



A Deep Learning-Based Method for Sleep Stage Classification Using Physiological Signal

Guanjie Huang¹, Chao-Hsien Chu^{1(✉)}, and Xiaodan Wu²

¹ Pennsylvania State University, University Park, PA 16802, USA
chu@ist.psu.edu

² Hebei University of Technology, Tianjin 300130, People's Republic of China

Abstract. A huge number of people suffers from different types of sleep disorders, such as insomnia, narcolepsy, and apnea. A correct classification of their sleep stage is a prerequisite and essential step to effectively diagnose and treat their sleep disorders. Sleep stages are often scored by experts through manually inspecting the patients' polysomnography which are usually needed to be collected in hospitals. It is very laborious for experts and discommodious for patients to go through the process. Accordingly, current studies focused on automatically identifying the sleep stages and nearly all of them need to use hand-crafted features to achieve a decent performance. However, the extraction and selection of these features are time-consuming and require domain knowledge. In this study, we adopt and present a deep learning approach for automatic sleep stage classification using physiological signal. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) of popular deep learning models are employed to automatically learn features from raw physiological signals and identify the sleep stages. Our experiments shown that the proposed deep learning-based method has better performance than previous work. Hence, it can be a promising tool for patients and doctors to monitor the sleep condition and diagnose the sleep disorder timely.

Keywords: Sleep stage classification · Deep learning · Physiological signal
Sleep disorders · Feature extraction

1 Introduction

Sleep, an indispensable physiological activity in our daily life, plays a critical role in keeping people's physical and mental health. An orderly sleep can help people remain vigorous in their daily works. However, a huge number of people suffer from different types of sleep disorders, such as insomnia, narcolepsy, and apnea, etc., which can seriously harm human health. The International Classification of Sleep Disorders (ICSD) has identified over 80 different sleep disorders with associated treatments [3]. A correct classification of the patients' sleep stage is a prerequisite and essential step to effectively diagnose and treat sleep disorders [6, 26, 39]. Because sleep-stage abnormality is correlated with the symptoms of sleep disorders, for instances, obstructive sleep apnea (OSA) decreases the temporal stability of non-rapid eye movement (NREM) and rapid eye movement (REM) sleep bouts [5]. Most obstructive sleep

apnea-hypopnea syndrome (OSAHS) are associated with decreased stage N3 sleep [4]. The study by [27] also claimed that the larger ratios of REM/N3 and N1/Wake are related to obstructive sleep apnea (OSA). Thus, sleep stage classification has attracted increased interest for research.

Sleep scoring, proposed by Rechtschaffen and Kales (R&K), is a gold standard used to classify sleep stage and diagnose sleep disorders. The method divided sleep into five stages: REM, and NREM stages 1, 2, 3, and 4 [18]. However, in practice, the rules are often difficult or impossible to follow and deviations are common. It also has some limitations, for instances, low temporal resolution, ignorance of spatial information, insufficient number of stages, low correspondence between electrophysiological activity and stages, and ignorance of other physiological parameters such as autonomous nervous system activity and body motility [17]. The American Academy of Sleep Medicine (AASM) updated and expanded R&K scheme to AASM scoring manuals [33]. According to AASM, the sleep consists of four distinctive stages: stage R sleep, stage N1, N2, and N3, where R corresponds to REM stage, N1 is analogs to stage 1, N2 is similar to stage 2, and N3 can be considered as stages 3 and 4. N1 is a transition stage between wakefulness and sleep, which usually take 1 to 5 min [31]. N2 follows N1 and usually acts as a “baseline” of sleep. N3 can be considered as “deep sleep” which is the most restorative stage of sleep. Stage REM is characterized by the rapid eye movement under eyelids and the occurrence of dream [10].

