

A Multi-scale Residual Convolutional Neural Network for Sleep Staging Based on Single Channel Electroencephalography Signal

QINGHUA ZHONG (✉ zhongqinghua@m.scnu.edu.cn)

South China Normal University Guangzhou Higher Education Mega Center <https://orcid.org/0000-0001-6985-1610>

Haibo Lei

South China Normal University Guangzhou Higher Education Mega Center

Qianru Chen

South China Normal University

Guofu Zhou

South China Normal University Guangzhou Higher Education Mega Center

Research Article

Keywords: Residual network, Sleep stage, Multiscale, Convolutional neural network, Deep learning, Electroencephalography

Posted Date: June 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-554671/v1>

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

[Read Full License](#)

RESEARCH

A Multi-scale Residual Convolutional Neural Network for Sleep Staging Based on Single Channel Electroencephalography Signal

Qinghua Zhong^{1,2*}, Haibo Lei², Qianru Chen² and Guofu Zhou^{1,3,4}

*Correspondence:

zhongqinghua@m.scnu.edu.cn

¹Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, P. R. China

Full list of author information is available at the end of the article

Abstract

Sleep disorder is a serious public health problem. Non hospital sleep monitoring system for monitoring sleep quality can effectively support the screening of sleep disorder related diseases. A new algorithm of multi-scale residual convolutional neural network (MS-RESCNN) was proposed to discover the feature of electroencephalography (EEG) signals detected with wearable system and staging the sleep stage. EEG signals were analyzed by this algorithm every 30 seconds, and then sleep staging results of wake-up (W), rapid eye movement sleep (REM) and non-rapid eye movement sleep (NREM) were outputted. NREM can also be subdivided into N1, N2 and N3 stages. 5-fold cross validation and independent subject cross validation were performed on the dataset with Kappa coefficients 0.7360 and 0.7001, respectively. The accuracy rates of those methods were 92.06% and 91.13%, respectively. Compared with the other methods, our proposed method can expediently obtain the information of sleep stages from single channel EEG signals without special feature extraction. It has a good performance and can provide support for clinical application based on automatic sleep staging.

Keywords: Residual network; Sleep stage; Multiscale; Convolutional neural network; Deep learning; Electroencephalography

1 Introduction

Sleep plays an important role in human health [1]. The monitoring of human sleep has significant implications for medical research and practice [2]. Sleep specialists usually obtain the quality of sleep by analyzing electrical activity signals from sensors connected to different parts of body in accordance with the Rechtschaffen & Kales Rules [3] or the American Academy of Sleep Medicine (AASM) sleep score manual [4]. In particular, polysomnography (PSG), which records EEG, electrooculogram (EOG), electrocardiograph (ECG), electromyography (EMG), respiratory effort, leg movement, and blood oxygen saturation over several nights in a sleep laboratory, is considered as a gold standard for evaluating sleep status of subjects [5]. In order to improve the efficiency of sleep monitoring, several effective sleep staging methods with multiple physiological parameters based on EEG, ECG and EMG signals have been proposed in recent years [6-10]. However, their signal acquisition mostly adopted the silver/silver chloride electrode with certain adhesive or conductive paste, and fixed in the positions to be tested, which affected the natural sleep of subjects and was not suitable for longterm sleep monitoring in the home environment.

There are two main types of current research on sleep monitoring and staging methods, one is to simplify and optimize the signal acquisition channel, the other is to explore change rules of other physiological signals during sleep [11-20]. Wearable measurements of physiological signals such as ECG, EOG and respiration, as well as body movements and other behaviors, offer the potential of low cost and easy operation for longterm dynamic sleep monitoring [12-15]. Rahman [15] used EOG signals to obtain up to 73% accuracy of sleep staging, but the classification effect of N1 stage still needs to be improved. In recent years, non-invasive or noncontact measurement of cardiac, respiratory and body movement signals has gradually gained the favor of researchers [16-20]. Ballistocardiograph (BCG) measurements based on pressure-sensitive sheets or bedmounted measurements have been used for sleep analysis and preliminary screening for diseases [16-18]. Methods such as video monitoring or microwave radar were also used in the detection of cardiac, respiratory and body movement of patients [19-20]. Of course, their performances mainly depends on the quality of signal acquisition.

Single channel EEG parameters in time domain, frequency domain, time-frequency domain can be used for sleep stage classification [21-22]. Channel selection is important for measurement convenience. Fp1, Fp2 and Fpz below hairline are suitable for wearing and data collection, F3 and F4 channels which near hairline also have operating space. However, these channels whose direct information can support for W, N1, N2, and N3 stages, which are different from most models based on sleep dataset of EEG position in the parietal lobe, and it can affect the credibility of staging results. Deep neural network models such as recurrent neural network (RNN), convolutional neural network (CNN) and deep belief network (DBN) were used to analyze single channel EEG signals, and a good overall average accuracy of sleep staging was obtained [23-29].

Multiscale convolutional neural network (MSCNN) was also proposed to perform multiscale feature extraction and classification simultaneously [30-31]. However, gradient dispersion or gradient explosion is likely to occur with the increase of the depth of CNN. The residual convolutional neural network (RESCNN) proposed by He [32] was aimed to solve the degeneration problem of network. The method has been applied to the machine fault detection and achieved good results [33]. The sleep stage classification based on residual network also achieved good results [34]. But central and occipital EEG from right hemisphere, left and right EOG, and chin EMG channels were extracted from each PSG study.

Considering that the multiscale RESCNN can capture detailed signal features required for pattern classification, the idea was adopted to single channel EEG sleep stage classification for realizing a practical wearable smart eye mask in our study. The research work included following aspects: apply deep MS-RESCNN where no special EEG feature extraction is required; train the network model by using PSG dataset; test the application effect of the new method in sleep stage classification by using cross validation. The new method used single channel EEG signals, dispensed with for special feature extraction process. It has good performance and application potential, which can provide support for clinical applications such as sleep disorder disease screening and diagnosis.

2 Method

Sleep staging is a problem of recognition or regression of time series signals. According to the sleep manual, experts divided the EEG, EOG, EMG and other signals obtained by sensors at intervals of 30 s into W, REM, N1, N2, N3, N4. In this paper, the multi-scale feature learning was integrated into the RESCNN to automatically learn sleep features of original physiological signals at different time scales, and classification in a parallel way from complex original physiological signals of sleep was achieved.

2.1 CNN

CNN is a unique artificial neural network inspired by the cerebral cortex, which uses convolution method to extract signal features and compress signal size. In order to retain valid information, it reduces the amount of input data by several orders of magnitude, which can form a easy network to optimize and reduce the risk of network overfitting. CNN contains two core layers: convolution and pooling. The purpose of convolution is to extract features at different levels from original data, and a certain number of filters are used to extract feature maps of input data. The pooling layer is periodically inserted between CNN's convolutional layers, and a subsampling operation is used to reduce the number of parameters. At the end of the algorithm, the ReLU activation function which can significantly improve the training speed is used.

