Computers in Biology and Medicine 196 (2025) 110655

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Efficient sleep apnea detection using single-lead ECG: A CNN-Transformer-LSTM approach

Duc Thien Pham *, Roman Mouček

Department of Computer Science and Engineering, University of West Bohemia in Pilsen, Pilsen, 30100, Czech Republic

ARTICLE INFO

Dataset link: https://physionet.org/content/apnea-ecg/1.0.0, https://physionet.org/content/ucddb/1.0.0/

Keywords:
Sleep apnea
CNN-Transformer-LSTM
Electrocardiogram
Detection
Deep learning

ABSTRACT

Background: Sleep apnea (SA), a prevalent sleep-related breathing disorder, disrupts normal respiratory patterns during sleep. This disruption can have a cascading effect on the body, potentially leading to complications in various organs, including the heart, brain, and lungs. Due to the potential for these complications, early and accurate detection of SA is critical. Electrocardiograms (ECG), due to their ability to continuously monitor heart rhythms and detect subtle changes in cardiac activity, such as heart rate variability and arrhythmias, which are often linked to sleep disruptions, have become crucial in identifying individuals at risk for SA.

Method: In this study, we propose a hybrid neural network model named CNN-Transformer-LSTM that uses a single-lead ECG signal to detect SA automatically. This method captures spatial and temporal features in the ECG data to improve classification performance. Our model utilizes RR intervals (RRI) and R-peak signals derived from ECG data as input and then classifies SA and normal states on a per-segment and per-recording basis. We evaluated the model using the Physionet Apnea-ECG dataset, consisting of 70 single-lead ECG recordings annotated by medical professionals, and the UCD St. Vincent's University Hospital's sleep apnea database (UCDDB) containing polysomnogram records from 25 patients.

Results: Our model achieved an accuracy of 91.6% for per-segment classification on the Physionet Apnea-ECG dataset using hold-out validation and the highest accuracy of 94.1% using five-fold cross-validation. As for per-recording classification, our model achieved an accuracy of 100% and the highest correlation coefficient value of 0.9996 using five-fold cross-validation. On the UCDDB dataset, our model achieved an accuracy of 99.37% on the reduced dataset excluding 4 patients and 98.34% on the full dataset. Compared to previous works, our model improved the per-segment classification accuracy by nearly 3% over the existing best result, thereby demonstrating that our model outperforms existing state-of-the-art methods in accurately detecting SA from a single-lead ECG signal.

Conclusion: These results highlight the effectiveness of the CNN-Transformer-LSTM model for SA detection and its potential to be used in SA detection devices for home health care and clinical settings.

1. Introduction

Sleep apnea (SA) is a prevalent sleep disorder characterized by repeated pauses in breathing during sleep. These interruptions in breathing, known as apneas, can occur numerous times throughout the night, leading to fragmented sleep and health complications if left untreated [1]. The two primary types of SA are obstructive sleep apnea (OSA) and central sleep apnea (CSA), with OSA being the more common form, typically caused by the relaxation of throat muscles. CSA, conversely, stems from a lack of respiratory effort initiated by the brain [2]. Despite their different etiologies, both types can severely impact an individual's overall health and quality of life. SA is a serious health condition that poses a significant risk for stroke. Some researchers believe it to be an independent risk factor, approximately doubling the likelihood of a stroke compared to those without SA [3]. This is alarming, especially considering that obstructive SA is widespread; an estimated 936 million adults worldwide (aged 30–69) have mild to severe forms of the condition, with 425 million of those cases considered moderate to severe [1]. These staggering numbers highlight the need for effective screening methods and timely interventions to address SA.

A definitive diagnosis of SA is determined by counting the number of apnea and hypopnea events over a specific period, such as a night of sleep. These counts are then averaged on a per-hour basis, leading to commonly used metrics like the Apnea/Hypopnea Index (AHI) or the

* Corresponding author.
E-mail addresses: ducthien@kiv.zcu.cz (D.T. Pham), moucek@kiv.zcu.cz (R. Mouček).

https://doi.org/10.1016/j.compbiomed.2025.110655
Received 27 June 2024; Received in revised form 2 May 2025; Accepted 17 June 2025
Available online 3 July 2025

Respiratory Disturbance Index (RDI) [4]. Standard diagnostic criteria classify the severity of apnea based on these indices. Accurate identification of apnea or hypopnea events requires direct measurement of upper airway airflow and respiratory effort, often necessitating the use of an esophageal balloon or a full-face mask. However, due to their invasive nature and potential to disrupt sleep, these methods are rarely utilized in clinical practice. Instead, polysomnography, a less invasive and more widely accepted diagnostic tool, is commonly employed [4].

The traditional SA diagnosis relies on overnight polysomnography (PSG) in sleep laboratories. PSG involves a comprehensive evaluation of various physiological data throughout the night, including brain activity (electroencephalogram, EEG), eye movements (electrooculogram, EOG), breathing patterns (airflow signal and respiratory rates), blood oxygen levels (arterial blood oxygen saturation), muscle activity (electromyogram, EMG), heart rhythm (electrocardiogram, ECG), body sounds, and body posture. These tests are conducted in dedicated sleep labs equipped with specialized monitoring systems and staffed by trained personnel. While PSG is considered the gold standard for diagnosing sleep disorders, its cost, inconvenience, and limited availability hinder widespread adoption [5]. In response to these challenges, researchers have explored alternative, non-invasive ECG-based methods, which can be applied more easily and reflect SA-related cardiovascular responses.

The key insight behind ECG-based SA detection is that apneas trigger changes in heart rate variability (HRV), a parameter that can be monitored using a single-lead ECG [6,7]. During an apnea event, the lack of airflow disrupts blood oxygen levels, triggering the cardiovascular system to compensate. Researchers have investigated various techniques to look for and extract informative features from the ECG signal. These features include parameters like RR intervals, which represent the time between consecutive heartbeats, and statistical measures of HRV. These features are then fed into machine learning (ML) or deep learning (DL) algorithms, like convolutional neural networks (CNN), to classify apnea events automatically. While ECG analysis offers a non-invasive and readily available way to monitor sleep, it faces challenges like potential interference from other cardiac conditions and further validation in diverse populations [8–10]. Moreover, balancing model complexity with computational efficiency poses a significant challenge, particularly for the potential deployment of these models in real-world or home healthcare applications.

This research aims to address the limitations of current ECG-based approaches, such as accuracy rates (typically less than 91%), by proposing a more complex yet efficient model that captures spatial and temporal features in ECG data to improve SA detection accuracy to over 91% on the Physionet Apnea-ECG dataset. In this study, we propose a novel CNN-Transformer-LSTM model for the automatic detection of SA using a single-lead ECG signal. Our approach leverages the strengths of each model component: CNNs are used to extract the spatial and channel features, Transformers capture long-term dependencies, and LSTMs model temporal sequences and long-term dependencies. This integrated architecture facilitates accurate classification of apnea events based on minimal physiological data, specifically relying solely on single-lead ECG signals without requiring additional physiological measurements such as airflow, oxygen saturation, or respiratory effort. This focus on single-lead ECG addresses critical limitations in feature extraction and generalization encountered by previous methodologies. The main contributions of this study are as follows:

- To propose an efficient hybrid neural network model named CNN-Transformer-LSTM that improves sleep apnea detection accuracy using a single-lead ECG signal.
- Experiment results show that the proposed method demonstrates better performance, achieving state-of-the-art results in accuracy and outperforming existing ECG-based methods.

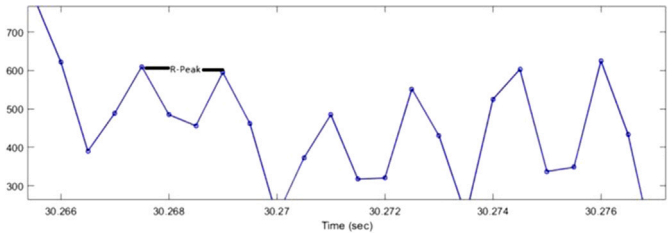


Fig. 1. RR interval of Apnea-ECG recording [12].

The remainder of the paper is structured as follows. Section 2 presents related works. Section 3 outlines the proposed SA detection method (the CNN-Transformer-LSTM model), the used dataset, the data preprocessing phase, and the performance evaluation process. Section 4 details the experimental results, and Section 5 provides the discussion. Finally, the conclusion is provided in Section 6.

2. Related work

Through analyzing breathing airflow signals, researchers have identified a link between these signals and SA diagnosis. The RR-intervals (time interval from one R-wave to the next R-wave) is a time interval between two successive R peaks [11,12], as shown in Fig. 1. Specifically, the amplitude of the R peaks in ECG signals closely correlates with changes in breathing airflow, as it tends to decrease during apneic events and recover during normal breathing. The RR intervals (RRI), reflecting heart rate variability, also strongly correlate with breathing airflow, often slowing during apneas and resuming normal rhythms as airflow stabilizes. Both R-peaks and RRI can be directly extracted from raw ECG data. Several studies support the effectiveness of R-peaks and RRI in SA detection [10,13–15]. As reported by Wang et al. [10], these studies often segment overnight ECG signals into one-minute intervals. Furthermore, these studies suggest that analyzing adjacent ECG segments can significantly improve SA detection performance.

Currently, SA detection methods utilizing single-lead ECG signals include models based on ML and models based on DL. Table 1 shows a literature review that summarizes the studies related to SA detection on two public datasets: the Physionet Apnea-ECG dataset and the St. Vincent's University Hospital/University College Dublin Sleep Apnea Database (UCDDB).

Building on prior works using unsupervised learning for feature extraction in SA detection, Feng et al. [16] employed a time-dependent, cost-sensitive (TDCS) approach. The frequential stacked sparse autoencoder (FSSAE) automatically extracts features using an unsupervised learning approach. To enhance classifier performance, the TDCS classification model combines the hidden Markov model (HMM) and the MetaCost algorithm, addressing both temporal dependencies and class imbalance. Their model achieved an 85.1% accuracy for per-segment classification on the Physionet Apnea-ECG dataset. Alternatively, Viswabhargav et al. [17] utilized a combination of features like EDR and sparse residual entropy (SRE) alongside fuzzy K-means clustering and support vector machines (SVM) for SA detection. The experimental results achieved accuracy, sensitivity, and specificity values of 78.07%, 78.01%, and 78.13%, respectively, using a Fourier dictionary with 10-fold cross-validation. In the subject-specific or leave-one-out validation scenario, the SVM classifier achieved sensitivity and specificity of 85.43% and 92.60%, respectively, when using SRE features with the Fourier dictionary.

