Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Wavelet transform and deep learning-based obstructive sleep apnea detection from single-lead ECG signals

Yuxing Lin
Xiamen University of Technology
Hongyi Zhang
Xiamen University of Technology
Wanqing Wu
Sun Yat-Sen University
Xingen Gao (2021000069@xmut.edu.cn)
Xiamen University of Technology https://orcid.org/0009-0000-7385-5825
Fei Chao
Xiamen University
Juqiang Lin
Xiamen University of Technology

Research Article

Keywords: Sleep apnea, ECG, Wavelet transform, Data balancing, Deep Learning

Posted Date: June 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3028322/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Physical and Engineering Sciences in Medicine on November 20th, 2023. See the published version at https://doi.org/10.1007/s13246-023-01346-0.

Wavelet transform and deep learning-based obstructive sleep apnea detection from single-lead ECG signals

Yuxing Lin[1], Hongyi Zhang[1*], Wanqing Wu[2], Xingen Gao[1*], Fei Chao[2], Juqiang Lin[1]

[1*]School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen, 361024, China.
[2]School of Biomedical Engineering, Sun Yat-Sen University, Guangzhou, 510006, China.
[3]School of Informatics, Xiamen University, Xiamen, 36100, China.

*Corresponding author(s). E-mail(s): zhanghongyi@xmut.edu.cn; 2021000069@xmut.edu.cn;

Abstract

Sleep apnea is a common sleep disorder. To address the characteristics of ECG signals, we introduce a coordinate attention mechanism and propose an automatic sleep apnea classification model (CA-EfficientNet) based on wavelet transform and lightweight neural network. One-dimensional signals were converted into two-dimensional images by wavelet transform and in put into the proposed model for classification. The effects of input time window, wavelet transform type and data balance on classification performance were considered. PhysioNet apnea ECG database was used for training and evaluation. The 3-minute Frequency B-Spline wavelets transform of ECG signal was carried out, and Dice Loss was used to train the classification model of sleep breathing. The classification accuracy was 93.44%, precision was 93.2%, sensitivity was 88.9%, specificity was 96.2%, F1 score was 91%, and most indexes were better than other related work. Wavelet transform and CA-EfficientNet model provide a feasible diagnostic method for automatic classification of sleep apnea.

Keywords: Sleep apnea,ECG,Wavelet transform,Data balancing,Deep Learning

1 Introduction

Sleep apnea is a common sleep breathing disorder characterized by repeated pauses and restarts in breathing during sleep, which includes three types of sleep apnea: Obstructive Sleep Apnea (OSA), Central Sleep Apnea (CSA), and combined sleep apnea [1, 2]. Sleep apnea can negatively impact a person's health and cause a variety of symptoms, such as excessive daytime sleepiness, snoring, and concentration difficulties. It also has been associated with an increased risk of certain health conditions, including high blood pressure, cardiac disease, and stroke. Thus, accurate sleep apnea detection is crucial to diagnosis and therapy.

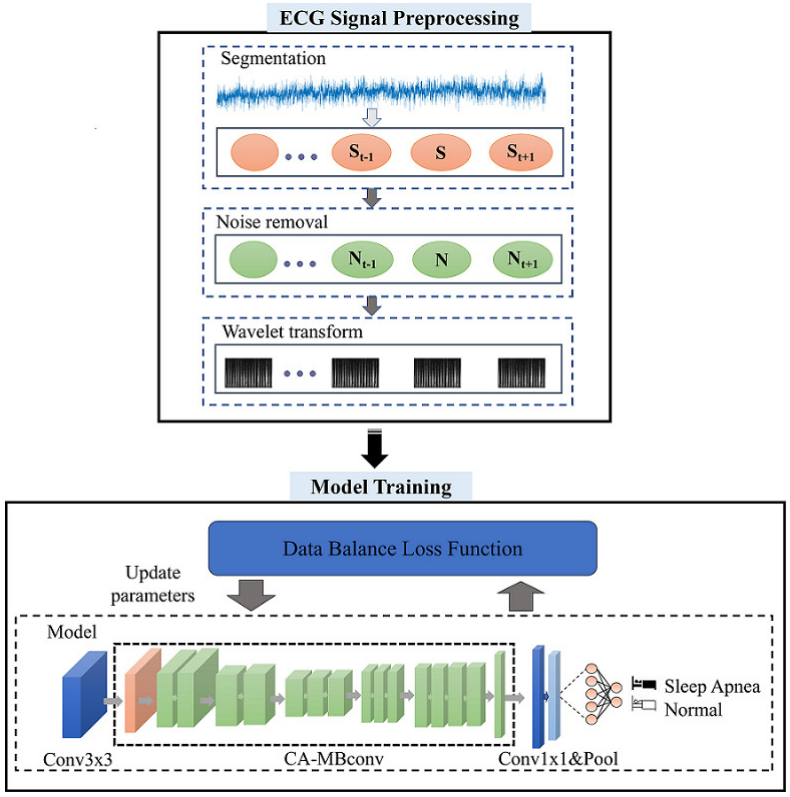
Polysomnography (PSG) is the "gold standard" for diagnosing sleep apnea disorders, such as sleep apnea, narcolepsy, and restless leg syndrome, and is applicable to individuals of all ages. PSG involves attaching multiple sensors to a patient's body during sleep and recording the electroencephalogram (EEG) signals, the blood oxygen level, and the electrocardiogram (ECG) signals. It is a relatively costly and time-consuming procedure that requires multiple specialized pieces of equipment and trained personnel to set up and monitor the patient overnight. In addition, patients may find it difficult or uncomfortable to sleep with sensors attached to their bodies, which may impact the reliability of the results. Previous studies have shown that ECG signals directly reflect the respiratory and circulatory systems of humans and have a strong correlation with sleep apnea. Moridani et al. [3] showed that sleep apnea can lead to irregular cardiac activity. Guilleminault et al. [4] demonstrated that sleep apnea syndrome is associated with alterations in the R-R interval of the ECG signal and hypoxia. ECG is a non-invasive, fast, and painless test that takes only a few minutes to perform and is completely painless. Moreover, ECG is a relatively inexpensive diagnostic tool compared to other diagnostic tests. ECG signals have been used in numerous works to classify sleep apnea. And these works are primarily divided into traditional machine learning methods and deep learning methods.

