



Deep Learning in Automatic Sleep Staging With a Single Channel Electroencephalography

Mingyu Fu¹, Yitian Wang¹, Zixin Chen², Jin Li³, Fengguo Xu⁴, Xinyu Liu¹ and Fengzhen Hou^{1*}

¹ School of Science, China Pharmaceutical University, Nanjing, China, ² College of Engineering, University of California, Berkeley, Berkeley, CA, United States, ³ College of Physics and Information Technology, Shaanxi Normal University, Xi'an, China, ⁴ Key Laboratory of Drug Quality Control and Pharmacovigilance, China Pharmaceutical University, Nanjing, China

OPEN ACCESS

Edited by:

Michael Döllinger,
University Hospital Erlangen, Germany

Reviewed by:

Guanghao Sun,
The University
of Electro-Communications, Japan
Kaare Bjarke Mikkelsen,
Aarhus University, Denmark

*Correspondence:

Fengzhen Hou
houfz@cpu.edu.cn

Specialty section:

This article was submitted to
Computational Physiology
and Medicine,
a section of the journal
Frontiers in Physiology

Received: 12 November 2020

Accepted: 01 February 2021

Published: 03 March 2021

Citation:

Fu M, Wang Y, Chen Z, Li J, Xu F,
Liu X and Hou F (2021) Deep
Learning in Automatic Sleep Staging
With a Single Channel
Electroencephalography.
Front. Physiol. 12:628502.
doi: 10.3389/fphys.2021.628502

This study centers on automatic sleep staging with a single channel electroencephalography (EEG), with some significant findings for sleep staging. In this study, we proposed a deep learning-based network by integrating attention mechanism and bidirectional long short-term memory neural network (AT-BiLSTM) to classify wakefulness, rapid eye movement (REM) sleep and non-REM (NREM) sleep stages N1, N2 and N3. The AT-BiLSTM network outperformed five other networks and achieved an accuracy of 83.78%, a Cohen's kappa coefficient of 0.766 and a macro F1-score of 82.14% on the PhysioNet Sleep-EDF Expanded dataset, and an accuracy of 81.72%, a Cohen's kappa coefficient of 0.751 and a macro F1-score of 80.74% on the DREAMS Subjects dataset. The proposed AT-BiLSTM network even achieved a higher accuracy than the existing methods based on traditional feature extraction. Moreover, better performance was obtained by the AT-BiLSTM network with the frontal EEG derivations than with EEG channels located at the central, occipital or parietal lobe. As EEG signal can be easily acquired using dry electrodes on the forehead, our findings might provide a promising solution for automatic sleep scoring without feature extraction and may prove very useful for the screening of sleep disorders.

Keywords: deep learning, single channel electroencephalography, automatic sleep staging, bidirectional long short-term memory, attention mechanism

INTRODUCTION

Sleep is important for the optimal functioning of the brain and the body (Czeisler, 2015). However, a large number of people suffer from sleep related disorders, such as sleep apnea, insomnia and narcolepsy (Ohayon, 2002). Effective and feasible sleep assessment is essential for recognizing sleep problems and making timely interventions.

Sleep assessment is generally based on the manual staging of overnight polysomnography (PSG) signals, including electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), blood oxygen saturation and respiration (Weaver et al., 2005), by trained and certified technicians. According to the American Academy of Sleep Medicine (AASM) manual (Iber et al., 2007), sleep can be staged as wakefulness (WAKE), rapid eye movement (REM) sleep and non-REM (NREM) sleep, which is further divided into three stages, N1, N2 and N3. Usually, it takes about 2–4 h for a technician to mark an overnight (lasting about 8 h)

PSG. The time-consuming nature of manual sleep staging hampers its application on very large datasets and limits related research in this field (Hassan and Bhuiyan, 2016a). Moreover, the inter-scorer agreement is less than 90% and its improvement remains a challenge (Younes, 2017). The multiple channels of PSG also present drawbacks preventing wider usage for the general population, due to complicated preparation and disturbance to participants' normal sleep. Therefore, the past decades have witnessed the growth of automatic sleep staging based on single-channel EEG (Liang et al., 2012; Ronzhina et al., 2012; Aboalayon et al., 2014; Radha et al., 2014; Zhu et al., 2014; Wang et al., 2015; Hassan and Bhuiyan, 2016a, 2017; Boostani et al., 2017; Phan et al., 2017; Silveira et al., 2017; Tian et al., 2017; Lngkvist and Loutfi, 2018; Seifpour et al., 2018; Sors et al., 2018; Tripathy and Acharya, 2018). These methods may eventually lead to a sufficiently accurate, robust, cost-effective and fast means of sleep scoring (Wang et al., 2015).

In the field of machine learning, deep networks are drawing more and more attention because they can learn from data directly without manual feature extraction (Lecun et al., 2015; Tsinalis et al., 2015; Dong et al., 2016; Supratak et al., 2017; Zhang and Wu, 2017; Bresch et al., 2018; Malafeev et al., 2018; Stephansen et al., 2018). There are many useful and well-established deep networks for the data mining of time series, such as the convolutional neural network (CNN) (Lecun and Bengio, 1997) and recurrent neural network (RNN) (Elman, 1990). Although CNN has mainly been applied in automated recognition of images, its application in the analysis of time series has also been notable (Chambon et al., 2018; Cui et al., 2018; Zhang and Wu, 2018; Yildirim et al., 2019). That said, it is generally demonstrated that RNN has better performance than CNN for the analysis of time series (Fiorillo et al., 2019). One of the most widely used RNN is the Long Short-Term Memory (LSTM) neural network, which is capable of capturing the long-term dependent information underlying the temporal structure of the time series (Hochreiter and Schmidhuber, 1997). Furthermore, bidirectional LSTM (BiLSTM), composed of two unidirectional LSTMs, can read data from both ends of the time series and is able to make full use of information embedded in both directions of the time series (Schuster and Paliwal, 1997). Moreover, the concept of attention is arguably one of the most powerful in the deep learning field nowadays. It is based on a common sense intuition that we "attend to" a certain part when processing a large amount of information. This simple yet powerful concept has led to many breakthroughs, not only in natural language processing tasks, such as speech recognition (Jo et al., 2010) and machine translation (Ferri et al., 2012; Karpathy and Fei-Fei, 2014; Hassan and Bhuiyan, 2017), but also in time series analysis. Recently, Zhang et al. (2019) proposed an attention-based LSTM model for financial time series prediction and a comparative analysis conducted by Hollis et al. (2018) further demonstrates that an LSTM with attention indeed outperforms a standalone LSTM for forecasting financial time series.

The application of deep neural networks for automatic sleep staging is soaring (Table 1). The PhysioNet Sleep-EDF Expanded (PSEE) dataset (Goldberger et al., 2000; Kemp et al., 2000) was

the most widely employed dataset in related studies. As shown in Table 1, Tsinalis et al. (2016) and Phan et al. (2019) reported an accuracy of 74.0% and 81.9% respectively, for 5-class sleep staging of the PSEE dataset with a CNN algorithm, while Supratak found that the combination of CNN and BiLSTM increased the accuracy to 82.4% (Supratak et al., 2017). There are also some datasets aside from PSEE that are routinely employed in studies of automatic sleep staging with a single-channel EEG and deep learning algorithms. Hsu et al. (2013) built an RNN model on the PhysioNet Sleep-EDF (PSE) dataset and achieved an accuracy of 87.2%. On the Montreal Archive of Sleep Studies (MASS) dataset, Phan et al. (2019) built a CNN model and achieved an accuracy of 83.6% while Supratak et al. (2017) built a CNN-LSTM model and obtained an accuracy of 86.2%. A CNN was also applied on the Sleep Heart Health Study (SHHS) dataset, yielding an accuracy of 87% (Sors et al., 2018). However, few works investigated whether the performance of sleep staging can be further improved by the combination of BiLSTM and the attention mechanism. Aside from that, there is a lack of comparison between the performance of deep learning based and conventional feature extraction based models.