The standard approach for segmenting the stages is to have the domain experts manually inspect every epochs of the patient’s polysomnography (PSG) data based on the R&K rules or AASM. In this study, we use the R&K standard, because the dataset [13] we used for exploration is labeled based on R&K. The PSG data often include electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), respiratory effort signals, blood oxygen saturation, and other measurements [29]. Typically, each epoch is obtained by dividing the entire time series data into epochs of 30 s subsequence. The EEG data from [13] are shown in Fig. 1. We can see that the EEG data in different stages have different patterns. For instances, there is a vertex sharp wave in stage 1 and there are some sawtooth waves in stage R. The expert can score the sleep stages through inspecting these small and big “contexts”. The architecture of our model is also inspired by it. However, inspecting epoch by epoch is time-consuming and sometimes involves personal subjective judgment. Hence, an automatic sleep stage classification would be a promising and valuable approach. There are four challenges for automatic sleep stage classification:

Challenge 1: Feature Extraction and Selection. Many studies have been done for automatic sleep stages classification, nearly all of them (e.g., [15, 16, 21, 24, 34, 36]) classified the stages using hand-crafted features. These appropriate features are picked carefully and computed manually based on the expert’s domain knowledge. However, these processes are highly labor-intensive and time-consuming [37]. We even need to consider the interaction between features. What’s worse, these hand-crafted features might be sub-optimal [22].

Challenge 2: Temporal Information. The classification of sleep stages follows a sequential order [2, 9]. That is, the classification depends on not only the information of current data, but also on the information from the past and the future [33]. However,

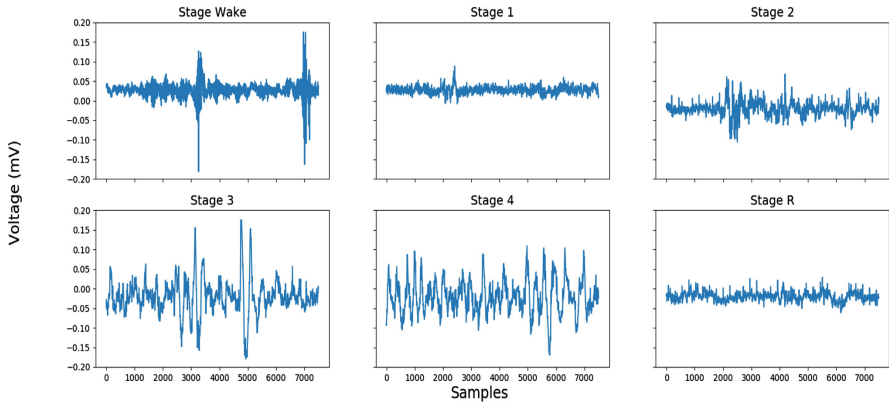


Fig. 1. The samples of 30 s EEG epochs in different sleep stages (sample rate is 250 Hz).

most of the studies (e.g., [16, 21, 24]) only consider the current epoch and did not take advantage of the temporal information from others.

Challenge 3: Patient Impact and Resistance. Though many previous studies (e.g., [7, 10, 11]) achieved good performance in sleep stage classification using multiple physiological data. In the real world, wearing too many sensors during sleep is obtrusive and uncomfortable, which may affect the sleep quality.

Challenge 4: Unlabeled Sleep Data. Currently, the amount of labeled sleep data is limited and thereby it is an obstacle for developing the method for classifying the sleep stages. In fact, there is a huge number of sleep data which is unlabeled. Much of meaningful information are contained in these unlabeled data.

In this study, we focus on tackling the first three challenges. We adopt a newer approach, deep learning, to automatically learn the useful feature representation and integrated it with classification step. The convolutional neural networks (CNN), a popular neural network model, is used to automatically learn the appropriate features with backpropagation. In order to consider the temporal information from the past and the future, the Bidirectional Recurrent Neural Networks (RNN) is adopted to build the classifier. In addition, single EEG is chosen as the biomarker for sleep stage classification, because it is more comfortable compared with PSG recording which needs multiple sensors. We compare its results with other's method [30] for benchmark.