Sharma et al. [18] proposed a method based on a biorthogonal antisymmetric wavelet filter bank (BAWFB) and using SVM for OSA classification, achieving average classification accuracy, sensitivity, specificity, and F-score of 90.11%, 90.87% 88.88% and 0.92, respectively, on the Physionet Apnea-ECG dataset. Pinho et al. [19] leveraged HRV

Table 1
A literature review summarizing the studies related to SA detection using ECG signal.

Dataset	References	Year	Methods	Acc (%)	Other parameters
Physionet Apnea-ECG	Sharma et al. [13]	2016	LS-SVM	83.8	Sen: 79.5, Spe: 88.4, AUC: 0.834
	Wang et al. [10]	2019	LeNet-5 CNN	87.6	Sen: 83.1, Spe: 90.3, AUC: 0.950
	Chang et al. [20]	2020	1D CNN	87.9	Sen: 81.1, Spe: 92.0, AUC: 0.935
	Feng et al. [16]	2021	TDCS	85.1	Sen: 86.2, Spe: 84.4
	Almutairi et al. [21]	2021	CNN-LSTM	90.9	Sen: 91.2, Spe: 90.4
	Shen et al. [22]	2021	MSDA-1DCNN	89.4	Sen: 89.8, Spe: 89.1, AUC: 0.964
	Qin et al. [23]	2022	1DCNN-RLM	91.1	Sen: 88.9, Spe: 92.4, AUC: 0.970
	Yang et al. [24]	2022	1D-SEResGNet	90.3	Sen: 87.6, Spe: 91.9, AUC: 0.965
	Tyagi et al. [25]	2023	FT-EDBN	89.1	Sen: 83.9, Spe: 92.3, AUC: 0.960
	Abasi et al. [26]	2023	MHBA+LeNet-5	91.3	Sen: 90.1, Spe: 93.6, AUC: 0.975
	Srivastava et al. [27]	2023	ALexNet + LSTM	90.9	Sen: 95.5, Spe: 83.4
UCDDB	Travieso et al. [28]	2014	SVM-HMM	98.64	–
	Hassan et al. [29]	2017	TQWT, RUSBoost	91.94	Sen: 90.35, Spe: 92.67
	Mashrur et al. [30]	2021	SCNN	81.86	Sen: 71.62, Spe: 86.05
	Zarei et al. [31]	2022	CNN-LSTM	93.7	Sen: 90.69, Spe: 95.82
	Chen et al. [32]	2022	CNN-BiGRU	92.3	Sen: 70.5, Spe: 93.9
	Kemidi et al. [33]	2024	OSAD-Net	98.0	Sen: 100, Spe: 95.08

and ECG-derived respiration (EDR) features combined with artificial neural networks (ANN) and SVM for SA detection, obtaining an accuracy of 82.12% on the Physionet Apnea-ECG dataset. While these methods achieved certain success, their performance heavily relied on the specific characteristics of the manually designed features. These features were carefully crafted for each model and were often limited in adaptability to other datasets, potentially reducing robustness and generalizability across different subjects and recording environments.

Hassan et al. [29] utilized the tunable-Q factor wavelet transform (TQWT) to extract features from single-lead ECG signals and employed a machine learning algorithm, random under-sampling boosting (RUSBoost), for SA classification. Their proposed method achieved 88.88% accuracy, 87.58% sensitivity, and 91.49% specificity on the Physionet Apnea-ECG dataset. For the UCDDB dataset, the accuracy, sensitivity, and specificity values are 91.94%, 90.35%, and 92.67%, respectively. Rachim et al. [34] decomposed ECG signals into five levels using wavelet decomposition, extracting 15 features from the detail coefficients (D3–D5). They then used principal component analysis (PCA) for feature dimension reduction, followed by SVM classification, with the specificity, sensitivity, and subject-based classification accuracy at 95.20%, 92.65%, and 94.3%, respectively.

Varon et al. [35] focused on extracting principal components from QRS complexes (Q-wave, R-wave, and S-wave) and respiratory information, followed by classification with a least-squares SVM. Accuracies of about 85% are achieved on a minute-by-minute basis for two independent datasets (the Physionet Apnea-ECG dataset and the Sleep Laboratory dataset from University Hospital Leuven (UZ Leuven) in Belgium), including both hypopneas and apneas. Additionally, apnea episodes were distinguished from normal recordings with 100% accuracy. Song et al. [36] addressed the temporal nature of SA by designing a combined SVM and hidden Markov model classifier. An accuracy of 86.2% was achieved for per-segment classification and 97.1% for per-recording OSA detection on the Physionet Apnea-ECG dataset. SVM classification also using manually crafted features has gained significant attention, as evidenced by the works of [13,37].

DL models for SA detection have received growing attention in recent years. Li et al. [38] presented an SA detection method that combines a sparse autoencoder and a hidden Markov model (HMM). This method leverages an unsupervised sparse autoencoder to learn features from ECG signals, followed by SVM classification. Their method achieved 85% classification accuracy for per-segment and 100% for per-recording SA detection on the Physionet Apnea-ECG dataset. Similarly, Urtnasan et al. [39] employed a DL approach with a six-layer CNN, including activation, pooling, dropout, Adam optimal performance, softmax binaries classification, and the fully connected to detect SA. For training and evaluating the CNN model, a single-lead ECG dataset was collected from 82 subjects with OSA. This dataset was split into a training set, comprising data from 63 patients with 34,281 events, and a testing set, comprising data from 19 patients with 8571 events. The CNN model achieved a precision of 99%, a recall of 99%, and an F1-score of 99% on the training dataset, and these metrics were each 96% when applied to the testing dataset. Singh et al. [40] conducted continuous wavelet transformation on individual 1 min segments of ECG signals to produce scalogram images. They subsequently introduced a classification framework utilizing CNN with an AlexNet pre-trained model. Their proposed CNN-based method achieved an accuracy of 86.22% with a sensitivity of 90% in per-minute segment classification of OSA and 100% accuracy for per-recording.

In other work, Erdenebayar et al. [41] explored various neural network architectures for modeling purposes, ultimately suggesting the adoption of 1DCNN and gated recurrent unit (GRU) for automated SA detection. The accuracy of the 1D CNN and GRU model was 99.0%. Wang et al. [10] employed 18 features for the diagnosis of SA, comprising 12 features derived from RR intervals and six features related to amplitude. The R-peaks were extracted using the Hamilton algorithm, followed by the calculation of RR distances. They proposed a modified LeNet-5 convolutional neural network with adjacent segments for SA detection. Subsequently, a fully connected layer was utilized to achieve a low-complexity network structure. Their modified LeNet-5 model achieved a specificity of 90.3%, sensitivity of 83.1%, accuracy of 87.6%, and AUC of 0.950 on the Physionet Apnea-ECG dataset.

Chen et al. [32] introduced an end-to-end spatio-temporal learning-based SA detection method that consists of multiple spatio-temporal blocks. Each block features a CNN layer, a max-pooling layer, and a bidirectional gated recurrent unit (BiGRU) layer. This repeated spatio-temporal block architecture can capture both morphological spatial and temporal features from ECG signals. Their proposed CNN-BiGRU achieved an accuracy of 91.22% and 97.10% for per-minute classification and per-recording classification, respectively, on the Physionet Apnea-ECG dataset and an accuracy of 91.24% on the UCDDB dataset. Qin et al. [23] proposed a representation learning model based on a one-dimensional convolutional neural network (1DCNN-RLM) to extract feature vectors from the RRI series. They then developed a temporal dependence model (BiGRU-TDM) to learn the transition patterns between states (SA/normal) across segments, facilitating the classification task. Their method achieved an accuracy of 91.1%, sensitivity of 88.9%, and specificity of 92.4% for per-segment detection, and an accuracy of 100% for per-recording detection.

Tyagi et al. [25] proposed a novel approach for the SA classification using a single-lead ECG signal called fine-tuned enhanced-Deep Belief Networks (FT-EDBN) method. The proposed FT-EDBN method obtained features from training data to distinguish between apnea and healthy episodes. The proposed method achieved 89.11% accuracy for per-segment detection, with specificity, sensitivity, and F1-score of 92.28%, 83.89%, and 0.913, respectively. For per-recording detection, it attained an accuracy of 97.17% and a correction coefficient of

0.938. Recently, two models have been introduced for SA detection using a single-lead ECG signal: the Restricted Attention Fusion Network (RAFNet) [42] and the Bottleneck Attention-based Fusion Network (BAFNet) [43]. RAFNet comprises a target segment feature extractor, an adjacent segment feature extractor (The term of the target segment and adjacent segments for target RR intervals/R peak and adjacent RR intervals/R peak segments in the subsequent sections), a restricted attention module, and the Fully Convolutional Network classifier. On the other hand, BAFNet is composed of an RRI stream network, an RPA stream network, global query generation, feature fusion, and a classifier. RAFNet achieved 91.40% accuracy on the Physionet Apnea-ECG dataset and 84.70% accuracy on the real clinical FAH-ECG dataset. BAFNet achieved 89.05%, 92.63%, and 91.25% in terms of sensitivity, specificity, and accuracy, respectively, for per-segment performance. As for per-recording performance, all sensitivity, specificity, and accuracy are 100% on the Physionet Apnea-ECG dataset.

Kemidi et al. [33] developed OSAD-Net, a network for detecting OSA, using an optimized bi-directional long short-term memory (OBi-LSTM) model combined with a random forest-based exhaustive feature selector (RF-EFS). Initially, a multi-layer convolutional neural network (MLCNN) was employed to extract deep features from an ECG-based OSA dataset. Subsequently, RF-EFS was applied to identify the most relevant features through multi-level decision-making. Finally, OBi-LSTM was trained on these optimized features to perform OSA detection. The model achieved an accuracy of 98.00%, an F1-score of 97.91%, a sensitivity of 100.0%, and a specificity of 95.08% on the UCDDB dataset.

The CNN-LSTM model was used by Almutairi et al. to detect SA on a per-segment basis and showed good performance with an accuracy of 90.9% [21]. While this result demonstrates the potential of deep learning in SA detection, their approach is limited by the model's reliance on features in the ECG signal.