In a CNN, receptor field is defined as the size of the region mapped on the input image by pixel points on the output feature map in each layer of CNN. In layman's terms, this is the area where the input feature is "seen" by the output feature point. The receptive field is calculated by Equation (1).

$$RF_i = (RF_{i+1} - 1) \times stride_i + kernel_i \quad (1)$$

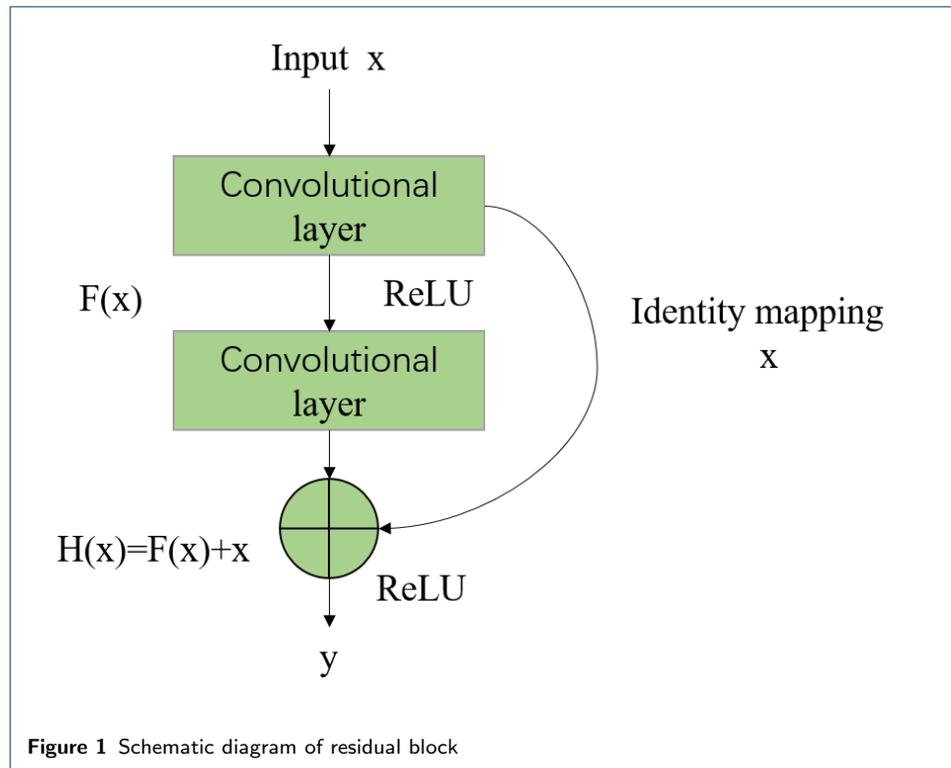
where RF_i is the receptive field of the i -th convolutional layer, RF_{i+1} is the receptive field of the $(i+1)$ -th layer, $stride_i$ is the convolution step size of layer i , $kernel_i$ is the size of the convolution kernel of i -th layer.

2.2 Residual block

In general, the more layers in convolutional neural network, the more diverse features can be extracted. But simply increasing the depth of the network, which can lead to gradient dispersion or gradient explosion. Although the CNN of dozens in layers can be trained by normalized initialization and batch normalization (BN), it is prone to degenerate as the number of layer increases. He [32] proposed the RESCNN to solve the degradation phenomenon: if the rear layer of a certain layer in the deep network is identity mapping, the model would degenerate into a shallow network.

For the network with an input x , the learning feature is denoted as $H(x)$, expected network learning residuals $F(x) = H(x) - x$. As shown in Figure 1, a residual block is formed by neural network with shortcuts connections. It contains two kinds of mapping: one is the identity mapping, which is the shortcut curve in the graph, and the other is the residual mapping. If the residual error is not equal to 0, the network

performance can still be improved by increasing the number of network layers. If the residual is 0, then the current layer is just an identity mapping which neither improves nor degrades, thus the network degradation problem can be solved.



The output y from the shortcut operation is given as Equation (2).

$$y = F(x, W_i) + x \quad (2)$$

where x and y are the input and output vectors of the layer under consideration, $F(x, W_i)$ is the residual mapping function to be learned, and it can be expressed by Equation (3).

$$F = W_2 \sigma(W_1 x) \quad (3)$$

where σ is the ReLU activation function.

2.3 MS-RESCNN

According to the above description, the expression of RESCNN unit constructed can be expressed by Equation (4) – (10).

$$y_1 = W \otimes x + b \quad (4)$$

$$y_2 = BN(y_1) \quad (5)$$

$$y_3 = ReLU(y_2) \quad (6)$$

$$y_4 = W \otimes y_3 + b \quad (7)$$

$$y_5 = BN(y_4) \quad (8)$$

$$y_6 = y_5 + x \quad (9)$$

$$y_7 = ReLU(y_6) \quad (10)$$

where \otimes is convolution operator, BN is a batch normalization operation used to improve generalization capability.

In order to extract features from different receiving domains, we used three RESCNN pathways constructed by multiple RESCNN units as shown in Fig. 2, and time series signal fragments were directly used as inputs of the network. Each path contained four RESCNN units with two convolutional layers and a shortcut. Each convolution layer was followed by the batch normalization and the activation function ReLU. The solid line shortcut meant that it can be added data directly, dashed line shortcuts indicated that they need to be added the 1×1 Conv to the same dimension, the result of each path was averaged by pooling 512 features. Each RESCNN block had a different convolution kernel when the core size was set to 1×3 , 1×5 and 1×7 .

EEG signals inputted into the network was proceeded by the convolution layer with convolution kernel length of 15, and it was transmitted to Batch Normalization and ReLU activation functions for carrying out maximum pooling. Then the output data was sent to three channels of different sizes of the convolution kernel for calculation. Finally, the characteristics of three channels were combined and connected to the full connection layer with 1536 neurons, and the network classification results were obtained by the Softmax function.

The receptive field of each convolution layer can be calculated according to Equation (1). The receptive field of the output characteristic graph of the last convolution layer in the network was 563 in the input data. EEG signals had a sampling frequency of 100 Hz, the effective frequency resolution of 563 data points was 0.35 Hz, which can meet the frequency resolution requirements of all rhythmic waves.

3 Experiments

3.1 Dataset

The dataset used in the experiment was Sleep-EDFX [36], which contains 153 SC* files and 44 ST* files. Records named SC * were EEG signals from healthy subjects during 24 hours of normal life, and records named ST * were signals from subjects with mild sleep disturbances who underwent nighttime sleep tests in hospitals. Signals in two subsets included EOG Fpz-Cz EEG and Pz-Oz EEG, both sampled at 100 Hz. The annotation files included sleep stages W, R, S1, S2, S3, S4, M (Movement time) and UNKNOWN, it consisted of a manual score by a skilled technician. In this study, we cropped the SC * files in the data set, and only signals in the period from 30 minutes before the beginning of sleep to 30 minutes after the end of sleep were retained, data from FPz-Cz channels was used. Stage M and UNKNOWN are deleted for their extremely small percentage. At the same time, according to the latest AASM sleep scoring manual, S3 and S4 were combined into N3.