2 Related Works

Many researchers have attempted to classify the sleep stages using extracted features of the physiological data automatically. Radha et al. [28] compared the performance of six different EEG signals using various signal processing feature sets including spectral-domain, time-domain and nonlinear features as data source, and used Random Forest (RF) and Support Vector Machine (SVM) to classify the sleep stages. Their results showed that the RF with spectral linear features of frontal EEG signal achieve the

optimal real-time online classification. Huang et al. [21] used Relevance Vector Machine (RVM) with the features extracted by short-time Fourier transform (STFT) to contrast with manual scoring knowledge. Hsu et al. [20] employed Elman recurrent neural classifier with six energy features from single channel EEG to classify five sleep stages. Tsinalis et al. [36] used stacked sparse autoencoders with time-frequency analysis-based features for automatic sleep stage scoring. He also addressed the problem of misclassification errors due to class imbalance using class-balanced random sampling. Silveira et al. [34] employed variance, kurtosis and skewness of the discrete wavelet transform (DWT) of single channel EEG as features and classify the sleep stages by RF.

Ebrahimi et al. [10] employed Pz-Oz channel EEG signal, extracted features based on Wavelet Transform, and built a three-layer feed-forward Artificial Neural Network (ANN) to classify the sleep stages. Similarly, Fraiwan et al. [12] presented a feed-forward ANN based on the techniques of time-frequency analysis, which includes Wigner–Ville distribution (WVD), Hilbert–Hough spectrum (HHS) and continuous wavelet transform (CWT) for automatic sleep stage classification in neonates. Lajnef et al. [24] extracted a wide range of time and frequency-domain features of EEG, EOG and EMG. And a standard sequential forward selection (SFS) was used to select an optimal feature subset in order to facilitate their method of Dendrogram-SVM (DSVM) to classify the sleep stages. Likewise, Chapotot and Becq [7] extracted a variety of features from EEG and EMG, such as Shannon entropy and relative power of a sub-band signal, and selected the effective ones using SFS algorithm. A three-layer feed-forward ANN was then adopted to classify the stages using the selected feature set. Sen et al. [33] conducted a comparative study on EEG-based sleep stage classification in terms of feature selection and classification algorithms. Their results showed that the RF with 12 features, which are selected from 41 features by Fisher score, achieved the best classification rate. Besides using EEG, some studies also employed ECG to identify the sleep stages. Yilmaz et al. [38] demonstrated that sleep stages classification using features of RR-interval of single-lead ECG is feasible. Fonseca et al. [11] selected 80 features from a set of 142 features from ECG and respiratory (RIP) according to SFS-based feature selection, and used a linear discriminant classifier to identify the sleep stages.

These studies all relied on a domain knowledge to design or extract appropriate features. It is quite challengeable as we described in the previous section. Moreover, due to the intricacy of various physiological data, the size of the feature space can become huge so that a feature selection step is always indispensable. Recently, a few studies started to explore the automatic feature learning for sleep stage classification using deep learning. The strength of deep learning methods is end-to-end learning, i.e., feature extraction, feature selection and classification are integrated into a single algorithm using only the raw data as input [36]. Tsinalis et al. [36] used CNNs to learn task-specific filters for sleep stage classification without any prior domain knowledge. Supratak et al. [35] proposed a DeepSleepNet for automatic sleep stage scoring based on raw single-channel EEG. DeepSleepNet consists of two cascaded parts: representation learning and sequence residual learning. Representation learning including two CNNs can extract time-invariant features from raw EEG. Sequence residual learning contains two layers of Bidirectional-Long Short-term Memory (Bidirectional-LSTM)

and a shortcut connection, and aims at learning transition rules among sleep stages. These deep-learning related studies are still in exploration stage.

3 Methodology

The objectives of our study are two-fold: (1) to save the efforts of computing the hand-crafted features and (2) to find a suitable method that use less number of signals to increase usability. The proposed method following a deep learning approach which integrates the data preprocessing, the feature extraction, feature selection, and classification into a single end-to-end algorithm. It takes the raw data as input rather than hand-crafted features, and consists of two main modules as shown in Fig. 2. The architecture of proposed deep learning model. the *time-invariant feature learning* and *temporal feature learning*.

