering*, *12*(1), 106–115.

Deviaene, M., Testelmans, D., Buyse, B., Borzée, P., Van Huffel, S., & Varon, C. (2018). Automatic screening of sleep apnea patients based on the spo 2 signal. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 607–617.

Faust, O., Barika, R., Shenfield, A., Ciaccio, E. J., & Acharya, U. R. (2021). Accurate detection of sleep apnea with long short-term memory network based on RR interval signals. *Knowledge-Based Systems*, http://dx.doi.org/10.1016/j.knosys.2020.106591.

Floras, J. S. (2015). Hypertension and sleep apnea. *Canadian Journal of Cardiology*, *31*(7), 889–897.

Hamilton, P. (2002). Open source ECG analysis. In *Computers in cardiology* (pp. 101–104). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034).

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, G., & Ma, F. (2021). Concad: Contrastive learning-based cross attention for sleep apnea detection. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 68–84). Springer.

Javaheri, S., Barbe, F., Campos-Rodriguez, F., Dempsey, J. A., Khayat, R., Javaheri, S., et al. (2017). Sleep apnea: types, mechanisms, and clinical cardiovascular consequences. *Journal of the American College of Cardiology*, *69*(7), 841–858.

John, A., Nundy, K. K., Cardiff, B., & John, D. (2021). Somnnet: An spo2 based deep learning network for sleep apnea detection in smartwatches. In *2021 43rd Annual international conference of the IEEE Engineering in Medicine & Biology Society* (pp. 1961–1964). IEEE.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980.

Li, Y., Pan, W., Li, K., Jiang, Q., & Liu, G. (2018a). Sliding trend fuzzy approximate entropy as a novel descriptor of heart rate variability in obstructive sleep apnea. *IEEE Journal of Biomedical and Health Informatics*, *23*(1), 175–183.

Li, K., Pan, W., Li, Y., Jiang, Q., & Liu, G. (2018b). A method to detect sleep apnea based on deep neural network and hidden Markov model using singlelead ECG signal. *Neurocomputing*, *294*, 94–101.

Mahmud, T., Khan, I. A., Mahmud, T. I., Fattah, S. A., Zhu, W.-P., & Ahmad, M. O. (2020). Sleep apnea event detection from sub-frame based feature variation in EEG signal using deep convolutional neural network. In *2020 42nd Annual international conference of the IEEE Engineering in Medicine & Biology Society* (pp. 5580–5583). IEEE.

Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., & Peter, J. H. (2000). The apnea-ECG database. In *Computers in cardiology 2000, vol. 27 (Cat. 00CH37163)* (pp. 255–258). IEEE.

Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W., & Hla, K. M. (2013). Increased prevalence of sleep-disordered breathing in adults. *American Journal of Epidemiology*, *177*(9), 1006–1014.

Rajesh, K. N., Dhuli, R., & Kumar, T. S. (2021). Obstructive sleep apnea detection using discrete wavelet transform-based statistical features. *Computers in Biology and Medicine*, http://dx.doi.org/10.1016/j.compbiomed.2020.104199.

Salari, N., Hosseinian-Far, A., Mohammadi, M., Ghasemi, H., Khazaie, H., Daneshkhah, A., et al. (2022). Detection of sleep apnea using machine learning algorithms based on ECG signals: A comprehensive systematic review. *Expert Systems with Applications*, http://dx.doi.org/10.1016/j.eswa.2021.115950.

Sharan, R. V., Berkovsky, S., Xiong, H., & Coiera, E. (2020). ECG-derived heart rate variability interpolation and 1-d convolutional neural networks for detecting sleep apnea. In *2020 42nd Annual international conference of the IEEE Engineering in Medicine & Biology Society* (pp. 637–640). IEEE.

Sharma, H., & Sharma, K. (2016). An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions. *Computers in Biology and Medicine*, *77*, 116–124.

Shen, Q., Qin, H., Wei, K., & Liu, G. (2021). Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–13.

Somboon, T., Grigg-Damberger, M. M., & Foldvary-Schaefer, N. (2019). Epilepsy and sleep-related breathing disturbances. *Chest*, *156*(1), 172–181.

Song, C., Liu, K., Zhang, X., Chen, L., & Xian, X. (2015). An obstructive sleep apnea detection approach using a discriminative hidden markov model from ECG signals. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1532–1542.

Surrel, G., Aminifar, A., Rincón, F., Murali, S., & Atienza, D. (2018). Online obstructive sleep apnea detection on medical wearable sensors. *IEEE Transactions on Biomedical Circuits and Systems*, *12*(4), 762–773.

Taran, S., Bajaj, V., Sinha, G., & Polat, K. (2021). Detection of sleep apnea events using electroencephalogram signals. *Applied Acoustics*, http://dx.doi.org/10.1016/j.apacoust.2021.108137.

Tripathy, R., Gajbhiye, P., & Acharya, U. R. (2020). Automated sleep apnea detection from cardio-pulmonary signal using bivariate fast and adaptive emd coupled with cross time–frequency analysis. *Computers in Biology and Medicine*, http://dx.doi.org/10.1016/j.compbiomed.2020.103769.

Van Steenkiste, T., Groenendaal, W., Deschrijver, D., & Dhaene, T. (2018). Automated sleep apnea detection in raw respiratory signals using long shortterm memory neural networks. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2354–2364.

Vimala, V., Ramar, K., & Ettappan, M. (2019). An intelligent sleep apnea classification system based on EEG signals. *Journal of Medical Systems*, *43*(2), 1–9.

Wang, T., Lu, C., Shen, G., & Hong, F. (2019). Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified lenet-5 convolutional neural network. *PeerJ*, http://dx.doi.org/10.7717/peerj.7731.

Wang, H., Parker, J. D., Newton, G. E., Floras, J. S., Mak, S., Chiu, K.-L., et al. (2007). Influence of obstructive sleep apnea on mortality in patients with heart failure. *Journal of the American College of Cardiology*, *49*(15), 1625–1631.

Ye, G., Yin, H., Chen, T., Chen, H., Cui, L., & Zhang, X. (2021). Fenet: A frequency extraction network for obstructive sleep apnea detection. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 2848–2856.

Yoon, H., Hwang, S. H., Choi, J.-W., Lee, Y. J., Jeong, D.-U., & Park, K. S. (2017). Slow-wave sleep estimation for healthy subjects and osa patients using R–R intervals. *IEEE Journal of Biomedical and Health Informatics*, *22*(1), 119–128.

Zarei, A., & Asl, B. M. (2018). Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ECG signal. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 1011–1021.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).