SingleChannelNet: A model for Automatic Sleep Stage Classification with Raw Single-Channel EEG

Dongdong Zhou, Guoqiang Hu, Jiacheng Zhang, Rui Yan, Fan Li, Qi Xu, Lauri Kettunen, Zheng Chang, *Senior Member, IEEE* and Fengyu Cong, *Senior Member, IEEE*

***Abstract*— In diagnosing sleep disorders, sleep stage classification is a very essential yet time-consuming process. Most of the existing state-of-the-art approaches rely on hand-crafted features and multi-modality polysomnography (PSG) data, where prior knowledge is compulsory and high computation cost can be expected. Besides, few studies are able to obtain high accuracy sleep staging using raw single-channel electroencephalogram (EEG). To overcome these shortcomings, this paper proposes an end-to-end framework with a deep neural network, namely SingleChannelNet, for automatic sleep stage classification based on raw single-channel EEG. The proposed model utilizes a 90s epoch as the textual input and employs two multi-convolution blocks and several max-average pooling layers to learn different scales of feature representations. To demonstrate the efficiency of the proposed model, we evaluate our model using different raw single-channel EEGs (C4/A1 and Fpz-Cz) on two different datasets (SHHS and Sleep-EDF datasets). Experimental results show that the proposed architecture can achieve better overall accuracy and Cohen's kappa (SHHS: 89.2%-84.8%, Sleep-EDF: 89.4%-85%) compared with state-of-the-art approaches. Additionally, the proposed model can learn features automatically for sleep stage classification using different single-channel EEGs with distinct sampling rates from different datasets without using any hand-engineered features.**

***Index Terms*— Sleep stage classification, Convolutional neural network, Raw single-channel EEG**

This work was support by National Natural Science Foundation of China (Grant No.91748105), National Foundation in China (No. JCKY2019110B009), Fundamental Research Funds for Central Universities [DUT2019] in Dalian University of Technology in China and the scholarships from China Scholarship Council (No.201806060164; No.201806060038; No.201606060227).

D. Zhou and R. Yan are with School of Biomedical Engineering, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China & Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland (e-mail: dongdong.w.zhou@student.jyu.fi; ruiyanmodel@foxmail.com).

G. Hu and F. Li are with School of Biomedical Engineering, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China (e-mail: guoqiang.hu@mail.dlut.edu.cn; lifandlpu@foxmail.com).

J. Zhang is with School of Information and Communication Engineering, Faculty of Electronic and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China (e-mail: 56best@mail.dlut.edu.cn).

Q. Xu is with College of Computer Science and Technology, Zhejiang University, 310027, Hangzhou, China (e-mail: xuqi123@zju.edu.cn).

L. Kettunen and Z. Chang are with Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland (e-mail: lauri.y.o.kettunen@jyu.fi; zheng.chang@jyu.fi).

F. Cong is with School of Biomedical Engineering & School of Artifical Intelligence, Faculty of Electronic and Electrical Engineering & Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province, Dalian University of Technology, 116024, Dalian, China & Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland (e-mail: cong@dlut.edu.cn).

I. Introduction

HUMANS spend about one-third time of life on sleeping, and high-quality sleep plays a vitally important role in the restoration of body and mind [1]. Whereas roughly 33% of the population in the world suffers from insomnia disorder [2]. Correctly identifying sleep stage using whole-night PSG data is essential to diagnose and treat sleep-related disorders [3]–[6]. The PSG recordings comprise of the EEG, electrocardiogram (ECG), electrooculogram (EOG), electromyogram (EMG) and other respiration signals [7].

According to the guidelines of the Rechtschaffen and Kales (R&K) [8] or American Academy of Sleep Medicine (AASM) [9], the PSG data should be first segmented into 30s epochs typically, then these sequential epochs are defined as different stages. Some sleep-related disorders have particular sleep structure, it is therefore beneficial to diagnose them with accurate sleep stage classification. Traditionally, the sleep stage classification task is conducted by experts manually following the R&K and AASM rule which is often time-consuming, labor-intensive and prone to subjective mistakes [6]. Hence, there is an urgent need for automatic sleep stage classification approach to assist the clinician's work and achieve reliable results.

Some methods based on machine learning have been proposed to identify the sleep stage. These approaches generally extract either time-domain features [3], [10], [11] or frequency-domain features [12]–[16] from the PSG signals and those pre-trained hand-crafted features are fed into the conventional classifier, such as support vector machine (SVM) [4], [14], [17], [18], k-nearest neighbors (KNN) [16], [19], [20], random forest [21]–[24] and so on. The performance tremendously relies on the categories and the number of features, which are extracted based on the characteristics of experimental datasets. Therefore, these approaches may not be robust enough to be generalized to different datasets because of the distinct properties between datasets.

