Biomedical Engineering Letters (2022) 12:413–420
https://doi.org/10.1007/s13534-022-00244-w

ORIGINAL ARTICLE

Automatic sleep stages classification using multi-level fusion

Hyungjik Kim[1] · **Seung Min Lee**[2] · **Sunwoong Choi**[2]

Received: 12 March 2022 / Revised: 12 July 2022 / Accepted: 25 July 2022 / Published online: 10 August 2022

Abstract
Sleep efficiency is a factor that can determine a person's healthy life. Sleep efficiency can be calculated by analyzing the results of the sleep stage classification. There have been many studies to classify sleep stages automatically using multiple signals to improve the accuracy of the sleep stage classification. The fusion method is used to process multi-signal data. Fusion methods include data-level fusion, feature-level fusion, and decision-level fusion methods. We propose a multi-level fusion method to increase the accuracy of the sleep stage classification when using multi-signal data consisting of electroencephalography and electromyography signals. First, we used feature-level fusion to fuse the extracted features using a convolutional neural network for multi-signal data. Then, after obtaining each classified result using the fused feature data, the sleep stage was derived using a decision-level fusion method that fused classified results. We used public datasets, Sleep-EDF, to measure performance; we confirmed that the proposed multi-level fusion method yielded a higher accuracy of 87.2%, respectively, compared to single-level fusion method and more existing methods. The proposed multi-level fusion method showed the most improved performance in classifying N1 stage, where existing methods had the lowest performance.

Keywords Multi-level fusion · Sleep stage classification · EEG · EMG · Convolutional neural network

1 Introduction

Sleep is the most important factor for a person to live a healthy life. Low-sleep efficiency can cause certain problems [1, 2] in human life. Major depressive disorder, bipolar disorder, and schizophrenia could occur by sleep efficiency problems. As these problems interfere with healthy life, it is important to maintain good sleep efficiency.

Sleep efficiency can be analyzed with polysomnography. Polysomnogram (PSG), which is a test result from polysomnography, is analyzed, sleep stages are classified, and then sleep efficiency is calculated. PSG includes information on electroencephalography (EEG), electromyography (EMG) and electrooculogram (EOG). We can analyze this information to classify sleep stages. Sleep stages are classified into five stages, Wake (W), Non-rapid eye movement (REM) Stage 1 to 3 (N1, N2, and N3), and REM, based on American Academy of Sleep Medicine (AASM) manual [3]. Each sleep stage is classified every 30 s from the information in PSG and each 30s interval is called an epoch [4].

As the sleep stage classification and sleep efficiency calculations are visually examined by an expert, sleep stage scoring is expensive, prone to human error and long time consuming [5–7]. Therefore, automated classification of sleep stages is needed and various studies are being conducted accordingly.

To classify sleep stages, there are studies using machine learning techniques (support vector machine, decision tree, and random forest) by extracting features from signals [8–11]. After analyzing frequency domain and time domain of the signals to extract features, machine learning techniques are used to classify sleep stages. Extractable features include EEG's Power, Zero Crossing, entropy, wavelet transform, and STFT. Additional time consuming occurs because various processing processes are required to extract features suitable for sleep phase classification. And the downside is that as the number of features to be extracted increases, the preprocessing becomes longer and heavier.

There are studies that classify sleep stages by extracting features using deep learning architectures such as CNN (convolutional neural network) or RNN (recurrent neural

✉ Sunwoong Choi
schoi@kookmin.ac.kr

[1] Department of Secured Smart Electric Vehicle, Kookmin University, 02707 Seoul, Korea

[2] Department of Electrical Engineering, Kookmin University, 02707 Seoul, Korea

network) [12–15]. A method to extract features from a single signal using CNN or RNN has a simpler architecture than methods, which is mentioned in the previous paragraph. However, as the information that can be obtained from a single signal is limited, various studies exist where the sleep stages were classified by extracting CNN or RNN and deep learning techniques or features from multiple signals [16–25]. As characteristics extracted from multiple signals are used, the accuracy of sleep stage classification can be enhanced, as compared to studies that use a single signal.

