Research Article

A Sleep Stage Classification Algorithm of Wearable System Based on Multiscale Residual Convolutional Neural Network

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

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, China

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

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

⁴Academy of Shenzhen Guohua Optoelectronics, Shenzhen 518110, China

Correspondence should be addressed to Qinghua Zhong; zhongqinghua@m.scnu.edu.cn

Received 27 September 2021; Accepted 30 November 2021; Published 16 December 2021

Academic Editor: Mu Zhou

Copyright © 2021 Qinghua Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sleep disorder is a serious public health problem. Unobtrusive home sleep quality monitoring system can better open the way of sleep disorder-related diseases screening and health monitoring. In this work, a sleep stage classification algorithm based on multiscale residual convolutional neural network (MRCNN) was proposed to detect the characteristics of electroencephalogram (EEG) signals detected by wearable systems and classify sleep stages. EEG signals were analyzed in each epoch of every 30 seconds, and then 5-class sleep stage classification, wake (W), rapid eye movement sleep (REM), and nonrapid eye movement sleep (NREM) including N1, N2, and N3 stages was outputted. Good results (accuracy rate of 92.06% and 91.13%, Cohen's kappa of 0.7360 and 0.7001) were achieved with 5-fold cross-validation and independent subject cross-validation, respectively, which performed on European Data Format (EDF) dataset containing 197 whole-night polysomnographic sleep recordings. Compared with several representative deep learning methods, this method can easily obtain sleep stage information from single-channel EEG signals without specialized feature extraction, which is closer to clinical application. Experiments based on CinC2018 dataset also proved that the method has a good performance on large dataset and can provide support for sleep disorder-related diseases screening and health surveillance based on automatic sleep staging.

1. Introduction

Sleep plays an important role in human health [1]. The sleep monitoring of human has significant implications for medical research and practice [2]. Sleep specialists usually evaluate the quality of sleep by analyzing signals from sensors connected to different parts of body in accordance with the Rechtschaffen and Kales Rules [3] or the American Academy of Sleep Medicine (AASM) sleep score manual [4]. In particular, polysomnography (PSG), which records EEGs, electrooculograms (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, sev-

eral effective sleep staging methods based on EEG, ECG, and EMG signals have been proposed in recent years [6]. However, wearing too many sensors during sleep is obtrusive and uncomfortable, the silver/silver chloride electrodes with certain adhesive or conductive paste the signal acquisition are adopted mostly, and the placement of them is demanded carefully in hairy regions of scalp to minimize movement-related noise, which affects the natural sleep of subjects and is not suitable for long-term sleep monitoring in home environment.

In recent years, noninvasive [7, 8] or noncontact [9, 10] measurement of cardiac, respiratory, and body movement signals which offers the potential of low cost and easy operation for long term dynamic sleep monitoring has gradually gained the favour of researchers. But its performances badly depend on the quality of signal acquisition and complex

signal processing. EEG recordings play a crucial role in the classification of sleep stages. In order to solve the problem of wearers' comfort, classification of sleep stages based on single-channel EEG signals has been extensively investigated [11, 12], because compared to complex PSG devices, the corresponding dual-electrode device has the advantages of wearable and less interference.

Compared with frontal electrodes, EEG recordings from central, occipital, and parietal electrodes could better detect sleep spindles, vertex waveforms, and rhythms [13]. However, channel selection is important for measurement convenience. Fp1, Fp2, and Fpz electrodes below hairline are suitable for wearing and data collection, so do F3 and F4 electrodes. Researchers have tried to address this problem by extracting the same relevant information from frontal electrodes [14]. But the information which can support for W, N1, N2, and N3 stages is different from that extracted from the parietal lobe, so it can affect the credibility of staging results.