Although deep learning algorithms have shown themselves promising in automatic sleep staging with a single-channel EEG, few studies investigated whether the performance of such algorithms is sensitive to the choice of EEG channel. Therefore, in this study, the PSEE dataset and the DREAMS Subjects (DRM-SUB) dataset (Devuyst, 2005) were used. Both datasets have more than one channel of EEG and the DRM-SUB dataset was involved in many automatic sleep staging studies with conventional feature extraction (Hassan and Bhuiyan, 2016a, 2017; Ghimatgar et al., 2019; Shen et al., 2019). A neural network named AT-BiLSTM was proposed, which uses the neural attention mechanism of the BiLSTM to classify sleep stages. For comparison, five other networks, CNN, LSTM, BiLSTM, the combination of CNN and LSTM (CNN-LSTM), and the combination of CNN and BiLSTM (CNN-BiLSTM) were also trained and tested. Our aims are threefold: first, to investigate whether AT-BiLSTM can achieve the highest performance among these networks; second, to confirm whether RNN algorithms (i.e., LSTM and BiLSTM) outperform CNN in sleep staging with single channel EEG; third, to explore whether the method of making hybrid networks further improves the performance of sleep staging.

MATERIALS AND METHODS

Datasets

The data analyzed in this study were obtained from two open-access datasets: the DRM-SUB dataset and the PSEE dataset. The DRM-SUB consists of 20 whole-night PSG recordings (lasting 7–9 h) obtained from 20 subjects (four males and 16 females, 20–65 years old). Three EEG channels located in different lobes (Cz-A1, Fp1-A1 and O1-A1) were included in DRM-SUB, with a sampling rate of 200 Hz. To investigate the impact of the choice of EEG derivations on the performance of automatic sleep staging, EEG

TABLE 1 | An overview of the application of deep networks on sleep staging.

Authors	Dataset	Channel	Model	Accuracy
Tsinalis et al.	PSEE	Fpz-Cz	CNN	74.0%
Phan et al.	PSEE	Fpz-Cz	CNN	81.9%
Supratak et al.	PSEE	Fpz-Cz	CNN-BiLSTM	82.4%
Hsu et al.	PSE	Fpz-Cz	RNN	87.2%
Phan et al.	MASS	C4-A1	CNN	83.6%
Supratak et al.	MASS	F4-EOG (Left)	CNN-BiLSTM	86.2%
Sors et al.	SHHS	C4-A1	CNN	87.0%

TABLE 2 | Data distribution of sleep stages in both datasets.

Dataset	Total epochs	WAKE (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)
PSEE	41663	19.2	6.6	42.2	13.4	18.5
DRM-SUB	20265	17.6	7.3	40.7	19.4	14.9

signals from all three channels were used separately for the following analysis.

Twenty healthy subjects (10 males and 10 females, 25–34 years old) from the PSEE dataset were also included. There are two EEG channels (Fpz-Cz and Pz-Oz) available in the PSEE dataset, with a sampling rate of 100 Hz. For each subject, two PSGs of about 20 h each were recorded during two subsequent day-night periods at the subjects' homes. In order to remain consistent with previous studies (Supratak et al., 2017), for each subject and each PSG, only the data from 30 min before sleep-onset (i.e., the first sleep epoch after light-off in the evening) and 30 min after the last sleep epoch in the morning were included. Both channels were investigated separately.

For both datasets, labels of sleep staging for each 30-s EEG epoch were provided by the data distributors according to AASM rules. Five staging classes, i.e., WAKE, N1, N2, N3, and REM were used in this study. The distribution of 30-s EEG epochs of both datasets is illustrated in **Table 2**.

Construction of the AT-BiLSTM Network

The proposed AT-BiLSTM network architecture for automatic sleep staging is illustrated in **Figure 1**. It is composed of two main components, three stacked BiLSTM layers for feature extracting and one attention layer to weight the most relevant parts of the input sequence. According to a preset parameter, called the input dimension m , each raw 30-s EEG epoch is divided into multiple vectors, which are fed into the BiLSTM part sequentially to construct a feature matrix. Then to emphasize the different importance of different vectors, an attention layer is applied in the intra-epoch feature learning and summarizes the outputs of the BiLSTM part with different weights. Finally, the probability of each sleep stage can be derived from a fully connected (FC) layer and a softmax layer.

Given a 30-s EEG epoch $X = [x_1, x_2, \dots, x_N]$ with N data points, a moving window with input dimension of m is applied to X without overlap, leading to the matrix form of X , as shown in Equation 1, where n equals to N/m and X_t represents the vector

in time step t .

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_t \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_{m+1} & x_{m+2} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{(t-1)m+1} & x_{(t-1)m+2} & \cdots & x_{tm} \\ \vdots & \vdots & \cdots & \vdots \\ x_{(n-1)m+1} & x_{(n-1)m+2} & \cdots & x_{nm} \end{bmatrix} t \in [1, n] \quad (1)$$

All the vectors are fed into the first BiLSTM layer, forward and backward respectively. For time step t , the output of the forward or backward network, denoted as h_t^f or h_t^b , can be obtained, respectively, according to Equation 2 or 3.

$$h_t^f = \sigma(W_{fx}x_t + W_{ff}h_{t-1}^f + b_f) \quad (2)$$

$$h_t^b = \sigma(W_{bx}x_t + W_{bb}h_{t-1}^b + b_b) \quad (3)$$

where σ is the logistic sigmoid function, W is the weight matrix (e.g., subscription “ fx ” in W represents the forward network of x_t) and b is the bias vector of the network (b_f and b_b represents the bias vector of forward and backward network, respectively).

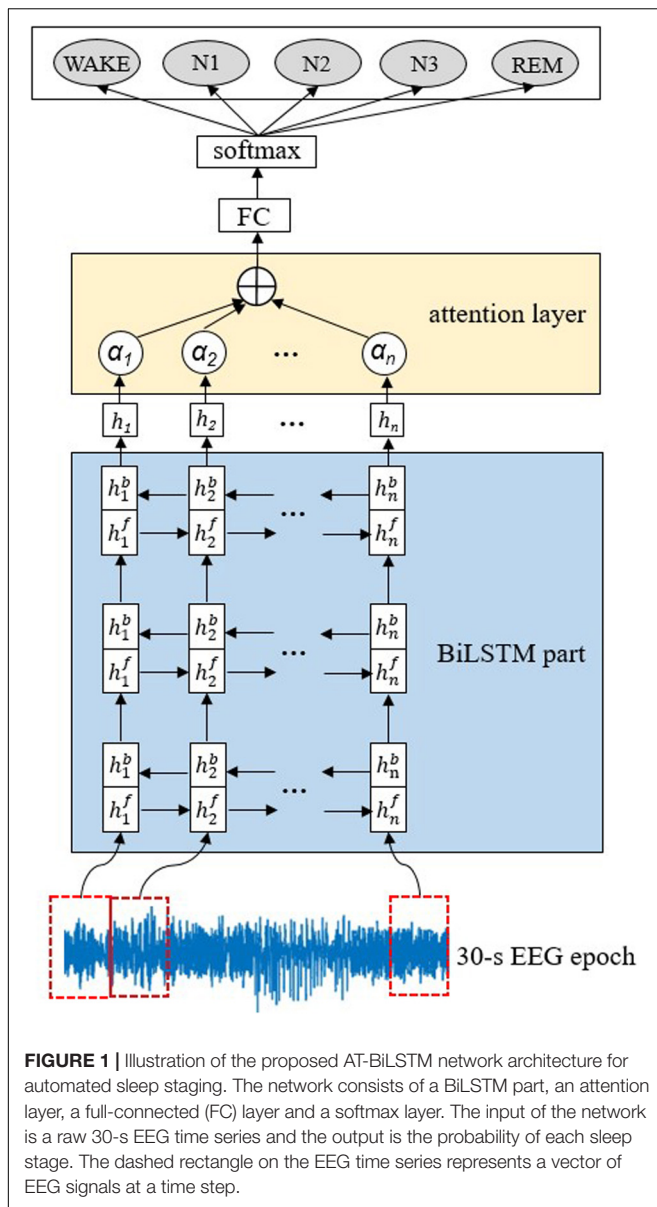
The weighted sum of h_t^f and h_t^b , denoted as h_t , is computed as the output of the first BiLSTM layer following Equation 4.