Among the various signals, the most important signal in classifying the sleep stage is EEG. The above-mentioned studies are also based on classifying sleep stages using EEG. EEG signals have different characteristics at each sleep stage [26]. Therefore, most of the automatic classification papers introduced previously have classified sleep stages using EEG. However, it is difficult to accurately determine all specific stages using only EEG signals, because there are sleep stages where EEG signals exhibit similar characteristics. EEG signals of REM, N1, and N2 sleep stages are especially very similar. Therefore, in addition to EEG signals, EMG signals are used to improve accuracy of sleep stage classification for N1 and REM stages. EMG signals also have different characteristics by each stage of sleep [27].

In this paper, we propose a multi-level fusion method to improve accuracy of sleep stages classification. We extract features using CNN layer for each EEG and EMG signals. We designed the method to perform fusion on the extracted feature data for each EEG and EMG signal at the feature level and blended the classified output back at the decision level to obtain the final result, the sleep stage. Finally, we compared the performance with other studies and confirmed that the proposed method showed good performance.

The rest of this paper is organized as follows: in Sect. 2, the dataset and metrics used in this paper are introduced; Sect. 3 describes introduction and analysis of fusion methods, a proposed multi-level fusion method and training processes; results are analyzed in Sect. 4; and finally, conclusions are outlined in Sect. 5.

2 Materials and metrics

2.1 Datasets

In this paper, we used public sleep dataset called Sleep-EDF dataset. The Sleep-EDF dataset was manually classified by well-trained experts according to Rechtschaffen and Kales manual (1968) [28–30]. The R&K manual classified Wake, N1, N2, N3, N4 and REM as 6 classes, but we classified Wake, N1, N2, N3, and REM as 5 classes according to the AASM guidelines. A combination of deep sleep stages of N3 stage and N4 stage into one henceforth denoted by N3 stage. We adopted the same method in [13]. We included 30 min of such periods just before and after the sleep periods, which retained only portions of wake epochs, because there are too many awake and wake stages in original dataset. We represented the configuration of the Sleep-EDF dataset in Table 1 after adjusting the W stage. Sleep-EDF dataset contains two EEG signals (Fpz-Cz, Pz-Oz) and one EMG signal. The sampling rate of the EEG signals was 100 Hz, and the sampling rate of the EMG signal was 1 Hz. To maintain processing consistency, we upsampled the EMG signal rate to 100 Hz. The 20 subjects of sleep data are used from "Sleep Cassette Study and Data" in Sleep-EDF dataset, of which data from Subject 0 to 19 (male: 10, female: 10) are used.

2.2 Metrics

We used accuracy and F1-score as metrics to evaluate the performance of our proposed multi-level fusion model. We classified it into five multi-classes and checked its performance with classified results. We used overall accuracy and F1-Score as indicators using TP, FP, FN and TN values, where TP refers to true positives, TN is true negatives, FP is false positives and FN is false negatives. The equations of overall accuracy and F1-Score are shown below. Overall accuracy is the ratio correctly predicted for the overall prediction [31, 32].

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

F1-Score is a measure used to properly balance precision and recall, it consists of weighted harmonic average values of precision and recall [33].

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2}$$

Table 1 Number of epochs for Sleep-EDF datasets

Datasets	Wake	N1	N2	N3	REM	Total
Sleep-EDF	6734 (16.6%)	2771 (6.9%)	17,624 (43.6%)	5628 (13.9%)	7702 (19.0%)	40,459

3 Method

3.1 Fusion methods with electroencephalography and electromyography

Three fusion methods were considered in this paper: data-level fusion, feature-level fusion, and decision-level fusion [34]. In three fusion methods, fusion used concatenate. Feature extraction used CNN layers. And for learning and classification, we used a Fully-connected layers (FC layers) and softmax. Data-level fusion method fuses input data before extracting its features of input data and provides output using fused data. The data-level fusion method is useful when extracting features using similar characteristics at once. When extracting features using techniques such as CNN, for example, adjusting the filter size allows you to extract features from multiple data together. Feature-level fusion extracts features of input data and, then, blends the extracted features. It is a way to receive output using fused features. The feature-level extraction method is similar to the data-level fusion method, but the features of the signals are extracted separately. You can use the values of the functions you extract independently and use different feature extraction methods. In decision-level fusion, each input data has its own independent deep learning structure. Decision-level function fuses obtained results from each softmax layer to receive output. Decision-level fusion is similar to the ensemble method. The final output is obtained by combining the results from each model as if they were voting. It has the advantage of being able to use various features because the models can be organized differently. Figure 1 shows flowcharts of fusion methods used in the paper.