The performance of most sleep stage classification methods relies on the selection of representative features for different sleep stages. Frequency domain [15], time-domain, and time-frequency domain [16] decompositions are the common steps for processing time signals and extracting features directly, and various mathematical models have been established in the process of discovering hidden features. After feature extraction, various machine learning algorithms are used for classification [17, 18], such as linear discriminant analysis, nearest neighbour classifier, decision tree, support vector machine, random forest, and ensemble learning. It also shows the good results with combinatorial machine learning models [19]. In recent years, deep learning methods such as convolutional neural network (CNN) [20], recurrent neural network, and other deep neural networks have become common in pattern recognition in biomedical signal processing. In [21], long short-term memory (LSTM) model which takes advantage of sequential data learning to optimize classification performance was proposed for automated sleep stage. Since feature-based approaches may not be suitable for a comprehensive description of subject heterogeneity, CNNs were also applied to learn multiple filters to extract time-invariant features from raw EEG channel [22].

To solve both the subject heterogeneity and temporal pattern recognition problems, the combination of CNN and LSTM has shown a good performance for the usage of precomputed spectrograms in [23]. However, most representative deep learning models rely heavily on hyperparameter tuning, which is challenging to extend. Although some studies such as [24] consider the temporal context, training must be divided into pretraining and fine-tuning. In addition, since the learning rate is set to a very low value during fine-tuning, it takes more time to reach the optimum or may not even be optimal. Therefore, the computation cannot be performed in parallel, which also prolongs the training process.

Increasing the network layer number of CNNs can improve the extraction effect of signal features. Multiscale CNNs were also proposed to perform multiscale feature

extraction and classification simultaneously [25]. However, gradient dispersion or gradient explosion is likely to occur when the depth of CNNs increases. The residual network (ResNet) proposed by He et al. [26] was aimed to solve the degeneration problem of network. The method was applied to the machine fault detection and achieved good results [27]. The sleep stage classification based on residual-based attention model was also adopted in [28], but only k -fold cross validation not subject cross-validation was performed, and in the meantime, the amount of test data was not enough.

Considering that multiscale convolutional neural network can capture the detailed signal features required for pattern classification, the idea was adopted to realize a wearable smart eye-mask in our prior study [29]. This method uses single-channel original EEG signals and omits the process of special feature extraction. It has good performance and application potential and can provide support for clinical applications such as screening and diagnosis of sleep disorders. The main contributions of this work are as follows:

- (1) A deep learning architecture of ResNet with different filter sizes was developed. Time-invariant features from original single-channel EEG signals can be extracted by training learning filters, so it can save the time of features computation, while residual can be trained to encode temporal information such as sleep phase transition rules into the model
- (2) A training algorithm that can effectively train the model end-to-end was developed. Single-channel EEG from forehead was adapted to reduce the patients impact and increase usability
- (3) The proposed approach is evaluated on two publicly available datasets: sleep-EDF-expanded [30] and CinC2018 [31] through subject cross-validation experiment with a strong robustness performance. The results are compared with state-of-the-art results in the field of sleep stage classification to demonstrate the superiority of our method, and it solves the problem that too little data may lead to poor generalization effect of sleep stage model

2. Methods

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, and N3. In this paper, the multiscale feature learning was integrated into a residual network (ResNet) 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. Residual Block. In general, the more layers in convolutional neural network, the more diverse features can be extracted. But previous experiments show that there exists a degradation problem in deep networks: when the network

depth of network increases, accuracy saturates and then degrades, and the addition of more layers can lead to an even higher training error, which is not caused by overfitting.

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 layers increases. Theoretically, if the additional layer of a certain layer in the deep network does not learn anything, but just copies the features of last layer, it is called identity mapping, and the training error should not increase. A deep residual learning framework was proposed to solve the degradation phenomenon [26]: if an identity mapping was optimal, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers.