Some works on automatic sleep apnea detection relied primarily on traditional machine learning, which employed a predefined set of features to train different classifiers. Bsoul et al. [5] developed a real-time method for detecting apnea and hypoventilation episodes using 18 features extracted from R-R intervals in the time and frequency domains. Hassan et al. [6] used statistical analysis for selecting features and Bootstrap aggregating for classification. Song et al. [7] extracted time and frequency features from ECG and respiratory signals and designed a classifier that combines support vector machine (SVM) and a hidden Markov model that takes into account the time dependence of apnea. Babaeizadeh et al. [8] performed heart rate variability (HRV) analysis to detect heart rate change features and incorporated them into classification algorithms for apnea. Additionally, Achmad Rizal et al. [9] extracted HRV features from ECG signals but used SVM for OSA classification. Zarei and Asl [10] proposed a framework for the automatic detection of OSA in which alphabet entropy, fuzzy/approximate, and sample entropy are calculated from ECG-derived respiration (EDR) and HRV signals. Tripath [11] extracted features from the intrinsic band functions of EDR and HRV signals and used a kernel limit learning machine for sleep apnea classification. Fatimah et al. [12] used Fourier decomposition to transform ECG signals into frequency bands and calculated their features, including mean absolute

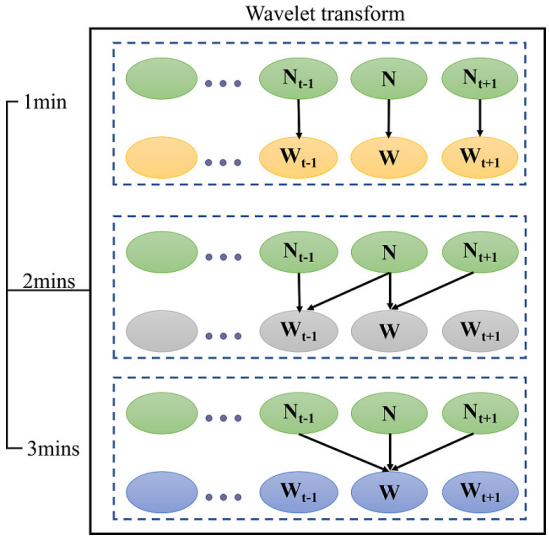
deviation and entropy, to classify ECG segments using an SVM classifier and a Gaussian kernel. Faal et al. [13] extracted ARIMA-EGARCH coefficients and used them as a feature vector to classify apneic and normal ECG segments. Using only eight features, the new ARIMA-EGARCH parameter-based method attained performance comparable to other methods. The aforementioned works utilized feature engineering, which requires a significant amount of trial and error to find the optimal set of features that will improve the performance of a model. In addition, if the features are not carefully selected, it can result in overfitting, where the model performs well on the training data but poorly on the test data.

In recent years, deep learning methods have found wide use in sleep apnea detection. Praveen et al. [14] proposed an approach of cascading two different types of Restricted Boltzmann Machine (RBM) in Deep Belief Networks (DBN) method for the sleep apnea classification. Note that this method needs to extract HRV and EDR signals from a 1-minute segmented ECG signal. Zarei et al. [15] developed an automatic feature extraction method by combining the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) recurrent networks. Additionally, fully connected layers are used to distinguish apnea events from normal segments. Haifa et al. [16] also developed a hybrid model comprised of CNN and LSTM networks, and used the PhysioNet Apnea-ECG database for training and evaluating. Jiang et al. [17] used Mel-spectrogram to extract deep spectral features from CNNs-LSTMs-DNNs for snoring determination, thus reflecting the severity of obstructive sleep apnea hypoventilation syndrome in patients. Singh et al. [18] generated two-dimensional images from single-lead ECG signals using continuous wavelet transform (CWT) and classified them with CNN, achieving good accuracy and sensitivity in per-minute segment OSA classification and demonstrating that extracting time-frequency features can improve sleep apnea detection performance. Mashrur et al. [19] proposed a novel convolutional neural network based on scale maps for apnea detection by converting single-lead ECG signals to hybrid scale maps through CWT and empirical modal decomposition. Niroshana et al. [20] utilized CWT and short-term Fourier transform (STFT) to fuse images and detect OSA. The fused images provided more information on apnea events when fed to a deep CNN model. Ahmad Ayatollahi et al. [21] filtered 2-second ECG segments and utilized the recursive graph (RP) algorithm to transform them into 2D images with high precision. Liu et al. [22] combined the current and adjacent 1-min ECG segments to form the 3-min contextual input segment and incorporated Transformer structures into CNN to improve the model's performance.

In addition, the number of instances in each category (whether the patient has sleep apnea or not) is not equal in the dataset on sleep apnea. Imbalanced data may result in degraded performance, i.e., a classifier trained on an imbalanced dataset may predict that each instance belongs to the majority class, leading to poor performance on the minority class. To date, however, data imbanlance in sleep apnea has been considered by just a few researchers. Shen et al. [23] combined weighted cross-entropy loss function and hidden Markov model to effectively alleviate the problem of data imbalance and improved the classification accuracy of the classifier. Sharan et al. [24] used a single-lead ECG signal and a one-dimensional residual neural network combined with a weighted cross-entropy loss for end-to-end sleep apnea detection.



(a) System Block Diagram



(b) Wavelet transform with different time lengths

Fig. 1: Overview of the proposed model architec.

In conclusion, despite the advances in sleep apnea detection, there are still some issues that must be resolved. Here are some remaining problems with sleep apnea detection:

1. Previous research has demonstrated that CWT is useful for classifying sleep apnea; however, the optimal mother wavelet for this task has not yet been determined.
2. Previous work has trained models with data of varying time durations; however, the relationship between data length and classification accuracy has not been studied.
3. The majority of sleep apnea studies ignore data imbalance and employ deep learning models directly; however, the optimal data balance approach for this task has not been determined.

To resolve the aforementioned problems, we take into account the ECG data length, wavelet transform type, data balance, and classification model. The following is a summary of the main contributions of this work:

1. Different mother wavelets were taken to generate different lengths of time-frequency maps, from which the most suitable mother wavelet and data length for the sleep apnea classification task were identified.
2. Introduce CA attention mechanism to obtain temporal information and improve the accuracy of EfficientNet.
3. Introduce a cost-sensitive algorithm to address the data imbalance problem and compare the effects of different loss functions on the model generalization performance.

2 Methodology

2.1 System overview

Using wavelet transform and CA-EfficientNet model, we developed in this work an automatic sleep apnea classification model based on ECG signals. The schematic of the procedure used to detect sleep apnea is shown in Figure 1. In the preprocessing process, the raw ECG signal is segmented and denoised, and then the ECG signal is transformed into 1-minute, 2-minute, and 3-minute time-frequency images by wavelet transform. CA-EfficientNet is automatically trained to extract features from the time-frequency maps and uses a coordinate attention mechanism to capture the temporal correlation of sleep time data. In addition, a cost-sensitive algorithm (data balance loss function) is introduced, which assigns weights based on the cost associated with each error and is used to calculate the overall loss of the model. The goal of the cost-sensitive loss function is to optimize the performance of the model by minimizing the total cost of error classification, rather than simply minimizing the number of errors. These components are explored further in the following sections.

2.2 Wavelet transform

Wavelet transform is a technique for time-frequency analysis that can extract the local characteristics of a signal in the time-frequency domain. The wavelet transform has the

property of multi-resolution analysis, which decomposes the signal into sub-signals of different frequencies to make the local characteristics of the signal more obvious, and it can be adapted to different types of signals by selecting different wavelet basis functions (mother wavelets), which is particularly useful for extracting characteristic information from non-stationary ECG signals. This paper therefore compares the effects of various wavelet bases on the time-frequency analysis of ECG signals. This work utilizes the following wavelet basis functions.