Performance analysis was conducted to determine which fusion method is appropriate for input data. The performance was confirmed by applying a method using multiple EEG, single EEG, and single EMG, with input data provided

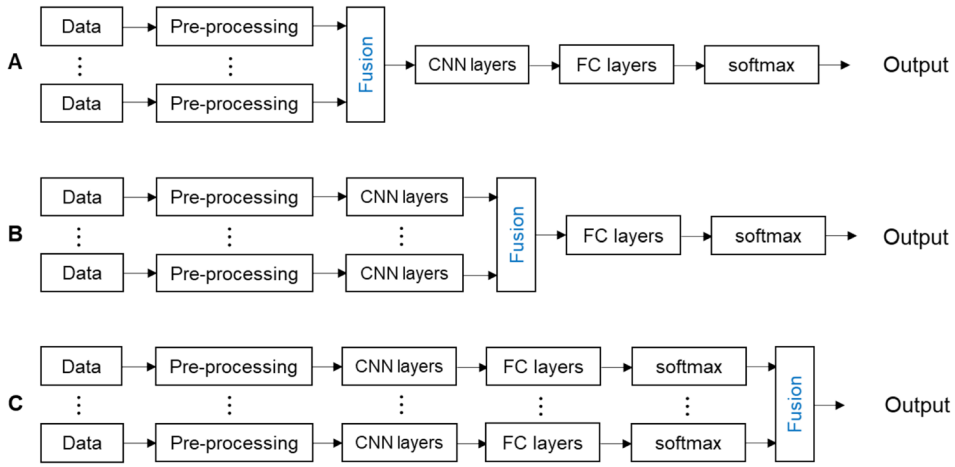

Fig. 1 Flowcharts of fusion methods: **A** is the Data-level fusion method; **B** is the Feature-level fusion method; **C** is the Decision-level fusion method

Table 2 Performance of different fusion methods with EEG and EMG

Datasets	Signals	Fusion method	Overall Metrics		Per-class F1-Score				
			Accuracy	F1-Score	W	N1	N2	N3	REM
Sleep-EDF	Multiple EEGs (Fpz-Cz, Pz-Oz)	Data-level Fusion	82.2	76.3	86.6	41.6	87.5	87.3	78.5
		Feature-level Fusion	81.9	75.9	85.9	40.4	87.3	86.9	79.2
		Decision-level Fusion	**83.5**	**77.5**	**88.1**	**42.9**	**88.2**	**88.1**	**80.0**
	EEG (Fpz-Cz) & EMG	Data-level Fusion	82.3	77.5	84.7	48.1	87.1	86.6	80.9
		Feature-level Fusion	**83.3**	**78.6**	**87.3**	**49.4**	**87.4**	**86.7**	**81.9**
		Decision-level Fusion	80.7	73.5	83.0	38.5	85.6	83.6	77.0
	EEG (Pz-Oz) & EMG	Data-level Fusion	81.3	75.9	85.1	**43.4**	85.9	83.4	81.9
		Feature-level Fusion	**81.7**	**76.4**	**86.0**	**43.4**	**86.2**	**84.0**	**82.1**
		Decision-level Fusion	79.2	72.7	81.8	36.8	84.9	82.2	77.9

to each fusion method. Results are shown in Table 2. The detailed structure and parameters are explained in the Sect. 3.3.

First, when using multiple EEG as input data, the result of decision-level fusion is shown as the highest. The results of data-level fusion and feature-level fusion are similar, and all the results of decision-level fusion are shown as the highest.

Next, when single EEG and single EMG as input data are used, the result of feature-level fusion is shown as the highest. And the result of decision-level fusion, which performed the highest when using multiple EEGs, was the lowest. EEG alone can classify all sleep stages; however, the result of decision-level fusion was the lowest using EMG alone, because EMG alone cannot classify sleep stages.

Lastly, it can be inferred that accuracy of N1 and REM stage classification can be enhanced when using EMG and EEG together rather than using only multiple EEGs as input data. When comparing the case of using only multiple EEGs and the case of using EMG and EEG together on Sleep-EDF dataset, it can be seen that the F1-score of N1 is higher when using EMG and EEG together. It can be inferred that the combination of EMG and EEG helps to improve accuracy of N1 and REM stage classification.

As a result, there are differences between using only EEG and using EEG and EMG together, and it is inferred that performance is further improved only when the appropriate fusion method is used. Since we use both multiple EEG and EMG, we designed a multi-level fusion method by selecting feature-level fusion and decision-level fusion.