For the network with an input x , the learning feature is denoted as $H(x)$, expected network learning residuals $F(x) = H(x) - x$. Because residual learning is easier than the traditional feature learning, the residual learning adopts every few stacked layers, as shown in Figure 1, a residual block is formed by neural network with shortcuts connections. It contains two kinds of mappings: 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, 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 the following equation.

$$y = F(x, W_i) + x, \quad (1)$$

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 the following equation.

$$F = W_2 \sigma(W_1 x), \quad (2)$$

where σ is the rectified linear unit (ReLU) activation function.

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

According to the above description, the expression of ResNet unit constructed can be expressed by the following equations.

FIGURE 1: Schematic diagram of residual block.

$$y_1 = W \otimes x + b, \quad (3)$$

$$y_2 = \text{BN}(y_1), \quad (4)$$

$$y_3 = \text{ReLU}(y_2), \quad (5)$$

$$y_4 = W \otimes y_3 + b, \quad (6)$$

$$y_5 = \text{BN}(y_4), \quad (7)$$

$$y_6 = y_5 + x, \quad (8)$$

$$y = \text{ReLU}(y_6), \quad (9)$$

where $\otimes$ is convolution operator, BN is used to improve generalization capability.

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 an optimal network 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 the following equation.

$$\text{RF}_i = (\text{RF}_{i+1} - 1) \times \text{stride}_i + \text{kernel}_i, \quad (10)$$

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 the i -th layer, and kernel_i is the size of the convolution kernel of i -th layer.

FIGURE 2: Flow of sleep staging method based on MRCNN.

The EEG signals input to the network are processed by the convolution layer with a convolution kernel length of 15 and transmitted to the BN and ReLU activation functions

for 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, and the effective frequency resolution of 563 data points was 0.35 Hz, which can meet the frequency resolution requirements of all rhythmic waves.

2.3. Network Training. In order to realize the loss calculation of multiclassification of sleep stage, cross entropy (CE) was used as the loss function, and its definition was shown as the following equation.

$$\text{loss}(\text{input}, \text{target}) = \text{weight}[\text{target}] \times \left(-\text{input}[\text{target}] + \log \left(\sum_j e^{\text{input}[j]} \right) \right), \quad (11)$$

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, and weight[target] is defined as the following equation.

$$\text{weight}[\text{target}] = \ln \left(\frac{1}{p(\text{target})} \right), \quad (12)$$

where $p(\text{target})$ is the proportion of the label target in total label.

For the loss function, the adaptive moment estimation solver (Adam) was used for optimization. Learning rate was set to 5×10^{-4} , and all other parameters were set to default. Each time the network training was 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. The MRCNN was built by using PyTorch 1.0, and it was trained by using GTX950M with Ubuntu 18.04 system. Other hardware configurations include Intel Core i7-4710MQ, 12 GB RAM.

2.4. Performance Evaluation. 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 the following equations.

$$\text{Re}_k = \frac{\text{TP}}{\text{TP} + \text{FN}} (\%), \text{Re}_\varepsilon = \frac{\sum_{k=1}^n \text{Re}_k}{n} (\%), \quad (13)$$

$$\text{Acc}_k = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} (\%), \text{Acc}_\varepsilon = \frac{\sum_{k=1}^n \text{Acc}_k}{n} (\%), \quad (14)$$

$$\text{Sp}_k = \frac{\text{TN}}{\text{TN} + \text{FP}} (\%), \text{Sp}_\varepsilon = \frac{\sum_{k=1}^n \text{Sp}_k}{n} (\%), \quad (15)$$

where TP, TN, FP, FN represent 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, $n = 5$, $k = 1, 2, 3, 4, 5$ represents 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 the following equation.

$$\text{err} = \frac{\sum (1 | Y_i \neq \text{PY}_i)}{\sum (1 | Y_i = Y_i)}, \text{kappa} = \frac{p_o - p_e}{1 - p_e}, p_e = \frac{\sum a_i b_i}{n^2}, \quad (16)$$

where Y_i represents the real label of the sample i , PY_i represents the predicted label of the sample i of the model, p_o presents the sum of the samples of each correct category divided by the total number of samples, $p_e = 1 - \text{err}$, a_i is the actual number of the sample i , b_i is the predicted number of the sample i , and n is the total number of samples.