$$h_t = W_{hf}h_t^f + W_{hb}h_t^b + b_h \quad (4)$$

The output of the previous BiLSTM layer is fed into the next layer in the same way. The third layer gives the final output of the BiLSTM part, which is weighted by the attention layer before feeding into the FC layer. Considering that EEG signal in different time steps should contribute differently to the classification task, it is rational to give strong weights to the more discriminative parts and vice versa. Formally, the attention weight a_t at the time step t is computed according to Formula (5) – (6).

$$u_t = \tanh(w_w h_t + b_w) \quad (5)$$

$$a_t = \frac{\exp(u_t^T u_w)}{\sum_t \exp(u_t^T u_w)} \quad (6)$$



In Formula (5)–(6), \mathbf{u}_t represents the state of the hidden layer obtained from a simple neural network, \mathbf{u}_w represents a weight vector randomly initialized, \mathbf{a}_t represents the similarity between \mathbf{u}_t and \mathbf{u}_w obtained by softmax function.

$$\mathbf{s}_t = \sum_t \mathbf{a}_t \mathbf{h}_t \quad (7)$$

By weighting and summing the output of the BiLSTM part, the attention vector, denoted as \mathbf{s}_t , can be obtained and fed into FC layer, preceding to the softmax layer which finally yields the probability of each sleep stage.

Construction of Baseline Networks

Apart from the proposed AT-BiLSTM network, we also constructed five baseline networks, including three single

networks, i.e., CNN, LSTM and BiLSTM, and two hybrid networks, i.e., CNN-LSTM and CNN-BiLSTM.

Single Networks

Figure 2A illustrated the CNN topology used in this study, which is fed with a matrix reconstructed from a raw 30-s EEG epoch according to Equation 1. It consists of three convolution blocks and three max pooling layers. Each convolutional block contains a one-dimensional convolutional layer and a rectified linear unit (ReLU) activation layer. The input matrix is padded with zeros to ensure that the number of rows in the matrix is constant during the convolutional process. The output of CNN is fed into a FC layer, then activated by softmax function to obtain the sleep stage probability.

Two scenarios were considered in single RNN network. In the first scenario, three layers of LSTMs were stacked, also followed by a FC layer and a softmax layer. The second scenario employed stacked BiLSTM layers instead of the LSTM layers.

Hybrid Networks With CNN and RNN

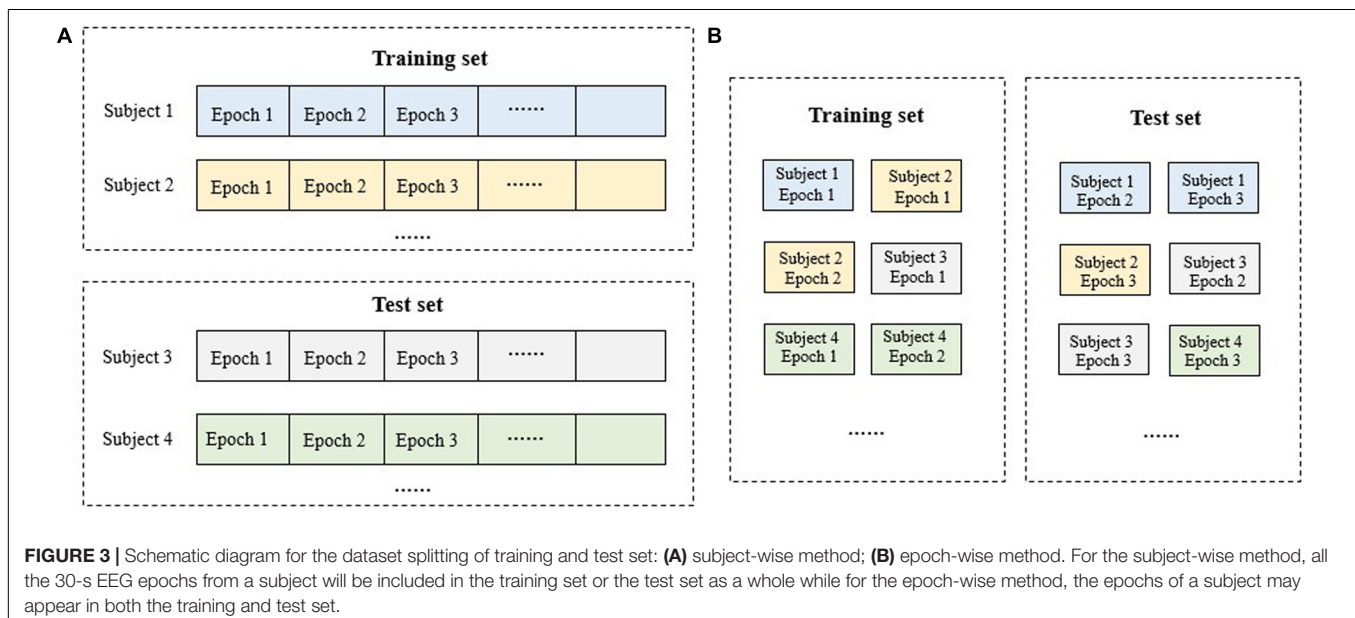
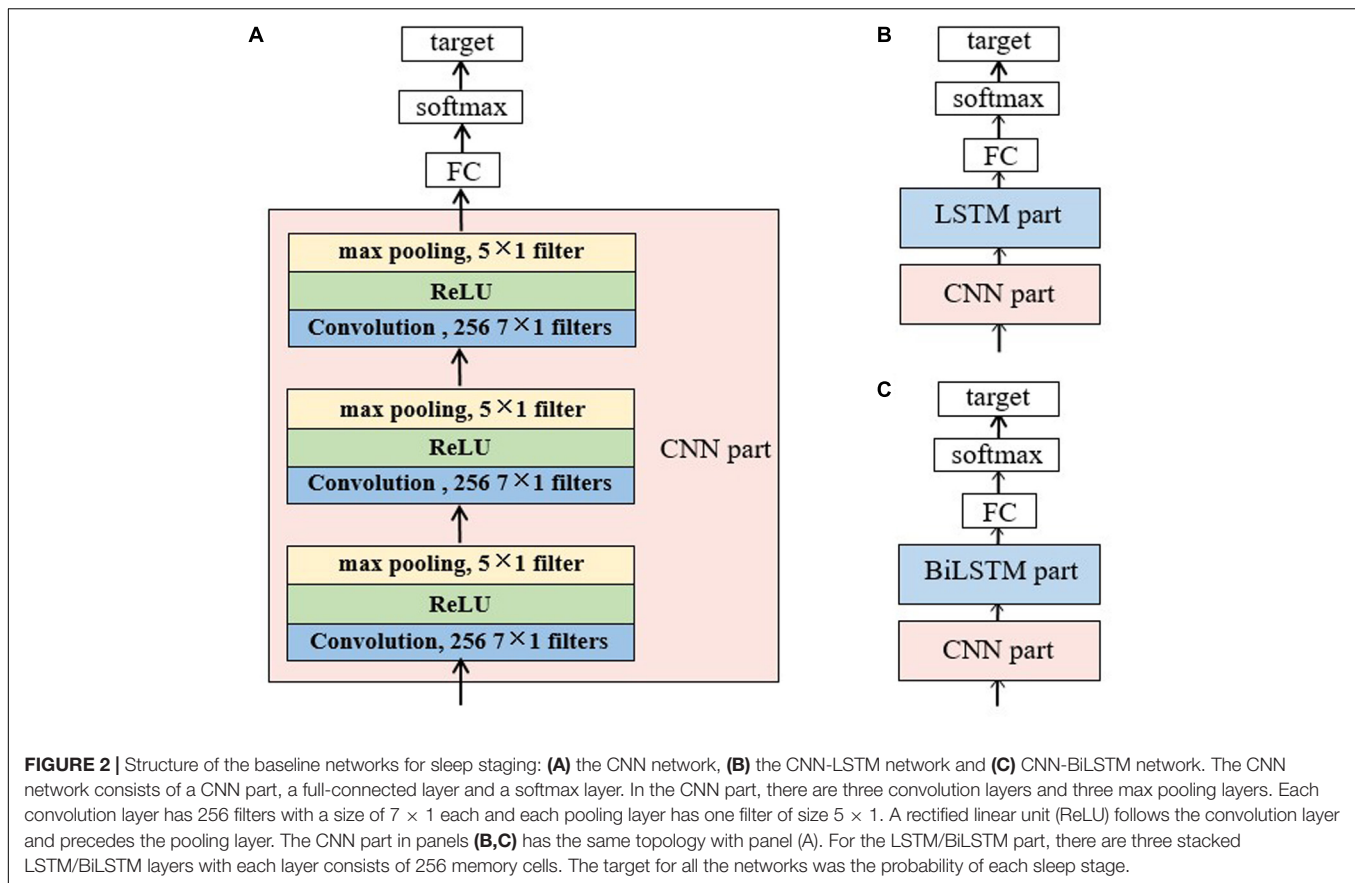
As shown in Figure 2B,C, a CNN part followed by an RNN part was adopted in the hybrid networks, in order to make use of RNN for further processing the features extracted by CNN. The structures of the CNN part and RNN part are the same with the single networks aforementioned.