3.2 The proposed architectures of multi-level fusion method

We propose a multi-level fusion method using a fusion method tailored to the characteristics of data. We used multiple EEG and EMG signals and built a method using the CNN.

Figure 2 shows flowcharts of a proposed multi-level fusion method in the paper. We have deployed independent CNN layers for each signal, as the characteristics that can be identified in each EEG and EMG signals are different. This can be expressed as a CNN architecture module using a single EEG and EMG signal, as represented by the red color squares in Fig. 2. The more diverse EEG signals are provided, the more the numbers of modules are available. As a result, the number of features extracted for each module increases, enabling the use of more information to derive sleep stage. Characteristics extracted for each method are fused at feature level. Then, the results of fused data are obtained using fully-connected layers and the Softmax layers. Lastly, a sleep stage was derived using a decision-level fusion, which fuses the results of each module. Fusing results of each module can increase the proportion of signals with certain characteristics that appear in sleep stages, resulting in a more accurate classification of sleep stages.

3.3 Training parameters and implementation

The detailed structure and parameters of the proposed method are explained as follows: parameters and values used in the method are described in Table 3. Input data of methods are 30 s of data from multiple EEG and EMG signals. Input data were downsampled to 50 Hz in preprocessing layers. EMG signals provided by Sleep-EDF data were upsampled to 50 Hz, as it was 1 Hz. Moreover, features were extracted by changing kernel sizes for each CNN layer. The CNN layer consists of three pairs of convolutional layers and max-pooling layers. Each CNN layer used batch normalization, and a ReLU function was used as an activation function. At the end of the method, two fully-connected layers and one Softmax layer were used.

In this paper, we constructed training set and test set using 20 cross-validation methods tailored to the number of subjects. The 19 subject's data were used as training sets and the 1 subject's data was used as a test set. The results from 20 iterations of the above process were combined to confirm the performance. The batch size of training data is 100 and the number of training epochs was 20. This experiment was performed on a server on CentOS 7.5 version. The server's

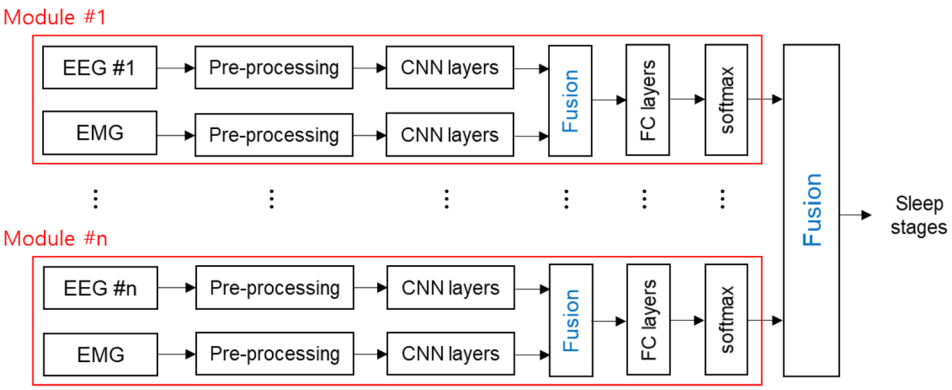

Fig. 2 Flowcharts of proposed multi-level fusion method: EEG and EMG extract features through their respective CNN layers in the module and then fuse feature levels. The results of each module are fused to the decision level to obtain the final sleep stage

CPU has 10 cores (Intel (R) Xeon (R) Silver 4114 CPU @ 2.20 GHz) and the GPU is an NVIDIA TITAN method.

4 Result

4.1 Performance of fusion methods

We evaluated the proposed multi-level fusion method using two EEG signals and one EMG signal from the Sleep-EDF dataset. The confusion matrix is shown in Table 4. The last three columns in each row indicate per-class performance metrics computed from the classification results.

We compared the accuracy of the proposed multi-level fusion method with other fusion methods. Methods used for the comparison are data-level fusion, feature-level fusion, and decision-level fusion methods. Results are shown in Table 5.