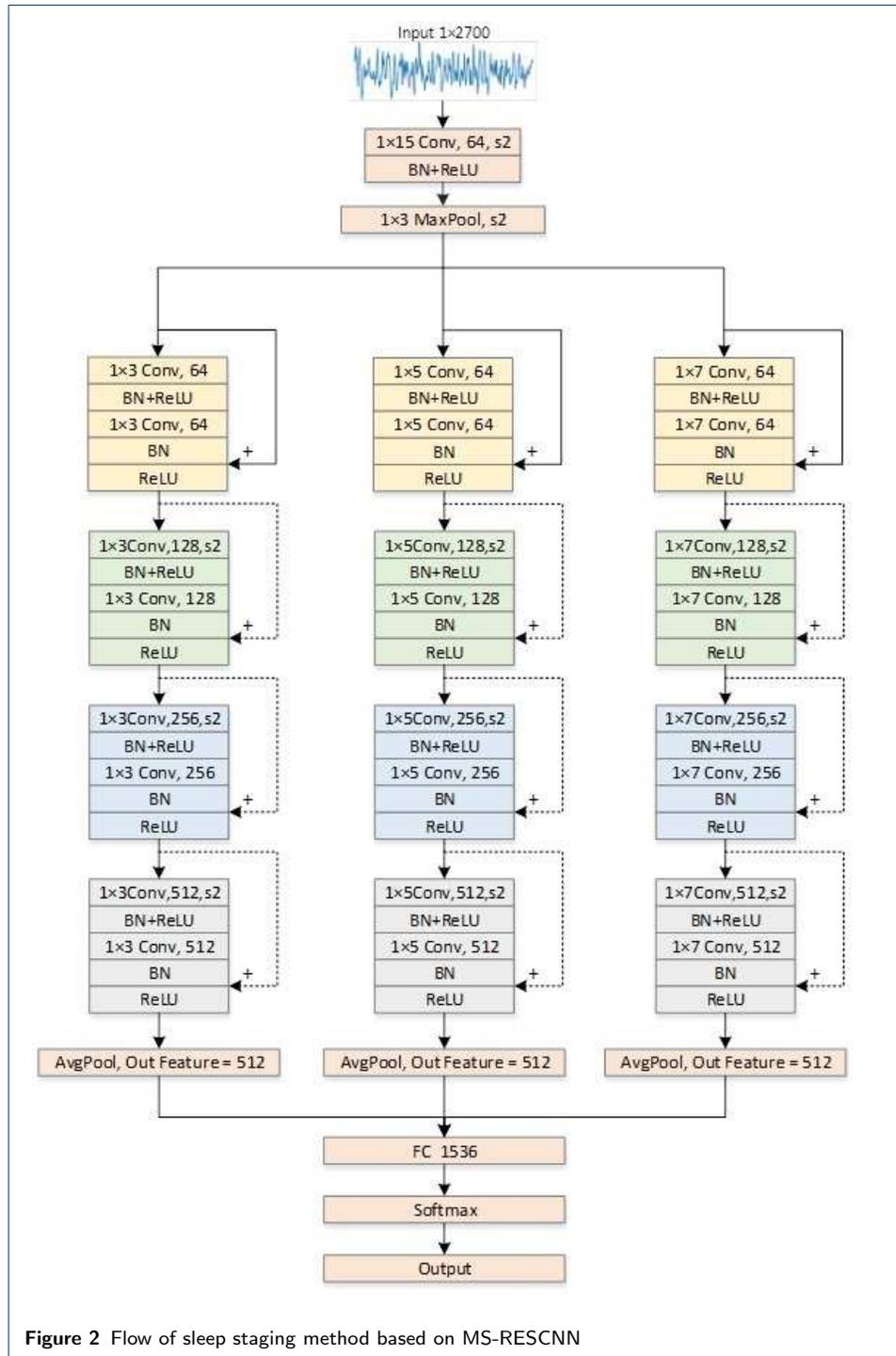


Figure 2 Flow of sleep staging method based on MS-RESCNN

3.2 Preprocessing

In order to adapt to different PSG acquisition devices and individual differences of different subjects, EEG signals of each subject were normalized by using the 5-th and 95-th quantiles [34], as shown in Equation (11).

$$x_{norm} = 2 \frac{x - Q_{0.05}(x)}{Q_{0.95}(x) - Q_{0.05}(x)} - 1 \quad (11)$$

where x_{norm} was the result of signal normalization, x was original signal, $Q_{0.05}(x)$ and $Q_{0.95}(x)$ were the 5% and 95% largest point in the signal, respectively.

In order to expand the dataset and improve the network generalization ability, each input data should be randomly clipped (the 3000 data points in each epoch were randomly clipped to 2700), the flip probability was 50%, and 0.01 times random noise was added. Finally, the preprocessed data were integrated, and then added batch size and channel number, it was converted to tensor data type, as shown in Equation (12).

$$x_{nct} \in R^{N \times C \times T} \quad (12)$$

where $N=16, C=1, T=2700$ represented batch size, channel number, data points of single epoch respectively. In order to input the sleep stages corresponding to the EEG of each Epoch into the network, sleep stages were mapped as Equation (13).

$$N3 \rightarrow 0, N2 \rightarrow 1, N1 \rightarrow 2, REM \rightarrow 3, W \rightarrow 4 \quad (13)$$

3.3 Network training

3.3.1 Loss function

In order to realize the loss calculation of multiclassification of sleep stage, cross entropy (CE) is used as the loss function, its definition is shown as Equation (14).

$$loss(input, target) = weight[target] \times (-input[target] + \log(\sum_j e^{input[j]})) \quad (14)$$

where input is one-dimensional array (the array consists of predicted probability values for each tag) after being processed by Softmax, target is the actual label. $input[j]$ represents the element in the input array whose ordinal number is j . $weight[target]$ is the actual label weight. Since the classification of sleep stages is an unbalanced classification task, this paper tried to balance differences in the amount of data of each label, $weight[target]$ is defined as Equation (15).

$$weight[target] = \ln\left(\frac{1}{p(target)}\right) \quad (15)$$

where $p(target)$ is the proportion of the label target in total label.

3.3.2 Optimizer

For the loss function, the adaptive moment estimation (Adam) solver was used for optimization. Learning rate is set to 5×10^{-4} , all other parameters are set to

default. Each time the network training is performed, the algorithm would be used to optimize all training set data for 20 times, and then the verification set would be used to obtain the system performance index results.

3.3.3 Operating environment

The multi-scale residual network was built by using PyTorch 1.0, and it was trained by using GTX950M with Ubuntu 18.04 system. Other hardware configurations included: Intel Core i7-4710MQ, 12GB RAM.

3.4 Performance evaluation

3.4.1 Cross validation

The cross validation used in this paper included K-fold cross validation and subject cross validation. The former randomly could divide the entire data set into k subsets with epoch as the smallest unit. Each subset was taken as the verification set, and the remaining k-1 subset was taken as the training set. Experiments were performed for k times, and the results of all verification sets were weighted and summarized to get the final result. In this paper, k is set to 5. The latter divided the data set into training set and verification set with a partition ratio of 8:2.

3.4.2 Performance

Recall (Re_k), accuracy (Acc_k) and specificity (Sp_k) were used to evaluate the results of sleep staging. Overall recall (Re_ε), overall accuracy (Acc_ε) and overall specificity (Sp_ε) were expressed by Equation (16) – (18).

$$Re_k = \frac{TP}{TP + FN}(\%), Re_\varepsilon = \frac{\sum_{k=1}^n Re_k}{n}(\%) \quad (16)$$

$$Acc_k = \frac{TP + TN}{TP + FN + TN + FP}(\%), Acc_\varepsilon = \frac{\sum_{k=1}^n Acc_k}{n}(\%) \quad (17)$$

$$Sp_k = \frac{TN}{TN + FP}(\%), Sp_\varepsilon = \frac{\sum_{k=1}^n Sp_k}{n}(\%) \quad (18)$$

where TP, TN, FP and FN represented the true cases, true negative cases, false positive cases and false negative cases formed by the classifier to judge the category, respectively. In these equations, n=5, k=1,2,3,4,5 represented five different stages of sleep. At the same time, the kappa coefficient which could describe overall performance of the system would also be calculated as Equation (19) – (20).

$$err = \frac{\sum(1|Y_i \neq PY_i)}{\sum(1|Y_i = Y_i)}(\%) \quad (19)$$

$$kappa = \frac{p_o - p_e}{1 - p_e}, p_e = \frac{\sum a_i b_i}{n^2} \quad (20)$$

where Y_i represented the real label of the sample i, PY_i represented the predicted label of the sample i of the model, and a_i was the actual number of the sample i, and a_i was the predicted label of the first sample. b_i was the number of predicted samples at last, and n was the total number of samples.