(1) Morlet wavelet (Morl) is a single frequency sinusoidal function under Gaussian envelope, which can effectively avoid phase distortion in image processing [25]. The mathematical expression is as follows:

$$\Psi(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} \left(e^{-\frac{t^2}{2}} - e^{-\frac{\omega_0^2}{2}} \right) \tag{1}$$

where ω_0 is a positive real number that controls the center frequency of the wavelet.

(2) Complex Morlet wavelet (Cmor) is a complex continuous wavelet, which is extended by Morlet wavelet and has good resolution in both time and frequency domains [26]. Its mathematical form is:

$$\Psi(t) = \pi^{-1/4} e^{i\omega_0 t} \left(e^{-\frac{t^2}{2}} - e^{-\frac{\omega_0^2}{2}} \right) \tag{2}$$

(3) Mexican hat wavelet (Mexh) is a symmetric, non-tightly supported non-orthogonal wavelet [27], which more closely resembles the spatial response properties of human eye vision. The mathematical expression is as follows:

$$\psi(t) = \frac{2}{\sqrt{3a}\pi^{1/4}} \left(1 - \frac{t^2}{a^2} \right) e^{-\frac{t^2}{2a^2}} \tag{3}$$

where a is a positive real number that controls the width of the wavelet.

(4) Shannon wavelet (Shan) is a complete orthogonal wavelet [28]. Since the Shannon wavelet is well-localized in both the time and frequency domains, it is ideally suited for analysing signals with both temporal and spectral characteristics. Its mathematical form is:

$$\psi[n] = \begin{cases} \sqrt{2} & n = 0 \\ \sqrt{2} \sin\left[\frac{\pi n}{2}\right] & n \neq 0 \end{cases} \tag{4}$$

(5) Frequency B-Spline wavelet (Fbsp) is a wavelet based on B-sample function [29], which has good smoothness and phase, and can handle signals such as Gaussian noise very well. Its mathematical form is:

$$\Psi(w) = \frac{1}{\sqrt{2}} \left(\frac{\sin(\frac{\pi w}{k})}{\frac{\pi w}{k}} \right)^k, \tag{5}$$

where k is a positive integer that controls the frequency resolution of the wavelet. During data preprocessing, the ECG signal is converted to a 2D image using wavelet

transform and fed into a deep convolutional neural network for classification. In the next section, the model used in this study is described.

2.3 Apnea classification model

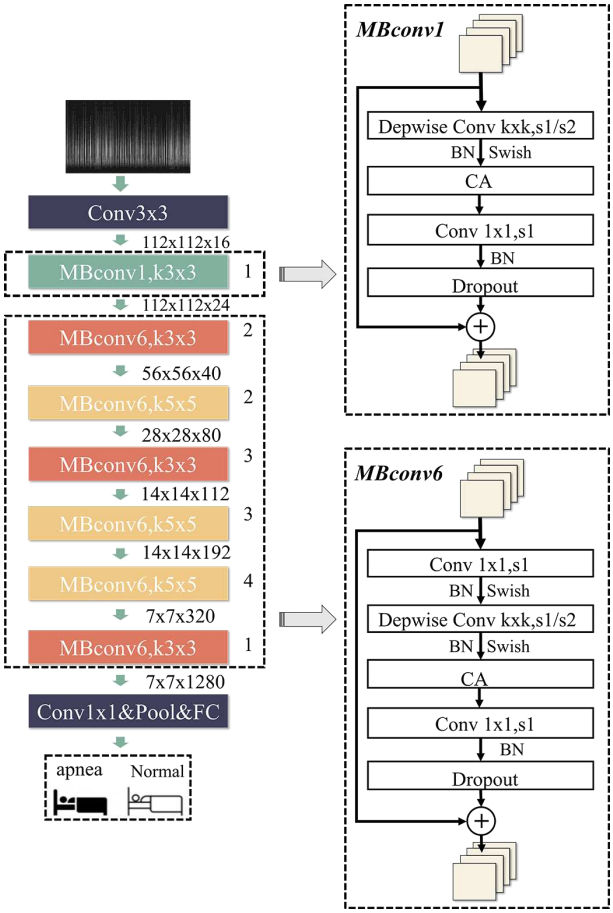


Fig. 2: CA-EfficientNet.

This section presents an improved CA-EfficientNet model based on EfficientNetB0 [30].To improve the accuracy of the model, a coordinated attention mechanism (CA) [31] is introduced. we replace the squeeze and excite (SE) module in MBConv with a CA module that captures remote dependencies and preserves location information. Figure 2 depicts the CA-EfficientNet network diagram, and the components are described in detail below.

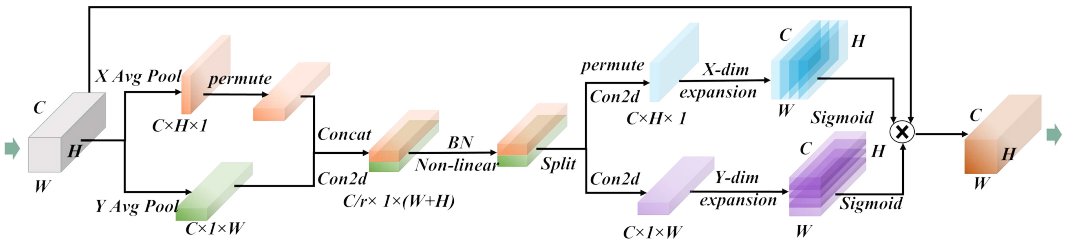


Fig. 3: Coordinate attention structure diagram.

2.3.1 EfficientNet

EfficientNet [30] is a family of convolutional neural networks that achieves high accuracy in image classification tasks while being more computationally efficient than comparable models. To optimize accuracy and efficiency, they are based on a method of compound scaling that balances network depth, width, and resolution by composite coefficients ϕ, which are denoted as

$$\begin{cases} \text{depth: } d = \alpha^{\phi} \\ \text{width: } w = \beta^{\phi} \\ \text{resolution:r } = \gamma^{\phi} \end{cases} \text{ s.t. } \begin{cases} \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ \alpha \geqslant 1, \beta \geqslant 1, \gamma \geqslant 1 \end{cases} \tag{6}$$

where α,β and γ are constants that can be determined by a small grid search. ϕ is a composite factor that controls how many resources are available for model scaling, while α,β and γ specify how these additional resources are allocated to network width, depth, and resolution, respectively.

EfficientNet is stacked by Mobile Inverted Bottleneck Convolution (MBConv). In order to achieve a good classification effect, EfficientNet-B0 is chosen as the base network in this paper.