Conventionally, ECG is a time series strongly characterized by long-term dependencies between its values at different time points. Our proposed approach addresses these limitations by integrating CNN, Transformer, and LSTM networks into a hybrid architecture. The CNN extracts the spatial and channel features of the ECG signal, while the Transformer enables the model to capture long-term dependencies, which are critical in detecting intermittent SA events. These events include temporary disruptions in breathing patterns, such as apneas (complete pauses in breathing) and hypopneas (partial reductions in airflow), which are characteristic of SA. The LSTM enhances the model's ability to model temporal sequences and long-term dependencies, ensuring the efficient classification of SA events across varying time scales. The proposed model automatically learns and extracts relevant features from the RRI and R-peak signals. This simplifies the workflow and improves the model's generalizability to diverse datasets. Moreover, we have also applied this hybrid architecture to classify sleep stages using a single-channel EEG signal and achieved state-of-the-art results [44], and classify motor imagery patterns using multi-channel EEG [45]. This highlights the versatility of our approach across different physiological signals and classification tasks, showing promise in addressing both SA detection and potentially other domains like sleep disorder detection.

3. Materials and method

3.1. Dataset

The PhysioNet Apnea-ECG dataset is provided by Philipps University [46,47] as an openly accessible public database. We used this dataset to validate and evaluate the performance of the proposed method, comparing it with existing literature. The dataset contains 70 single-lead ECG signal recordings (released set: 35 recordings, withheld set: 35 recordings); the sampling rate is 100 Hz, and the duration of each ECG signal ranges between 401 and 578 min. The released set is used for model building and parameter estimation, and the withheld set is used for testing the model. The withheld set includes 23 recordings of SA subjects and 12 recordings of normal subjects, according to the American Academy of Sleep Medicine (AASM) standard.

Table 2
PhysioNet Apnea-ECG dataset description.

Dataset	SA	Normal	Total
Released set	6514	10,611	17,125
Withheld set	6552	10,751	17,303
Total	13,066	21,362	34,428

An expert annotated each one-minute ECG signal recording segment in the dataset: segments with an apnea event were labeled SA, while those without were labeled normal. The annotation did not distinguish between hypopnea and apnea, categorizing all events as either OSA or mixed, with no inclusion of CSA. The total number of labeled segments for the released and withheld sets were 17,125 (6514 SA and 10,611 normal) and 17,303 (6552 SA and 10,751 normal), respectively. The limitation of the PhysioNet Apnea-ECG dataset is that it is imbalanced with 6514 SA segments and 10,611 healthy segments. Consequently, OSA segments represent less than 38% of the total dataset. Table 2 describes the PhysioNet Apnea-ECG dataset.

The UCDDB dataset includes PSG recordings from 25 patients (21 males and 4 females) [47]. For this study, we utilize ECG signals sampled at 128 Hz. Sleep segments are classified as apnea or non-apnea based on annotations provided by sleep experts. To address the class imbalance, the training and validation sets are balanced by oversampling the minority class (SA events). Given the class imbalance in the UCDDB dataset, we conducted experiments using both the full dataset (25/25 patients) and a reduced version that excluded patients without any SA events. In the reduced set (21/25 patients), data from patients ucddb008, ucddb011, ucddb013, and ucddb018 were omitted.

3.2. Data preprocessing

We initially used the Hamilton algorithm [48] to detect the R peaks and then calculated the RRI (intervals between the R peaks) and extracted the R peak values based on their positions. The raw ECG signal was normalized and filtered using the Finite Impulse Response (FIR) bandpass filter.

In this study, we chose a median filter as recommended by Chen, Zhang, and Song [49] to address data spikes. Unlike a simple moving average filter, the median filter more effectively mitigates outliers without distorting RRI trends, making it a suitable choice for ECG signals with abrupt changes. A median filter was applied as we expected to have non-interpretable points that were generated from RRI. This provided a robust estimate of the generated value from RRI. The variations of generated values of RRI led to marking them as suspect RRI, which happened because of false R peak detections or missed R peaks. Cubic interpolation was employed to detect false R peaks by comparing the sum of adjacent RRI with a robust estimate of expected RRI values. When this robust estimation was applied, false R peaks were numerically closer to the sum of adjacent RRI than the individual RRI [21,50]. Then we obtained 900 points of RRI and 900 points for R amplitudes within each 5 min segment [10].

Following preprocessing, we removed 774 segments from the PhysioNet Apnea-ECG dataset due to false R-peak detections. This resulted in a total of 33,654 segments: 16,709 in the released set (6473 SA and 10,236 normal) and 16,945 in the withheld set (6490 SA and 10,455 normal). Table 3 provides an overview of the PhysioNet Apnea-ECG dataset post-preprocessing. Fig. 2 shows the preprocessing steps of the ECG signals as suggested by Wang et al. [10].

For the UCDDB dataset, following previous research [51,52] implemented with detection resolutions on a second-by-second basis, the window size was fixed at 11 s as an apnea event is defined by a

Table 3
PhysioNet Apnea-ECG dataset description after preprocessing.

Dataset	SA	Normal	Total
Released set	6473	10,236	16,709
Withheld set	6490	10,455	16,945
Total	12,963	20,691	33,654

breathing cessation of at least 10 s. To capture such events effectively, overlapping windows of 11 s were generated with a 10-s overlap. Each window was labeled as apnea or non-apnea based on the state of the 2nd second within that window.

3.3. Method

In our proposed method for SA detection using a single-lead ECG signal, we use the model we developed for sleep stage classification using a single-channel EEG signal [44] and for motor imagery classification using multi-channel EEG [45]. The CNN-Transformer-LSTM model included three units: the Convolutional Neural Network (CNN), the Transformer Encoder, and the Long-Short-Term Memory (LSTM) network. Finally, a Softmax activation function is used for accurate SA classification per-segment and per-recording. Instead of using nine layers [44], we reduced it to three layers. This simplifies the model, reduces the number of parameters, and decreases the training time [45]. A schematic diagram for a proposed method for SA detection is shown in Fig. 3.

3.3.1. CNN

CNN is deployed to extract the spatial and channel features from the ECG signal. This stage compresses the signal information while retaining the critical characteristics relevant to SA detection. In this study, the CNN model comprises three Convolution 1D layers (Filter: 64, 128, 128, Kernel Size: 7, Stride: 1) and three Max-Pooling 1D layers (Pool size: 4), and the last layer is a dropout (0.5). Fig. 4 presents the architecture of the CNN.

3.3.2. Transformer encoder

Unlike traditional recurrent models, the Transformer enables the model to capture long-term dependencies and temporal dynamics without sequential processing [53]. Its architecture comprises two sublayers organized around a Multi-Head Attention and Feed-forward Neural Network (FNN). Each sub-layer incorporates residual connections and is followed by layer normalization. Specifically, the output of each sub-layer is computed as $LayerNorm(x + Sublayer(x))$, where $Sublayer(x)$ is the function implemented by the sub-layer itself [54]. Fig. 5 illustrates the structure of the Transformer Encoder architecture.

3.3.3. LSTM

Hochreiter and Schmidhuber first introduced the LSTM network, a type of RNN, in their work published in 1997 [55]. The LSTM consists of sequential cells, each passing a cell state and a hidden state to the next. Each cell features three gates (forget, input, and output) that control information flow. The forget gate decides whether to retain or discard the cell's current memory, the input gate regulates how much new information is added to the memory, and the output gate determines whether the memory cell should influence the output at the current time step. An additional input node, usually activated by tanh, further modulates the cell state. LSTM can solve the problem of long-term dependencies, as they implement gates to control the process of memorization [56]. Fig. 6 illustrates the LSTM network.

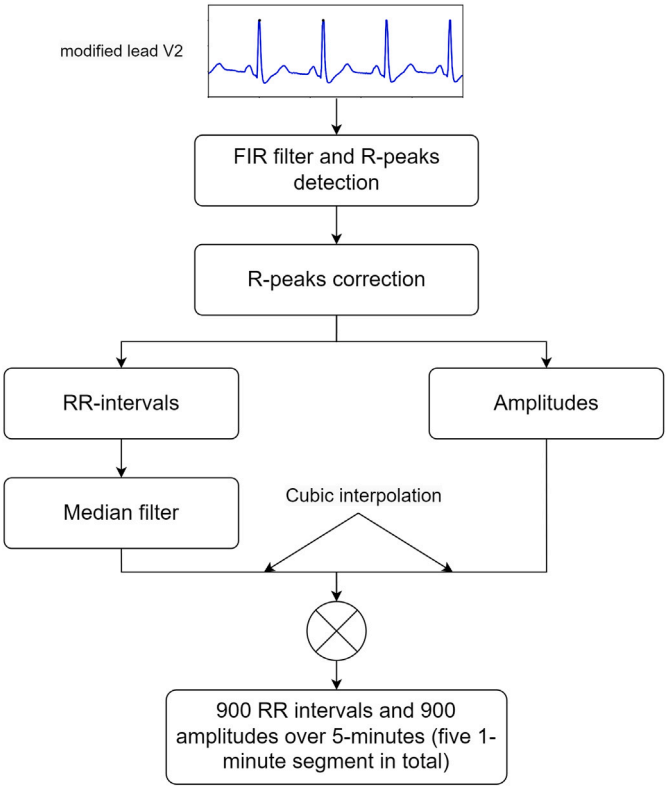


Fig. 2. PhysioNet Apnea-ECG dataset preprocessing scheme.

3.4. Ablation study

To assess the effectiveness of each component in the proposed model, we performed ablation experiments. An ablation study is an additional experimental analysis designed to evaluate the impact of individual components on model performance under different conditions. It aims to provide insights into the algorithm's behavior rather than relying on clinical knowledge of SA. In this experiment, we incorporated three key components: CNN, Transformer Encoder, and LSTM. Moreover, to explore the effect of varying CNN configurations, we adjusted the filter sizes in the CNN layers using two different settings: (64, 128, 128) and (64, 128, 256). This variation resulted in 12 distinct model configurations labeled from M1 to M12. A comprehensive overview of each model, including specific parameter settings, is presented in Table 4.

3.5. Training

The model was optimized using the "Adam" optimizer with a "binary_crossentropy" loss function. Based on experimental tuning, we set the initial learning rate to 0.001 and used the 'ReduceLROnPlateau' function to automatically lower the learning rate when model performance plateaued. Early stopping was applied to prevent overfitting, terminating training if the validation loss did not improve over 30 consecutive epochs. For the PhysioNet Apnea-ECG dataset, we employed both hold-out and five-fold cross-validation to evaluate the model's performance, ensuring both robustness and generalizability. With hold-out validation, the training set was further divided into an 80% training and 20% validation split. In contrast, for the UCDDB dataset, we adopted a hold-out validation, splitting the data into training, validation, and test sets in an 8:1:1 ratio. This comprehensive evaluation approach ensured that the reported metrics reliably reflected the model's ability to generalize.