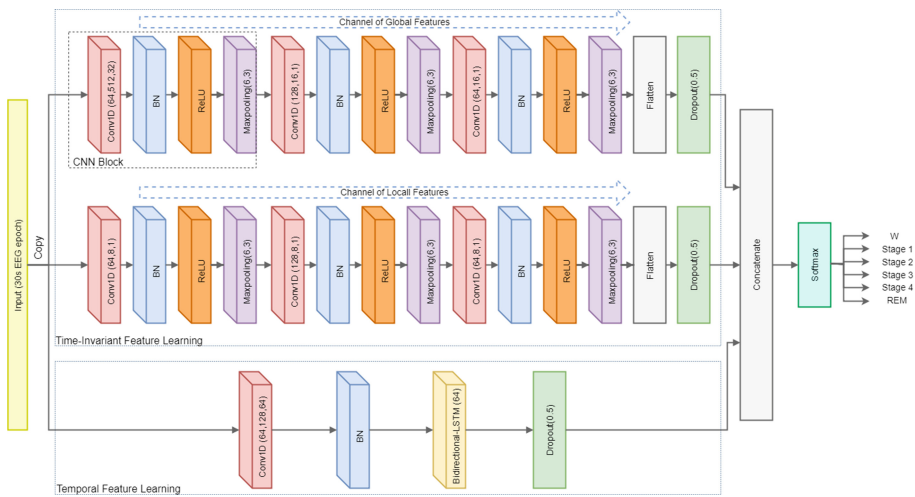


Fig. 2. The architecture of proposed deep learning model.

3.1 The Time-Invariant Feature Learning

The time-invariant feature learning module is made up of multiple Convolutional Neural Network (CNN) blocks. There are two channels in the module of time-invariant feature learning. Each channel of the time-invariant feature learning is mainly composed of three similar CNN blocks. One channel with smaller size of feature maps is designed for capturing the “small contexts” of the signal which are considered as *local features*. Another channel with larger size of feature maps is used to capture the “big contexts” of the signal which we refer them as *global features*.

The CNN block consists of four cascading layers: convolution layer, batch normalization layer, activation layer and pooling layer. After the CNN blocks, one flatten layer and one dropout layer are used. The motivation for constructing such architecture

is that the examination of physiological data for sleep stage classification is usually determined in terms of two aspects: the narrow and sharp waves, and the larger trends of the slow change. The smaller feature maps can be used to recognize the narrow and sharp waves which can be considered as local features, such as the occurrence of vertex sharp waves in stage N1. And the larger receptive fields are used to learn the larger context of waves which also can be considered as global features, such as the occurrence of slow waves in stage N3 and N4.

In order to explain how these two channels work, it is important to know the receptive field (RF) and the effective receptive field (ERF). RF represents the area of previous layer which is directly connected to a current layer's neuron [25]. ERF denotes the area of the original input data which is indirectly connected to a current layer's neuron. Both RF and ERF can be affected by the convolutional layer and the pooling layer. Then we can easily see that the channel of local features was obtained by computing the 8 points of input EEG data of in the first layer, which is a very small region. And the channel of local features was obtained by computing the 512 points of input EEG data of the first layer, which is a relatively big region. For the subsequent layers, RF is not clear to understand anymore because it always depends on its previous layer. So, we need to compute the ERF of them. For instance, the ERF of the last layer of the channel of local feature is 157, and the ERF of the last layer of the channel of global feature is 8352. That is, in the last layer, each neuron of local feature is computed from 157 data points of the original input EEG, and each neuron of global feature is computed from 8352 data points of the original input EEG. In addition, since the length of each 30 s EEG data is 7500, each neuron of global feature is actually computed from entire input data. Hence, the channel with small receptive field can learn the local features and the channel with big receptive field can learn the global features.

3.2 The Temporal Feature Learning

The LSTM [19] is one kind of Recurrent Neural Network (RNN) which employs a memory cell to store information temporally so that it is better to utilize the information in a long period of time. However, it is hard to handle very long sequential data due to the gradient issue and memory issue. So, we adopt a convolutional layer with medium size of feature maps to extract a shorter representation of the input data and a batch normalization layer to maintain the stability as shown in Fig. 2.