3. Experiments

3.1. Datasets. In order to ensure the robustness and reproducibility of the results, two public datasets were conducted experiments. The first dataset used in the experiment was sleep-EDF-expanded and contains two different groups of subjects, named as Sleep Cassette (SC) group and Sleep Telemetry (ST) group, respectively. The annotation files included sleep stages W, REM, Stage 1 (S1), Stage 2 (S2), Stage 3 (S3), Stage 4 (S4), movement time (M), and UNKNOWN, and it consisted of a manual score by a skilled technician. Stage M and UNKNOWN were deleted for their extremely small percentage. At the same time, according to the latest AASM sleep scoring manual, S1 and S2 were corresponded to N1 and N2, respectively, and S3 and S4 were combined into N3. In this study, we cropped the SC* files in the dataset, 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.

The second dataset used in the experiment was provided by the CinC2018, which contains 1,985 samples. The sleep stages of each sample were labeled by Massachusetts General Hospital clinical staff and divided into six stages: W, N1, N2, N3, REM, and Undefined. For research and application consideration, data from F4-M1 channels was selected here. The data was divided into a training set ($n = 994$) and a test set ($n = 989$). We randomly selected 500 subjects from the training set as the dataset and deleted the undefined period with a small proportion. To adapt to the AASM sleep scoring manual, the EEG signals of datasets were divided into 30 s as an epoch. After processing, the sleep stage statistics of the two datasets are shown in Table 1.

TABLE 1: The number and proportion of epochs of various sleep stages.

Dataset	W	N1	N2	N3	REM	Total
Sleep-EDF-expanded (197 samples)	71197 (29.8%)	25169 (10.5%)	88975 (37.2%)	19454 (8.1%)	34181 (14.3%)	238976 (100%)
CinC2018 (500 samples)	82368 (17.95%)	66243 (14.4%)	197338 (43.0%)	54318 (11.835%)	58672 (12.78%)	458939 (100%)

TABLE 2: Subject cross-validation statistics in sleep-EDF-expanded dataset.

	Samples	W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
Total	197	71199 (29.8%)	25174 (10.5%)	88983 (37.2%)	19454 (8.1%)	34184 (14.3%)
Train	157	53014 (27.7%)	20236 (10.6%)	73213 (38.3%)	16235 (8.5%)	28497 (14.9%)
Eval	40	18185 (38.0%)	4938 (10.3%)	15770 (33.0%)	3219 (6.7%)	5687 (11.9%)

TABLE 3: Subject cross-validation statistics in CinC2018 dataset.

	Samples	W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
Total	500	82368 (18.0%)	66243 (14.4%)	197338 (43.0%)	54318 (11.8%)	58672 (12.8%)
Train	400	64477 (17.8%)	48227 (13.3%)	155927 (43.1%)	45090 (12.5%)	47942 (13.3%)
Eval	100	17891 (18.4%)	18016 (18.5%)	41411 (42.6%)	9228 (9.5%)	10730 (11.0%)

TABLE 4: 5-fold cross-validation confusion matrix and performance in sleep-EDF-expanded.

	W	N1	N2	N3	REM	Re _k (%)	Acc _k (%)	Sp _k (%)	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	0.7360

TABLE 5: Subject cross-validation confusion matrix and performance in sleep-EDF-expanded.

	W	N1	N2	N3	REM	Re _k (%)	Acc _k (%)	Sp _k (%)	Kappa
W	16097	1483	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	0.7001

TABLE 6: Performance comparison of different sleep staging models based on subject cross-validation.

