2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

LightSleepNet: A Lightweight Deep Model for Rapid Sleep Stage Classification with Spectrograms

Dongdong Zhou[1,2], Qi Xu[3,4,*], Jian Wang[1,2], Jiacheng Zhang[5], Guoqiang Hu[1], Lauri Kettunen[2], Zheng Chang[2], *Senior Member, IEEE* and Fengyu Cong[1,2,3,6,*], *Senior Member, IEEE*

***Abstract*— Deep learning has achieved unprecedented success in sleep stage classification tasks, which starts to pave the way for potential real-world applications. However, due to its enormous size, deployment of deep neural networks is hindered by high cost at various aspects, such as computation power, storage, network bandwidth, power consumption, and hardware complexity. For further practical applications (e.g., wearable sleep monitoring devices), there is a need for simple and compact models. In this paper, we propose a lightweight model, namely LightSleepNet, for rapid sleep stage classification based on spectrograms. Our model is assembled by a much fewer number of model parameters compared to existing ones. Furthermore, we convert the raw EEG data into spectrograms to speed up the training process. We evaluate the model performance on several public sleep datasets with different characteristics. Experimental results show that our lightweight model using spectrogram as input can achieve comparable overall accuracy and Cohen's kappa (SHHS100: 86.7%-81.3%, Sleep-EDF: 83.7%-77.5%, Sleep-EDF-v1: 88.3%-84.5%) compared to the state-of-the-art methods on experimental datasets.**

I. INTRODUCTION

High quality sleep plays an important role in humans' health. It has a significant influence on diagnosing and treating sleep-related disorders (e.g., insomnia) through correct sleep stage classification [1]–[4]. In order to accomplish the sleep scoring task, overnight polysomnography (PSG) data need to be recorded by several sensors attaching to different parts of the body. The PSG recordings mainly comprise electroencephalogram (EEG), electromyogram (EMG), electrocardiogram (ECG), electrooculogram (EOG) and so on [5]. In clinical practice, the PSG data are usually split into 30s segments sequentially. Each 30s epoch is further classified into different sleep stages by experienced clinicians manually according to sleep manuals. Specifically, sleep stages include six stages: Wake (W), Rapid Eye Movement (REM), Non-REM1 (N1), Non-REM2 (N2), Non-REM3 (N3) and Non-REM4 (N4) based on the Rechtschaffen and Kales (R&K) standard [6].

Nevertheless, the manual sleep stage classification is not only prone to subjective error but also time-consuming. Therefore, there is an urgent need for an effective sleep scoring method to release the workload of clinicians and obtain reliable performance. Recently, deep learning has been applied to automatic sleep stage classification successfully due to its powerful learning ability of feature extraction in a data-driven way. Whereas, various approaches based on convolutional neural network (CNN) have too much complex structures with millions of parameters [1] leading to probable overfitting issue and the high demand for computing resources. This drawback hinders those methods from further practical application (e.g., portable sleep monitor devices), which require light but efficient methods on resource-constrained devices, the time costs of model training should be also considered. Compared to previous studies using PSG data time series ($30 \times f_s$, channel) as input, CNNs are more efficient to process static imagery or matrix structure, which means CNNs are good at processing static images through their powerful feature extraction structure.

To solve mentioned problems, we propose the LightSleepNet (LSNet), a lightweight model for automatic sleep stage classification based on single-channel EEG. Transforming the raw time-series EEG signals to static spectrograms as the input, this model could be implemented and trained in a more efficient way which is suitable for rapid sleep stage classification tasks. The main contributions of this work are as follows:

i) We propose a light but efficient model with much fewer model parameters for automatic sleep stage classification.
ii) To speed up the training process, we utilize the spectrograms through short-time Fourier transform as the model input rather than the long time series EEG signals.
iii) The results demonstrate that our model can achieve comparable performance on different single-channel EEGs (C4/A1, Fpz-Cz) on experimental datasets with different characteristics.

This work was support by National Natural Science Foundation of China (Grant No.91748105), National Foundation in China (No. JCKY2019110B009, 2020-JCJQ-JJ-252), Fundamental Research Funds for Central Universities [DUT2019, DUT20LAB303] in Dalian University of Technology in China and the scholarships from China Scholarship Council (No.201806060164, No.202006060226), CAAI-Huawei MindSpore Open Fund (CAAIXSJLJJ-2020-024A).