In recent years, the deep networks show great capacity for automatic features learning from data, and it can avoid the

reliance on hand-engineered features. Meanwhile, a series of deep learning methods are applied to sleep stage classification. Here, we categorize these approaches into multi-channel [6], [25]–[30] versus single-channel schemes [31]–[36] based on the number of input channels. Following the multi-channel scheme, Phan *et al*. [6] first transformed the raw signals into the time-frequency image through the short-time Fourier transform as the input of the proposed convolutional neural network (CNN). The overall accuracy achieved was equal to 82.3%, which is not state-of-the-art. In addition, the time-frequency image relies much on many preprocessing steps, it would be time-consuming and in need of prior knowledge of signal processing. Aiming at this, Chambon *et al*. [28] proposed a novel network architecture of low computational cost adopting multivariate and multimodal time series from EEG, EMG and EOG, but the classification performance is not good enough with the accuracy of 80% compared to state-of-the-art methodologies. One important reason is that the convolutional layers with fixed filter size were stacked sequentially, which can not learn multiscale features simultaneously. A promising approach was proposed by Zhang *et al*. [30], who employed the CNN and recurrent neural network (RNN) to capture temporal and spatial information simultaneously from the PSG data. The architecture attained an accuracy of 87%. Although the combination of CNNs and RNNs can enhance the model performance to some extent, the high computational cost of RNNs should be taken into consideration. To the best of our knowledge, the training speed of CNNs would be dozens of times faster than that of RNNs under the same GPU acceleration when implementing long time-series input. To sum up, despite the fact that multi-channel PSG data can provide additional referenced information compared to single-channel EEG, there is also some irrelevant information being introduced. Furthermore, multi-channel recordings can limit the practical application on account of more complex operation and equipment costs.

Compared to the multi-channel scheme, the single-channel scheme can reduce the related cost and be much easier for data acquisition. Under the single-channel scheme, Supratak *et al*. [31] introduced a deep learning model called DeepSleepNet. DeepSleepNet utilizes the capacity of deep learning to extract time-invariant features automatically, the proposed model can be adapted to different datasets. However, the accuracy obtained from DeepSleepNet was 82%, which can not outperform the state-of-the-art approaches. A promising CNN model was proposed by Sors *et al*. [32], who used raw single-channel EEG to classify the sleep stage without any preprocessing. The architecture attained an accuracy of 87%, whereas the model complexity is a bit high with 12 convolutional layers. Furthermore, the filer size was chosen among 7, 5, 3, the performance of larger size filters should be compared considering the long length of input (1.5×10^4).

To tackle these problems, this paper proposes the SingleChannelNet (SCNet), a model for automatic sleep stage classification based on raw single-channel EEG, which can learn different scales features simultaneously. We aim to automate the sleep stage classification completely by utilizing the capabilities of the proposed model. The main contributions of this work are as follows:

i) We propose a new deep learning model with low model complexity for sleep stage classification using 90s raw single-channel EEG.
ii) We implement two multi-convolution (MC) blocks with different filter sizes in our model. In addition, the max-average (M-Apooling) layer is applied to take place of the conventional max-pooling layer. Two strategies are used for capturing more feature representations from different scales to enhance the capacity of the feature extraction.
iii) The results demonstrate that our model can obtain promising performance on different raw single-channel EEGs (C4/A1, Fpz-Cz) from SHHS and Sleep-EDF datasets, without modifying the architecture and hyper-parameters of model and training algorithm. Moreover, all features are learned by the proposed model automatically.

The rest of this paper is organized as follows. We represent the experimental datasets in Sec. II. Sec. III describes the structure of the SCNet model and the training algorithm. In Sec. IV, the experimental results are represented. The final discussion and conclusion are included in Sec. V.

II. Experimental Datasets

Two public datasets are employed to evaluate the performance of the proposed framework in this work, namely Sleep Heart Health Study (SHHS) [37] and Sleep-EDF Database Expanded (Sleep-EDF, 2018 version) [38]. It should be noted that all hypnograms of both datasets are manually scored according to the R&K manual rather than the AASM rule.