Method	W	N1	Re _k (%) N2	N3	REM	Acc _ε (%)	Re _ε (%)	Kappa
LSTM	83	39	80	74	68	80	68.8	0.64
CNN	80	40	81	69	64	78	66.8	0.58
CNN-LSTM	84	45	82	83	82	86	75.2	0.71
ResNet18	89	49	71	81	82	91	74.4	0.70
Our MRCNN	89	60	70	83	77	91	75.8	0.70

TABLE 7: Subject cross-validation confusion matrix and performance in CINC2018.

	W	N1	N2	N3	REM	$Re_k(\%)$	$Acc_k(\%)$	$Sp_k(\%)$	Kappa
W	14518	2442	427	62	440	81.15	92.49	95.05	\
N1	3131	8588	2670	33	3590	47.67	83.57	91.00	\
N2	582	4416	29880	4077	2454	72.16	83.75	91.62	\
N3	22	9	1469	7663	65	83.04	94.33	95.47	\
REM	193	973	537	6	9918	84.04	92.04	92.96	\
Overall						73.61	89.24	93.22	0.6224

FIGURE 3: Comparison of effects of automatic sleep staging and expert manual staging in sleep-EDF-expanded. (a) EEG spectrum, (b) manual staging of expert, and (c) automatic staging.

3.2. Preprocessing. In order to adapt to different PSG devices and individual differences of subjects, EEG signals were normalized by using the 5-th and 95-th quantiles [32], as shown in the following equation.

$$x_{\text{norm}} = 2 \frac{x - Q_{0.05}(x)}{Q_{0.95}(x) - Q_{0.05}(x)} - 1, \quad (17)$$

where x_{norm} is the result of signal normalization, x is original signal, and $Q_{0.05}(x)$ and $Q_{0.95}(x)$ are the 5% and 95% largest value of the signals.

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, and it was converted to tensor data type, as shown

in the following equation.

$$x_{\text{nct}} \in R^{N \times C \times T}, \quad (18)$$

where $N = 16$, $C = 1$, $T = 2700$ represented batch size, channel number, and 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 the following equation.

$$N3 \longrightarrow 0, N2 \longrightarrow 1, N1 \longrightarrow 2, \text{REM} \longrightarrow 3, W \longrightarrow 4. \quad (19)$$

3.3. Performance Evaluation. The cross-validation used in this paper included k -fold cross-validation and subject cross-validation. The former randomly divided the entire dataset into k subsets with epoch as the smallest unit. Each subset was taken as the verification set, and the remaining $k - 1$ subsets 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.

FIGURE 4: Output results of the first convolution layer in CinC2018 dataset. (a) primitive EEG in W stage. (b) Fast Fourier transform results of EEG in W stage. (c) Output results of the first convolution layer in W stage. (d) Primitive EEG in N3 stage. (e) Fast Fourier transform results of EEG in N3 stage. (f) Output results of the first convolution layer in N3 stage.

In order to evaluate the performance of proposed method, 5-fold cross-validation and subject cross-validation were completed on sleep-EDFx dataset, and data of 197 subjects from FPz-Cz channels was used. Subject cross-validation divided the dataset into training set and verification set with a partition ratio of 8:2, and statistical information is shown in Tables 2 and 3. In order to evaluate the effect in practical application, subject cross-validation was also completed on CinC2018 dataset, and the data of 500 subjects from F4-M1 channel was used.

4. Results and Discussion

4.1. Cross-Validation. The results of 5-fold cross-validation are shown in Table 4. The table contains 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, and kappa coefficient are 78.94%, 92.06%, 95.13%, 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 is found by the confusion matrix that N1 is easily misjudged as N2 and REM, which is consistent with the results in [33]. This may be due to the relatively small proportion of sleep time in N1, resulting in less training data. In the meantime, N1 is a transitional period between waking state and sleep state and contains both α and β waves, which makes classification difficult.