4 Results and discussion

4.1 Cross validation

4.1.1 5-fold cross validation

The 5-fold cross validation was completed on Sleep-EDFX dataset, and the results were shown in Table 1. The table contained the classification performance index of each sleep stage and the overall performance index of the original confusion matrix obtained after cross validation. The overall recall rate, accuracy rate, specificity,

Table 1 5-fold cross validation confusion matrix and performance

	W	N1	N2	N3	REM	Re(%)	Acc(%)	Sp(%)	err(%)	Kappa
W	63232	6813	254	58	840	88.81	95.93	98.38	-	-
N1	1871	16228	3869	141	3060	64.48	86.91	89.55	-	-
N2	345	10401	68533	5401	4295	77.02	87.87	94.30	-	-
N3	30	93	2566	16738	27	86.04	96.51	97.44	-	-
REM	470	5036	1864	25	26786	78.36	93.46	95.99	-	-
Overall						78.94	92.06	95.13	0.1986	0.7960

error rate and Kappa coefficient were 78.94%, 92.06%, 95.13%, 19.86% and 0.7360, respectively. It can be seen that the network proposed in this paper can provide good classification performance with high resolution for the period W, but poor resolution for N1. In particular, it was found by the confusion matrix that N1 was easily misjudged as N2 and REM, which was consistent with the results in [24]. This may be due to the relatively small proportion of sleep time in N1, resulting in less training data.

4.1.2 Subject cross validation

The results of subject cross validation on the dataset were shown in Table 2. The overall recall rate, accuracy rate, specificity, error rate and Kappa coefficient of classification were 75.81%, 91.13%, 94.65%, 22.19% and 0.7001, respectively. Memar

Table 2 Subject cross validation confusion matrix and performance

	W	N1	N2	N3	REM	Re(%)	Acc(%)	Sp(%)	err(%)	Kappa
W	16097	1486	63	17	517	88.56	94.13	97.55	-	-
N1	532	2971	668	14	753	60.17	85.78	88.73	-	-
N2	78	2494	11064	1006	1141	70.10	86.83	95.07	-	-
N3	6	78	450	2670	16	82.92	96.67	97.67	-	-
REM	110	778	397	6	4399	77.31	92.22	94.24	-	-
Overall						75.81	91.13	94.65	22.19	0.7001

[26] used random forest to classify W, S1+S2, S3+S4 and REM, and the recall rates were 97.99%, 79.62%, 75.39% and 59.62%, respectively. For comparison, N1 and N2 stages were combined and recalculated according to the confusion matrix, and recall rates of each sleep stage were 88.56%, 83.00%, 82.92% and 77.31% respectively. Except for W, recall rates of the other three stages were much higher than those in [26]. Tsinalis [24] used the convolutional neural network to perform 20-fold subject cross validation on the Sleep-EDFX dataset. According to the confusion matrix given in this paper, we calculated the overall recall rate, accuracy rate, specificity, error rate and Kappa coefficient, which were 73.7%, 89.91%, 93.62%, 25.23% and 0.6535 respectively, and the system performance of the method was slightly lower than that of the method proposed in this paper.

4.2 Performance comparison with different residual networks

Resnet18 [37], Multi-Scale 1D Convolutional Neural Network(Multi-scale-1d-Resnet) [35] and our proposed MS-RESCNN were compared in same circumstances, and 5-fold cross validation results were shown in Table 3-5. In order to adapt to the input of one-dimensional data, all two-dimensional layers in the Resnet18 network structure were modified to one-dimensional layers. The results showed that

Table 3 Confusion Matrix and Performance for Resnet18

	W	N1	N2	N3	REM	Re(%)	Acc(%)	Sp(%)	Kappa
W	7729	190	7	9	37	96.95	98.12	99.44	-
N1	18	368	60	3	149	61.53	95.46	96.87	-
N2	3	120	2968	296	195	82.85	93.82	97.73	-
N3	0	7	148	1133	3	87.76	96.88	97.73	-
REM	18	135	99	3	1342	84.03	95.75	97.14	-
Overall						82.63	96.01	97.69	0.847

Table 4 Confusion Matrix and Performance for Multi-Scale-1d-Resnet

	W	N1	N2	N3	REM	Re(%)	Acc(%)	Sp(%)	Kappa
W	7729	127	9	3	33	97.84	98.43	99.23	-
N1	33	353	77	3	134	58.33	95.77	97.30	-
N2	5	95	3170	176	138	88.37	94.45	96.35	-
N3	1	6	221	1061	2	82.18	97.25	98.67	-
REM	15	158	110	1	1314	82.22	96.07	97.71	-
Overall						81.89	96.41	97.86	0.8611

Table 5 Confusion Matrix and Performance for our proposed MS-RESCNN

	W	N1	N2	N3	REM	Re(%)	Acc(%)	Sp(%)	Kappa
W	7717	214	2	2	27	96.92	98.18	99.60	-
N1	13	436	45	4	102	72.67	94.80	95.72	-
N2	5	153	2961	335	143	82.31	94.03	97.19	-
N3	1	10	114	1159	3	90.05	96.88	97.52	-
REM	9	241	110	0	1244	78.04	95.84	97.95	-
Overall						84.00	95.95	97.70	0.8452

there was no significant difference in the overall accuracy of the three networks, but in terms of overall recall rate, the network model proposed in this paper had an improvement of about 2% compared with the other two networks. In terms of N1 resolution, the three network structures have great differences. The classification recall rate of our proposed network for N1 was 72.67%, which was more than 10% better than that of Resnet18 (61.53%) and Multi-scale-1d-Resnet (58.33%).