Figure 3 shows a single cell of LSTM. It contains an input gate i , an output gate o , a forget gate f and a memory cell. The input gate controls which new input feeds into the cell, the forget gate decides which information stored in the cell, and the output gate determines which information is used to compute the output. These gates are connected with each other, and some of the connection are recurrent. Bidirectional LSTM [32] is an extended version of LSTM. It contains two LSTMs, a forward one and a backward one. And hence, it has capability to exploit information from the past and the future.

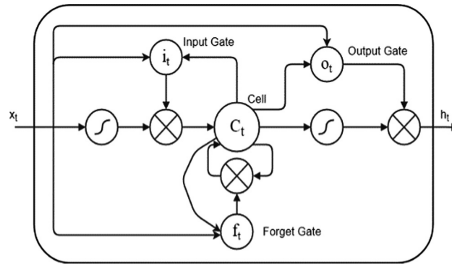


Fig. 3. Long short-term memory cell.

4 Experiment and Results

4.1 Dataset Used

We use the data sets from the MIT-BIH polysomnographic database [13] to examine the relative performance of our proposed model. The database contains over 80 h' worth of four-, six-, and seven-channel polysomnographic recordings during sleep, with sampling rate 250 Hz. Sleep stages were annotated at 30-s intervals based on the R&K rules. It has seven stages: Stages 1, 2, 3, 4, REM, W and MT, where the first five stages are introduced in Sect. 1, W denotes the wake stage, and MT represents the movement time. MT is not presented because it no longer affects the scoring of sleep stages according to AASM. There are 18 subjects in total. Among them, 17 subjects are male, aged from 32 to 56, with weights ranging from 89 to 152 kg. The last one is also male but his age and weight were unknown. The details of the dataset are summarized in Table 1. The numbers in the table denotes the amount of the 30-s EEG epochs. We can easily see that the class distributions of sleep stages are imbalanced. So, we need to balance the dataset in order to learn unbiased features. And the strategy to handle imbalance data will be described in the next section.

4.2 Experiment Setup

In order to leverage the temporal information and improve the efficiency of the training process of the model, the dataset is split up as shown in Fig. 4. Though we have 18 subjects in the dataset, we only utilized 5 of them, i.e., slp01a, slp01b, slp32, slp37 and slp41 in the first experiment. There are two reasons to do so: (1) these data are collected by the same position of sensor, i.e. C4-A1; (2) our benchmark study [30] only picked these subjects and thereby it is easy for us to compare the performance. In Fig. 4, EEG records are divided into epochs of 30 s according to the R&K manual and the annotation of the dataset. Every 32 epochs of EEG record are put in one big chunk following the sequential order of the original data. So the total data can be divided into 2717 epochs and these epochs can be divided into 87 chunks. And then 10-fold cross validation is applied. The reason of using the cross-validation is that, if we split data into training, validation, testing sets, the training set may be too small to learn effective and generalized features.

Table 1. Sleep stages in the MIT-BIH polysomnographic database

Subject	EEG	Sleep stages						Total
		W	1	2	3	4	R	
slp01a	C4-A1	7	1	105	47	66	13	239
slp01b	C4-A1	180	0	128	0	27	25	360
slp02a	O2-A1	43	18	206	5	2	77	351
slp02b	O2-A1	107	14	119	0	0	29	269
slp03	C3-O1	151	105	312	78	0	74	720
slp04	C3-O1	162	60	442	33	0	23	720
slp14	C3-O1	321	188	126	30	12	36	713
slp16	C3-O1	316	108	181	22	2	65	694
slp32	C4-A1	394	27	159	43	17	0	640
slp37	C4-A1	75	21	591	0	0	11	698
slp41	C4-A1	229	230	218	13	0	90	780
slp45	C3-O1	119	54	399	51	52	81	756
slp48	C3-O1	213	241	272	2	0	31	759
slp59	C3-O1	140	105	98	50	30	35	458
slp60	C3-O1	286	344	49	0	0	31	710
slp61	C3-O1	124	88	326	103	0	79	720
slp66	C3-O1	175	143	116	5	0	0	439
slp67x	C3-O1	72	41	40	1	0	0	154
Total		3114	1789	3889	486	212	700	10180