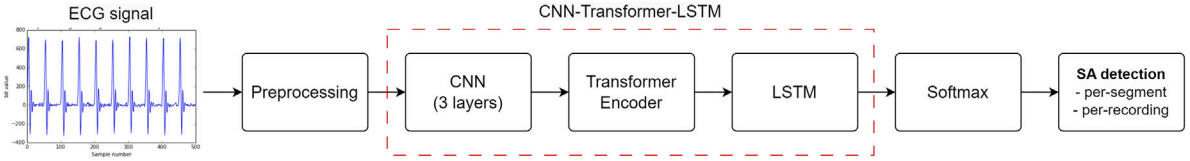


Fig. 3. The architecture of the proposed method.

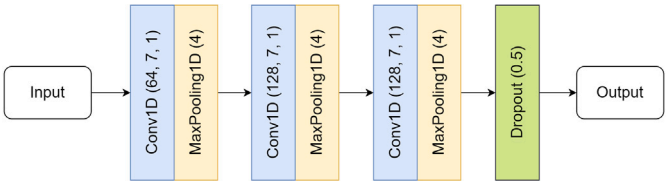


Fig. 4. The architecture of the CNN.

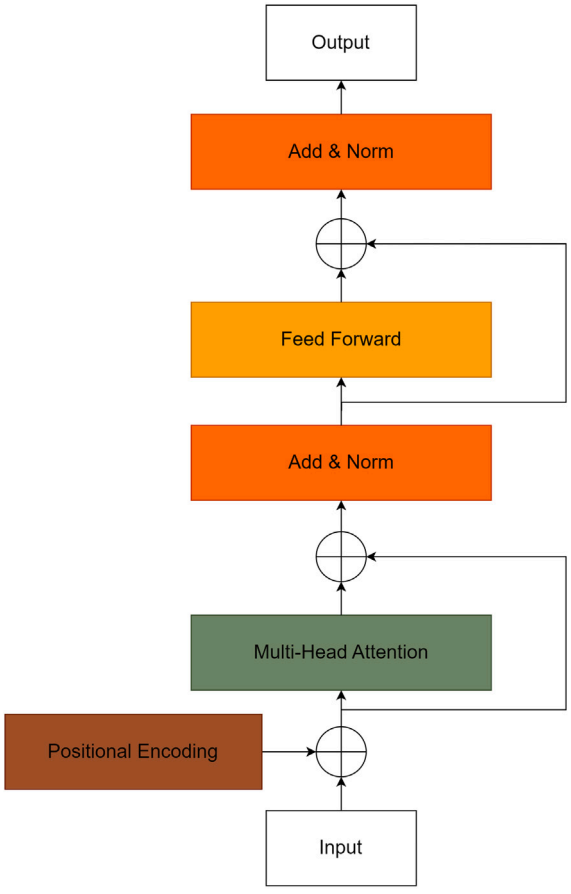


Fig. 5. The architecture of the Transformer Encoder.

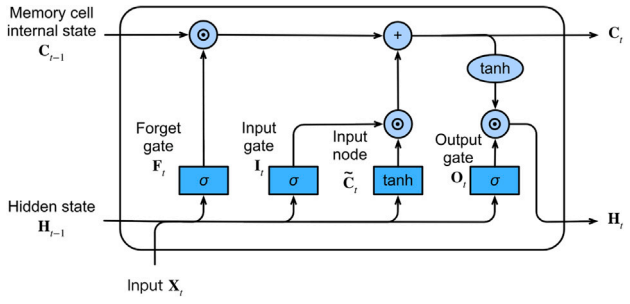


Fig. 6. The architecture of the LSTM [57].

3.6. Performance metrics

3.6.1. Per-segment metrics

Accuracy (Acc), sensitivity (Sen), specificity (Spe), precision (Pre), F1-score, the area under the receiver operating characteristic curve (AUC), and Cohen's kappa (k) are used as the evaluation metrics for per-segment SA detection. These metrics are defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Sen = \frac{TP}{TP + FN} \tag{2}$$

$$Spe = \frac{TN}{TN + FP} \tag{3}$$

$$Pre = \frac{TP}{TP + FP} \tag{4}$$

$$F1 - score = 2 \times \frac{Pre \times Sen}{Pre + Sen} \tag{5}$$

$$p_e = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \tag{6}$$

$$k = \frac{Acc - p_e}{1 - p_e} \tag{7}$$

where TP (True Positive) represents the number of SA segments correctly predicted as SA; TN (True Negative) denotes the number of normal segments correctly predicted as normal; FP (False Positive) represents the number of normal segments incorrectly predicted as SA; FN (False Negative) denotes the number of SA segments incorrectly predicted as normal; p_e represents the expected agreement, or how much agreement would get by chance.

3.6.2. Per-recording metrics

Per-recording metrics include Acc, Sen, Spe, AUC, and the Pearson correlation coefficient (Corr). To determine Acc, Sen, and Spe, defining whether an overnight recording is classified as SA is crucial. According to the AASM guidelines, a recording is diagnosed as SA if AHI exceeds 5; otherwise, it is classified as normal [58]. The AHI for each recording is computed based on the results of per-segment SA detection, which is defined as follows:

$$AHI = \frac{60}{T} \times N \tag{8}$$

where T denotes the total number of one-minute ECG segment signals, and N is the number of corresponding one-minute-long SA segments.

The Pearson correlation coefficient is employed to assess the performance of the proposed method in per-recording SA detection, measuring the correlation between the experimentally estimated AHI and the actual AHI. This metric ensures the reliability of the comparison [13]. The Pearson correlation coefficient is defined as:

$$Corr = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}} \tag{9}$$

where X is the list of actual AHI values, Y is the list of estimated AHI values, $\bar{X}$ and $\bar{Y}$ are mean values of X and Y, respectively.

Table 4
The models and their corresponding parameters used in the ablation study.

Model	Name	Parameters	CNN			Transformer	LSTM
M1	CNN	403K	64	128	128	–	–
M2	CNN	747K	64	128	256	–	–
M3	Transformer	1,9M	–	–	–	128	–
M4	LSTM	2,0M	–	–	–	–	128
M5	CNN-LSTM	534K	64	128	128	–	128
M6	CNN-LSTM	1,2M	64	128	256	–	256
M7	CNN-Transformer	469K	64	128	128	128	–
M8	CNN-Transformer	946K	64	128	256	256	–
M9	CNN-LSTM-Transformer	601K	64	128	128	128	128
M10	CNN-LSTM-Transformer	1,4M	64	128	256	256	256
M11	CNN-Transformer-LSTM	601K	64	128	128	128	128
M12	CNN-Transformer-LSTM	1,4M	64	128	256	256	256

Table 5
Per-segment and per-recording classification results using hold-out validation for 12 models on the Physionet Apnea-ECG dataset.

Model	Time	Per-segment						Per-recording				
	(s/epoch)	Acc (%)	Sen (%)	Spe (%)	F1-score	AUC	k	Acc (%)	Sen (%)	Spe (%)	AUC	Corr
M1	6	91.1	88.8	92.4	0.884	0.968	0.811	97.14	100	95.65	1	0.978
M2	6	90.7	86.7	93.2	0.877	0.966	0.802	97.14	100	95.65	1	0.972
M3	16	85.7	81.0	88.7	0.813	0.928	0.697	94.29	91.67	95.65	0.992	0.936
M4	45	86.5	77.2	92.3	0.814	0.936	0.709	91.67	97.14	95.65	0.993	0.959
M5	12	91.2	86.2	94.3	0.885	0.968	0.812	97.14	100	95.65	1	0.977
M6	13	91.1	92.5	90.2	0.888	0.971	0.814	97.14	91.67	100	0.996	0.985
M7	8	91.4	90.6	91.9	0.890	0.969	0.819	100	100	100	1	0.984
M8	9	91.0	87.9	92.8	0.881	0.967	0.808	97.14	100	95.65	0.993	0.979
M9	13	91.3	90.1	92.0	0.888	0.970	0.816	97.14	100	95.65	1	0.985
M10	15	91.0	91.2	90.9	0.886	0.970	0.812	97.14	100	95.65	1	0.985
M11	13	**91.6**	**88.8**	**93.4**	**0.890**	**0.969**	**0.822**	**100**	**100**	**100**	**1**	**0.982**
M12	15	90.8	89.2	91.8	0.881	0.967	0.806	100	100	100	1	0.982

4. Results

4.1. Per-segment classification

Table 5 presents the classification results of models M1–M12 for per-segment and per-recording analysis using hold-out validation on the Physionet Apnea-ECG dataset. Among these models, the CNN-Transformer-LSTM (M11) achieved the highest per-segment classification accuracy (91.6%), along with strong performance across sensitivity (88.8%), specificity (93.4%), F1-score (0.890), AUC (0.969), and k (0.822). Subsequently, when evaluated using five-fold cross-validation, the model's accuracy improved from 91.6% to 94.1%, with enhanced performance in sensitivity (95.4%), specificity (93.2%), F1-score (0.925), AUC (0.986), and k (0.876), further confirming its robustness and generalization capability.

As shown in Table 7, the per-segment performance comparison on the Physionet Apnea-ECG dataset demonstrates that the CNN-Transformer-LSTM model achieves the highest accuracy among state-of-the-art approaches using both hold-out and five-fold cross-validation. Figs. 7 and 8 illustrate the training and validation accuracy and loss curves for both the hold-out and five-fold cross-validation methods. In the hold-out setting, the model exhibited stable performance between epochs 10 and 40, with early stopping halting the training at epoch 40. For five-fold cross-validation, training continued until epoch 50 before convergence. The confusion matrix in Fig. 9 summarizes the per-segment classification results for both validation strategies, evaluated on the withheld test dataset comprising 35 subjects, including 10,455 normal and 6490 SA-affected samples.

Based on the results in Table 5, we selected models M1, M5, M7, M9, and M11 for evaluation on the UCDDB dataset. This selection was primarily due to their superior performance compared to models using the 64-128-256 architecture (e.g., M2, M6, M8, M10, M12). Table 6 presents the classification results of the five selected models on the UCDDB dataset, demonstrating consistently high performance across all metrics. M5 (CNN-LSTM) achieved the best overall performance, with accuracies of 99.61% on the full dataset (25/25 patients) and 98.69% on the reduced dataset excluding four patients (21/25). M11 (CNN-Transformer-LSTM) also performed exceptionally well, achieving accuracies of 99.37% and 98.34% on the full and reduced datasets, respectively, along with high scores across all other evaluation metrics.