Datasets Splitting Strategy

Machine learning algorithms require independent training and test sets for model training and performance evaluation. Also, k-fold cross validation is preferred in application. Generally, there are two types of training data partitioning for clinic data: subject-wise and epoch-wise (Figure 3). For the subject-wise method, all the subjects were split into k folds equally and onefold is taken as the test set in turn while the remains as the training set. For the epoch-wise method, all the 30-s EEG epochs from all the subjects were merged and then split into k equal folds for each stage randomly. That is, for each sleep stage, all the 30-s EEG epochs from all the subjects were collected and divided into k folds. Consequently, the epochs of a subject may appear in both the training and test set, violating the independence between the training and test set and contributing to a virtual high performance. Thus, in the present study, the subject-wise method with fivefold cross validation was adopted. The model was trained using the training set and evaluated using the test set. Finally, all evaluation results were combined.

Experimental Setting and Network Optimization

Using the first fold as the test set, the network parameters, such as the input dimension, the number of hidden units in each LSTM/BiLSTM/convolutional layer, and the filter/stride size of each convolutional layer and pooling layer, were determined by a grid-search to minimize the errors of networks with Python 3.6 and TensorFlow v1.15.0 (Abadi et al., 2016). The standard cross-entropy loss was used as the loss function in model training due to its good performance in measuring the errors of networks with



discrete targets (Boer et al., 2005). Each network was trained for 30 epochs with a mini batch size of 64 sequences. As a result, the input dimension m was set as 5, the number of hidden units as 256, and the stride size for both convolution layers and max pooling layers as 1×1 . The filter size of each convolutional

layer and max pooling layer in CNN were set to 1×7 and 1×5 respectively.

For backpropagation, the adaptive moment estimation (ADAM) algorithm was adopted because it solves the optimization problem in non-stationary conditions and

works faster than the standard gradient descent algorithm and the root mean square propagation (Kingma and Ba, 2017). The main hyper-parameters used for ADAM algorithm were set as: learning rate ($\alpha = 0.001$), gradient decay factor ($\beta_1 = 0.9$), squared gradient decay factor ($\beta_2 = 0.999$), and epsilon ($\epsilon = 10^{-8}$) for numerical stability. Moreover, a dropout layer before the last FC layer was used to avoid over-fitting and its dropout rate was set to 0.2, leading to 20% of the weights dropped during the training phase.

Performance Metrics

Overall metrics, including accuracy, macro F1-score (MF1) and Cohen's kappa (κ) were used to evaluate the performance of each model. Performance on individual sleep stages was also assessed *via* class-wise precision and sensitivity.

Cohen's kappa coefficient is a statistical measure of inter-rater agreement for categorical items (Cohen, 1960). When two persons (algorithms or raters) try to evaluate the same data, Cohen's Kappa coefficient, κ , is used as a measure of agreement between their decisions. In this study, it measures the amount of agreement between the output of the proposed algorithm and the provided labels of sleep stages.

Another metric used for performance evaluation here is the area under the receiver operating characteristics (ROC) curve, called AUC. The ROC curve is a graphical tool and demonstrates the classification performance by plotting the true positive rate (TPR) against the false positive rate (FPR) at different classification thresholds (Zweig and Campbell, 1993). Furthermore, it provides a convenient way for selecting the threshold that provides the maximum classification TPR while not exceeding a maximum allowable FPR level (Kim et al., 2019). For an n -class classification task, n ROC curves can be obtained by splitting the task into n binary classification tasks. For each binary classification task, its AUC value can be used as a class-wise measure of performance and the macro-average AUC of these tasks can be regarded as an overall metric for the performance evaluation.

RESULTS

Table 3 shows the overall performance of different networks on the PSEE dataset. The proposed AT-BiLSTM network outperforms other networks with overall accuracy, κ , MF1

and MAUC of 83.78%, 0.766, 82.14% and 97.45% on channel Fpz-Cz, respectively and an overall accuracy, κ , MF1 and MAUC of 80.79%, 0.731, 79.27% and 96.33% on channel Pz-Oz, respectively. The AT-BiLSTM network performs better than the other networks overall. For the single networks, the RNN-based networks outperform the CNN network while the results of BiLSTM and LSTM are comparable. The hybrid networks further improve the overall performance compared to the single models. Moreover, AT-BiLSTM achieves better precision and sensitivity on N3 and REM than the hybrid networks with CNN and RNN, although they have a comparable performance on stages Wake, N1 and N2. Furthermore, better performance is found in Fpz-Cz than Pz-Oz channel, regardless of the network topology used, indicating EEG derived from the frontal lobe is more valuable than those from the parietal lobe in sleep staging.

Table 4 shows the performance of different networks on the DRM-SUB dataset. The AT-BiLSTM network still outperforms other networks, suggesting its good generalization in sleep staging. Consistent with the results in PSEE dataset, the frontal EEG channel (Fp1-A1 here) achieves the best performance. The results are in line with a recent work, which found that EEG signals from an Fp1-A1 channel yielded higher accuracy values in automatic sleep staging than those of a Cz-A1 or O1-A1 channel (Ghimatgar et al., 2019).

Figure 4 illustrates the hypnograms labeled manually by a clinical technician of sleep and by the trained AT-BiLSTM model. The corresponding EEG recoding was obtained from the first person in PSEE dataset (SC4001E0), who spent 7 h during sleep. Noting that the subject is located in the test set for the trained model. The accuracy of the automatic sleep staging for this subject is 87.30%, showing considerable reliability of the proposed AT-BiLSTM network. Most of the wrong classifications were made during the transitions from one stage to another.