2.3.2 Coordinate Attention Module

Coordinate Attention (CA) mechanism [31] is a self-attention mechanism that encodes both spatial and channel information. As shown in Figure 3, it divides channel attention into two parallel one-dimensional feature encoding processes and incorporates spatial coordinate information into the feature maps. The coordinate attention mechanism can capture long-range dependencies along one spatial direction, while retaining precise location information along another spatial direction. The expressions of CA

are given by

$$\begin{cases} z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \\ z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \\ \mathbf{f} = \delta\left(F_1\left(\left[\mathbf{z}^h, \mathbf{z}^w\right]\right)\right) \\ \mathbf{g}^h = \sigma\left(F_h\left(\mathbf{f}^h\right)\right) \\ \mathbf{g}^w = \sigma\left(F_w\left(\mathbf{f}^w\right)\right) \\ y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \end{cases} \tag{7}$$

The coordinate attention first encodes the input feature map of size $C \times H \times W$ using global average pooling (GAP) with a pooling kernel of size $(H, 1)$ or $(1, W)$ along the horizontal and vertical coordinates, respectively, and decomposes the channel attention into two one-dimensional feature encodings to generate feature maps of size $C \times H \times 1$ and $C \times 1 \times W$, respectively. The location information is embedded in the channel attention, and the concatenate operation is performed on $z_c^h(h)$ and $z_c^w(w)$ feature maps, followed by the F_1 operation (using 1×1 convolution for dimensionality reduction to obtain two tensors with the same number of channels), and then nonlinear activation operation is performed to generate a feature map f of size $C/r \times 1 \times (W+H)$. Along the spatial dimension, the split operation is performed on f to generate the feature maps f^w of size $C/r \times 1 \times W$ and f^h of size $C/r \times 1 \times H$. Then f^w and f^h are are both expanded to the same size as the input feature map by using 1×1 convolution. Using the sigmoid activation function, the attention weight in two directions is then determined. And finally the output of $y_c(i, j)$ is obtained.

2.4 Loss function

In this paper, cost-sensitive algorithms are introduced to alleviate the problem of data imbalance. The selection of an appropriate loss function is crucial, therefore, this paper compares the impact of different loss functions on model performance.

(1) Cross entropy loss is a loss function used in machine learning to quantify the difference between a classification problem's predicted probability distribution and its actual probability distribution. It is computed by taking the negative logarithm of the predicted likelihood of the true class label. The formula is as follows:

$$L_{\mathrm{CE}}(t_i, p_i) = -\sum_{i=1}^{n} t_i \log\left(p_i\right), \text{ for n classes} \tag{8}$$

where t_i is the truth label and p_i is the softmax probability for the i^{th} class.

(2) Label smoothing cross entropy [32] is an extension of the standard cross-entropy loss that tries to enhance the generalization performance of the model by reducing overfitting and label noise sensitivity. It is defined as:

$$\begin{cases} q'(k) = (1 - \epsilon) q(k) + \frac{\epsilon}{K} \\ L_{\mathrm{LSCE}} = -\sum_{k=1}^{K} \log p(k) q'(k) \end{cases} \tag{9}$$

where ϵ is a weight factor, $\epsilon \in [0, 1]$, and k is the number of label categories.

(3) Focal loss [33] dynamic adjustment loss function is to dynamically adjust the cross-entropy loss according to the confidence level, and to add a moderating factor to reduce the weight of easy-to-classify samples and focus on the training of difficult samples on the basis of the balanced cross-entropy loss function, which is given by:

$$L_{Focal}\left(p_{\mathrm{t}}\right)=-\alpha_{\mathrm{t}}\left(1-p_{\mathrm{t}}\right)^{\gamma}\log\left(p_{\mathrm{t}}\right) \tag{10}$$

where p_t denotes the probability of target label prediction, γ is modulating factor, and α is a weight that handles the category imbalance.

(4) Dice loss [34] can also be utilised when the classes are unbalanced. It measures the dissimilarity between the predicted and ground truth labels using the Dice similarity coefficient, which is defined as follows:

$$L_{Dice}(y_i,\hat{y}_i)=1-\frac{2\sum_{i=1}^{N}y_i\hat{y}_i}{\sum_{i=1}^{N}y_i+\sum_{i=1}^{N}\hat{y}_i} \tag{11}$$

where $y_i$and $\hat{y}_i$ denote the labeled and predicted values of pixels, respectively, and N is the total number of pixel points, which is equal to the number of pixels in a single image multiplied by the batch size.

(5) Poly loss [35] allows the importance of different polynomial bases to be easily adjusted depending on the targeting tasks and datasets. The key to Poly loss is the decomposition of the loss function by a Taylor expansion into a series of weighted polynomial bases, which is given by:

$$L_{\mathrm{Poly}}=\sum_{j=1}^{\infty}\alpha_j\left(1-p_t\right)^j, \tag{12}$$

where α_j denotes the weight of polynomial loss, p_t denotes the probability of target label prediction, and γ is the modulation factor. Poly loss provides simple and effective Poly-1 loss functions to improve the commonly used cross-entropy loss and focal loss, respectively, which are given by:

$$L_{\mathrm{Poly}-1}^{\mathrm{CE}}=L_{\mathrm{CE}}+\epsilon_1(1-p_t), \tag{13}$$

$$L_{\mathrm{Poly}-1}^{\mathrm{Focal}}=L_{\mathrm{Focal}}+\epsilon_1(1-p_t)^{\gamma+1}, \tag{14}$$

where ϵ_1 is the perturbation coefficient of the first polynomial term.

3 Results

In this section, we report the experiments and results used to evaluate the proposed approach on the Apnea-ECG dataset. The experimental dataset, training setup, and model evaluation metrics are first described, followed by the experimental results.

3.1 Apnea-ECG data set and pre-processing

PhysioNet Apnea-ECG database [36] served as the source for the dataset used in this study. There are 70 records in the database, including a learning set of 35 records (a01 to a20, b01 to b05, and c01 to c10) and a test set of 35 records (x01 to x35). These records ranged in duration from less than 7 hours to nearly 10 hours. Each record contains a continuous single-lead ECG signal with a 100 Hz sampling rate.ECG signals often contain various types of noise that can affect the accuracy of the signal analysis. Therefore, we first splited the original continuous ECG signals into 1-minute ECG signal segments and then filtered the noise using first-order Butterworth filtering and p-smooth wavelets. Finally, three time-frequency images for 1, 2 and 3 minutes were generated by wavelet transform. For the 2-minute ECG segments, the current 1-minute segment is combined with the following 1-minute segment to create a complete sample. For the 3-minute ECG segments, same as Liu et al.Liu et al. [22], the current segment is combined with the adjacent (both previous and following) 1-minute segments to form a complete sample.

3.2 Model training setup

Pytorch was used as the training framework. The models were trained with the learning set of the Apnea-ECG database, which was divided into training and validation sets in a ratio of 8:2. The trained models was evaluated using the test set, from which all metrics were obtained.