Table 6
Classification results using 5 models on the UCDDB dataset.

Model	Time (s/epoch)	Acc (%)	Sen (%)	Spe (%)	F1-score	AUC	k
Full dataset (25/25)							
M1	160	98.25	96.58	98.12	0.694	0.995	0.685
M5	219	**98.69**	**98.58**	**98.69**	**0.774**	**0.999**	**0.767**
M7	231	97.52	96.31	97.55	0.638	0.995	0.626
M9	289	98.13	89.72	98.33	0.685	0.991	0.676
M11	287	98.34	87.60	98.59	0.706	0.995	0.697
Reduced dataset excluding 4 patients (21/25)							
M1	132	98.94	95.41	99.04	0.827	0.997	0.822
M5	189	**99.61**	**95.41**	**99.72**	**0.929**	**0.999**	**0.927**
M7	203	98.29	96.63	98.34	0.751	0.995	0.742
M9	260	98.88	95.77	98.97	0.820	0.996	0.815
M11	258	99.37	97.35	99.42	0.891	0.999	0.888

4.2. Per-recording classification

As shown in Table 5, the CNN-Transformer-LSTM (M11 and M12) model achieved perfect per-recording classification performance on the Physionet Apnea-ECG dataset using hold-out validation, with 100% accuracy, sensitivity, and specificity, an AUC of 1, and Corr of 0.982. When evaluated using five-fold cross-validation, the model maintained the same perfect scores for accuracy, sensitivity, specificity, and AUC, while the Corr increased to 0.9996. The model has correctly classified 23 apnea subjects and 12 normal subjects on the dataset.

As for the per-recording performance comparison on the Physionet Apnea-ECG dataset shown in Table 8, our per-recording classification results surpass all other studies. Eight out of the 18 comparative studies achieved similar results [22–24,38,42,43,59,65].

Table 7
Comparison of the proposed method with existing methods for per-segment classification on the PhysioNet Apnea-ECG dataset.

References	Year	Methods	Validation	Acc (%)	Sen (%)	Spe (%)	F1-score	AUC	*k*
Sharma et al. [13]	2016	LS-SVM	Ten-fold	83.8	79.5	88.4	–	0.834	–
Li et al. [38]	2018	DNN-HMM	Hold-out	84.7	88.9	82.1	0.869	0.869	–
Wang et al. [10]	2019	LeNet-5 CNN	Ten-fold	87.6	83.1	90.3	–	0.950	–
Chang et al. [20]	2020	1D CNN	Hold-out	87.9	81.1	92.0	–	0.935	–
Feng et al. [16]	2021	TDCS	Two-fold	85.1	86.2	84.4	–	–	–
Huang et al. [14]	2021	ConCAD	Hold-out	91.2	–	–	–	–	–
Almutairi et al. [21]	2021	CNN-LSTM	Ten-fold	90.9	91.2	90.4	0.928	–	–
Shen et al. [22]	2021	MSDA-1DCNN	Ten-fold	89.4	89.8	89.1	–	0.964	–
Chen et al. [59]	2021	SE-MSCNN	Hold-out	90.6	86.0	93.5	–	–	–
Yeh et al. [60]	2022	1D CNN	Hold-out	88.6	83.8	91.5	–	–	–
Fang et al. [61]	2022	ResNet + Multiscale	Hold-out	86.0	84.1	87.1	0.821	0.931	–
Qin et al. [23]	2022	1DCNN-RLM	Ten-fold	91.1	88.9	92.4	0.883	0.970	–
Yang et al. [24]	2022	1D-SEResGNet	Ten-fold	90.3	87.6	91.9	–	0.965	–
Bahrami et al. [62]	2022	ZFNet-BiLSTM	five-fold	88.1	81.5	92.3	0.840	–	–
Chen et al. [32]	2022	CNN-BiGRU	Hold-out	91.2	86.5	94.2	0.883	–	–
Tyagi et al. [25]	2023	FT-EDBN	Ten-fold	89.1	83.9	92.3	0.913	0.960	–
Mohan et al. [63]	2023	1D deep CNN	–	87.9	81.1	92.0	–	0.94	–
Chen et al. [42]	2023	RAFNet	Hold-out	91.4	88.8	93.1	–	–	–
Chen et al. [43]	2023	BAFNet	Hold-out	91.3	89.0	92.3	–	–	–
Abasi et al. [26]	2023	MHBA+LeNet-5	Ten-fold	91.3	90.1	93.6	–	0.975	–
Srivastava et al. [27]	2023	ALexNet + LSTM	Hold-out	90.9	95.5	83.4	–	–	–
Abasi et al. [64]	2024	BBHBA3	Ten-fold	90.1	89.2	92.8	–	0.964	–
Pham et al. [65]	2025	CNN-Transformer	Hold-out	91.4	90.6	91.9	0.890	0.969	–
		CNN-Transformer-LSTM	**Hold-out**	**91.6**	**88.8**	**93.4**	**0.890**	**0.969**	**0.822**
		CNN-Transformer-LSTM	**Five-fold**	**94.1**	**95.4**	**93.2**	**0.925**	**0.986**	**0.876**

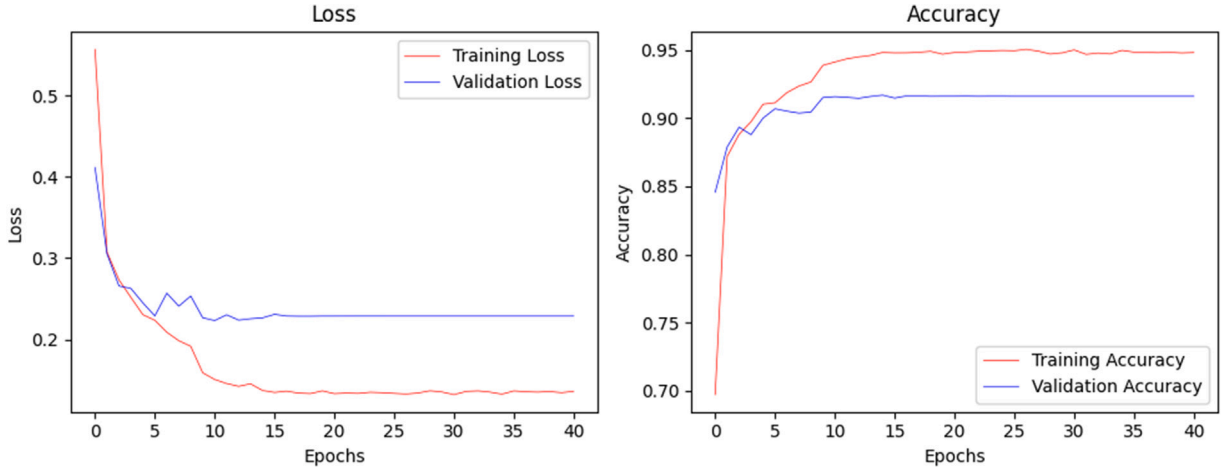


Fig. 7. The accuracy and loss of training and validation sets for hold-out validation on the PhysioNet Apnea-ECG dataset.

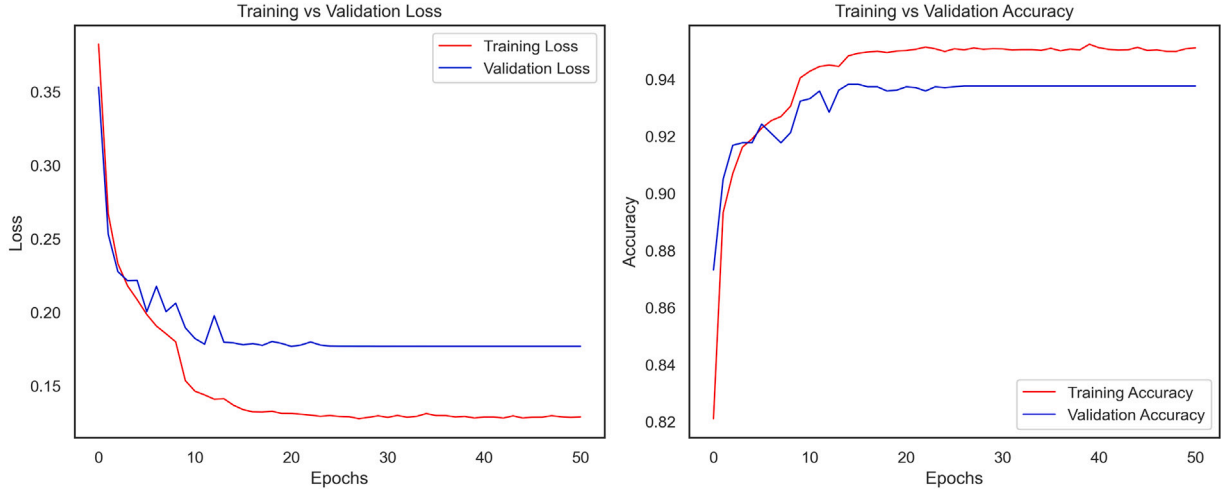
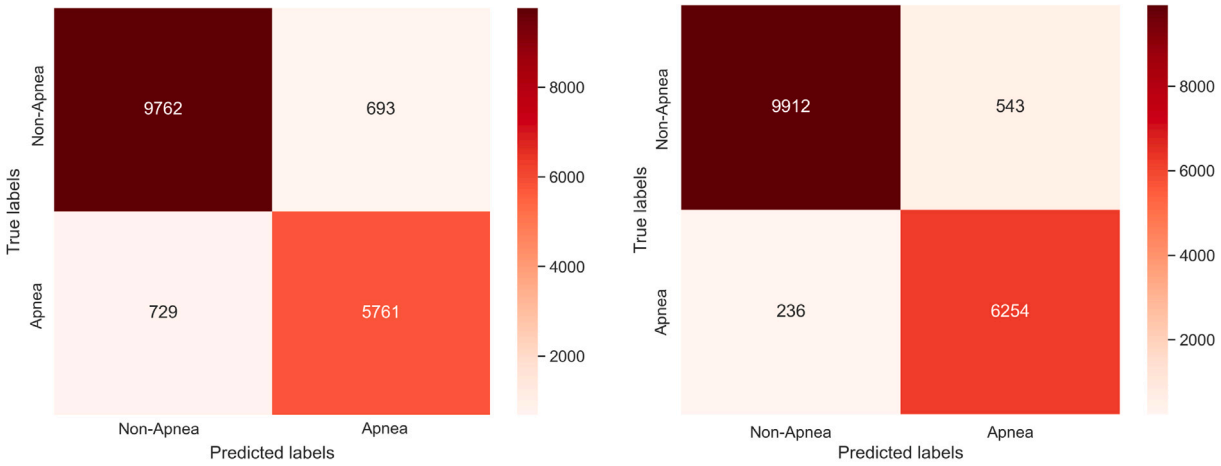


Fig. 8. The accuracy and loss of training and validation sets for five-fold cross validation on the PhysioNet Apnea-ECG dataset.