[1]School of Biomedical Engineering, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China (*corresponding authors: xuqi123@zju.edu.cn, cong@dlut.edu.cn).

[2]Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland.

[3]School of Artifical Intelligence, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China.

[4]3College of Computer Science and Technology, Zhejiang University, 310027, Hangzhou, China.

[5]School of Information and Communication Engineering, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China.

[6]Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province, Dalian University of Technology, 116024, Dalian, China.

TABLE I
THE DISTRIBUTION OF EACH SLEEP STAGE OF EACH DATASET

Dataset	W	N1	N2	N3	REM	Total
SHHS-100	23708	3010	41207	14306	14989	97220
Sleep-EDF	69518	21522	69132	13039	25835	199046
Sleep-EDF-v1	10917	2804	17799	5703	7717	44220

II. MATERIALS AND METHODS

A. Data Description

We evaluate the performance of proposed model employing three public PSG datasets: Sleep Heart Health Study (SHHS), Sleep-EDF Database (Sleep-EDF-v1, version 2013) and Sleep-EDF Database Expanded (Sleep-EDF, version 2018). The corresponding hypnograms of three datasets were scored by the well-trained clinicians following the R&K rule. For all employed datasets, we adopt single-channel EEG which can benefit to further reduce the computational cost and simply the scheme of data acquisition.

The SHHS dataset includes two subsets: initial PSG (SHHS1) and second PSG (SHHS2). Unlike the computer vision research, it is difficult to acquire abundant PSG samples to train the model. In order to better show the few-shot learning ability of the proposed model in sleep stage classification tasks, we adopt 100 near-normal subjects from the SHHS1 (i.e., SHHS-100) with the standard of the respiratory disturbance index 3 percent (RDI3P) < 15 and no reported high pressure, cardiopathy or stroke. Single-channel EEG C4 sampled at 125 Hz is utilized for evaluating the proposed model as the suggestion of AASM manual. Detailed information of the SHHS dataset can be found in [7].

In the Sleep-EDF database, a total of 78 subjects with 153 whole-night PSG recordings from the sleep-cassette (SC) subset are selected. In addition, we also conduct the experiments on the first version of the Sleep-EDF dataset (Sleep-EDF-v1) before the expansion to make a fair comparison with the existing methods. We employ the single-channel EEG Fpz-Cz with 100 Hz in our experiments. To keep the same f_s, the EEG Fpz-Cz is resampled at 125 Hz. The [8] presents the detailed description of Sleep-EDF dataset.

For both datasets, we merge the N3 and N4 stages into stage N3 according to the latest AASM manual [9]. Hence each 30s epoch is labeled as one of five sleep stages (i.e., W, N1, N2, N3 and REM). In this paper, we use three successive EEG epochs (90s epoch) rather than the conventional 30s epoch as the contextual input of model. It is considered that experts classify the sleep stage depend not only on the current epoch but also the preceding and succeeding epochs. Also, the contextual input can enhance the model's learning ability of the transition information between epochs. The corresponding label of the 90s epoch is the label of current 30s epoch. As shown in Table I, we reveal the distribution of each stage from experimental datasets.

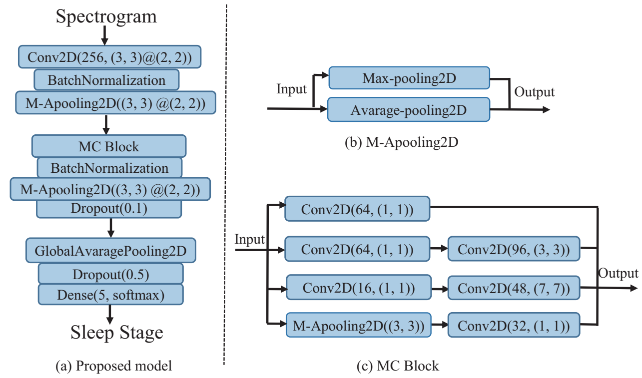

Fig. 1. The overall architecture of proposed model.

B. Data Preprocessing

The raw EEG data are filtered by a notch filter, a high-pass filter and a low-pass filter to eliminate the effect of noise and artifacts. To get the power spectrum of each 90s epoch, we adopt the short-time Fourier transform (STFT) with a window size of two seconds and 50% overlap. Moreover, Hamming window and 256 points Fast Fourier Transform (FFT) are conducted. The effective frequency band is set to 0.5-30 Hz and the obtained power spectrum is then converted to the log-power spectrum of size of $F \times T$, where F = 61, T = 89.