Table 5 shows the class-wise performances obtained on the PSEE dataset. For most stages, better performance is achieved by the AT-BiLSTM model than the baseline networks and Fpz-Cz channel outperforms the Pz-Oz one. Although the classification accuracy of stage N1 is significantly lower than that of the other stages, which might be due to the small percentage of N1 during sleep, it is higher than those reported in previous studies (Hsu et al., 2013; Supratak et al., 2017). Similar findings can be found on the DRM-SUB (**Table 6**).

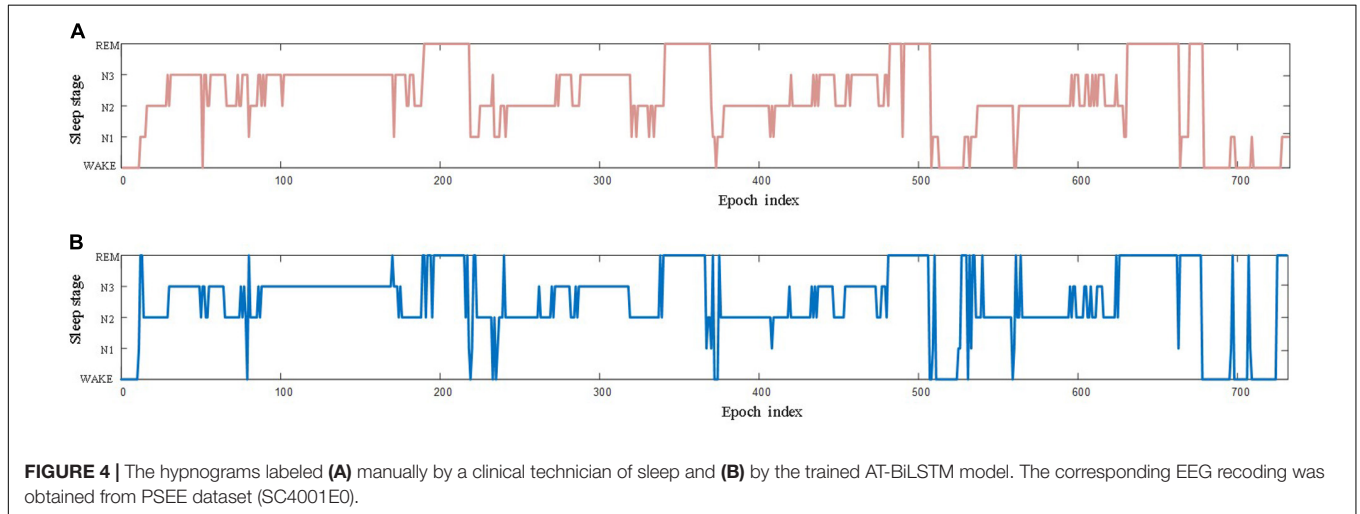
Furthermore, ROC curves were used to compare the performances of the proposed AT-BiLSTM model

TABLE 3 | The overall performance of different networks on the PSEE dataset (value in bold represents for the best among all the networks).

Networks	Fpz-Cz				Pz-Oz			
	Acc.	κ	MF1	AUC	Acc.	κ	MF1	AUC
AT-BiLSTM	83.78	0.766	82.14	96.08	80.79	0.731	79.27	93.63
CNN	78.84	0.706	76.10	92.89	76.45	0.669	74.56	89.91
LSTM	81.59	0.747	79.25	95.36	79.02	0.706	75.92	92.14
BiLSTM	81.48	0.740	80.13	93.78	78.95	0.707	77.44	91.81
CNN-LSTM	82.58	0.759	80.40	93.96	79.51	0.718	76.44	92.36
CNN-BiLSTM	82.58	0.759	81.15	94.67	79.37	0.710	77.92	92.70

TABLE 4 | The overall performance of different networks on the DRM-SUB dataset (value in bold represents for the best among all the networks).

Networks	Fp1-A1				Cz-A1				O1-A1			
	Acc.	κ	MF1	AUC	Acc.	κ	MF1	AUC	Acc.	κ	MF1	AUC
AT-BiLSTM	81.72	0.751	80.74	94.99	81.62	0.749	80.76	95.25	77.09	0.685	75.98	94.91
CNN	77.84	0.732	67.17	89.90	75.82	0.664	73.98	91.00	72.45	0.617	70.84	90.59
LSTM	80.19	0.738	70.96	94.29	80.53	0.733	80.23	94.65	74.13	0.641	72.32	92.97
BiLSTM	80.31	0.739	70.41	93.84	80.41	0.733	79.58	94.66	74.22	0.644	72.98	92.85
CNN-LSTM	80.55	0.738	71.96	94.43	80.87	0.736	79.24	94.50	75.78	0.665	74.52	93.73
CNN-BiLSTM	80.61	0.737	71.62	93.87	80.83	0.738	80.71	94.68	75.94	0.666	74.42	93.79

**TABLE 5** | The class-wise performance obtained on the PSEE dataset (value in bold represents for the best among all the networks).

EEG signal	Networks	Precision					Sensitivity					Class-wise AUC				
		W	N1	N2	N3	REM	W	N1	N2	N3	REM	W	N1	N2	N3	REM
Fpz-Cz	AT-BiLSTM	86.38	45.06	87.82	88.88	76.84	89.42	25.76	89.18	89.62	82.18	97.58	88.85	97.31	99.34	97.33
	CNN	84.68	34.32	84.78	82.66	67.08	83.88	19.34	84.98	82.63	77.22	95.98	82.00	94.18	97.99	94.32
	LSTM	84.72	43.28	85.98	88.68	71.92	87.74	20.68	88.16	85.12	80.14	98.72	84.85	97.25	99.08	96.94
	BiLSTM	81.34	41.23	86.95	86.71	69.90	88.32	20.84	86.88	89.42	81.43	96.70	81.18	96.54	99.04	95.45
	CNN-LSTM	85.72	45.16	88.46	87.22	71.04	90.14	12.46	86.54	88.46	79.22	96.52	81.29	96.77	99.17	96.09
	CNN-BiLSTM	85.16	42.51	87.56	87.42	75.72	88.72	25.34	87.42	88.64	79.76	97.19	83.26	96.96	99.22	96.74
Pz-Oz	AT-BiLSTM	82.58	40.24	84.64	84.84	71.76	82.58	40.24	84.64	84.84	71.76	96.23	82.54	96.08	98.76	94.52
	CNN	78.48	24.18	81.16	79.42	63.28	78.48	24.18	81.16	79.42	63.28	94.59	75.96	92.70	95.17	91.14
	LSTM	79.84	41.82	82.94	82.36	64.82	79.84	41.82	82.94	82.36	64.81	95.34	78.34	94.80	98.35	93.88
	BiLSTM	80.64	42.94	83.78	82.26	66.70	80.64	42.94	84.28	82.26	66.74	95.19	76.55	95.16	98.46	93.69
	CNN-LSTM	80.95	42.65	83.95	82.55	69.75	80.95	42.65	83.95	82.55	69.75	95.31	78.72	95.26	98.48	94.04
	CNN-BiLSTM	79.52	44.62	84.26	83.04	70.37	79.55	44.62	84.26	83.04	70.38	95.79	80.53	95.16	98.39	93.63

for different sleep stages with the frontal channels in both datasets (Figure 5). As shown in Figure 5, AT-BiLSTM is sufficient to identify WAKE, N3 and REM, but insufficient to identify N1.