A. Sleep Heart Health Study (SHHS)

The SHHS dataset comprises of overnight PSG recordings from 6411 individuals aged 40 years or older, and it is used to investigate the relationship between sleep-related breathing and the increased risk of cardiovascular diseases. The dataset is composed of two rounds of PSG recordings: initial PSG (SHHS1) and second PSG (SHHS2). Recorded PSG data include two-channel EEGs (C3/A2 and C4/A1 EEGs, sampled at 125 Hz), two-channel EOGs (right and left EOGs, sampled at 50 Hz), single-channel EMG (sampled at 125 Hz), single-channel ECG (sampled at 125 Hz for most SHHS1 and 250 Hz for SHHS2), heart rate (HR, sampled at 1 Hz), light sensor data, airflow sensor data and body position data. We adopt 100 subjects from the SHH1 with the criteria of the respiratory disturbance index 3 percent (RDI3P) < 15. Moreover, the selected subjects do not suffer from high pressure, cardiopathy or stroke. Each 30s epoch is manually labeled by experts according to the R&K standard into several stages: Wake, Rapid Eye Movement (REM), Non-REM1 (N1), Non-REM2 (N2), Non-REM3 (N3) and Non-REM4 (N4). In this work, single-channel EEG C4/A1 is selected.

B. Sleep-EDF Database Expanded (Sleep-EDF)

The Sleep-EDF dataset consists of two subsets: sleep-cassette (SC) contains 77 healthy Caucasians aged from 25

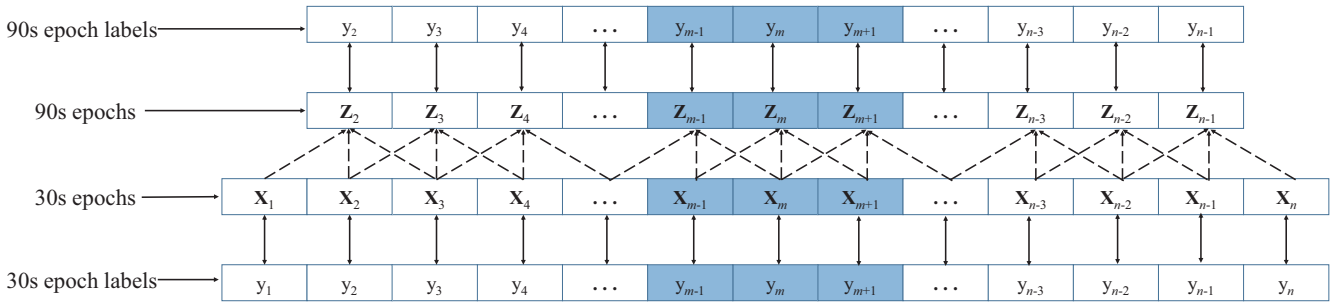

Fig. 1. Illustration of 90s epochs and labels used in this paper, n donates the number of 30s epochs for a subject, $\mathbf{Z}_m$ is comprised of $\mathbf{X}_{m-1}$, $\mathbf{X}_m$ and $\mathbf{X}_{m+1}$, $2 \leq m \leq n-1$.

TABLE I
NUMBER OF 90S EPOCHS FOR EACH SLEEP STAGE FROM TWO DATASETS

Dataset	Wake	N1	N2	N3/N4	REM	Total
SHHS	23708	3010	41207	14306	14989	97220
Sleep-EDF	54661	11428	45282	10600	16634	138605

to 101 years and sleep-telemetry (ST) comprises 22 Caucasians receiving temazepam treatment. Each participant was recorded two subsequent night PSG data except the subject 36 from the SC subset who had only a one-night record. Each epoch of recordings is manually labelled by clinicians according to the R&K rule into Wake, N1, N2, N3, N4, REM, MOVEMENT and UNKNOWN stages respectively. In addition, MOVEMENT and UNKNOWN are excluded, as they do not belong to the six stages. The PSG data include two-channel EEGs (Fpz-Cz and Pz-Oz), single-channel EOG, single-channel EMG and the event marker (sampled at 1 Hz). The sampling rate f_s of EEG, EOG, and EMG is 100 Hz. A total 100 of whole-night PSG recordings from 61 subjects are employed and single-channel EEG Fpz-Cz is adopted in our experiments. For both SHHS and Sleep-EDF datasets, stages N3 and N4 are merged into stage N3 which is consistent with the AASM manual. Additionally, resampling operation is not applied to C4/A1 and Fpz-Cz EEGs.