C. The Proposed Model

Different from our prior work [10] in which the SCNet is trained with the raw EEG data, the proposed model here, namely LSNet, is designed for handling with the spectrogram input. We show in Fig. 1 the overall architecture of proposed model. The first two-dimensional convolutional (Conv2D) layer with 256 filters of size 3×3 and the stride of 2 points is used to attain the feature map from the spectrum input ($61 \times 89 \times 1$) and the activation function is rectified linear unit (ReLU). Additionally, we apply the batch normalization to normalize the output of the first Conv2D layer. The M-Apooling2D layer is the concatenation of max-pooling2D and average-pooling2D layers that can learn feature representation from two scales.

We construct a multi-convolution (MC) block containing three different sizes of filters (1×1, 3×3, 7×7) to obtain multiscale features simultaneously. To be specific, the small filter is better to learn the temporal context, while the large filter is prone to capture the frequency information. Similar to the first Conv2D layer, the MC block is followed by the batch normalization and M-Aplooing2D layers. Besides, a dropout layer with the probability of 0.1 is applied to avoid the overfitting problem. It should be noticed that the size of stride in the MC block is set as 1×1. Aiming to flat the previous output, we implement the global average pooling layer before the decision layer. A dropout layer with a drop rate of 0.5 can further prevent the overfitting issue. Except for the dropout method, another solution for overfitting applying in this work is the L2 regularization,

TABLE II
PARAMETERS OF THE PROPOSED MODEL

Layer	Layer Type	Filters	Size	Stride	Activation	Output dimension
1	Input	-	-	-	-	$(61 \times 89 \times 1)$
2	Conv2D	256	(3,3)	(2,2)	Relu	$(31 \times 45 \times 256)$
3	M-Apooling2D	-	(3,3)	(2,2)	-	$(16 \times 23 \times 512)$
4	MC Block	-	-	(1,1)	Relu	$(16 \times 23 \times 240)$
5	M-Apooling2D	-	(3,3)	(2,2)	-	$(8 \times 12 \times 480)$
6	Dropout(0.1)	-	-	-	-	$(8 \times 12 \times 480)$
7	GAP	-	-	-	-	480
8	Dropout(0.5)	-	-	-	-	480
9	Dense	-	-	-	Softmax	5

which adds a squared magnitude of coefficient as penalty term to the loss function. The regularization rate, lambda, is chosen as 10^{-3} based on the experimental results of four lambda values (10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}). The final output is achieved by a dense layer whose activation function is the softmax to determine the probability of each stage, the stage with maximum probability is considered as the predicted sleep stage. The detailed parameters of proposed model are illustrated in Table II.

We also assemble a baseline model, in which time series as the input, for making a comparison with the proposed model in terms of the training computational cost for each iteration. The baseline is consistent with the structure and training setup except for the filter sizes. To be specific, the filter size of $N \times N$ is replaced by the filer size of N. For instance, 3×3 in the LSNet model should be converted into size of 3 in the baseline model.

D. Training Setup

We use 5-fold cross-validation to assess our model performance on all databases. The whole subjects are divided into training and test sets with a ratio of 4 to 1 using the subject-wise and epoch-wise schemes independently. For the subject-wise approach, the division of training and test sets is based on subjects. Nevertheless, we divide the whole epochs into training and test sets following the epoch-wise scheme. In each fold, we further employ 20% of the training set as the validation set to validate the training model. The model with the best overall accuracy is kept for evaluation on the test set. The model is trained by the Adam optimizer in 30 epochs, where the learning rate (lr), $beta1$ and $beta2$ are set as 10^{-3}, 0.9 and 0.999 respectively. Moreover, ReduceLROnPlateau of Callback in Keras is implemented to tune the lr dynamically. The lr would be reduced to half of it when the validation accuracy does not increase within 3 epochs. We choose the categorical cross-entropy as the loss function of the model. In addition, batch sizes of 32, 64, 128 and 256 are tested to determine the final batch size of 64 for training. Furthermore, we evaluate the performance of proposed model using the accuracy (ACC) and Cohen's kappa coefficient (K), which are defined as follows:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}. \quad (1)$$

$$K = \frac{\frac{\sum_{i=1}^{n} x_{ii}}{N} - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}{1 - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}. \quad (2)$$

Where TP, FP, TN and FN represent the true positive, false positive, true negative and false negative respectively. The N is the number of 90s epoch in the test set, n is the number of classes. x_{ii} represents the diagonal value of the confusion matrix. It is noteworthy that we optimize the hyper-parameters of our model on the SHHS-100 database. Once obtaining the optimal model, there is no need to tune the architecture and hyper-parameters on the Sleep-EDF and Sleep-EDF-v1 datasets.