During the training process, our models are trained for 30 epochs using Stochastic Gradient Descent (SGD) with a learning rate of 0.01, a batch size of 16, and L2 regularization. Besides, cross-entropy loss was used as the baseline loss function in the experiments for different inputs.

3.3 Model evaluation metrics

In the experiments, the accuracy (ACC), sensitivity (SEN), specificity (SPE), and precision (PRE), as well as the F1 score, were used to assess the efficacy of the models. They are defined follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall/\text{Sensitivity} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Pecall}} \tag{19}$$

where TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative, respectively.

We demonstrate the apnea detection performance under varying classification thresholds using the Receiver Operating Characteristic (ROC) curve, which is a plot of the true positive rate (TPR) versus the false positive rate (FPR) at different classification thresholds. In addition, the area under the ROC curve (AUC) is calculated to evaluate the overall performance of the classifiers.

3.4 Experiments and results for different inputs

In this study, three time intervals of input were evaluated: 1, 2, and 3 minutes. In addition, the wavelet transform can be adapted to various signal types by selecting various wavelet basis functions, which is useful for extracting characteristic information from non-stationary ECG signals. Therefore, different wavelet basis functions (mother wavelets) were used to convert ECG signals into time-frequency maps of different time lengths. To evaluate the effects of distinct wavelet basis functions and input time durations, we trained EfficientNet with the different time-frequency map datasets.

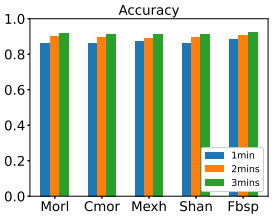
Table 1: Comparisons with different wavelet basis functions and different input durations on test set

Type	Input	ACC(%)	PRE(%)	SEN(%)	SPE(%)	F1(%)	AUC(%)
Morl	1 min	86.28	84.0	79.0	90.8	81.4	84.9
	2 mins	89.84	92.4	79.5	96.0	85.5	87.8
	3 mins	91.63	94.7	82.1	**97.3**	88.0	89.7
Cmor	1 min	86.22	86.4	75.7	92.7	80.7	84.2
	2 mins	89.61	89.5	81.9	94.2	85.5	88.0
	3 mins	91.42	94.5	81.7	97.2	87.6	89.4
Mexh	1 min	87.43	83.4	83.6	89.8	83.5	86.7
	2 mins	88.98	90.8	78.7	95.2	84.3	86.9
	3 mins	91.49	89.5	**87.4**	93.9	88.4	90.7
Shan	1 min	86.2	89.0	72.7	94.5	80.0	83.6
	2 mins	89.27	93.2	77.1	96.6	84.4	86.8
	3 mins	91.26	93.7	82.1	96.7	87.5	89.4
Fbsp	1 min	88.37	84.2	85.4	90.2	84.8	87.8
	2 mins	90.78	91.0	83.7	95.0	87.2	89.4
	3 mins	**92.51**	**94.8**	84.5	**97.3**	**89.4**	**90.9**

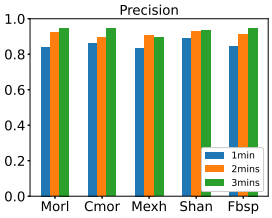
Bold indicates best result (highest value) in each column.

In this experiment, the same Apnea-ECG dataset and model were used. Figure 4 shows the effect and ROC curves of different mother wavelet apnea classifications at one-minute, two-minute and three-minute time input windows, respectively, while Table 1 compares the model performance using different wavelet basis functions to generate temporal frequency maps of different durations.

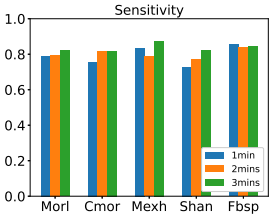
Figures 4 in (a)-(f) show that the overall performance of the model trained on the 3-minute segment data is the best, regardless of the wavelet basis function used.



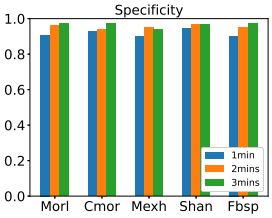
(a) ACC



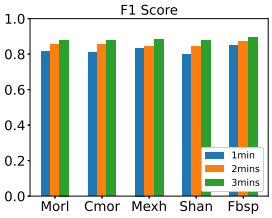
(b) PRE



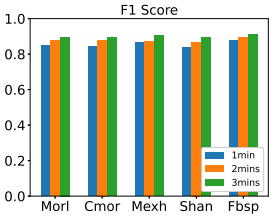
(c) SEN



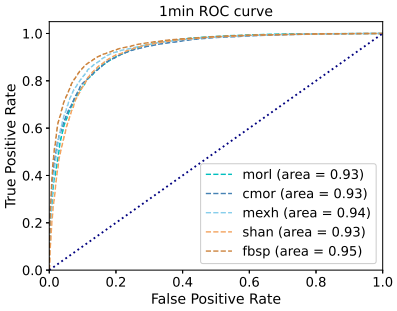
(d) SPE



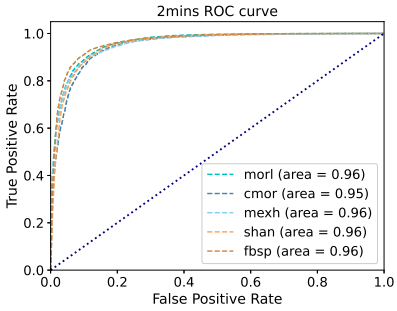
(e) F1



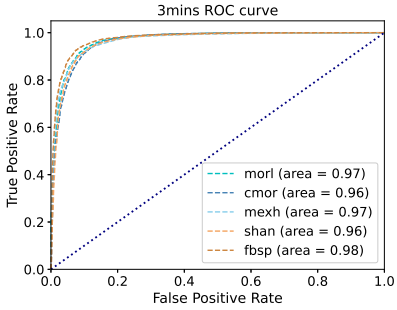
(f) AUC



(g) 1 min ROC curve



(h) 2 mins ROC curve



(i) 3 mins ROC curve

Fig. 4: Experimental results of different input data.

From the ROC curves shown in (g)-(h) in Figure 4 and Table 1, it can be seen that the Shan and Cmor wavelet bases perform relatively poorly and the Fbsp wavelet base performs best on the 1-minute, 2-minute, and 3-minute test sets. Specifically, the Fbsp mother wavelet outperforms the other mother wavelets in four evaluation indexes, including accuracy, sensitivity, F1 score and AUC value, under the 1-minute and 2-minute input time windows, while when the Fbsp mother wavelet is trained on the 3-minute segmentation data, the accuracy is 92. 51%, the accuracy is 94.8%, the sensitivity is 84.5%, the specificity is 97.3%, and the F1 score was 89.4%, and the AUC was 90.9%, which was significantly higher than the 1-minute and 2-minute time windows.