(a) Hold-out validation (b) Five-fold cross validation

Fig. 9. The confusion matrix of (a) Hold-out and (b) Five-fold cross validation on the PhysioNet Apnea-ECG dataset.

Table 8
Comparison of the proposed method with existing methods for per-recording classification on the PhysioNet Apnea-ECG dataset.

References	Year	Methods	Validation	Acc (%)	Sen (%)	Spe (%)	AUC	Corr
Sharma et al. [13]	2016	LS-SVM	Ten-fold	97.1	95.8	100	0.978	0.841
Li et al. [38]	2018	DNN-HMM	Hold-out	100	100	100	–	–
Wang et al. [10]	2019	LeNet-5 CNN	Ten-fold	97.1	100	91.7	–	0.943
Chang et al. [20]	2020	1D CNN	Hold-out	97.1	95.7	100	–	0.865
Feng et al. [16]	2021	TDCS	Two-fold	97.1	95.7	100	–	–
Shen et al. [22]	2021	MSDA-1DCNN	Ten-fold	100	100	100	–	–
Chen et al. [59]	2021	SE-MSCNN	Hold-out	100	100	100	–	0.979
Yeh et al. [60]	2022	1D CNN	Hold-out	94.3	100	83.3	–	–
Fang et al. [61]	2022	ResNet + Multiscale	Hold-out	97.1	100	91.7	1	0.956
Qin et al. [23]	2022	1DCNN-RLM	Ten-fold	100	100	100	–	0.968
Yang et al. [24]	2022	1D-SEResGNet	Ten-fold	100	100	100	–	0.985
Chen et al. [32]	2022	CNN-BiGRU	Hold-out	97.1	95.7	100	–	0.984
Tyagi et al. [25]	2023	FT-EDBN	Ten-fold	97.1	100	91.7	–	0.938
Mohan et al. [63]	2023	1D deep CNN	–	97.1	95.7	100	–	–
Chen et al. [42]	2023	RAFNet	Hold-out	100	100	100	–	0.985
Chen et al. [43]	2023	BAFNet	Hold-out	100	100	100	–	0.986
Srivastava et al. [27]	2023	ALexNet + LSTM	Hold-out	97.1	–	–	–	–
Pham et al. [65]	2025	CNN-Transformer	Hold-out	100	100	100	1	0.984
		CNN-Transformer-LSTM	**Hold-out**	**100**	**100**	**100**	**1**	**0.982**
		CNN-Transformer-LSTM	**Five-fold**	**100**	**100**	**100**	**1**	**0.9996**

5. Discussion

This study utilized 70 ECG recordings from the PhysioNet Apnea-ECG dataset and the UCDDB dataset to evaluate the CNN-Transformer-LSTM method for classifying SA and normal states from a single-lead ECG signal. Previous studies, such as Sharma et al. [13] and Li et al. [38], have highlighted the importance of feature engineering, a process that relies on prior knowledge of SA and is often labor-intensive and time-consuming. Feature computation and engineering are typically the most crucial and resource-demanding aspects of machine learning techniques. We employed DL to automatically extract and select relevant features from ECG signals for classifying apnea and normal events to address this challenge. The proposed method combines the strengths of each model component: CNNs for spatial feature extraction, Transformers for capturing long-term dependencies, and LSTMs for modeling temporal sequences and long-term dependencies. Our model automatically learns and extracts features from the RRI and R-peak signals, streamlining the workflow and enhancing generalizability across diverse datasets. This approach simplifies the process and significantly reduces human effort. We compared our results with current state-of-the-art methods.

Table 5 presents the classification results using 12 models per-segment and per-recording using hold-out validation on the Physionet Apnea-ECG dataset. The CNN-Transformer-LSTM (M11) model achieves the best per-segment accuracy, sensitivity, specificity, F1-score, AUC, and k of 91.6%, 88.8%, 93.4%, 0.890, 0.969, and 0.822, respectively. Other models, such as CNN-LSTM (M5), CNN-LSTM-Transformer (M9), and CNN-Transformer (M7), also demonstrated competitive accuracy (91.2%, 91.3%, and 91.4%, respectively). However, they exhibited slightly lower sensitivity or specificity compared to the CNN-Transformer-LSTM, suggesting that the combined CNN-Transformer-LSTM approach provides a more balanced performance. Transformer-based model alone (M3) showed weaker results (85.7% accuracy), likely due to their reliance on global attention mechanisms, which may struggle to extract fine-grained local features without the hierarchical feature extraction of CNNs. Similarly, the standalone LSTM model (M4) had a lower accuracy (86.5%), possibly due to its sequential nature, which, while effective for capturing temporal dependencies, may not efficiently model spatial features present in ECG signals. For per-recording classification, both CNN-Transformer-LSTM models (M11 and M12) achieved perfect accuracy (100%), highlighting the robustness of this architecture when aggregating predictions over entire recordings. Notably, the CNN-Transformer (M7) was the only other model to reach 100% accuracy per-recording. In contrast, other models, including CNN-LSTM (M5) and CNN-LSTM-Transformer (M9), fell short of this perfect classification, reinforcing the advantage of the CNN-Transformer-LSTM architecture for both per-segment and per-recording

Table 9
Comparison of the proposed method with existing methods on the UCDDB dataset.

References	Year	Methods	Dataset	Resolution	Acc (%)	Sen (%)	Spe (%)	F1-score	AUC	*k*
Travieso et al. [28]	2014	SVM-HMM	25/25	1 min	98.64	–	–	–	–	–
Hassan et al. [29]	2017	TQWT, RUSBoost	25/25	1 min	91.94	90.35	92.67	–	–	–
Wang et al. [10]	2019	LeNet-5 CNN	25/25	1 min	71.8	26.6	86.9	–	–	–
Fatimah et al. [66]	2020	FDM	25/25	1 min	80.44	68.89	87.59	–	–	–
Mashrur et al. [30]	2021	SCNN	25/25	1 min	81.86	71.62	86.05	0.696	–	–
John et al. [51]	2021	1D-CNN	21/25	1 s	99.56	96.05	99.66	–	–	–
Zarei et al. [31]	2022	CNN-LSTM	25/25	30 s	93.7	90.69	95.82	–	–	–
Fang et al. [61]	2022	ResNet + Multiscale	25/25	–	72.4	36.5	83.6	–	–	–
Chen et al. [32]	2022	CNN-BiGRU	21/25	1 min	92.3	70.5	93.9	0.760	–	–
Kemidi et al. [33]	2024	OSAD-Net	25/25	–	98.0	100	95.08	–	–	–
		CNN-LSTM	**25/25**	**1 s**	**98.69**	**98.58**	**98.69**	**0.774**	**0.999**	**0.767**
		CNN-LSTM	**21/25**	**1 s**	**99.61**	**95.41**	**99.72**	**0.929**	**0.999**	**0.927**
		CNN-Transformer-LSTM	**25/25**	**1 s**	**98.34**	**87.60**	**98.59**	**0.706**	**0.995**	**0.697**
		CNN-Transformer-LSTM	**21/25**	**1 s**	**99.37**	**97.35**	**99.42**	**0.891**	**0.999**	**0.888**

tasks. Beyond accuracy, CNN-Transformer-LSTM (M11) also offers a balance of performance and computational efficiency. M11 requires 13 s per epoch, making it notably more efficient than the LSTM-only model (M4), which demands 45 s per epoch while delivering significantly lower accuracy (86.5%). Moreover, M11 achieves better accuracy than CNN-Transformer (M7) [65], despite M7's slightly faster training time (8 s per epoch). This suggests that the additional LSTM component in M11 enhances temporal feature extraction without significantly increasing computational cost, ultimately leading to a more robust classification model. To further validate the generalizability of the CNN-Transformer-LSTM model, we also applied five-fold cross-validation. Under this more rigorous evaluation, it demonstrated an improvement in per-segment classification accuracy from 91.6% to 94.1%. These highlight the importance of combining CNN, Transformer, and LSTM layers, as seen in M11 (CNN-Transformer-LSTM), which achieved the highest per-segment and per-recording classification accuracy by leveraging the strengths of each component.

Table 7 provides a comparison of our proposed CNN-Transformer-LSTM model against existing methods for the per-segment classification on the PhysioNet Apnea-ECG dataset. Our model achieved an accuracy of 91.6% using hold-out validation and 94.1% using five-fold cross-validation, which is among the highest reported. In terms of accuracy, our model outperforms other approaches by achieving the highest accuracy score among the compared methods, demonstrating its ability to correctly classify SA segments. This high accuracy is attributed to the combination of CNN, Transformer, and LSTM components in the proposed model, which effectively capture both spatial features and temporal dependencies within the ECG signals. This reflects the proposed model's strength in effectively distinguishing between SA and non-SA events, contributing to reliable overall performance. While Srivastava et al. [27] achieves the highest sensitivity (95.5%), it suffers from a significantly lower specificity of 83.4%, suggesting an over-detection of apnea events. Although high sensitivity is crucial in clinical applications to minimize missed apnea events and ensure timely intervention, the trade-off with lower specificity may lead to an increased number of false positives, potentially causing unnecessary clinical evaluations or interventions. When considering specificity, Chen et al. [32] achieved the highest specificity (94.2%) but lower sensitivity (86.5%), indicating a tendency to miss apnea events. Our model provides a more balanced trade-off between sensitivity (95.4%) and specificity (93.2%) using five-fold cross validation, ensuring robust detection with minimal false positives and false negatives. The best F1-score shown by Almutairi et al. [21] is 0.928, but their performance metrics show lower results than ours (such as accuracy of 90.9%, specificity of 90.4%). Moreover, our model achieves an AUC of 0.986, which is the highest reported. However, the studies did not report kappa values, making it difficult to assess their classification consistency. Our model achieved kappa values of 0.822 and 0.876, indicating strong agreement between predicted and actual labels, reinforcing its reliability and robustness. Additionally, our model maintains a well-balanced performance across all metrics, ensuring high sensitivity for correctly identifying apnea events while maintaining high specificity to minimize false positives. This balance makes it an ideal choice for SA detection, offering both accuracy and stability in real-world applications.