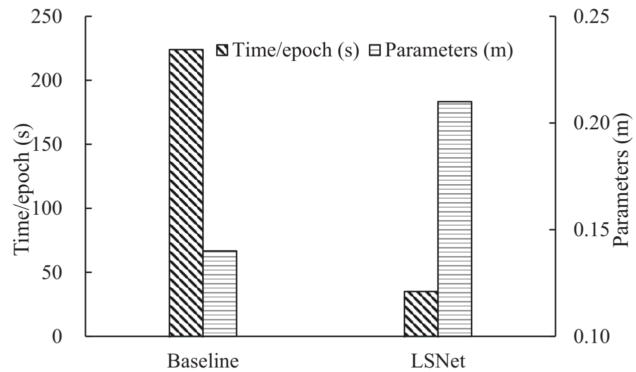

Fig. 2. The comparison between the baseline and LSNet in terms of the model parameters and computational cost for each epoch. The black diagonal stripe represents the training time of each iteration, the gray cross stripe denotes the number of model parameters.

III. EXPERIMENTAL RESULTS

To show the efficiency of spectrogram input, Fig.2 demonstrates the number of parameters and training computational cost for each iteration in the baseline and LSNet models. The number of parameters of the baseline is about 0.14 million (m) and time cost in each iteration ups to 224 s. By contrast, even if the LSNet has more parameters (around 0.21 m), the training speed is more than 6 times faster (35 s) for each epoch.

In Table III, we further make the performance comparison between our framework (epoch-wise and subject-wise) and other state-of-the-art methods using the same dataset across the ACC and K. The values of ACC are more than 83% on all datasets, which show the proposed LSNet model can achieve comparable performance compared to the existing ones. More importantly, the number of parameters of LSNet is much less than that of compared models. Besides, the K demonstrates that our model can reach perfect (0.81 to 1) and substantial (0.61 to 0.8) agreement with the sleep experts.

IV. DISCUSSION

In this paper, we propose a lightweight but effective CNN based model for rapid automatic sleep stage classification, named LSNet. Different with taking time series EEG signals as input, the proposed LSNet transforms those dynamic data

TABLE III
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD (LSNET) AND PREVIOUS METHODS ON THE SHHS, SLEEP-EDF AND SLEEP-EDF-V1 DATASETS

Study	Dataset	Method	Input channel	Input type	Parameters ($\times 10^6$)	Subjects	$ACC(\%)$	$K(\%)$
Proposed (epoch-wise)	SHHS-100	Deep CNN	C4-A1	Spectrogram	0.2	100	86.7	81.3
Proposed (subject-wise)	SHHS-100	Deep CNN	C4-A1	Spectrogram	0.2	100	85.6	79.4
Sors *et al*. [11]	SHHS	Deep CNN	C4-A1	Time series	2.2	5728	**87**	81
Seo *et al*. [16]	SHHS	CNN + LSTM	C4-A1	Time series	-	5791	86.7	79.8
Zhang *et al*. [12]	SHHS	CNN + LSTM	2EEG + 2EOG + EMG	Spectrogram	1.3	5793	**87**	**82**
Proposed (epoch-wise)	Sleep-EDF	Deep CNN	Fpz-Cz	Spectrogram	0.2	78	**83.7**	**77.5**
Proposed (subject-wise)	Sleep-EDF	Deep CNN	Fpz-Cz	Spectrogram	0.2	78	83.4	76.7
Supratak *et al*. [13]	Sleep-EDF	CNN + LSTM	Fpz-Cz	Time series	1.3	78	83.1	77
Mousavi *et al*. [15]	Sleep-EDF	CNN + LSTM	Fpz-Cz	Time series	21	78	80.0	73
Proposed (epoch-wise)	Sleep-EDF-v1	Deep CNN	Fpz-Cz	Spectrogram	0.2	20	**88.3**	**84.5**
Proposed (subject-wise)	Sleep-EDF-v1	Deep CNN	Fpz-Cz	Spectrogram	0.2	20	86.1	81.0
Supratak *et al*. [14]	Sleep-EDF-v1	CNN + LSTM	Fpz-Cz	Time series	21	20	82.0	76
Supratak *et al*. [13]	Sleep-EDF-v1	CNN + LSTM	Fpz-Cz	Time series	1.3	20	85.4	80
Mousavi *et al*. [15]	Sleep-EDF-v1	CNN + LSTM	Fpz-Cz	Time series	2.6	20	84.3	79