C. Contextual input

In previous works, most schemes use a single 30s epoch as the classifier input [7], [32], [36] and then produce a single output label. Although being straightforward, this classification method ignores the existing correlation and dependency between surrounding epochs. It is considered that the sleep stage classification depends not only on the local epoch, but also on the prior and following temporal features [6], [9]. For this reason, an extension of single 30s epoch input is conducted by combining it with its neighboring epochs to make a contextual input. Furthermore, we employ 90s epoch ($\mathbf{Z}_m$) as contextual input of the proposed model, and it contains three sequential epochs: prior 30s epoch ($\mathbf{X}_{m-1}$), 30s epoch ($\mathbf{X}_m$) and subsequent 30s epoch ($\mathbf{X}_{m+1}$). The ground truth label of $\mathbf{Z}_m$ is y_m which also denotes $\mathbf{X}_m$'s label. As in

$$\mathbf{Z}_m = (\mathbf{X}_{m-1}, \mathbf{X}_m, \mathbf{X}_{m+1}) \mapsto y_m. \tag{1}$$

Details are illustrated in Fig. 1. As shown in Table I, we summarize the number of 90s epochs for each sleep stage from SHHS and Sleep-EDF datasets in our experiments. The distribution of the number of five stages is imbalanced. For both datasets, Wake and N2 stages account for more than 60% of all 90s epochs. By contrast, the proportion of stages N1 and N3/N4 is the smallest.

III. PROPOSED SCNET

Fig. 2 shows the overall architecture of the SCNet. The convolution block performs three operations sequentially: one-dimensional convolutional layer (Conv1D), batch normalization and M-Apooling1D. Similarly, each MC Block is followed by batch normalization, M-Apooling1D and Dropout layer in sequence. In our model, we employ the concatenation of max-pooling and average-pooling to take place of the max-pooling for capturing more representable features. Similar to the inception module [39], the MC block contains different sizes of convolutional filters to capture the corresponding information. Besides, we use the GAP layer to replace the traditional fully connected layer, and it is proved to be more robust spatial translations of the input without parameter optimization [40].

A. Model Specification

In Table II, we relate detailed parameters of the proposed model. The size of the model's input is $(90 \times f_s, 1)$, where f_s is the sampling rate. To be specific, the f_s of EEG C4/A1 and Fpz-Cz is 125 Hz and 100 Hz, respectively. Here, the SCNet does not restrict the length of input which can be applied to different datasets.

The first convolutional layer with 128 filters of size 128 and a stride of 2 is applied to obtain time-invariant features from raw single-channel EEG. The activation function of this layer is rectified linear unit (ReLU) which is defined as the positive part of its argument:

$$f(x) = \max(0, x) \tag{2}$$

where x is the input of a neuron. To normalize the prior layer input, we apply the batch normalization technique. Besides, the M-Apooling layer can get the combination of maximum and average values from each of a cluster of neurons at the previous layer.