The results show 83.6% accuracy of data-level fusion, 84.4% accuracy of feature-level fusion, 82.7% accuracy of decision-level fusion, and 87.3% of accuracy of the proposed multi-level fusion method. The combination of EEG and EMG shows that the proposed method has the highest results. And as we saw earlier, EEG and EMG can see that using feature-level method is best except for the proposed method. When the EMG signal is used alone, it is difficult to classify the sleep phase, so it shows the worst result of the decision-level fusion method. It can be inferred that the proposed multi-level fusion method shows higher accuracy than the single-level fusion method.

Next, we compared the proposed multi-level fusion method with other fusion methods using F1-score per class. It can be inferred that the results of the proposed method show the highest in all stages in Table 5. In particular, it can be inferred that the growth rates of F1-Score in N1 stage and REM stage are the highest.

Table 3 Hyper-parameters of proposed multi-level fusion method

Layer	Layer type	Modules	Kernel size	Stride size	Output Dimension
Input					(N, 3000)
Pre-processing					(N, 1500)
1st CNN layer	Convolutional	5	(1, 100)	(1, 1)	(N, 1500, 5)
	Max-pooling		(1, 2)	(1, 2)	(N, 750, 5)
2nd CNN layer	Convolutional	10	(1, 100)	(1, 1)	(N, 750, 5)
	Max-pooling		(1, 2)	(1, 2)	(N, 375, 5)
3rd CNN layer	Convolutional	20	(1, 50)	(1, 1)	(N, 375, 5)
	Max-pooling		(1, 2)	(1, 2)	(N, 188, 5)
1st FC layer		1024			(N, 1024)
2nd FC layer		512			(N, 512)
Softmax		5			(N, 5)

Table 4 Confusion matrix of the proposed multi-level fusion method

	Predicted					Per-class metrics		
	W	N1	N2	N3	REM	Precision	Recall	F1-Score
W	5972	555	124	21	62	91.8	88.7	90.2
N1	315	1963	307	14	172	55.1	70.8	62.0
N2	108	787	15,730	435	564	90.2	89.3	89.7
N3	31	17	570	4987	23	91.0	88.6	89.8
REM	82	239	716	22	6643	89.0	86.3	87.6

Table 5 Performance of different fusion methods

Dataset	Method	Overall metrics		Per-class F1-Score				
		Accuracy	F1-Score	W	N1	N2	N3	REM
Sleep-EDF	Data-level fusion	83.6	78.8	86.5	48.9	87.9	87.7	83.2
	Feature-level fusion	84.4	79.9	88.6	51.5	88.0	87.6	84.0
	Decision-level fusion	82.7	77.4	86.2	45.3	87.1	86.3	82.1
	Proposed	87.3	83.8	90.2	62.0	89.7	89.8	87.6

Table 6 Performance of different methods on Sleep-EDF dataset

Method	Signals	Overall Metrics		Per-class F1-Score				
		Accuracy	F1-Score	W	N1	N2	N3	REM
Supratak et al. [13]	Fpz-Cz	82.0	76.9	84.7	46.6	85.9	84.8	82.4
Tsinalis et al. [14]	Fpz-Cz	78.9	73.7	71.6	47.0	84.6	84.0	81.4
Mousavi et al. [15]	Fpz-Cz	84.3	79.7	89.2	52.2	86.8	85.1	85.0
Tianqi et al. [16]	Fpz-Cz Pz-Oz EOG EMG	85.8	81.2	**92.3**	54.5	87.8	85.4	85.9
Xiaoqing et al. [17]	Fpz-Cz Pz-Oz EOG EMG	83.6	78.1	86.4	49.8	88.7	84.5	81.6
Phan et al. [18]	Fpz-Cz EOG	84.6	79.0	82.6	50.0	87.8	86.2	88.4
Phan et al. [19]	Fpz-Cz EOG	86.4	80.9	–	–	–	–	–
Guillot et al. [20]	Fpz-Cz Pz-Oz EOG	-	79.1	–	–	–	–	–
Proposed	**Fpz-Cz Pz-Oz EMG**	**87.2**	**83.8**	90.2	**62.0**	**89.7**	**89.8**	**87.6**

4.2 Comparison of performance with other studies

We compared several different methods using Sleep-EDF dataset to verify the performance of the proposed multi-level fusion method. Comparison results are shown in Table 6. It can be seen that the results of the proposed multi-level fusion method are the highest in overall metrics. As overall accuracy results reflect on the metrics, the proposed multi-level fusion method recorded the highest accuracy of 87.3%. The highest result of F1-score was 83.8 in the proposed multi-level fusion method. From F1-scores for each sleep stage, it can be seen that the result of the proposed method is the highest except for W stage. In particular, it could be inferred that the classification accuracy of N1 stage has increased the most. The F1-score of N1 stage is 25.4 units higher than [13] and 7.5 units higher than [16].