to static spectrograms. The proposed LSNet also employs a light structure with few parameters compared to over-parameterized CNN models which lead to the proposed model could be trained in a more efficient way. As the result shows that even though there are roughly 1.5 times as many model parameters of the LSNet as of the baseline model, our model can realize rapid sleep stage classification with more than 6 times speed promotion.

On the other hand, among the previous studies that we compare, the numbers of model parameters in Sors *et al* [11], Zhang *et al* [12], Supratak *et al* [13], [14] and Mousavi *et al* [15] are about 2.2 m, 1.3 m, 1.3 m, 21 m and 2.6 m respectively, which are at least 6 times larger than our model (0.21 m). We do not list the number of model parameters in [16] as it cannot be calculated from the literature. Besides, we do not sacrifice the model performance to achieve the purpose of reducing the model parameters. Experimental results show that our model can attain similar performance compared to the state-of-the-art methods on adopted datasets, which indicate the desirable generalization of the LSNet model. Considering the computing resources and time delay for real-time application, the lightweight model we propose for rapid sleep stage classification maybe more easily adaptable to clinical or wearable devices applications. In future works, we will explore more brain-inspired models (e.g., spiking neural networks) [17], [18] to realize energy-efficient implementation on sleeping scoring tasks.

ACKNOWLEDGMENT

This study is to memorize Prof. Tapani Ristaniemi from University of Jyväskylä for his great help.

REFERENCES

[1] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 485–496, 2006.

[2] G. Zhu, Y. Li, and P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel eeg signal," *IEEE J. Biomed. Health. Inf.*, vol. 18, no. 6, pp. 1813–1821, 2014.

[3] R. Yan *et al.*, "Automatic sleep scoring: A deep learning architecture for multi-modality time series," *J. Neurosci. Methods.*, vol. 348, p. 108971, 2021.

[4] H. Phan *et al.*, "Joint classification and prediction cnn framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, 2018.

[5] H. Dong *et al.*, "Mixed neural network approach for temporal sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 324–333, 2017.

[6] A. Rechtschaffen, "A manual of standardized terminology and scoring system for sleep stages of human subjects," *Electroencephalogr. Clin. Neurophysiol.*, vol. 26, no. 6, p. 644, 1969.

[7] G. Zhang *et al.*, "The national sleep research resource: towards a sleep data commons," *J. Am. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, 2018.

[8] B. Kemp *et al.*, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, 2000.

[9] C. Iber *et al.*, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, vol. 1. Westchester, IL, USA: Amer. Acad. Sleep Med., 2007.

[10] D. Zhou *et al.*, "Singlechannelnet: A model for automatic sleep stage classification with raw single-channel eeg," *bioRxiv*, 2021.

[11] A. Sors *et al.*, "A convolutional neural network for sleep stage scoring from raw single-channel eeg," *Biomed. Signal Process. Control.*, vol. 42, pp. 107–114, 2018.

[12] L. Zhang *et al.*, "Automated sleep stage scoring of the sleep heart health study using deep neural networks," *Sleep.*, vol. 42, no. 11, p. zsz159, 2019.

[13] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel eeg," in *Proc. IEEE Eng. Med. Biol. Soc (EMBC).*, pp. 641–644, 2020.

[14] A. Supratak *et al.*, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, 2017.

[15] S. Mousavi, F. Afghah, and U. R. Acharya, "Sleepeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLOS ONE.*, vol. 14, no. 5, pp. 1–15, 2019.

[16] H. Seo *et al.*, "Intra-and inter-epoch temporal context network (iitnet) using sub-epoch features for automatic sleep scoring on raw single-channel eeg," *Biomed. Signal Process. Control.*, vol. 61, p. 102037, 2020.

[17] Q. Xu *et al.*, "Deep CovDenseSNN: A hierarchical event-driven dynamic framework with spiking neurons in noisy environment," *Neural Netw.*, vol. 121, pp. 512–519, 2020.

[18] Z. Yu *et al.*, "Emergent inference of hidden markov models in spiking neural networks through winner-take-all," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1347–1354, 2018.