Table 7 illustrates the results of a comparison between the proposed AT-BiLSTM model and the state-of-the-art works using the same dataset of DRM-SUB (Hassan and Bhuiyan, 2016a,b; Hassan and Subasi, 2017; Ghimatgar et al., 2019; Shen et al., 2019). With the same dataset, same EEG channel and

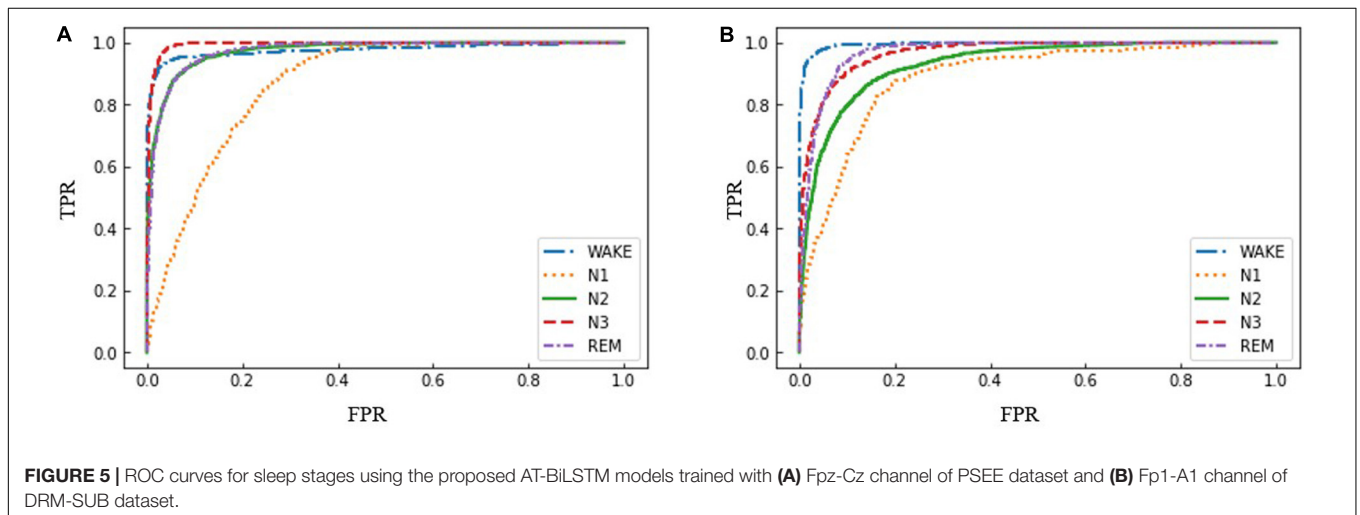
same dataset splitting strategy, the proposed AT-BiLSTM model achieves the highest accuracy.

DISCUSSION

In this study, we proposed an AT-BiLSTM network for automatic sleep staging with single-channel EEG. The main findings were: (1) the frontal EEG derivations contribute to better

TABLE 6 | The class-wise performance obtained on the DRM-SUB dataset (value in bold represents for the best among all the networks).

EEG signal	Networks	Precision					Sensitivity					Class-wise AUC				
		W	N1	N2	N3	REM	W	N1	N2	N3	REM	W	N1	N2	N3	REM
Fp1-A1	AT-BiLSTM	88.48	45.92	84.98	89.08	68.26	89.34	25.88	85.12	85.18	83.06	99.50	88.80	93.05	96.69	96.93
	CNN	83.54	40.82	79.76	87.14	63.02	84.22	11.54	85.92	78.52	78.96	97.85	82.77	87.05	86.85	94.96
	LSTM	83.63	44.16	82.58	88.26	68.14	89.92	19.18	84.82	84.74	75.28	99.35	87.06	92.33	96.40	96.33
	BiLSTM	85.24	43.82	83.28	86.82	66.28	86.78	17.36	85.31	86.83	77.74	98.90	85.83	92.01	96.39	96.07
	CNN-LSTM	87.12	42.26	84.52	86.26	66.44	87.04	23.34	84.52	87.66	79.91	99.26	88.22	91.94	96.61	96.13
	CNN-BiLSTM	86.62	46.92	83.88	86.74	66.38	89.98	20.74	84.54	85.88	79.86	99.21	86.30	92.04	96.28	95.54
Cz-A1	AT-BiLSTM	88.02	42.02	84.98	87.56	69.22	90.96	22.22	86.32	86.28	80.92	99.40	89.24	93.62	97.10	96.87
	CNN	83.68	23.44	75.18	86.14	65.12	86.55	8.92	85.36	73.28	72.68	98.18	82.81	88.39	90.85	94.77
	LSTM	87.24	43.88	83.44	87.76	69.26	89.54	22.34	86.12	86.26	77.04	99.14	88.59	92.39	97.06	96.09
	BiLSTM	86.46	39.68	83.82	88.58	67.12	89.24	20.62	85.36	83.38	78.42	99.27	88.56	92.57	96.77	96.14
	CNN-LSTM	84.34	48.72	81.44	89.58	70.64	91.34	14.16	87.66	82.46	78.06	98.99	88.43	92.42	96.49	96.15
	CNN-BiLSTM	86.22	42.96	82.32	87.98	70.88	90.36	23.68	86.78	84.26	76.58	99.30	88.68	92.48	96.66	96.29
O1-A1	AT-BiLSTM	88.86	46.58	78.76	83.20	62.56	91.36	18.53	81.94	78.67	72.74	99.55	89.76	92.40	96.67	96.18
	CNN	86.21	38.78	73.48	82.92	52.02	84.38	7.56	80.62	72.40	65.38	98.25	82.82	87.94	92.30	91.63
	LSTM	83.66	28.76	74.96	83.98	55.44	90.44	8.12	81.88	75.96	62.66	99.17	84.52	91.31	96.09	93.75
	BiLSTM	89.92	39.72	75.06	80.52	56.02	87.72	13.14	79.64	81.02	62.96	99.24	84.52	91.28	95.89	93.33
	CNN-LSTM	90.52	42.81	76.32	82.25	59.18	88.76	13.56	82.34	78.86	68.73	99.23	87.46	91.39	95.13	95.46
	CNN-BiLSTM	88.04	45.96	76.26	82.44	60.84	89.92	11.72	82.18	79.62	67.68	99.26	86.43	91.41	96.40	95.44

**FIGURE 5** | ROC curves for sleep stages using the proposed AT-BiLSTM models trained with (A) Fpz-Cz channel of PSEE dataset and (B) Fp1-A1 channel of DRM-SUB dataset.**TABLE 7** | Comparison of sleep staging performance on the DRM-SUB dataset between the proposed method and previous works based on conventional feature extraction.

Authors	Year	Methodology	Dataset splitting strategy	Channel	Accuracy
Hassan and Bhuiyan, 2016a	2016	Tunable Q-factor wavelet transform, random forest (Hassan and Bhuiyan, 2016a)	Epoch-wise	Fp1-A1	72.28%
Hassan and Bhuiyan, 2016b	2016	Implementation of ensemble empirical mode decomposition in conjunction with random under sampling boosting (Hassan and Bhuiyan, 2016a)	Epoch-wise	Fp1-A1	74.59%
Hassan and Subasi, 2017	2017	Tunable Q-factor wavelet transform, bagging (Hassan and Subasi, 2017)	Epoch-wise	Fp1-A1	78.95%
Shen et al.	2019	Essence features extraction method (Shen et al., 2019)	Subject-wise	Cz-A1	80.90%
Ghimatgar et al.	2019	Features in time domain, frequency domain, cepstral domain, wavelet features, autoregressive coefficients and non-linear features with Hidden Markov Model (Ghimatgar et al., 2019)	Subject-wise	Fp1-A1	81.22%
Proposed method		Raw EEG signal and AT-BiLSTM	Subject-wise	Fp1-A1	81.72%

performance of sleep staging than those located in the central, occipital or parietal lobe; (2) the proposed AT-BiLSTM network outperforms the other networks based on CNN or RNN; (3) The proposed deep learning network achieves higher accuracy than conventional feature extraction methods.