The results of subject cross-validation on the dataset are shown in Table 5. The overall recall rate, accuracy rate, specificity, and kappa coefficient of classification are 75.81%, 91.13%, 94.65%, and 0.7001, respectively. Results compared with the existed sleep staging methods are shown in

Table 6. It can maintain better system performance for the data of invisible subjects in the case of large amount of data and reach the similar effect of CNN-LSTM [21]. This cross-validation method can effectively prove the generalization ability of the system on unknown subjects and has high practical application value. In terms of N1 resolution, the recall rate of our proposed network for N1 is 60%, which is much more than other deep learning models.

In [29], several methods based on residual network were compared by using sleep-EDF-expanded dataset, and it can be seen that our proposed MRCNN performs better than other residual networks also in terms of N1 resolution. In this paper, ResNet18 [26] and MRCNN were compared in same circumstances. In order to adapt to the input of one-dimensional data, all two-dimensional layers in the network structure of ResNet18 were modified to one-dimensional layers. It can also be seen that our proposed MRCNN performs better than ResNet18 in terms of N1 resolution. From the data shown in Table 6, it can be also seen that the proposed network can provide poor resolution for N2. In particular in Table 5, it was found by the confusion matrix that N2 was easily misjudged as N1 or N3. This may be due to N2 is a transitional period between N1 and N3 and contains both sleep spindles and *K*-complex waves, which makes classification difficult.

4.2. Performance Evaluation in Wearable Application. Especially, the channel used in the experiments is F4-M1, which is suitable for wearable application. Training set of 400 samples and test set of 100 samples from CinC2018 were used for subject cross-validation. The results of performance are shown in Table 7. Even though the performance of the system decreased a little when it applied to large amount of data, it still demonstrated the generalization ability of the system in the presence of unknown subjects.

4.3. Comparison of Automatic and Artificial Sleep Staging. Automated sleep staging was performed by using the trained model and compared with manual scoring results of expert, as shown in Figure 3, the automatic staging results are close to the manual staging results.

4.4. Automatic Extraction of EEG Features by MRCNN. The results of MRCNN's extraction of effective EEG features are shown here. EEG data in W stage and N3 stage from CinC2018 dataset were fed into the trained network model, respectively, and then the output of the first convolution layer of the model was visualized. It contains 64 convolution kernels, corresponding to the output of 64 channels, and the final results are shown in Figure 4. It can be seen from the figure that for the same input signal, different convolution kernels have different outputs. In addition, by comparing the convolution layer output results of W stage and N3 stage, it can be found that there are significant differences in the output of the same convolution kernel of different input signals, which fully demonstrates that the convolutional neural network can automatically extract the features of EEG signals.

5. Conclusions

In this paper, a new sleep staging method based on multiscale 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 uses a single-channel EEG, and it does not require complex data preprocessing and specialized feature extraction processes to achieve better system performance, which provides the 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.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The research was supported by the Science and Technology Program of Guangzhou (2019050001), Program for Chang Jiang Scholars and Innovative Research Teams in Universities (no. IRT_17R40), Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (2017B030301007), Guangzhou Key Laboratory of Electronic Paper Displays Materials and Devices (201705030007), and MOE International Laboratory for Optical Information Technologies and the 111 Project. "A multiscale residual convolutional neural network for sleep staging based on single channel electroencephalography signals" as the preliminary study has been presented as a preprint according to the following link: <https://www.researchsquare.com/article/rs-554671/v1>.

References

- [1] L. Xie, H. Kang, Q. Xu et al., "Sleep drives metabolite clearance from the adult brain," *Science*, vol. 342, no. 6156, pp. 373–377, 2013.
- [2] A. M. Bianchi, M. O. Martin, and S. Cerutti, "Processing of signals recorded through smart devices: sleep-quality assessment," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 741–747, 2010.
- [3] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Electroencephalography & Clinical Neurophysiology*, vol. 26, no. 2, pp. 644–644, 1969.