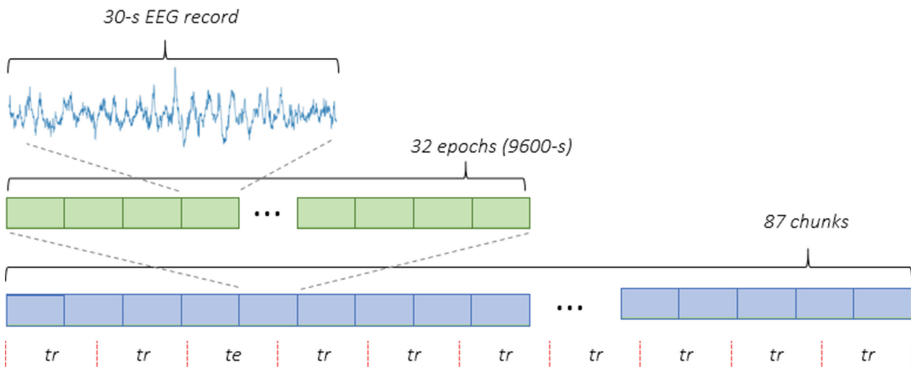


Fig. 4. The illustration of the assembly of the training and testing sets.

All models were implemented using Keras [8] with the TensorFlow [1] backend. They are trained with Adam optimization [23] with learning rate 0.00001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$, where β_1 and β_2 are used for decaying and ϵ is used to prevent any division by zero. The reason of using Adam optimizer is that it is robust to the choice of hyper-parameters to some extent [14] and it usually works very well empirically. In addition, in order to prevent the model from overfitting into the noise, the L2 regularization is applied in the first convolutional layer of every channel

according to the study by [35]. The categorical cross entropy is adopted as the loss function for the task of classifying the sleep stages.

Since the classes of the dataset are imbalance, a class weight (a.k.a., misclassification cost) scheme was used to fix this issue. Through using the class weight, the minority classes become more important. Specifically, its errors would cost more than those of the other classes. In addition, we have also tried random oversampling of the minority class and random down-sampling of the majority class to balance the dataset separately. However, the oversampling would cause overfitting problem on the data of minority class. And the down-sampling might lose information so that the model cannot correctly learn the features of the majority class.

4.3 Results and Discussion

The computational results of 10-fold cross-validation are summarized in Table 2, which compares our results and the results from [30], which utilized hand-crafted features, in terms of recall (a.k.a., True Positive Rate (TPR)), precision (a.k.a., Positive Predictive Value (PPV)) and F1 score. As shown in Table 2, we can see that if EEG signals are used to predict the sleep stages our proposed method, which uses deep learning method (CNN & LSTM), performed better than [30], which used Hidden Markov Model with hand-crafted features. Though they have carefully designed to extract and select the features, it might be suboptimal due to the difficulty of feature extraction. We used deep learning model to automatically learn the most effective features and use them to classify the stages. So, we considered the deep learning learned a better set of features than [30] in this scenario.

Table 2. Comparison of performance (%) between the proposed method with [30].

		[30]			Proposed method			Method with pretrain		
		TPR	PPV	F1	TPR	PPV	F1	TPR	PPV	F1
Sleep stages	W	80.99	92.13	86.20	81.24	81.16	81.42	86.44	83.24	84.81
	1	54.55	38.24	44.96	51.63	37.71	43.59	51.30	38.11	43.73
	2	32.79	65.79	43.76	65.45	87.72	74.96	69.27	84.38	76.09
	3	44.23	21.70	29.11	52.43	28.88	37.24	51.45	37.60	43.44
	4	80.95	55.74	66.02	74.82	49.08	55.49	65.06	55.67	60.00
	R	70.31	34.09	45.92	74.82	46.02	56.98	70.50	60.49	65.12
Avg.		60.14	51.28	52.27	64.90	58.28	54.22	65.67	59.91	62.20

Overall, in terms of average results, the proposed method with pretrain performed the best, followed by the proposed method. Specifically, all evaluation metrics of Stages 2, 3, and R are better than the benchmark model. The performance of the classification of Stage W is comparable. The classification of Stages 1 and 4 are less accurate than the benchmark model. We consider the model makes sensible mistakes here.