4.3 Discussion

Based on the above results, it can be seen that the performance of the proposed multi-level fusion method is the highest. Therefore, it is important to use multi-signal data to identify the characteristics of the data to be used and to use an appropriate fusion method when designing deep learning structures such as CNNs. Additionally, the overall accuracy of classification in sleep stages is also important, but it is important to check the performance of each class as shown in Tables 5 and 6. The ratio of N2 was the largest, and the ratio of the other stages was relatively low. For Sleep-EDF dataset, the ratio of N1 stage was 6.85% of the total epoch, and the N2 stage was 6.3 times more than N1 stage. Classes with relatively small numbers, such as N1 and REM stage, have less impact on the overall accuracy; therefore, the performance by each class should be checked. The overall accuracy is also important, but it is important to improve the performance of classes that are difficult to classify.

A single fusion method has different results depending on the characteristics of input data. As shown in the results of Table 2, when using a single fusion method, the results differ depending on data characteristics. Likewise, it is important to use different fusion methods depending on data characteristics. Therefore, in this paper, the best performance was achieved by fusing a decision-level fusion method that is suitable for using multiple EEG signals and a feature-level fusion method that can be used with additional features of the EMG signal.

5 Conclusions

Sleep is one of the most important indicators for humans to live a healthy life. It is important to accurately analyze and evaluate sleep efficiency through accurate sleep phase classification. As classifying sleep stage takes a long time, many studies discussing automatic sleep stage classification have been conducted.

In this paper, we proposed a method that automatically classifies sleep stages using a multi-level fusion method. We

improved the accuracy of sleep stages classification using multi-signal EEG and EMG. We found that the proposed method performs better than single-level fusion and other methods. In particular, we have further improved the accuracy of the N1 stage, whose accuracy was relatively low.

It is not easy to develop a method that can be applied to all data, because the quality and characteristics of sleep vary depending on people's health and age. And the results of the sleep stage automatic classification method vary significantly depending on the deep learning structure and the signals used for classification. For this reason, various studies are ongoing to improve accuracy of the classification of sleep stages. In the future, we will continue our research to improve accuracy of classification of sleep stages. We will increase the accuracy of the sleep stage by combining EOG signals that were not used in this paper. Then we will study the changes in the structure of multi-level fusion models in response to the additional signals.