Table 8 summarizes the per-recording SA detection performance metrics from various studies compared to our proposed method. Yeh et al. [60] reported the lowest result with 94.3% accuracy; 10 studies did not achieve absolute performance (100% accuracy). Although Chen et al. [32] reported a strong correlation coefficient of 0.984, they misclassified one apnea subject, preventing them from reaching perfect performance. Our model achieved the highest correlation coefficient of 0.9996 using five-fold cross-validation, demonstrating superior consistency in classification.

For the UCDDB dataset, Table 9 compares the proposed method with existing models. Our CNN-Transformer-LSTM model achieved an accuracy of 99.37% on the reduced dataset excluding four patients (21/25) and 98.34% on the full dataset (25/25), demonstrating strong overall classification performance. In contrast, the CNN-LSTM model achieved the best overall performance, with an accuracy of 99.61% on the full dataset and 98.69% on the reduced dataset excluding four patients. This demonstrates that the model performs consistently well, even when a subset of the data is excluded. Although John et al. [51] reported a slightly higher accuracy of 99.56%, their model had lower sensitivity. Furthermore, our CNN-LSTM model attains an even higher accuracy of 99.61% and the highest specificity among all methods. Although CNN-LSTM achieves slightly better accuracy than CNN-Transformer-LSTM, the CNN-Transformer-LSTM model shows a better balance between accuracy, sensitivity, and specificity. While Kemidi et al. [33] achieved a perfect sensitivity of 100%, their accuracy and specificity were lower than that of our model. The CNN-LSTM model proposed by Zarei et al. [31], which achieved an accuracy of 93.7%, sensitivity of 90.69%, and specificity of 95.82%. While their model also utilizes a CNN-LSTM architecture (with different parameters), our approach significantly improves accuracy and sensitivity. This superiority can be attributed to our finer detection resolution, which operates on a second-by-second basis rather than the 30-s epochs used by Zarei et al. The finer granularity allows our model to capture subtle variations in ECG signals associated with SA events, leading to more precise detection. Additionally, our CNN-LSTM model achieves a kappa value of 0.927 on the reduced dataset, whereas our CNN-Transformer-LSTM model reaches 0.888. Both values indicate strong agreement between predicted and actual labels, but the higher kappa on the reduced dataset suggests that the model performs more consistently when tested on a smaller, more homogeneous group of patients. The exclusion of patients without SA events helps mitigate class imbalance, leading to fewer misclassifications and improving the model's ability to capture patterns in the ECG data associated with SA events. On the full dataset, the kappa values are slightly lower, reflecting the increased complexity of generalizing across a broader, more diverse set of patients, some of whom may have less clear or varied ECG patterns.

Despite this, the results from both datasets highlight the robustness of our approach, demonstrating its ability to balance performance across different patient groups while achieving stronger results on the reduced dataset.

In summary, our proposed method achieves an accuracy of 91.6% using hold-out validation and 94.1% using five-fold cross-validation for per-segment classification, along with 100% accuracy for per-recording classification on the PhysioNet Apnea-ECG dataset. On the UCDDB dataset, our method achieves an accuracy of 99.37% on the reduced dataset (21/25) and 98.34% on the full dataset (25/25), outperforming many existing approaches in the literature. A key advantage of this approach is its automated feature extraction, which reduces the need for manual feature engineering and improves classification accuracy for SA detection from ECG signals. The combination of CNN, Transformer, and LSTM models enhances the ability to capture both spatial and temporal patterns in ECG signals, further enhancing classification performance. Nevertheless, combining CNN, Transformer, and LSTM models can increase the computational load, which could pose challenges for real-time deployment. Furthermore, additional studies are needed to validate its applicability across diverse populations, including those with varying health conditions, demographic factors, and environmental settings.

6. Conclusion

This study introduces a novel SA detection method using a CNN-Transformer-LSTM architecture applied to single-lead ECG signals. The derived RR intervals and R peaks in the ECG signals are extracted as the model input by analyzing the physiological mechanism of SA. Our proposed method subsequently conducts feature extraction and SA classification on a per-segment and per-recording basis. The experimental results on the Physionet Apnea ECG dataset and the UCDDB dataset demonstrate that CNN-Transformer-LSTM outperforms state-of-the-art methods, achieving an accuracy of 91.6% using hold-out validation and 94.1% using five-fold cross-validation for per-segment classification, and 100% for per-recording classification on the PhysioNet Apnea-ECG dataset. With the UCDDB dataset, our model achieves an accuracy of 99.37% on the reduced dataset (21/25) and 98.34% on the full dataset (25/25). This approach holds the potential for developing SA detection devices leveraging ECG signals in both clinical and home healthcare environments.

Despite these promising results, certain limitations require further investigation. Our current approach classifies SA events as either apnea or normal without distinguishing between different types of apnea. Although obstructive sleep apnea (OSA) is the most prevalent form, differentiating between OSA, central sleep apnea (CSA), and mixed sleep apnea (MSA) could provide more clinically relevant insights.

Future research could address several limitations to enhance the robustness and applicability of this method. First, exploring strategies to reduce the computational load of the CNN-Transformer-LSTM model could improve its suitability for real-time applications and embedded systems. Second, expanding validation studies across datasets representing diverse populations with different demographics, health conditions, and environmental factors would enhance the model's generalizability. Additionally, investigating the model's potential for detecting other cardiovascular conditions (such as arrhythmias, conduction disturbances, acute coronary syndromes, and cardiac chamber hypertrophy) could broaden its utility in cardiac monitoring. We also suggest balancing the dataset by using oversampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE). Finally, integrating data from supplementary sensors, including EEG, EMG, and SpO2, could improve accuracy and enable a comprehensive approach to sleep monitoring and overall health management.

CRediT authorship contribution statement

Duc Thien Pham: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Roman Mouček:** Writing – review & editing, Validation, Supervision, Investigation.

Code availability

The source code is available at: https://github.com/ducthien512/Sleep-apnea-detection.

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.

Acknowledgment

This work was supported by the University specific research project SGS-2025-022 New Data Processing Methods in Current Areas of Computer Science (project SGS-2025-022).

Data availability

The datasets used in this study are publicly available at https://physionet.org/content/apnea-ecg/1.0.0 and https://physionet.org/content/ucddb/1.0.0/.

References

[1] A.V. Benjafield, N.T. Ayas, P.R. Eastwood, R. Heinzer, M.S.M. Ip, M.J. Morrell, C.M. Nunez, S.R. Patel, T. Penzel, J.-L. Pépin, P.E. Peppard, S. Sinha, S. Tufik, K. Valentine, A. Malhotra, Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis, Lancet Respir. Med. 7 (2019) 687–698, http://dx.doi.org/10.1016/S2213-2600(19)30198-5.

[2] V.K. Kapur, D.H. Auckley, S. Chowdhuri, D.C. Kuhlmann, R. Mehra, K. Ramar, C.G. Harrod, Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: An American academy of sleep medicine clinical practice guideline, J. Clin. Sleep Med. 13 (2017) 479–504, http://dx.doi.org/10.5664/jcsm.6506.

[3] O.D. Lyons, C.M. Ryan, Sleep apnea and stroke, Can. J. Cardiol. 31 (2015) 918–927, http://dx.doi.org/10.1016/j.cjca.2015.03.014.

[4] American Academy of Sleep Medicine, Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force, Sleep 22 (5) (1999) 667–689, http://dx.doi.org/10.1093/sleep/22.5.667.

[5] L.J. Epstein, D. Kristo, P.J. Strollo, N. Friedman, A. Malhotra, et al., Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults, J. Clin. Sleep Med. : JCSM : Off. Publ. Am. Acad. Sleep Med. 5 (2009) 263–276.

[6] 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, http://dx.doi.org/10.1007/s11325-020-02249-0.

[7] R. Statello, S. Rossi, F. Pisani, M. Bonzini, R. Andreoli, A. Martini, M. Puligheddu, P. Cocco, M. Miragoli, Nocturnal heart rate variability might help in predicting severe obstructive sleep-disordered breathing, Biology 12 (4) (2023) 533, http://dx.doi.org/10.3390/biology12040533.

[8] J. Zhu, A. Zhou, Q. Gong, Y. Zhou, J. Huang, Z. Chen, Detection of sleep apnea from electrocardiogram and pulse oximetry signals using random forest, Appl. Sci. 12 (2022) 4218, http://dx.doi.org/10.3390/app12094218.

[9] T. Paul, O. Hassan, K. Alaboud, H. Islam, M.K.Z. Rana, S.K. Islam, A.S.M. Mosa, ECG and SpO2 signal-based real-time sleep apnea detection using feed-forward artificial neural network, AMIA Jt. Summits Transl. Sci. Proc. 2022 (2022) 379–385.

[10] T. Wang, C. Lu, G. Shen, F. Hong, Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network, PeerJ 7 (2019) http://dx.doi.org/10.7717/peerj.7731.