The above results show that for the weakness of poor anti-interference and susceptibility to noise of ECG signal, Fbsp mother wavelet has good smoothness and phase, better adaptability to noise, and superior generalization performance, which is more suitable for sleep apnea classification task. In addition, the impact of time window input length on classification performance is more significant, and longer time windows for input model training can capture more classification features.

3.5 Experimental results of CA-EfficientNet model

To evaluate the effects of the CA module, we replaced the SE module in the MBConv module with the CA module on the basis of EfficientNet-B0, and trained it using cross entropy loss and 3-minute time-frequency maps of ECG signals as inputs. The experimental results are shown in Table 2.

As shown in Table 2, the model with CA module has reached 92.80% accuracy, 86.3% sensitivity, 89.9% F1 score, and 91.5% AUC, which is 0.29%, 1.8%, 0.5%, and 0.6% higher than the original model, respectively. It can be indicated that the CA module had greater attention learning ability than the SE module and that the CA module was better suited for the sleep apnea classification task, proving the model improvement's feasibility and efficacy.

Table 2: Experimental results of CA-EfficientNet model based on 3mins Fbsp wavelet bases

MBconv Type	ACC(%)	PRE(%)	SEN(%)	SPE(%)	F1(%)	AUC(%)
MBconv-SE	92.51	**94.8**	84.5	**97.3**	89.4	90.9
MBconv-CA	**92.80**	93.9	**86.3**	96.7	**89.9**	**91.5**

Bold indicates best result (highest value) in each column.

3.6 Experiments and results for different loss functions

To evaluate the effects of various loss functions on classification performance, the CA-EfficientNet model was trained using various loss functions, including cross entropy loss, label smoothing cross entropy loss, focal loss, poly loss, and dice loss.

Focal loss, Poly-1 focal loss, and Dice loss are loss functions optimized to reduce the model's focus on simple samples. As shown in Table 3, Poly-1 Focal Loss achieved 93% of accuracy, 94.7% of precision, 97.1% of specificity, 90.1% of F1 score, 91.6% of AUC. They are 0.2%, 0.4%, 0.5%, 0.2% and 0.1% higher than the original model respectively, which can slightly improve the effect of data imbalance. Dice loss accuracy was 93.44%, sensitivity was 88.9%, F1 score was 91.0%, and AUC was 92.6%, which were improved by 0.64%, 2.6%, 1.1% and 1.1% compared with the original model, respectively. Compared with poly-1 focal loss, dice loss could better improve the effect of class imbalance. It has higher model robustness.

Table 3: Classification results using different loss functions under CA-EfficientNet model based on 3-minute fbsp wavelet basis function

Loss Function	ACC(%)	PRE(%)	SEN(%)	SPE(%)	F1(%)	AUC(%)
Cross Entropy Loss	92.80	93.9	86.3	96.7	89.9	91.5
Label Smoothing Cross Entropy	92.77	94.2	85.9	96.9	89.9	91.4
Focal Loss	91.06	93.2	82.1	96.4	87.3	89.2
Poly-1 Cross Entropy Loss	92.52	94.4	85.0	97.0	89.5	91.0
Poly-1 Focal Loss	93.0	**94.7**	86.0	**97.1**	90.1	91.6
Dice Loss	**93.44**	93.2	**88.9**	96.2	**91.0**	**92.6**

Bold indicates best result (highest value) in each column.

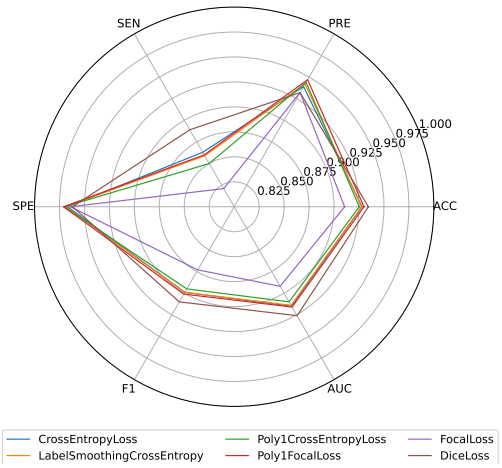


Fig. 5: Performance comparison in different loss functions.

The focus loss is a function designed for unbalanced data sets, but it is intuitively evident from the figure 5 that it is the least effective in detecting sleep apnea compared

to other loss functions because it is difficult to adjust the parameters in practice. Label smoothing loss and multi-1 cross-entropy loss optimize the cross-entropy loss, but their metrics are mostly lower than the original cross-entropy and are not suitable for the classification task of sleep breathing disorders. Experimental results show that the dice loss is the most suitable for the apnea classification task, and it can greatly improve the effect of data imbalance.

To further verify the improvement effect of Diceloss loss on the classification performance of CA-EfficientNet, a 5-fold cross-validation was used, and the box line plots of each evaluation index are shown in Figure 8, which yielded an average accuracy of 92.87%, an average precision of 93.18%, an average sensitivity of 87.64%, an average specificity of 96.18%, an average f1 score of 90.3%, and the average AUC was 91.94%. It can be seen that after 5-fold cross-validation, the accuracy, F1 score and AUC distribution are relatively concentrated, and the precision, sensitivity and specificity are stable within a reasonable range, and the above results indicate that Diceloss is effective and has good robustness.

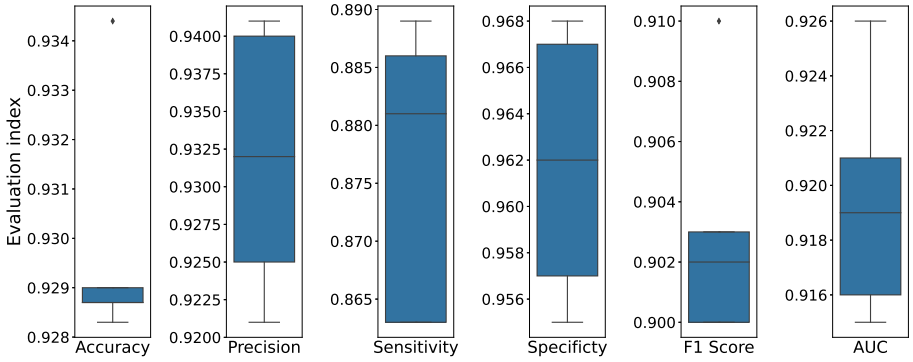


Fig. 6: Five-fold cross validation results of CA-EfficientNet using Diceloss loss.

4 Discussion

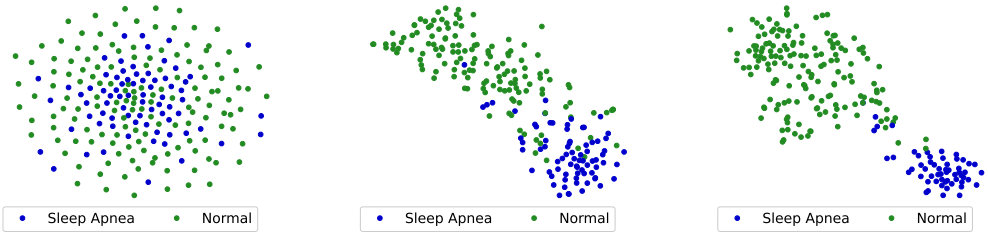
Previous research has demonstrated that classifying sleep apnea is facilitated by converting one-dimensional signals to two-dimensional images using CWT. However, the optimal wavelet basis function for this task has not yet been determined. In addition, the relationship between data length and classification accuracy has not been thoroughly investigated. Moreover, the majority of studies on sleep apnea have neglected to account for data imbalance.