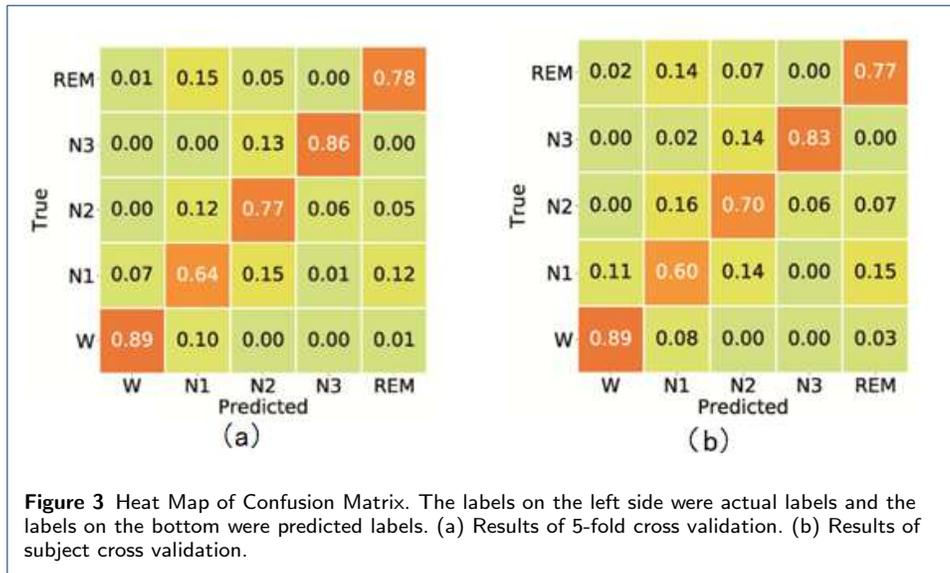
4.3 Visual demonstration

4.3.1 Confusion matrix visualization processing

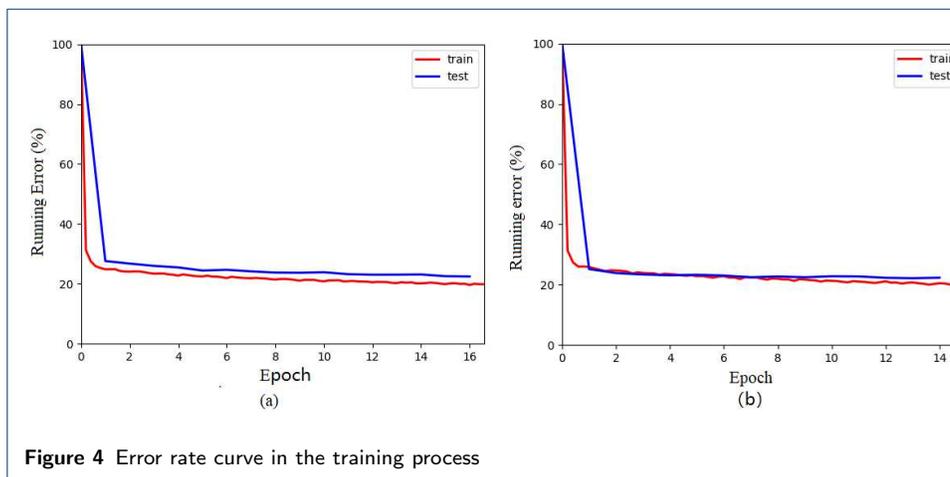
The confusion matrix obtained by the two cross validation methods was normalized according to the actual number of labels, and then a heat map was generated. The result was shown in Fig. 3. The heat map can be used to evaluate the method proposed in this paper intuitively.

4.3.2 Overfitting phenomenon in training

In order to observe the training of the network, our program output the error rate of the network in training set and verification set in real time. The curve of the error rate in 5-fold cross validation and Subject cross validation is shown in Fig. 4.

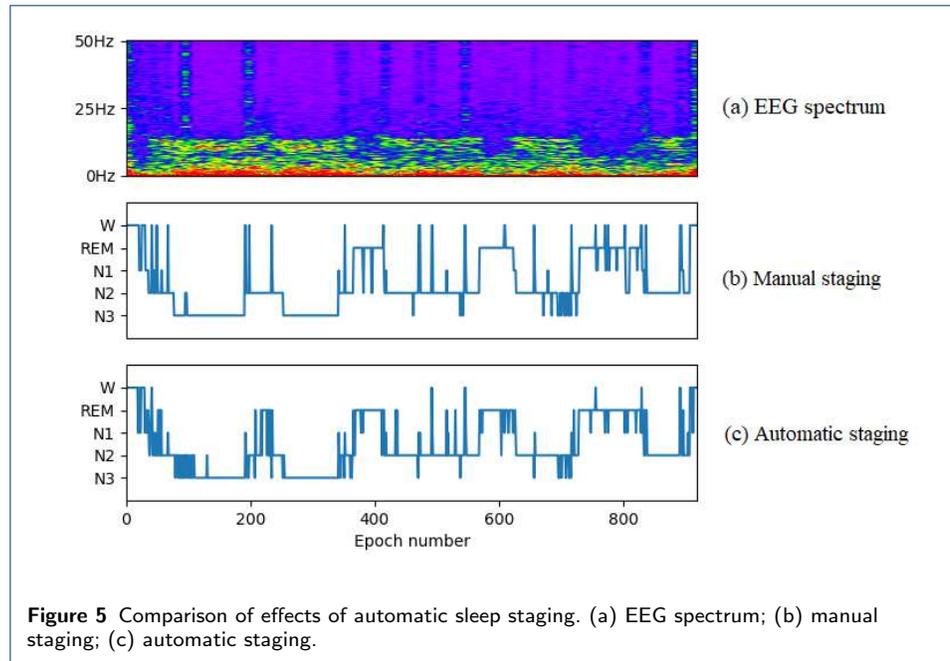


In the two graphs, we can find that serious overfitting could still occur in the actual training of the network although the data set, which has been expanded in the process of data preprocessing, i.e., the performance of the network in the training set was better than that of the verification set. In particular, the error rate (red line) of validation set still decreased with the increase of iteration times in 5-fold cross validation although severe overfitting occurred. However, the error rate of training set still decreased when the error rate of validation set had converged in subject cross validation. This phenomenon showed that unknown subject data would have a great impact on system performance from another perspective.



4.4 Comparison of automatic and artificial sleep staging

Automated sleep staging was performed by using the trained model and compared with manual scoring results, as shown in Fig. 5, the automatic staging result was close to the manual staging result.



5 Conclusion

In this paper, a new sleep staging method based on multi-scale residual network was proposed. It can automatically extract useful information from original single channel EEG signals and classify sleep stages. By the cross validation of datasets, the system performance can be maintained for the data of invisible subjects in the case of large data volume. Compared with other deep learning methods, our method only used a single channel EEG, it could not require a complex data preprocessing and specialized feature extraction process for obtaining a better system performance, which provided a possibility for the clinical application of automated sleep staging. In addition, the multiscale residual network could be further deepened when the computing capacity was enough, and then, a larger amount of data could be used for training, so that the model with better robustness and system performance can be obtained theoretically.

Acknowledgements

Not applicable.

Funding

The research was supported by Science and Technology Program of Guangzhou (2019050001), Program for Chang Jiang Scholars and Innovative Research Teams in Universities (*No.*IRT₁7R40), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (2017B030301007), Guangzhou Key Laboratory of Electronic Paper Displays Materials and Devices (201705030007), MOE International Laboratory for Optical Information Technologies and the 111 Project.

Abbreviations

MS-RESCNN: Multi-scale Residual Convolutional Neural Network; EEG: Electroencephalography; W: Wake-up; REM: Rapid Eye Movement sleep; NREM: Non Rapid Eye Movement sleep; AASM: American Academy of Sleep Medicine; PSG: Polysomnography; EOG: Electrooculogram; ECG: Electrocardiograph; EMG: Electromyography; BCG: Ballistocardiograph; RNN: Recurrent Neural Network; CNN: Convolutional Neural Network; DBN: Deep Belief Network; MSCNN: Multi-scale Convolutional Neural Network; RESCNN: Residual Convolutional Neural Network; BN: Batch Normalization;

Availability of data and materials

The database that supports the conclusions of this article is available in :The Sleep-EDF Database[Expanded]. <https://physionet.org/physiobank/database/sleep-edfx/>

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The main idea of this paper was proposed by QZ. The algorithm design and experimental environment construction were completed by HL and QC. Guidance was given by GZ. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Qinghua Zhong.

Authors' information

¹Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, P. R. China.

²School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, P. R. China.

³Shenzhen Guohua Optoelectronics Tech. Co. Ltd., Shenzhen 518110, P. R. China

⁴Academy of Shenzhen Guohua Optoelectronics, Shenzhen 518110, P. R. China

Author details

¹Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, P. R. China. ²School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, P. R. China. ³Shenzhen Guohua Optoelectronics Tech, Co. Ltd., Shenzhen 518110, P. R. China. ⁴Academy of Shenzhen Guohua Optoelectronics, Shenzhen 518110, Guangzhou 510006, P. R. China.