- [4] R. B. Berry, R. Budhiraja, D. J. Gottlieb et al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events," *Journal of Clinical Sleep Medicine*, vol. 8, no. 5, pp. 597–619, 2012.
- [5] H. T. Wu, R. Talmon, and Y. L. Lo, "Assess sleep stage by modern signal processing techniques," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1159–1168, 2015.
- [6] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels," *Expert Systems with Applications*, vol. 40, no. 17, pp. 7046–7059, 2013.
- [7] M. Takano and A. Ueno, "Noncontact in-bed measurements of physiological and behavioral signals using an integrated fabric-sheet sensing scheme," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 618–630, 2019.
- [8] J. Lomaliza, H. Park, and M. Billingham, "Combining photoplethysmography and ballistocardiography to address voluntary head movements in heart rate monitoring," *IEEE Access*, vol. 8, pp. 226224–226239, 2020.
- [9] G. Scebbba, G. Poian, and W. Karlen, "Multispectral video fusion for non-contact monitoring of respiratory rate and apnea," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 350–359, 2021.
- [10] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using WiFi signals," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2071–2084, 2018.
- [11] S. Liang, C. Kuo, Y. Hu, Y. H. Pan, and Y. H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1649–1657, 2012.
- [12] D. Jiang, Y. Lu, Y. Ma, and Y. Wang, "Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement," *Expert Systems with Applications*, vol. 121, pp. 188–203, 2019.
- [13] G. Zhu, Y. Li, and P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, pp. 1813–1821, 2014.
- [14] C. Huang, C. Lin, L. Ko, S. Y. Liu, T. P. Su, and T. P. Lin, "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels," *Frontiers in Neuroscience*, vol. 8, 2014.
- [15] Y. Hsu, Y. Yang, J. Wang, and C. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, vol. 104, no. 1, pp. 105–114, 2013.
- [16] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Annals of Biomedical Engineering*, vol. 44, no. 5, pp. 1587–1597, 2016.
- [17] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 77–91, 2017.
- [18] T. Lajnef, S. Chaibi, P. Ruby et al., "Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines," *Journal of Neuroscience Methods*, vol. 250, pp. 94–105, 2015.
- [19] I. Fernandez-Varela, E. Hernandez-Pereira, D. Alvarez-Estevéz, and V. Moret-Bonillo, "Combining machine learning models for the automatic detection of EEG arousals," *Neurocomputing*, vol. 268, no. 11, pp. 100–108, 2017.
- [20] A. Sors, S. Bonnet, S. Mirekci, L. Vercueil, and J. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.
- [21] N. Michielli, U. Acharya, U. Rajendra, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in Biology and Medicine*, vol. 106, pp. 71–81, 2019.
- [22] M. Sokolovsky, F. Guerrero, S. Paisarnsrisomsuk, C. Ruiz, and S. A. Alvarez, "Deep learning for automated feature discovery and classification of sleep stages," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 17, no. 6, pp. 1835–1845, 2020.
- [23] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.
- [24] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 324–333, 2018.
- [25] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2019.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, CVPR, 2016.
- [27] R. Liu, F. Wang, B. Yang, and S. Qin, "Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3797–3806, 2020.
- [28] W. Qu, Z. Wang, H. Hong et al., "A residual based attention model for EEG based sleep staging," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2833–2843, 2020.
- [29] Q. Zhong, H. Lei, Q. Chen, and G. Zhou, "A multiscale residual convolutional neural network for sleep staging based on single channel electroencephalography signals.
- [30] "The sleep-EDF database [Expanded]," <http://www.physionet.org/physiobank/database/sleep-edfx/>.
- [31] The PhysioNet Computing in Cardiology Challenge, 2018, <https://www.physionet.org/content/challenge-2018/1.0.0/>.
- [32] A. N. Olesen, P. Jennum, P. Peppard, E. Mignot, and H. B. D. Sorensen, "Deep residual networks for automatic sleep stage classification of raw polysomnographic waveforms," in *2018 IEEE 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018.
- [33] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 84–95, 2018.