Two EEG datasets, i.e., PSEE and DRM-SUB, with different EEG derivations were used in our study. To clarify the influence of the EEG channel on automatic sleep staging, here we applied the proposed method to all the EEG channels in both datasets. The results obtained from both datasets are similar: the model adopting frontal derivation behaved better than those from other lobes. Such a finding indicated that the performance of sleep scoring was sensitive to the selection of EEG channel and the derivations from the frontal region are the optimal choices. Physiologically, the prefrontal cortex is deactivated and reactivated during the sleep cycle, indicating its involvement in the wake-sleep cycle (Maquet et al., 1996). With the development of wearable EEG devices, EEG signals can be easily obtained using dry electrodes on the forehead (Hassan and Bhuiyan, 2016a); the proposed method would be promising in supporting people monitoring sleep.

In recent years, many automated sleep staging methods based on deep neural networks used CNNs for feature extraction and RNNs to capture temporal information. These approaches have significantly improved the accuracy of sleep staging (Hassan and Bhuiyan, 2016a; Boostani et al., 2017; Sors et al., 2018). In general, for the sequence-to-label model based on RNN, only the output vector at the last time step is retained for classification, e.g., *via* a softmax layer (Phan et al., 2017). However, it is reasonable to combine the output vectors of different time steps by some weighting schemes. Intuitively, those parts of the input sequence which are essential to the classification task at hand should be associated with strong weights, and those with less importance should be weighted correspondingly less. Ideally, these weights should be automatically learned by the network. This can be accomplished with an attention layer (Luong et al., 2015). Besides, previous works demonstrated that the performance of classification or regression can be further improved by stacking multiple BiLSTM in neural networks (Liu et al., 2017; Wang et al., 2018; Liu et al., 2018). Aside from that, we found the overall performance of the RNN based model to be better than that of the CNN models in automatic sleep staging, which might indicate that the RNNs are promising in capturing the temporal nature of an EEG time series. From such a perspective, the highest performance achieved by the proposed AT-BiLSTM might further confirm the role of stacking layers and attention mechanism in feature extracting of time series.

In this study, all experiments were performed on a server configured with 64 CPUs [Intel(R) Xeon(R) CPU @ 2.10 GHz], 64 GB memory, a GPU (NVIDIA GeForce GTX 1,080 Ti] and a Windows Server 2016 system. A CNN

network has the lowest computational cost as its training time for each batch was 0.16 s on average. LSTM and CNN-LSTM networks take similar times (8.46 and 8.60 s respectively) for each batch in training. The computational cost of BiLSTM based networks is twice that of LSTM based networks because they must calculate the input sequence in two directions and set up double parameters. Moreover, approximately 1.3 s more is required for each batch with the attention layer.

Our study demonstrated that a deep learning approach without manual feature extraction can also provide sufficient accuracy for sleep staging, which is even better than conventional methods based on manual feature extraction. Therefore, the proposed method is a promising choice for computer-aided detection of sleep stages and similar 1-D signal classification problems. In conclusion, our findings provide a possible solution for automatic sleep scoring without manual signal preprocessing and feature extraction. With the development of wearable EEG devices, such a solution would be valuable in the screening of sleep disorders at home for the general population.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Dreams Subjects: <https://zenodo.org/record/2650142#.X6tbymgzZdg>. Sleep-EDF Database Expanded: <https://physionet.org/content/sleep-edfx/1.0.0/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of two open-access datasets, i.e., the Sleep-EDF Expanded dataset available at Physionet and the DREAMS Subjects dataset. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FH, XL, FX, and JL designed this study. MF and YW analyzed the data. MF, FH, and ZC wrote the article. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (Grant No. 11974231) and the Double First-Class University Project of China.

REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1603.04467> (accessed February 15, 2020).
- Aboalayon, K. A., Ocbagabir, H. T., and Faezipour, M. (2014). Efficient sleep stage classification based on EEG signals. *IEEE LISAT 2014*, 978–983. doi: 10.1109/LISAT.2014.6845193
- Boer, P., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Ann. Operat. Res.* 134, 19–67. doi: 10.1007/s10479-005-5724-z
- Boostani, R., Karimzadeh, F., and Nami, M. J. (2017). A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput. Methods Prog. Biomed.* 140, 77–91. doi: 10.1016/j.cmpb.2016.12.004
- Bresch, E., Groekathfer, U., and Garcia-Molina, G. (2018). Recurrent deep neural networks for real-time sleep stage classification from single channel EEG. *Front. Comput. Neurosci.* 12:85. doi: 10.3389/fncom.2018.00085
- Chambon, S., Galtier, M., Arnal, P. J., and Wainrib, G. (2018). A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 758–769. doi: 10.1109/TNSRE.2018.2813138
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Cui, Z. H., Zheng, X. W., Shao, X. X., and Cui, L. Z. (2018). Automatic sleep stage classification based on convolutional neural network and fine-grained segments. *Complexity* 2018, 13. doi: 10.1155/2018/9248410
- Czeisler, C. A. (2015). Duration, timing and quality of sleep are each vital for health, performance and safety. *Sleep Health* 1, 5–8. doi: 10.1109/Trustcom.2015.524
- Devuyst, S. (2005). *The DREAMS Databases and Assessment Algorithm*. Genève: Zenodo. doi: 10.5281/zenodo.2650142
- Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P. M., and Guo, Y. (2016). Mixed neural network approach for temporal sleep stage classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 324–333. doi: 10.1109/TNSRE.2017.2733220
- Elman, J. (1990). Finding structure in time. *Trends Cogn. Sci.* 14, 179–211. doi: 10.1016/0364-0213(90)90002-E
- Ferri, R., Rundo, F., Novelli, L., and Terzano, M. G. (2012). A new quantitative automatic method for the measurement of non-rapid eye movement sleep electroencephalographic amplitude variability. *J. Sleep Res.* 21, 212–220. doi: 10.1111/j.1365-2869.2011.00981.x
- Fiorillo, L., Puiatti, A., Papandrea, M., and Ratti, P. L. (2019). Automated sleep scoring: a review of the latest approaches. *Sleep Med. Rev.* 48, 101204–101204. doi: 10.1016/j.smrv.2019.07.007
- Ghimatgar, H., Kazemi, K., Helfroush, M. S., and Aarabi, A. (2019). An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov model. *J. Neurosci. Methods* 324:108320. doi: 10.1016/j.jneumeth.2019.108320
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, E215–E220. doi: 10.1161/01.cir.101.23.e215
- Hassan, A. R., and Bhuiyan, M. I. (2016a). A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *J. Neurosci. Methods* 271, 107–118. doi: 10.1016/j.jneumeth.2016.07.012
- Hassan, A. R., and Bhuiyan, M. I. (2016b). Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybern. Biomed. Eng.* 1, 248–255. doi: 10.1016/j.bbe.2015.11.001
- Hassan, A. R., and Bhuiyan, M. I. (2017). Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Comput. Methods Prog. Biomed.* 140, 201–210. doi: 10.1016/j.cmpb.2016.12.015
- Hassan, A. R., and Subasi, A. (2017). A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowl. Based Syst.* 128, 115–124. doi: 10.1016/j.knsys.2017.05.005
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Hollis, T., Viscardi, A., and Yi, S. E. (2018). A comparison of LSTMs and attention mechanisms for forecasting financial time series. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1812.07699> (accessed March 7, 2020).
- Hsu, Y. L., Yane, Y. T., Wane, T. S., and Hsu, C. (2013). Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing* 104, 105–114. doi: 10.5555/2438096.2438127
- Iber, C., Israel, S. A., Chesson, A. L., and Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: AASM.
- Jo, H. G., Park, J. Y., Lee, C. K., and An, S. K. (2010). Genetic fuzzy classifier for sleep stage identification. *Comput. Biol. Med.* 40, 629–634. doi: 10.1016/j.cmpbiomed.2010.04.007
- Karpathy, A., and Fei-Fei, L. (2014). Deep visual-semantic alignments for generating image descriptions. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1412.2306v2> (accessed March 20, 2020).
- Kemp, B., Zwiderman, A. H., Tuk, B., Kamphuisen, H. A. C., and Obery, J. J. L. (2000). Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* 47, 1185–1194. doi: 10.1109/10.867928
- Kim, J., ElMoaqet, H., Tilbury, D. M., Ramachandran, S. K., and Penzel, T. (2019). Time domain characterization for sleep apnea in oronasal airflow signal: a dynamic threshold classification approach. *Physiol. Meas.* 40:5. doi: 10.1088/1361-6579/aaf4a9
- Kingma, D. P., and Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv [Preprint]*. Available online at: <http://arxiv.org/abs/1412.6980> (accessed March 20, 2020).
- Lecun, Y., and Bengio, Y. (1997). *Convolutional Networks for Images, Speech, and Time-Series*. New York, NY: ACM.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Liang, S. F., Kuo, C. E., Hu, Y. H., Pan, Y. H., and Wang, Y. H. (2012). Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Trans. Instrum. Meas.* 61, 1649–1657. doi: 10.1109/TIM.2012.2187242
- Liu, T., Yu, S., Xu, B., and Yin, H. (2018). Recurrent networks with attention and convolutional networks for sentence representation and classification. *Appl. Intell.* 48, 3797–3806. doi: 10.1007/s10489-018-1176-4
- Liu, Z., Yang, M., and Wang, X. (2017). Entity recognition from clinical texts via recurrent neural network. *BMC Med. Inform. Decis.* 17:67. doi: 10.1186/s12911-017-0468-7
- Lngkvist, M., and Loutfi, A. (2018). A deep learning approach with an attention mechanism for automatic sleep stage classification. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1805.05036> (accessed March 20, 2020).
- Luong, M. T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1508.04025> (accessed March 20, 2020).
- Malafeev, A., Laptev, D., Bauer, S., Omlin, X., Wierzbicka, A., Wichniak, A., et al. (2018). Automatic human sleep stage scoring using deep neural networks. *Front. Neurosci.* 12:781. doi: 10.3389/fnins.2018.00781
- Maquet, P., Péters, J.-M., Aerts, J. L., Delfiore, G., Eguedre, C. D., and Luxen, A. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature* 383, 163–166. doi: 10.1038/383163a0
- Ohayon, M. M. (2002). Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Med. Rev.* 6, 97–111. doi: 10.1053/smrv.2002.0186
- Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and Vos, M. D. (2019). Joint classification and prediction CNN framework for automatic sleep stage classification. *IEEE Trans. Bio Med. Eng.* 66, 1285–1296. doi: 10.1109/TBME.2018.2872652
- Phan, H., Koch, P., Katzberg, F., Maass, M., Mazur, R., and Mertins, A. (2017). Audio scene classification with deep recurrent neural networks. *arXiv [preprint]*. Available online at: <https://arxiv.org/abs/1703.04770> (accessed April 5, 2020).
- Radha, M., Garcia-Molina, G., Poel, M., and Tononi, G. (2014). Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal. *IEEE EMB 2014*, 1876–1880. doi: 10.1109/EMBC.2014.6943976