References

- Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., *et al.*: Sleep drives metabolite clearance from the adult brain. *science* **342**(6156), 373–377 (2013)
- Bianchi, A.M., Mendez, M.O., Cerutti, S.: Processing of signals recorded through smart devices: sleep-quality assessment. *IEEE Transactions on Information Technology in Biomedicine* **14**(3), 741–747 (2010)
- Wolpert, E.A.: A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Archives of General Psychiatry* **20**(2), 246–247 (1969)
- Iber, C., Ancoliisrael, S., Cheeson, A., Quan, S.: The aasm manual for scoring of sleep associated events: rules, terminology and technical specifications (2007)
- Wu, H.T., Talmon, R., Lo, Y.L.: Assess sleep stage by modern signal processing techniques. *IEEE transactions on bio-medical engineering* **62**(4), 1159–68 (2015)
- Samet, Mounir, Lajnef, Tarek, Eichenlaub, Jean-Baptiste, Kachouri, Abdennaceur, Chaibi, Sahbi: Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. *Journal of Neuroscience Methods* (2015)
- Fernández-Varela, I., Hernández-Pereira, E., Álvarez-Estévez, D., Moret-Bonillo, V.: Combining machine learning models for the automatic detection of eeg arousals. *Neurocomputing* **268**, 100–108 (2017)
- Chambon, S., Galtier, M.N., Arnal, P.J., Wainrib, G., Gramfort, A.: A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(4), 758–769 (2018)
- Cui, Z., Zheng, X., Shao, X., Cui, L.: Automatic sleep stage classification based on convolutional neural network and fine-grained segments. *Complexity* **2018** (2018)
- Sokolovsky, M., Guerrero, F., Paisarnrisomsuk, S., Ruiz, C., Alvarez, S.A.: Deep learning for automated feature discovery and classification of sleep stages. *IEEE/ACM transactions on computational biology and bioinformatics* **17**(6), 1835–1845 (2019)
- Khalighi, S., Sousa, T., Pires, G., Nunes, U.: Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. *Expert Systems with Applications* **40**(17), 7046–7059 (2013)
- Surrel, G., Aminifar, A., Rincón, F., Murali, S., Atienza, D.: Online obstructive sleep apnea detection on medical wearable sensors. *IEEE transactions on biomedical circuits and systems* **12**(4), 762–773 (2018)
- Jeon, Y.J., Kang, S.J.: Wearable sleepcare kit: Analysis and prevention of sleep apnea symptoms in real-time. *IEEE Access* **7**, 60634–60649 (2019)
- Kuo, C.E., Liu, Y.C., Chang, D.W., Young, C.P., Shaw, F.Z., Liang, S.F.: Development and evaluation of a wearable device for sleep quality assessment. *IEEE Transactions on Biomedical Engineering*, 1–1 (2016)
- Virkkala, J., Hasan, J., V?Rri, A., Himanen, S.L., Müller, K.: Automatic sleep stage classification using two-channel electro-oculography. *Journal of Neuroscience Methods* **166**(1), 109–115 (2007)
- Samy, L., Huang, M., Liu, J.J., Xu, W., Sarrafzadeh, M.: Unobtrusive sleep stage identification using a pressure-sensitive bed sheet. *IEEE Sensors Journal* **14**(7), 2092–2101 (2014)
- Takano, M., Ueno, A.: Noncontact in-bed measurements of physiological and behavioral signals using an integrated fabric-sheet sensing scheme. *IEEE journal of biomedical and health informatics* **23**(2), 618–630 (2018)
- Lomaliza, J.P., Park, H., Billingham, M.: Combining photoplethysmography and ballistocardiography to address voluntary head movements in heart rate monitoring. *IEEE Access* **PP**(99), 1–1 (2020)
- Scebba, G., Da Poian, G., Karlen, W.: Multispectral video fusion for non-contact monitoring of respiratory rate and apnea. *IEEE Transactions on Biomedical Engineering* **68**(1), 350–359 (2020)
- Liu, J., Chen, Y., Wang, Y., Chen, X., Cheng, J., Yang, J.: Monitoring vital signs and postures during sleep using wifi signals. *IEEE Internet of Things Journal* **5**(3), 2071–2084 (2018)

21. Liang, S.F., Kuo, C.E., Hu, Y.H., Pan, Y.H., Wang, Y.H.: Automatic stage scoring of single-channel sleep eeg by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation Measurement* **61**(6), 1649–1657 (2012)
22. Diyykh, M., Li, Y., Wen, P.: Eeg sleep stages classification based on time domain features and structural graph similarity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**(11), 1159–1168 (2016)
23. Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., De Vos, M.: Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(3), 400–410 (2019)
24. Tsinalis, O., Matthews, P.M., Guo, Y.: Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Annals of Biomedical Engineering* **44**(5), 1587–1597 (2015)
25. Zhang, J., Wu, Y.: A new method for automatic sleep stage classification. *IEEE transactions on biomedical circuits and systems* **11**(5), 1097–1110 (2017)
26. Memar, P., Faradji, F.: A novel multi-class eeg-based sleep stage classification system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(1), 84–95 (2017)
27. Li, X., Cui, L., Tao, S., Chen, J., Zhang, X., Zhang, G.: Hyclclass: A hybrid classifier for automatic sleep stage scoring. *IEEE Journal of Biomedical and Health Informatics* **22**(2), 375–385 (2018)
28. Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P.M., Guo, Y.: Mixed neural network approach for temporal sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(2), 324–333 (2017)
29. Supratak, A., Dong, H., Wu, C., Guo, Y.: Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(11), 1998–2008 (2017)
30. Qian, B., Xiao, Y., Zheng, Z., Zhou, M., Zhuang, W., Li, S., Ma, Q.: Dynamic multi-scale convolutional neural network for time series classification. *IEEE Access* **8**, 109732–109746 (2020)
31. Jiang, G., He, H., Yan, J., Xie, P.: Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Transactions on Industrial Electronics* **66**(4), 3196–3207 (2018)
32. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: *European Conference on Computer Vision*, pp. 630–645 (2016). Springer
33. Olesen, A.N., Jørgen Jennum, P., Mignot, E., Sorensen, H.B.D.: Automatic sleep stage classification with deep residual networks in a mixed-cohort setting. *Sleep* **44**(1), 161 (2021)
34. Qu, W., Wang, Z., Hong, H., Chi, Z., Feng, D.D., Grunstein, R., Gordon, C.: A residual based attention model for eeg based sleep staging. *IEEE journal of biomedical and health informatics* **24**(10), 2833–2843 (2020)
35. Liu, R., Wang, F., Yang, B., Qin, S.J.: Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions. *IEEE Transactions on Industrial Informatics* **16**(6), 3797–3806 (2019)
36. The sleep-edf database [expanded]. <https://physionet.org/physiobank/database/sleep-edfx/>
37. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)