Funding This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (Nos. 2016R1A5A1012966 and 2021R1F1A1062285).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Khald AI, Helen TO, Miad F. Efficient sleep stage classification based on EEG signals. In: Proceedings of the IEEE Long Island Systems, Applications and Technology (LISAT) Conference. Farmingdale, NY, USA; 2–2 May 2014. p. 1–6.
2. Wulff K, Gatti S, Wettstein G, Foster G. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. Nat Rev Neurosci. 2010;11:589–99.
3. Iber C, Ancoli-Israel S, Chesson AL, Quan SF. The AASM Manual for the Scoring of Sleep and Associated Events. Westchester: American Academy of Sleep Medicine; 2007.
4. Terzano MG, Parrino L, Smerieri A, Chervin R, Chokroverty S, Guilleminault C, Hirshkowitz M, Mahowald M, Moldofsky H, Rosa A, Thomas R, Walters A. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. Sleep Med. 2002;3:187–99.
5. Koley B, Dey D. An ensemble system for automatic sleep stage classification using single channel EEG signal. Comput Biol Med. 2012;42:1186–95.
6. Ronzhina M, Janoušek O, Kolář̆rová J, Nováková M, Honzík P. Provazník, I. Sleep scoring using artificial neural networks. Sleep Med Rev. 2012;16:251–63.
7. Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. Feature selection for sleep/wake stages classification using data driven methods. Biomed Signal Process Control. 2007;2:171–9.
8. Liu Z, Sun J, Zhang Y, Rolfe P. Sleep staging from the EEG signal using multi-domain feature extraction. Biomed Signal Process Control. 2016;30:86–97.
9. Sharma R, Pachori RB, Upadhyay A. London. Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. In: Neural Computing and Applications. U.K, Springer; 2017. p. 1–20.
10. Hassan AR, Subasi A. A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowl Based Syst. 2017;128:115–24.
11. Hsu YL, Yang YT, Wang JS, Hsu CY. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing. 2013;104:105–14.
12. Cui Z, Zheng X, Shao X, Cui L. Automatic sleep stage classification based on convolutional neural network and fine-grained segments. Complexity. 2018 Oct 8;2018.
13. Supratak A, Dong H, Wu C, Guo Y, DeepSleepNet:. A Model for Automatic Sleep Stage Scoring Based on Single-Channel EEG. IEEE Trans Neural Syst Rehabil Eng. 2017;25(11):1998–2008.
14. Tsinalis O, Matthews PM, Guo Y. Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders. Ann Biomed Eng. 2016;44:1587–97.
15. Mousavi S, Afghah F, Acharya UR, SleepEEGNet. Automated sleep stage scoring with sequence to sequence deep learning approach. PLoS ONE. 2019.
16. Tianqi Z, Wei L, Feng Y. Multi-branch convolutional neural network for automatic sleep stage classification with embedded stage refinement and residual attention channel fusion. Sensors. 2020;20(22):6592.
17. Xiaoqing Z, Mingkai X, Yanru L, Minmin Su, Ziyao X, Chunyan W, Dan K, Hongguang L, Xin M, Xiu D, Wen X, Xingjun W, Demin H. Automated multi-model deep neural network for sleep stage scoring with unfiltered clinical data. Sleep and Breathing. 2020;24:581–90.
18. Phan H, Chén OY, Koch P, Lu Z, McLoughlin I, Mertins A. Towards More Accurate Automatic Sleep Staging via Deep Transfer Learning. IEEE Trans Biomed Eng. 2021;68(6):1787–98.
19. Phan H, Chén OY, Tran MC, Koch P, Mertins A, Vos MD, XSleepNet. Multi-view sequential model for automatic sleep staging. IEEE Transactions on Pattern Analysis and Machine Intelligence, early access; Mar. 2021.
20. Guillot A, Thorey V, RobustSleepNet. Transfer Learning for Automated Sleep Staging at Scale. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1441–51.
21. Fernandez-Blanco E, Rivero D, Pazos A. Convolutional neural networks for sleep stage scoring on a two channel EEG signal. Methodol Appl. 2019;24:2067–4079.
22. Jun S, Xiao L, Yan L, Qi Z, Yingjie L, Shihui Y. Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning. J Neurosci Methods. 2015;254:91–101.
23. Dihong J, Yu M, Yuanyuan W. Sleep stage classification using covariance features of multi-channel physiological signals on Riemannian manifolds. Comput Method Progr Biomed. 2019;178:19–30.
24. Fernando A, Huy P, Navin C, Christine L, Michele TMH, Maarten DV. Multichannel Sleep Stage Classification and Transfer Learning using Convolutional Neural Networks, 2018, 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
25. Huy P, Oliver YC, Philipp K, Alfred M, Maarten DV. Fusion of End-to-End Deep Learning Models for Sequence-to-Sequence Sleep Staging, 2019, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

26. Pan ST, Kuo CE, Zeng JH, Liang SF. A transition constrained discrete hidden Markov model for automatic sleep staging. BioMedical Eng OnLine. 2012;11:52–71.
27. Daniel L, Erik L, Luigi S, Andrea G, Philip W, Chris B. Comparison of EMG power during sleep from the submental and frontalis muscles. Nat Sci Sleep. 2018;10:431–7.
28. Kemp B, Zwinderman AH, Tuk B, Kamphuisen HAC, Oberyé JJL. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave micro continuity of the EEG. IEEE Trans Biomed Eng. 2000;47:1185–94.
29. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
30. The Sleep-EDF Database. https://www.physionet.org/content/sleep-edfx/1.0.0/ (19,10,2020).
31. Hassan AR, Bhuiyan MIH. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. Biomed Signal Process Control. 2016;24:1–10.
32. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Comput Methods Programs Biomed. 2012;108:10–9.
33. Chinchor N MUC-4 evaluation metrics. In: Proceedings of the 4th conference on Message understanding, June, 1992.
34. Farzan MN, Michael R, Md ZU, Jim T. Human activity recognition from multiple sensors data using multi-fusion representations and CNNs. ACM Trans Multimedia Comput Commun Appl. 2020;16(2):1–9.