[11] L. Almazaydeh, K. Elleithy, M. Faezipour, Obstructive sleep apnea detection using SVM-based classification of ECG signal features, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 4938–4941, http://dx.doi.org/10.1109/EMBC.2012.6347100.
[12] X. Wang, M. Cheng, Y. Wang, S. Liu, Z. Tian, F. Jiang, H. Zhang, Obstructive sleep apnea detection using ecg-sensor with convolutional neural networks, Multimedia Tools Appl. 79 (23) (2020) 15813–15827, http://dx.doi.org/10.1007/s11042-018-6161-8.
[13] H. Sharma, K.K. Sharma, An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions, Comput. Biol. Med. 77 (2016) 116–124, http://dx.doi.org/10.1016/j.compbiomed.2016.08.012.
[14] G. Huang, F. Ma, ConCAD: Contrastive learning-based cross attention for sleep apnea detection, in: Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track, ECML PKDD, 2021, pp. 68–84, http://dx.doi.org/10.1007/978-3-030-86517-7_5.
[15] R.V. Sharan, S. Berkovsky, H. Xiong, E. Coiera, 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, EMBC, IEEE, 2020, pp. 637–640, http://dx.doi.org/10.1109/EMBC44109.2020.9175998.
[16] 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 (2021) 1–12, http://dx.doi.org/10.1109/TIM.2020.3017246.
[17] C. Viswabhargav, R. Tripathy, U.R. Acharya, Automated detection of sleep apnea using sparse residual entropy features with various dictionaries extracted from heart rate and EDR signals, Comput. Biol. Med. 108 (2019) 20–30, http://dx.doi.org/10.1016/j.compbiomed.2019.03.016.
[18] M. Sharma, S. Agarwal, U.R. Acharya, Application of an optimal class of antisymmetric wavelet filter banks for obstructive sleep apnea diagnosis using ECG signals, Comput. Biol. Med. 100 (2018) 100–113, http://dx.doi.org/10.1016/j.compbiomed.2018.06.011.
[19] A. Pinho, N. Pombo, B.M. Silva, K. Bousson, N. Garcia, Towards an accurate sleep apnea detection based on ECG signal: The quintessential of a wise feature selection, Appl. Soft Comput. 83 (2019) 105568, http://dx.doi.org/10.1016/j.asoc.2019.105568.
[20] H.Y. Chang, C.Y. Yeh, C.T. Lee, C.C. Lin, A sleep apnea detection system based on a one-dimensional deep convolution neural network model using single-lead electrocardiogram, Sensors 20 (2020) 1–15, http://dx.doi.org/10.3390/s20154157.
[21] H. Almutairi, G.M. Hassan, A. Datta, Classification of Obstructive Sleep Apnoea from single-lead ECG signals using convolutional neural and Long Short Term Memory networks, Biomed. Signal Process. Control. 69 (2021) 102906, http://dx.doi.org/10.1016/j.bspc.2021.102906.
[22] Q. Shen, H. Qin, K. Wei, G. Liu, Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal, IEEE Trans. Instrum. Meas. 70 (2021) 1–13, http://dx.doi.org/10.1109/TIM.2021.3062414.
[23] H. Qin, G. Liu, A dual-model deep learning method for sleep apnea detection based on representation learning and temporal dependence, Neurocomputing 473 (2022) 24–36, http://dx.doi.org/10.1016/j.neucom.2021.12.001.
[24] 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, http://dx.doi.org/10.1016/j.compbiomed.2021.105124.
[25] 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) http://dx.doi.org/10.1016/j.bspc.2022.104401.
[26] A.K. Abasi, M. Aloqaily, M. Guizani, Optimization of CNN using modified Honey Badger Algorithm for Sleep Apnea detection, Expert Syst. Appl. 229 (2023) http://dx.doi.org/10.1016/j.eswa.2023.120484.
[27] 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) http://dx.doi.org/10.1016/j.bspc.2023.104754.
[28] C.M. Travieso, J.B. Alonso, M. Del Pozo, J.R. Ticay, G. Castellanos-Dominguez, Building a Cepstrum-HMM kernel for Apnea identification, Neurocomputing 132 (2014) 159–165, http://dx.doi.org/10.1016/j.neucom.2013.04.048.
[29] A.R. Hassan, M.A. Haque, An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting, Neurocomputing 235 (2017) 122–130, http://dx.doi.org/10.1016/j.neucom.2016.12.062.
[30] F.R. Mashrur, M.S. Islam, D.K. Saha, S.R. Islam, M.A. Moni, SCNN: Scalogram-based convolutional neural network to detect obstructive sleep apnea using single-lead electrocardiogram signals, Comput. Biol. Med. 134 (2021) 104532, http://dx.doi.org/10.1016/j.compbiomed.2021.104532.
[31] A. Zarei, H. Beheshti, B.M. Asl, Detection of sleep apnea using deep neural networks and single-lead ECG signals, Biomed. Signal Process. Control. 71 (2022) 103125, http://dx.doi.org/10.1016/j.bspc.2021.103125.
[32] 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) http://dx.doi.org/10.3389/fnins.2022.972581.
[33] M. Kemidi, D.R. Marur, C.K. Reddy, Obstructive sleep apnea detection using optimized Bi-LSTM with random forest based exhaustive feature selector, Multimedia Tools Appl. (2024) 1–23, http://dx.doi.org/10.1007/s11042-024-18837-1.
[34] V.P. Rachim, G. Li, W.-Y. Chung, Sleep apnea classification using ECG-signal wavelet-PCA features, Bio-Med. Mater. Eng. 24 (2014) 2875–2882, http://dx.doi.org/10.3233/BME-141106.
[35] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, S.V. Huffel, A novel algorithm for the automatic detection of sleep apnea from single-lead ECG, IEEE Trans. Biomed. Eng. 62 (2015) 2269–2278, http://dx.doi.org/10.1109/TBME.2015.2422378.
[36] C. Song, K. Liu, X. Zhang, L. Chen, X. Xian, An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals, IEEE Trans. Biomed. Eng. 63 (2016) 1532–1542, http://dx.doi.org/10.1109/TBME.2015.2498199.
[37] G. Surrel, A. Aminifar, F. Rincon, S. Murali, D. Atienza, Online obstructive sleep apnea detection on medical wearable sensors, IEEE Trans. Biomed. Circuits Syst. 12 (2018) 762–773, http://dx.doi.org/10.1109/TBCAS.2018.2824659.
[38] K. Li, W. Pan, Y. Li, Q. Jiang, G. Liu, A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal, Neurocomputing 294 (2018) 94–101, http://dx.doi.org/10.1016/j.neucom.2018.03.011.
[39] E. Urtnasan, J.-U. Park, E.-Y. Joo, K.-J. Lee, Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network, J. Med. Syst. 42 (2018) 104, http://dx.doi.org/10.1007/s10916-018-0963-0.
[40] S.A. Singh, S. Majumder, A novel approach OSA detection using single-lead ECG scalogram based on deep neural network, J. Mech. Med. Biol. 19 (2019) 1950026, http://dx.doi.org/10.1142/S021951941950026X.
[41] U. Erdenebayar, Y.J. Kim, J.-U. Park, E.Y. Joo, K.-J. Lee, Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram, Comput. Methods Programs Biomed. 180 (2019) 105001, http://dx.doi.org/10.1016/j.cmpb.2019.105001.
[42] Y. Chen, H. Yue, R. Zou, W. Lei, W. Ma, X. Fan, RAFNet: Restricted attention fusion network for sleep apnea detection, Neural Netw. 162 (2023) 571–580, http://dx.doi.org/10.1016/j.neunet.2023.03.019.
[43] X. Chen, W. Ma, W. Gao, X. Fan, BAFNet: Bottleneck attention based fusion network for sleep apnea detection, IEEE J. Biomed. Heal. Inform. (2023) http://dx.doi.org/10.1109/JBHI.2023.3278657.
[44] D.T. Pham, R. Mouček, Automatic sleep stage classification by CNN-transformer-LSTM using single-channel EEG signal, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2023, pp. 2559–2563, http://dx.doi.org/10.1109/BIBM58861.2023.10385687.
[45] D.T. Pham, R. Mouček, Automatic motor imagery classification by CNN-transformer-LSTM using multi-channel EEG, in: ECAI 2024, IOS Press, 2024, pp. 4555–4562, http://dx.doi.org/10.3233/FAIA241048.
[46] T. Penzel, G. Moody, R. Mark, A. Goldberger, J. Peter, The Apnea-ECG Database, IEEE, 2000, pp. 255–258, http://dx.doi.org/10.1109/CIC.2000.898505.
[47] 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) E215–20, http://dx.doi.org/10.1161/01.cir.101.23.e215.
[48] P. Hamilton, Open source ECG analysis, IEEE, 2002, pp. 101–104, http://dx.doi.org/10.1109/CIC.2002.1166717,
[49] L. Chen, X. Zhang, C. Song, An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram, IEEE Trans. Autom. Sci. Eng. 12 (2015) 106–115, http://dx.doi.org/10.1109/TASE.2014.2345667.
[50] P. De Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, M. O'Malley, Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea, IEEE Trans. Biomed. Eng. 50 (6) (2003) 686–696, http://dx.doi.org/10.1109/TBME.2003.812203.
[51] A. John, B. Cardiff, D. John, A 1D-CNN based deep learning technique for sleep apnea detection in IoT sensors, in: 2021 IEEE International Symposium on Circuits and Systems, ISCAS, 2021, pp. 1–5, http://dx.doi.org/10.1109/ISCAS51556.2021.9401300.
[52] A. John, K.K. Nundy, B. Cardiff, D. John, Multimodal multiresolution data fusion using convolutional neural networks for IoT wearable sensing, IEEE Trans. Biomed. Circuits Syst. 15 (6) (2021) 1161–1173, http://dx.doi.org/10.1109/TBCAS.2021.3134043.
[53] S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, W. Pedrycz, A comprehensive survey on applications of transformers for deep learning tasks, Expert Syst. Appl. (2023) 122666, http://dx.doi.org/10.1016/j.eswa.2023.122666.
[54] A. Vaswan, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaise, I. Polosukhin, Attention is all you need, 2017,
[55] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[56] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J.A. Rogers, A.K. Katsaggelos, N. Maglaveras, Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets, Biomed. Signal Process. Control. 63 (2021) 102194, http://dx.doi.org/10.1016/j.bspc.2020.102194.

[57] A. Zhang, Z.C. Lipton, M. Li, A.J. Smola, Long Short-Term Memory (LSTM). URL: https://d2l.ai/chapter_recurrent-modern/lstm.html.
[58] R.B. Berry, R. Budhiraja, D.J. Gottlieb, D. Gozal, C. Iber, V.K. Kapur, C.L. Marcus, R. Mehra, S. Parthasarathy, S.F. Quan, et al., Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine, J. Clin. Sleep Med. 8 (5) (2012) 597–619.
[59] X. Chen, Y. Chen, W. Ma, X. Fan, Y. Li, 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, BIBM, 2021, pp. 1276–1280, http://dx.doi.org/10.1109/BIBM52615.2021.9669358.
[60] 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 22 (2022) http://dx.doi.org/10.3390/s22020510.
[61] H. Fang, C. Lu, F. Hong, W. Jiang, T. Wang, Sleep apnea detection based on multi-scale residual network, Life 12 (2022) http://dx.doi.org/10.3390/life12010119.
[62] 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) http://dx.doi.org/10.1109/TIM.2022.3151947.
[63] M. Mohan, G.N. Veena, U.S. Pavitha, H.C. Vinod, Analysis of ECG data to detect sleep apnea using deep learning, J. Surv. Fish. Sci. 10 (2023) 371–376.
[64] A.K. Abasi, M. Aloqaily, M. Guizani, Bare-bones based honey badger algorithm of CNN for Sleep Apnea detection, Clust. Comput. (2024) http://dx.doi.org/10.1007/s10586-024-04309-6.
[65] D.T. Pham, R. Mouček, Sleep apnea detection from single-lead ECG signal using hybrid deep CNN, in: International Conference on Brain Informatics, Springer, 2025, pp. 110–120, http://dx.doi.org/10.1007/978-981-96-3294-7_9.
[66] 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.