As shown in Fig. 5, most of the misclassification of Stage 4 are estimated as Stage 3. Stage 3 is defined by 20%–50% of the epoch consists of high voltage ($>75 \mu\text{V}$), low

frequency (<2 Hz) activity, and Stage 4 is low frequency (<2 Hz) activity where more than 50% of the epoch consists of high voltage (>75 μ V) according to R&F. That is, they are indistinguishable to each other. Hence, AASM proposed to merge them in to one stage.

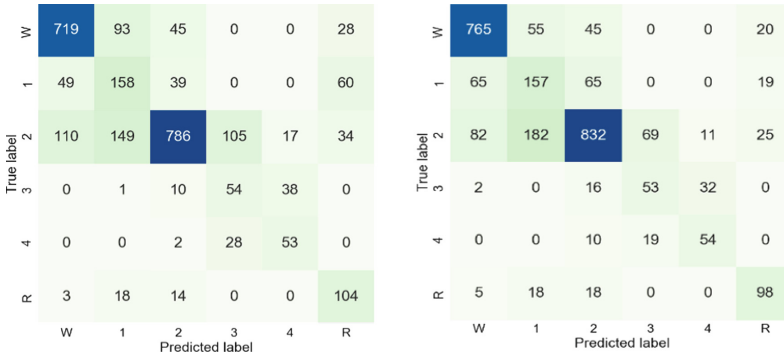


Fig. 5. The confusion matrix - the proposed method (left) and the method with pretrain (right).

However, since the data are limited, the effective features is difficult to be learned. Pretraining on similar data is one approach to overcome this issue. Hence, we pretrain the model on the remaining 13 records of MIT-BIH polysomnographic database [13] before we apply 10-fold cross-validation. The results of the proposed model with pretraining are also shown in Table 2. We can see that the recalls of Wake and Stage 2 are getting better while the others are getting worse. These can be considered as a reasonable change in its performance and there are two sensible reasons: (1) the amount of the pretraining dataset is still relatively small which cannot offer a thorough knowledge of sleep stages, and (2) there is a bias of the data of Stages 1, 3, and 4 between the training/testing dataset and pretraining dataset. In terms of the overall performance, the pretraining helps improve the method.

5 Conclusion and Future Work

In this paper, we propose a deep learning model for automatic sleep stage classification based on raw single-channel EEG without any hand-crafted features. Besides better performance, it also can save a lot of time and efforts of designing, computing and selecting the features manually. We employed CNN to learn the time-invariant features and used Bidirectional-LSTM to learn the temporal information from the past and the future. On the other side, the proposed method only needs one single EEG sensor data rather than using complete PSG recording which requires bunches of sensors. Accordingly, it can also improve the patient experience to some extent.

For future work, we plan to modify and deploy our method on some easily-collected data so that the sleep stages identification will not be so obtrusive any more.

Patient can monitor their sleep condition more easily and comfortably. Furthermore, tackling the unlabeled sleep data is also promising. Because current labeled data are limited and deep learning model usually prefer a huge amount of data to achieve a decent performance. We also plan to compare our results with results obtained from other deep learning models such as [35, 36] as they are not using the same data sets and takes time to compute using our data set.