In this study, the influence of different wavelet basis functions, input time windows, and data imbalance loss functions was investigated. We propose an automatic classification model for sleep apnea based on EfficientNet and the CA attention mechanism,

Table 4: Comparison of performances between the proposed method and other methods.

Study	Input type	Classifier	ACC(%)	SEN(%)	SPE(%)
Varon et al. [37]	QRS and HRV	LS-SVM	83.8	79.5	88.4
Li et al. [38]	RR interval	DNN+HMM	84.7	88.9	82.1
Singh et al. [18]	Time-frequency diagram	CNN	86.22	90	83.8
Pombo et al. [39]	HRV and EDR	ANN	82.1	88.4	72.3
Shen et al. [23]	RR interval	CNN	89.4	89.8	89.1
Feng et al. [40]	RR interval	FSSAE+HMM	85.1	86.2	84.4
Haifa et al. [16]	ECG	CNN+LSTM	90.92	**91.24**	90.36
Bahrami et al. [41]	RR interval and R-peak	CNN+BiLSTM	88.1	81.5	92.3
Yeh etal. [42]	Raw ECG	CNN	85.8	80.1	89.4
Liu et al. [22]	Raw ECG	CNN+Transformer	88.2	78.5	94.1
Ours	Time-frequency diagram	CNN	**93.44**	88.9	**96.2**

Bold indicates best result (highest value) in each column.



(a) Raw data (b) EfficientNet (c) CA-EfficientNet

Fig. 7: t-SNE visualization analysis.

which is trained using Dice loss and 3-minute time-frequency maps of ECG signals as inputs.

On the basis of the same dataset, the efficacy of our model is compared to that of other relevant works.As shown in Table 4, our proposed sleep apnea classifier outperforms all other related efforts on most metrics. The accuracy was 93.44% and the specificity was 96.2%. As shown, Figure 7.(a) shows the raw data, in which apnea samples are mixed with normal samples without clear distinction. Figure 7.(b) shows the classification features extracted from the original effentnet model, in which a portion of apnea samples were misclassified as normal samples. Figure 7.(c) visually shows that Dice loss improves the classification performance of the CA-EfficientNet model for sleep apnea samples and enhances the robustness of the model.

The proposed method reduces complex feature processing and computation, takes full advantage of the automatic feature extraction and structural flexibility of neural networks, and effectively alleviates the problem of sleep data imbalance.

There is still room for improvement in this work. Only the single-lead ECG sleep apnea dataset was utilized without a more detailed classification of the type and

severity of sleep apnea. Therefore, future studies could be based on multimodal physiological signals for apnea severity studies and combined with sleep staging to further detect abnormal breathing events during sleep.

5 Conclusion

In this study, we propose an automatic sleep apnea classification model based on CA-EfficientNet, which is trained using Dice loss and uses a 3-minute time-frequency map of the ECG signal as input. Using a 3-minute input time window and the Fbsp wavelet basis function, the proposed method achieves an accuracy of 93.44%. This work takes into account the input time window, type of wavelet transform, and data balance. Experimental results indicate that the Fbsp wavelet basis function is the most appropriate for sleep apnea classification, that the best results are obtained for 3-minute ECG signals, and that the use of Dice loss can address the imbalance problem and significantly enhance the classification performance of the model.

Statements and Declarations

5.1 Funding

This work was supported in part by the High-level Talent Project of Xiamen University of Technology (No. YKJ22028R and YKJ22029R).

5.2 Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

References

[1] Khandoker, A.H., Palaniswami, M.: Modeling respiratory movement signals during central and obstructive sleep apnea events using electrocardiogram. ANNALS OF BIOMEDICAL ENGINEERING **39**(2), 801–811 (2011) https://doi.org/10.1007/s10439-010-0189-x

[2] Guijarro-Berdinas, B., Hernandez-Pereira, E., Peteiro-Barral, D.: A mixture of experts for classifying sleep apneas. EXPERT SYSTEMS WITH APPLICATIONS **39**(8), 7084–7092 (2012) https://doi.org/10.1016/j.eswa.2012.01.037

[3] Moridani, M.K., Heydar, M., Jabbari Behnam, S.S.: A reliable algorithm based on combination of emg, ecg and eeg signals for sleep apnea detection : (a reliable algorithm for sleep apnea detection). In: 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), pp. 256–262 (2019). https://doi.org/10.1109/KBEI.2019.8734992

[4] Guilleminault, C., Winkle, R., Connolly, S., Melvin, K., Tilkian, A.: Cyclical variation of the heart rate in sleep apnoea syndrome: Mechanisms, and usefulness of 24 h electrocardiography as a screening technique. The Lancet **323**(8369), 126–131

(1984) https://doi.org/10.1016/S0140-6736(84)90062-X . Originally published as Volume 1, Issue 8369

[5] Bsoul, M., Minn, H., Tamil, L.: Apnea medassist: Real-time sleep apnea monitor using single-lead ecg. IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE **15**(3), 416–427 (2011) https://doi.org/10.1109/TITB.2010.2087386

[6] Hassan, A.R., Haque, M.A.: Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating. BIOCYBERNETICS AND BIOMEDICAL ENGINEERING **36**(1), 256–266 (2016) https://doi.org/10.1016/j.bbe.2015.11.003

[7] Song, C., Liu, K., Zhang, X., Chen, L., Xian, X.: An obstructive sleep apnea detection approach using a discriminative hidden markov model from ecg signals. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING **63**(7), 1532–1542 (2016) https://doi.org/10.1109/TBME.2015.2498199

[8] Babaeizadeh, S., White, D.P., Pittman, S.D., Zhou, S.H.: Automatic detection and quantification of sleep apnea using heart rate variability. Journal of Electrocardiology **43**(6), 535–541 (2010) https://doi.org/10.1016/j.jelectrocard.2010.07.003

[9] Rizal, A., Siregar, F.D.A.A., Fauzi, H.T.: Obstructive sleep apnea (osa) classification based on heart rate variability (hrv) on electrocardiogram (ecg) signal using support vector machine (svm). TRAITEMENT DU SIGNAL **39**(2), 469–474 (2022) https://doi.org/10.18280/ts.390208

[10] Zarei, A., Asl, B.M.: Automatic classification of apnea and normal subjects using new features extracted from hrv and ecg-derived respiration signals. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **59** (2020) https://doi.org/10.1016/j.bspc.2020.101927