Figures

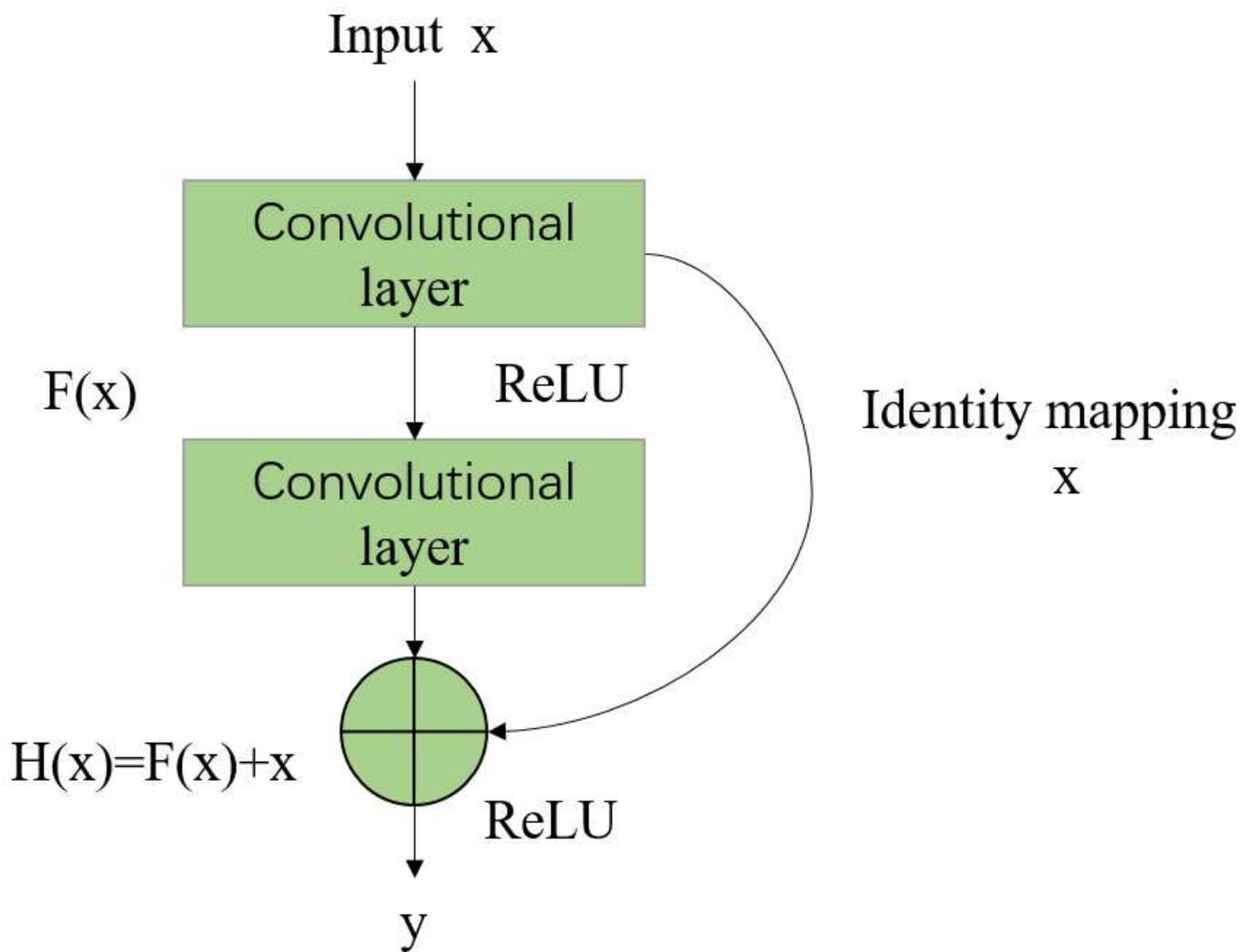


Figure 1

Schematic diagram of residual block

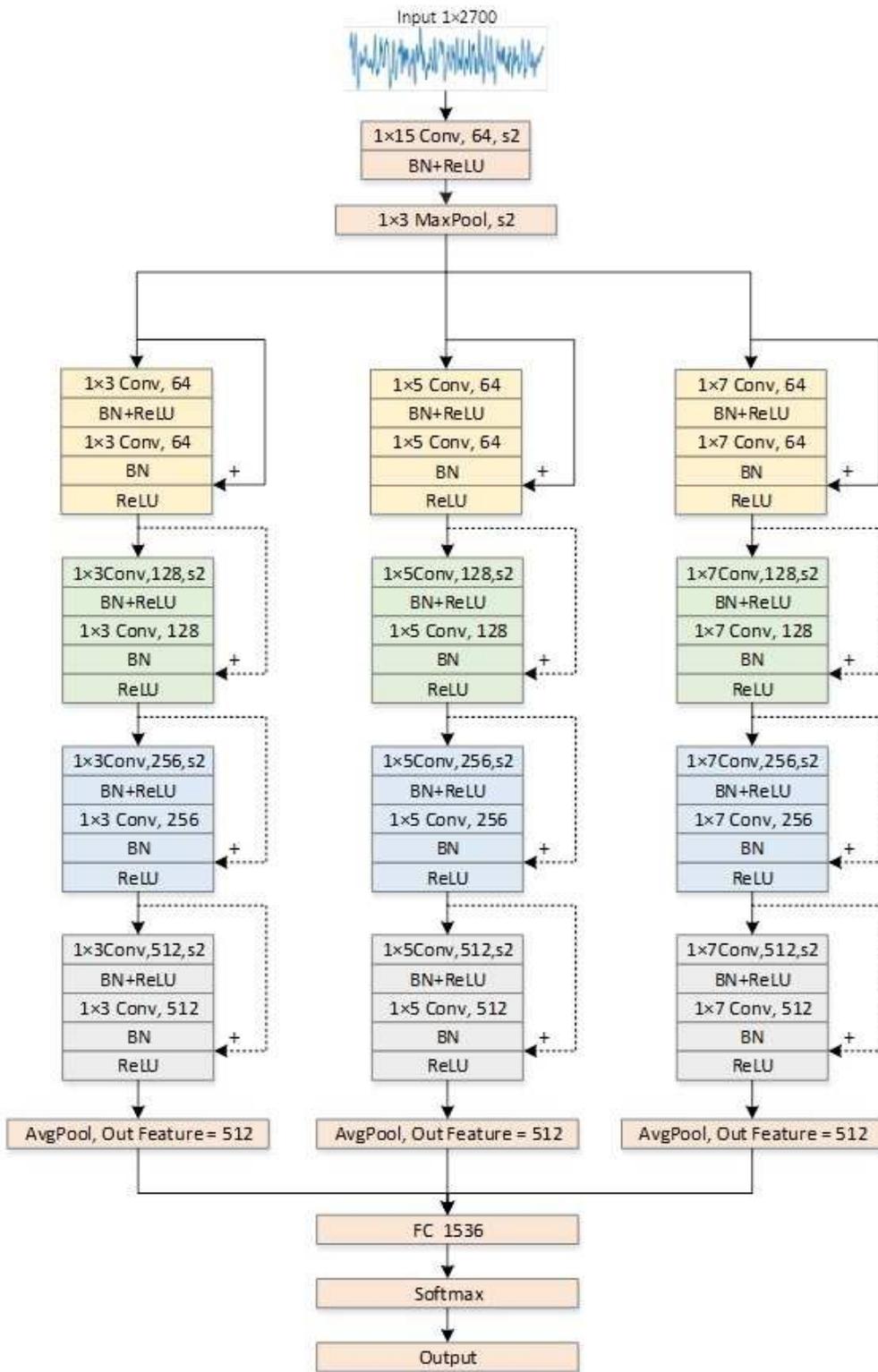


Figure 2

Flow of sleep staging method based on MS-ResCNN

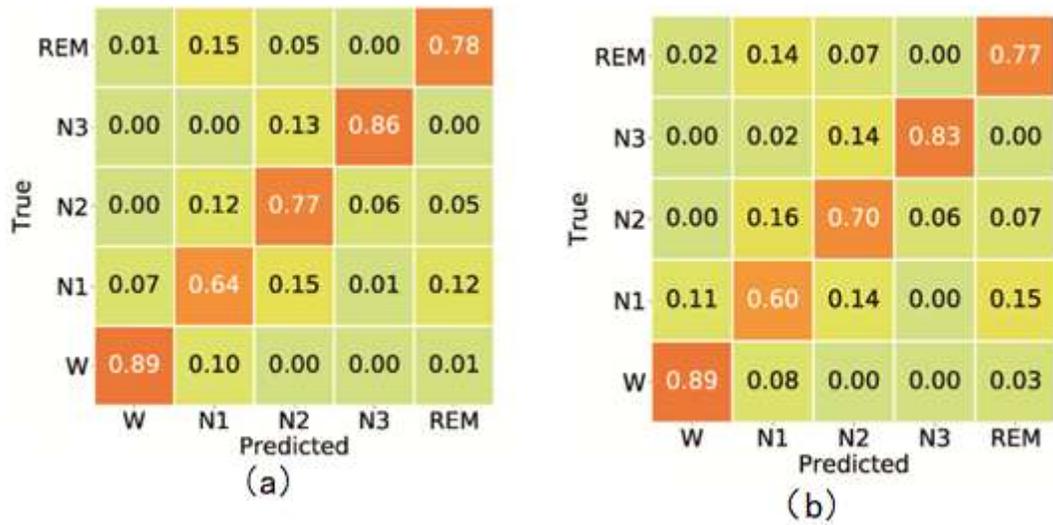


Figure 3

Heat Map of Confusion Matrix. The labels on the left side were actual labels and the labels on the bottom were predicted labels. (a) Results of 5-fold cross validation. (b) Results of subject cross validation.