References

1. Abadi, M., et al.: TensorFlow: a system for large-scale machine learning. In: OSDI, pp. 265–283 (2016)
2. Ancoli-Israel, S., et al.: The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. American Academy of Sleep Medicine, Westchester (2007)
3. American Sleep Disorders Association, et al.: The international classification of sleep disorders: diagnostic and coding manual. American Sleep Disorders Association (1990)
4. Basunia, M., et al.: Relationship of symptoms with sleep-stage abnormalities in obstructive sleep apnea-hypopnea syndrome. *J. Community Hosp. Intern. Med. Perspect.* **6**(4), 32170 (2016)
5. Bianchi, M.T., et al.: Obstructive sleep apnea alters sleep stage transition dynamics. *PLoS One* **5**(6), e11356 (2010)
6. Carskadon, M.A., Rechtschaffen, A.: Monitoring and staging human sleep. *Princ. Pract. Sleep Med.* **3**, 1197–1215 (2000)
7. Chapotot, F., Becq, G.: Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. *Int. J. Adapt. Control. Signal Process.* **24**(5), 409–423 (2010)
8. Chollet, F., et al.: Keras (2015)
9. Dong, H., et al.: Mixed neural network approach for temporal sleep stage classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**(2), 324–333 (2018)
10. Ebrahimi, F., et al.: Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, pp. 1151–1154 (2008)
11. Fonseca, P., et al.: Sleep stage classification with ECG and respiratory effort. *Physiol. Meas.* **36**(10), 2027 (2015)
12. Fraiwan, L., et al.: Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates. *J. Med. Syst.* **35**(4), 693–702 (2011)
13. Goldberger, A.L., et al.: Physiobank, physiotoolkit, and physionet. *Circulation* **101**(23), e215–e220 (2000)
14. Goodfellow, I., et al.: Deep learning. In: ICML2013 Tutor, pp. 1–800 (2011)
15. Güneş, S., et al.: Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Syst. Appl.* **37**(12), 7922–7928 (2010)
16. Hassan, A.R., et al.: Automatic classification of sleep stages from single-channel electroencephalogram, pp. 1–6, November 2015
17. Himanen, S.-L., Hasan, J.: Limitations of rechtschaffen and kales. *Sleep Med. Rev.* **4**(2), 149–167 (2000)

18. Hobson, J.A.: A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. In: Rechtschaffen, A., Kales, A. (eds.) 58 p. Public Health Service, US Government Printing Office, Washington, DC (1968)). (Electroencephalogr. Clin. Neurophysiol. **26**(6), 644 (1969))
19. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
20. Hsu, Y.-L., et al.: Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing* **104**, 105–114 (2013)
21. Huang, C.-S., et al.: Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels. *Front. Neurosci.* **8**, 263 (2014)
22. Humphrey, E.J., et al.: Feature learning and deep architectures: new directions for music informatics. *J. Intell. Inf. Syst.* **41**(3), 461–481 (2013)
23. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv Prepr.* [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
24. Lajnef, T., et al.: Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *J. Neurosci. Methods* **250**, 94–105 (2015)
25. Le, H., Borji, A.: What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks? *arXiv Prepr.* [arXiv:1705.07049](https://arxiv.org/abs/1705.07049)(2017)
26. Li, X., et al.: Hyclass: a hybrid classifier for automatic sleep stage scoring. *IEEE J. Biomed. Health Inform.* **22**(2), 375–385 (2018)
27. Ng, A.K., Guan, C.: Impact of obstructive sleep apnea on sleep-wake stage ratio. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4660–4663 (2012)
28. Radha, M., et al.: Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1876–1880 (2014)
29. Roebuck, A., et al.: A review of signals used in sleep analysis. *Physiol. Meas.* **35**(1), R1 (2013)
30. Rossow, A.B., et al.: Automatic sleep staging using a single-channel EEG modeling by Kalman filter and HMM. In: Biosignals and Biorobotics Conference (BRC), 2011 ISSNIP, pp. 1–6 (2011)
31. Sanei, S., Chambers, J.A.: *EEG Signal Processing*. Wiley, Hoboken (2013)
32. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **45**(11), 2673–2681 (1997)
33. Şen, B., et al.: A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *J. Med. Syst.* **38**(3), 18 (2014)
34. da Silveira, T.L., et al.: Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med. Biol. Eng. Comput.* **19**, 19 (2016)
35. Supratak, A., et al.: DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(11), 1998–2008 (2017)
36. Tsinalis, O., et al.: Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. *arXiv Prepr.* [arXiv:1610.01683](https://arxiv.org/abs/1610.01683) (2016)
37. Wang, K., et al.: Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Sci. Program.* **2016**, 9 (2016)
38. Yilmaz, B., et al.: Sleep stage and obstructive apneic epoch classification using single-lead ECG. *Biomed. Eng. Online* **9**(1), 39 (2010)
39. Zhang, X., et al.: Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device. *arXiv Prepr.* [arXiv:1711.00629](https://arxiv.org/abs/1711.00629) (2017)