- Ronzhina, M., Janoušek, O., Kolářová, J., and Nováková, M. (2012). Sleep scoring using artificial neural networks. *Sleep Med. Rev.* 16, 251–263. doi: 10.1016/j.smrv.2011.06.003
- Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans. Signal. Process.* 45, 2673–2681. doi: 10.1109/78.650093
- Seifpour, S., Niknazar, H., Mikaeli, M., and Nasrabadi, A. M. (2018). A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal. *Expert Syst. Appl.* 104, 277–293. doi: 10.1016/j.eswa.2018.03.020
- Shen, H. M., Xu, M. H., Guez, A., and Li, A. (2019). An accurate sleep stages classification method based on state space model. *IEEE Access.* 4, 1–12. doi: 10.1109/ACCESS.2019.2939038
- Silveira, T. L., Kozakevicius, A. J., and Rodrigues, C. R. (2017). Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med. Biol. Eng. Comput.* 55, 343–352. doi: 10.1007/s11517-016-1519-4
- Sors, A., Bonnet, S., Mirek, S., and Vercueil, L. (2018). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomed. Signal. Proces.* 42, 107–114. doi: 10.1016/j.bspc.2017.12.001
- Stephansen, J. B., Olesen, A. N., Olsen, M., Ambati, A., Leary, E. B., Moore, H. E., et al. (2018). Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat. Commun.* 9:1. doi: 10.1038/s41467-018-07229-3
- Supratak, A., Dong, H., Wu, C., and Guo, Y. (2017). “DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG,” in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (Piscataway, NJ: IEEE), 99. doi: 10.1109/TNSRE.2017.2721116
- Tian, P., Hu, J., Qi, J., Ye, X., Che, D., Ding, Y., et al. (2017). A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture. *Biocybern. Biomed. Eng.* 37, 263–271. doi: 10.1016/j.bbe.2017.01.005
- Tripathy, R., and Acharya, U. R. (2018). Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework. *Biocybern. Biomed. Eng.* 38, 890–902. doi: 10.1016/j.bbe.2018.05.005
- Tsinalis, O., Matthews, P. M., and Guo, Y. (2015). Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann. Biomed. Eng.* 44, 1587–1597. doi: 10.1007/s10439-015-1444-y
- Tsinalis, O., Matthews, P. M., Guo, Y., and Zafeiriou, S. (2016). Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1610.01683> (accessed May 10, 2020).
- Wang, C., Yang, H., and Meinel, C. (2018). Image captioning with deep bidirectional LSTMs and multi-task learning. *ACM Trans. Multim. Comput.* 14, 1–20. doi: 10.1145/3115432
- Wang, Y., Loparo, K. A., Kelly, M. R., and Kaplan, R. F. (2015). Evaluation of an automated single-channel sleep staging algorithm. *Nat. Sci. Sleep* 7, 101–111. doi: 10.2147/NSS.S77888
- Weaver, E. M., Woodson, B. T., and Steward, D. L. (2005). Polysomnography indexes are discordant with quality of life, symptoms, and reaction times in sleep apnea patients. *Otolaryngol. Head Neck Surg.* 132, 255–262. doi: 10.1016/j.otohns.2004.11.001
- Yildirim, O., Baloglu, U. B., and Acharya, U. R. (2019). A deep learning model for automated sleep stages classification using PSG signals. *Int. J. Environ. Res. Public Health* 16, 599–599. doi: 10.3390/ijerph16040599
- Younes, M. (2017). The case for using digital EEG analysis in clinical sleep medicine. *Sleep Sci. Prac.* 1:2. doi: 10.1186/s41606-016-0005-0
- Zhang, J., and Wu, Y. (2017). A new method for automatic sleep stage classification. *IEEE Trans. Biomed. Circuits Syst.* 5, 1097–1110. doi: 10.1109/TBCAS.2017.2719631
- Zhang, J., and Wu, Y. (2018). Complex-valued unsupervised convolutional neural networks for sleep stage classification. *Comput. Methods Prog. Biomed.* 164, 181–191. doi: 10.1016/j.cmpb.2018.07.015
- Zhang, X., Liang, X., Zhiyuli, A., and Zhang, S. (2019). AT-LSTM: an attention-based LSTM model for financial time series prediction. *IOP Conf. Ser. Mater. Sci. Eng.* 569:052037. doi: 10.1088/1757-899X/569/5/052037
- Zhu, G., Li, Y., and Wen, P. P. (2014). Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J. Biomed. Health* 18, 1813–1821. doi: 10.1109/JBHI.2014.2303991
- Zweig, M. H., and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39, 561–577. doi: 10.1093/clinchem/39.4.561

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Fu, Wang, Chen, Li, Xu, Liu and Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.