[11] Tripathy, R.K.: Application of intrinsic band function technique for automated detection of sleep apnea using hrv and edr signals. BIOCYBERNETICS AND BIOMEDICAL ENGINEERING **38**(1), 136–144 (2018) https://doi.org/10.1016/j.bbe.2017.11.003

[12] Fatimah, B., Singh, P., Singhal, A., Pachori, R.B.: Detection of apnea events from ecg segments using fourier decomposition method. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **61** (2020) https://doi.org/10.1016/j.bspc.2020.102005

[13] Faal, M., Almasganj, F.: Obstructive sleep apnea screening from unprocessed ecg signals using statistical modelling. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **68** (2021) https://doi.org/10.1016/j.bspc.2021.102685

[14] Tyagi, P.K., Agrawal, D.: Automatic detection of sleep apnea from single-lead ecg signal using enhanced-deep belief network model. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **80**(2) (2023) https://doi.org/10.1016/j.bspc.2022.104401

[15] Zarei, A., Beheshti, H., Asl, B.M.: Detection of sleep apnea using deep neural networks and single-lead ecg signals. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **71**(A) (2022) https://doi.org/10.1016/j.bspc.2021.103125

[16] Almutairi, H., Hassan, G.M., Datta, A.: Classification of obstructive sleep apnoea from single-lead ecg signals using convolutional neural and long short term memory networks. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **69** (2021) https://doi.org/10.1016/j.bspc.2021.102906

[17] Jiang, Y., Peng, J., Zhang, X.: Automatic snoring sounds detection from sleep sounds based on deep learning. PHYSICAL AND ENGINEERING SCIENCES IN MEDICINE **43**(2), 679–689 (2020) https://doi.org/10.1007/s13246-020-00876-1

[18] Singh, S.A., Majumder, S.: A novel approach osa detection using single-lead ecg scalogram based on deep neural network. JOURNAL OF MECHANICS IN MEDICINE AND BIOLOGY **19**(4) (2019) https://doi.org/10.1142/S021951941950026X

[19] Mashrur, F.R., Islam, M.S., Saha, D.K., Islam, S.M.R., Moni, M.A.: Scnn: Scalogram-based convolutional neural network to detect obstructive sleep apnea using single-lead electrocardiogram signals. COMPUTERS IN BIOLOGY AND MEDICINE **134** (2021) https://doi.org/10.1016/j.compbiomed.2021.104532

[20] Niroshana, S.M.I., Zhu, X., Nakamura, K., Chen, W.: A fused-image-based approach to detect obstructive sleep apnea using a single-lead ecg and a 2d convolutional neural network. PLOS ONE **16**(4) (2021) https://doi.org/10.1371/journal.pone.0250618

[21] Ayatollahi, A., Afrakhteh, S., Soltani, F., Saleh, E.: Sleep apnea detection from ecg signal using deep cnn-based structures. EVOLVING SYSTEMS **14**(2), 191–206 (2023) https://doi.org/10.1007/s12530-022-09445-1

[22] Liu, H., Cui, S., Zhao, X., Cong, F.: Detection of obstructive sleep apnea from single-channel ecg signals using a cnn-transformer architecture. BIOMEDICAL SIGNAL PROCESSING AND CONTROL **82** (2023) https://doi.org/10.1016/j.bspc.2023.104581

[23] Shen, Q., Qin, H., Wei, K., Liu, G.: 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 (2021) https://doi.org/10.1109/TIM.2021.3062414

[24] Sharan, R.V., Berkovsky, S., Xiong, H., Coiera, E.: End-to-end sleep apnea detection using single-lead ecg signal and 1-d residual neural networks. JOURNAL OF MEDICAL AND BIOLOGICAL ENGINEERING **41**(5, SI), 758–766 (2021) https://doi.org/10.1007/s40846-021-00646-8

[25] Morlet, J.: Wave propagation and sampling theory. Geophysics **47**, 203–236 (1982)

[26] Flandrin, P., Rilling, G., Goncalves, P.: Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters **11**(2), 112–114 (2004) https://doi.org/10.1109/LSP.2003.821662

[27] Marr, D., Hildreth, E.: Theory of edge detection. Proceedings of the Royal Society of London **207**(1167), 187–217 (1980)

[28] Shannon, C.E.: Communication in the presence of noise. Proceedings of the IRE **86**(1), 10–21 (1949)

[29] Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing. part i. theory. IEEE Transactions on Signal Processing **41**(2), 821–833 (1993)

[30] Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. CoRR **abs/1905.11946** (2019) 1905.11946

[31] Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. CoRR **abs/2103.02907** (2021) 2103.02907

[32] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. CoRR **abs/1512.00567** (2015)

[33] Lin, T., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. CoRR **abs/1708.02002** (2017)

[34] Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J.: Dice loss for data-imbalanced NLP tasks. CoRR **abs/1911.02855** (2019)

[35] Leng, Z., Tan, M., Liu, C., Cubuk, E.D., Shi, X., Cheng, S., Anguelov, D.: PolyLoss: A Polynomial Expansion Perspective of Classification Loss Functions (2022)

[36] Penzel, T., Moody, G.B., Mark, R.G., Goldberger, A.L., Peter, J.H.: The apnea-ecg database. In: Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163), pp. 255–258 (2000). https://doi.org/10.1109/CIC.2000.898505

[37] Varon, C., Caicedo, A., Testelmans, D., Buyse, B., Van Huffel, S.: A novel algorithm for the automatic detection of sleep apnea from single-lead ecg. IEEE Transactions on Biomedical Engineering **62**(9), 2269–2278 (2015) https://doi.org/10.1109/TBME.2015.2422378

[38] Li, K., Pan, W., Li, Y., Jiang, Q., Liu, G.: A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ecg signal. NEUROCOMPUTING **294**, 94–101 (2018) https://doi.org/10.1016/j.neucom.2018.03.011

[39] Pombo, N., Silva, B.M.C., Pinho, A.M., Garcia, N.: Classifier precision analysis for sleep apnea detection using ecg signals. IEEE Access **8**, 200477–200485 (2020) https://doi.org/10.1109/ACCESS.2020.3036024

[40] Feng, K., Qin, H., Wu, S., Pan, W., Liu, G.: A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. IEEE Transactions on Instrumentation and Measurement **70**, 1–12 (2021) https://doi.org/10.1109/TIM.2020.3017246

[41] Bahrami, M., Forouzanfar, M.: Sleep apnea detection from single-lead ecg: A comprehensive analysis of machine learning and deep learning algorithms. IEEE Transactions on Instrumentation and Measurement **71**, 1–11 (2022) https://doi.org/10.1109/TIM.2022.3151947

[42] Yeh, C.-Y., Chang, H.-Y., Hu, J.-Y., Lin, C.-C.: Contribution of different subbands of ecg in sleep apnea detection evaluated using filter bank decomposition and a convolutional neural network. SENSORS **22**(2) (2022) https://doi.org/10.3390/s22020510