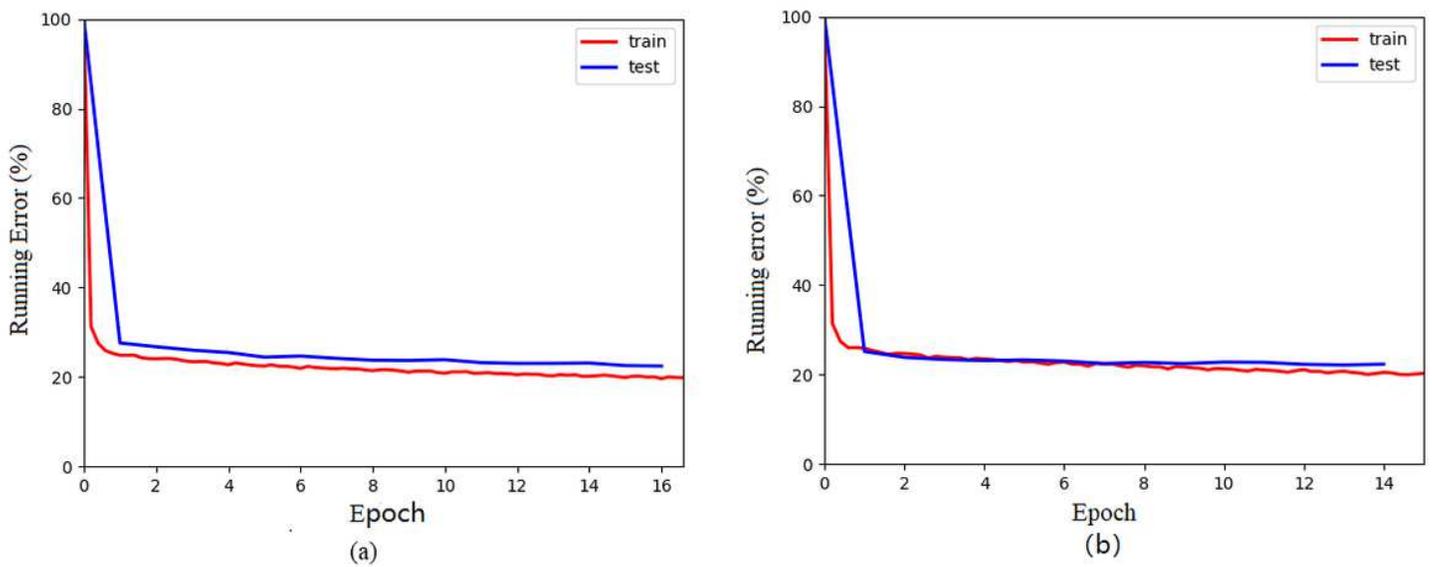


Figure 4

Error rate curve in the training process

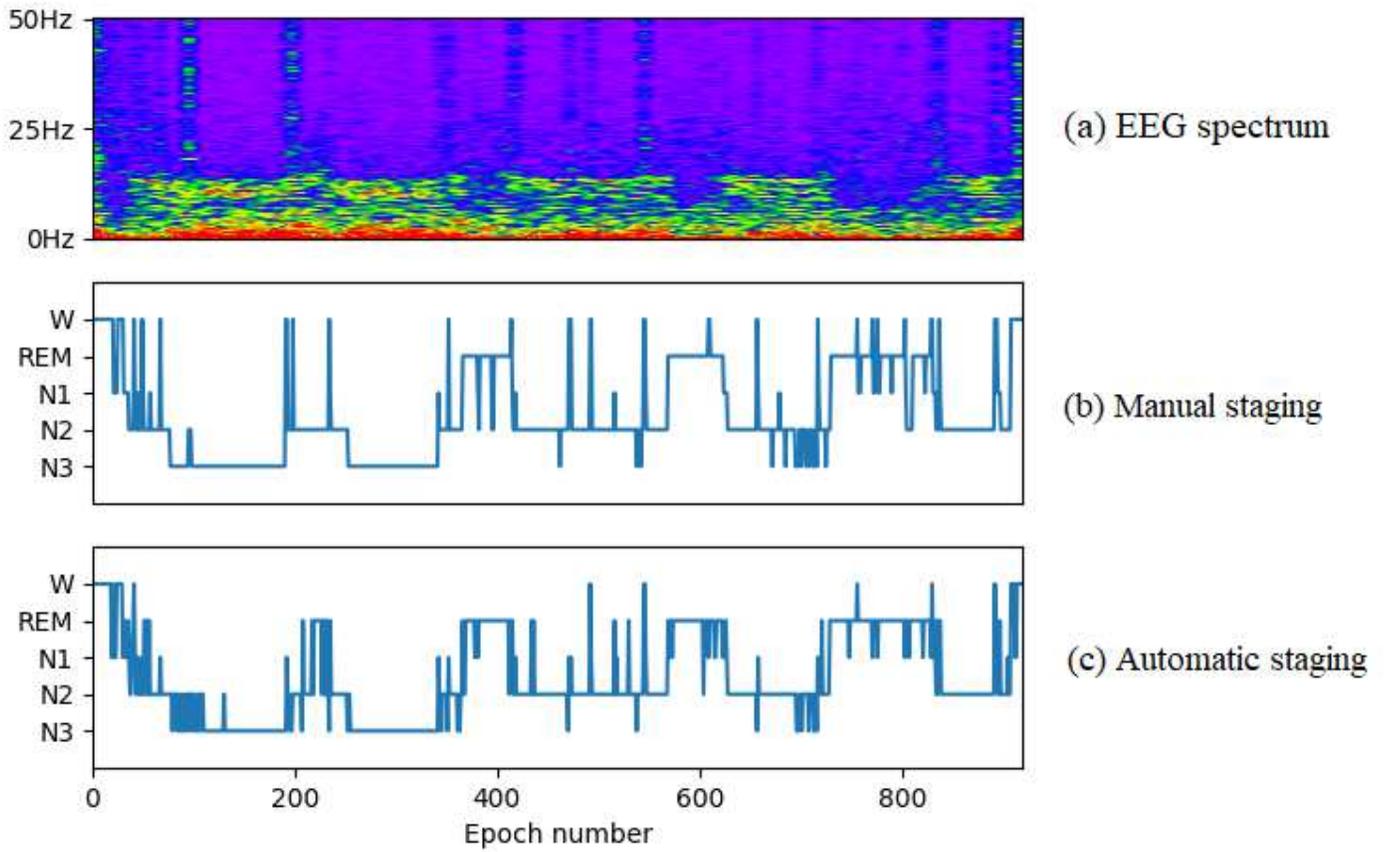


Figure 5

Comparison of effects of automatic sleep staging. (a) EEG spectrum; (b) manual staging; (c) automatic staging.