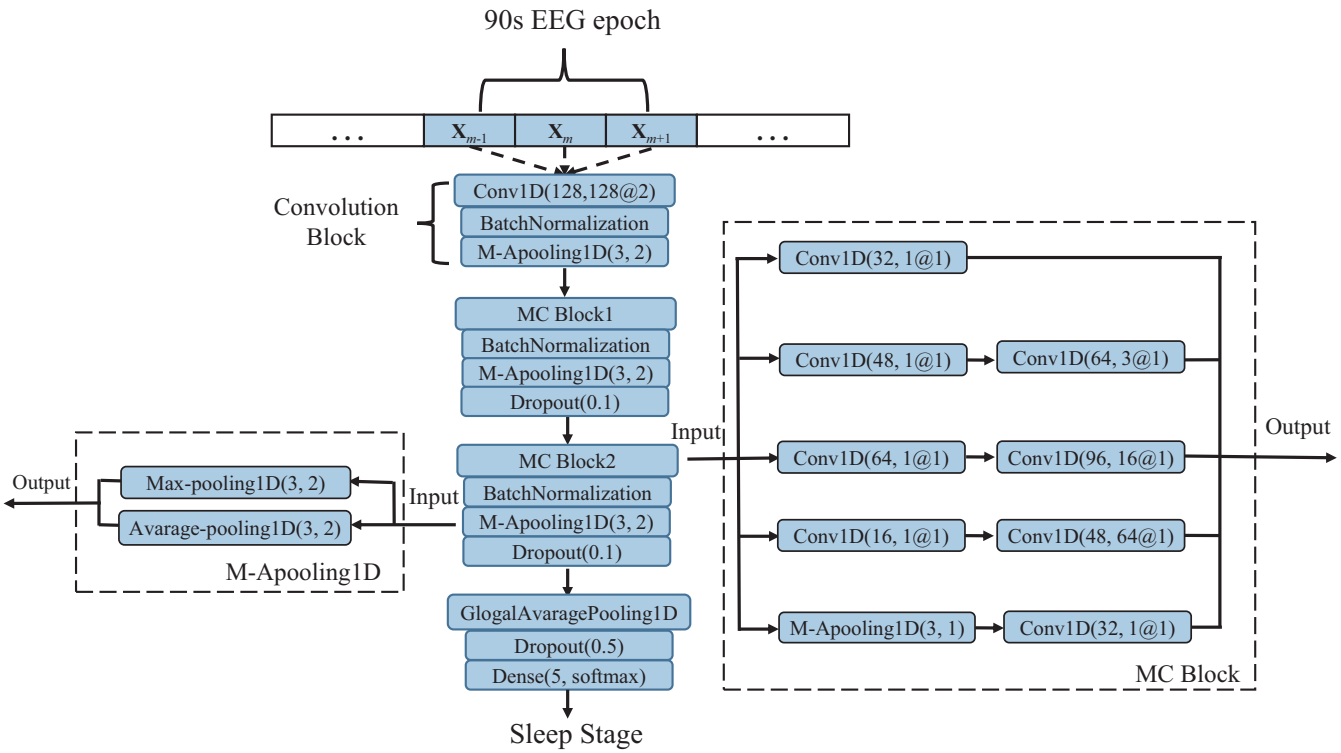

Fig. 2. An overall architecture of the proposed SCNet.

We implement two MC blocks in our model, and the filter sizes are selected among 1, 3, 16 and 64 to obtain multiscale representative features. More specifically, the small filter is prone to learn temporal information, while the large filter is better to capture frequency information. The main functions of Conv1D with size of 1 are to improve the nonlinearity of the network and reduce the dimension of previous layer output. Furthermore, after concatenating the output of all convolutional layers, the dimension of the MC block1 output is ($\lceil 45 \times f_s/2 \rceil$, 272). The following M-Apooling layer can get ($\lceil 45 \times f_s/4 \rceil$, 544) dimension feature map. Each MC block is followed by a batch normalization layer, a M-Apooling layer with size of 3 and a dropout layer with the probability of 0.1. The GAP layer is applied to flat the previous output before the final decision layer. Through a drop layer with drop rate of 0.5, the dense layer using softmax as the activation function makes the final decision. Softmax function can calculate the probabilities of five stages, the stage with maximum probability is as the consequence of the predicted sleep stage.

TABLE II
PARAMETERS OF THE PROPOSED MODEL

Layer	Layer Type	Filters	Size	Stride	Activation	Output dimension
1	Input	-	-	-	-	$(90 \times f_s, 1)$
2	Conv1D	128	128	2	Relu	$(45 \times f_s, 128)$
3	M-Apooling1D	-	3	2	-	$(\lceil 45 \times f_s/2 \rceil, 256)$
4	MC Block1	-	-	-	Relu	$(\lceil 45 \times f_s/2 \rceil, 272)$
5	M-Apooling1D	-	3	2	-	$(\lceil 45 \times f_s/4 \rceil, 544)$
6	Dropout(0.1)	-	-	-	-	$(\lceil 45 \times f_s/4 \rceil, 544)$
7	MC Block2	-	-	-	Relu	$(\lceil 45 \times f_s/4 \rceil, 272)$
8	M-Apooling1D	-	3	2	-	$(\lceil 45 \times f_s/8 \rceil, 544)$
9	Dropout(0.1)	-	-	-	-	$(\lceil 45 \times f_s/8 \rceil, 544)$
10	GAP	-	-	-	-	544
11	Dropout(0.5)	-	-	-	-	544
12	Dense	-	-	-	Softmax	5

B. Regularization

We adopt two regularization approaches to help prevent the overfitting problem. The first technique is L2 regularization that adds squared magnitude of coefficient as penalty term to the loss function. It is important to choose a proper regularization rate (lambda), if lambda is very large, it would add too much weight causing an underfitting issue. By contrast, a very small lambda would make the model more complex, then the model would learn too much about the particularities of the training data, L2 regularization therefore has little effect on avoiding overfitting. Hence, we test four lambda values: 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}, the results show that 10^{-3} achieves the best performance. The L2 regularization is applied to all convolutional layers, including the MC block.

Another regularization method is dropout, which randomly drops units from the model during training with a specific probability from 0 to 1. It is noteworthy that the dropout layers is not used for testing. Dropout layers with probability of 0.1 and 0.5 are employed for the MC block and GAP layer, respectively.

C. Training Setup

We select Adam as the network optimizer whose parameters ((learning rate) lr, $beta1$ and $beta2$) are set to 10^{-3}, 0.9 and

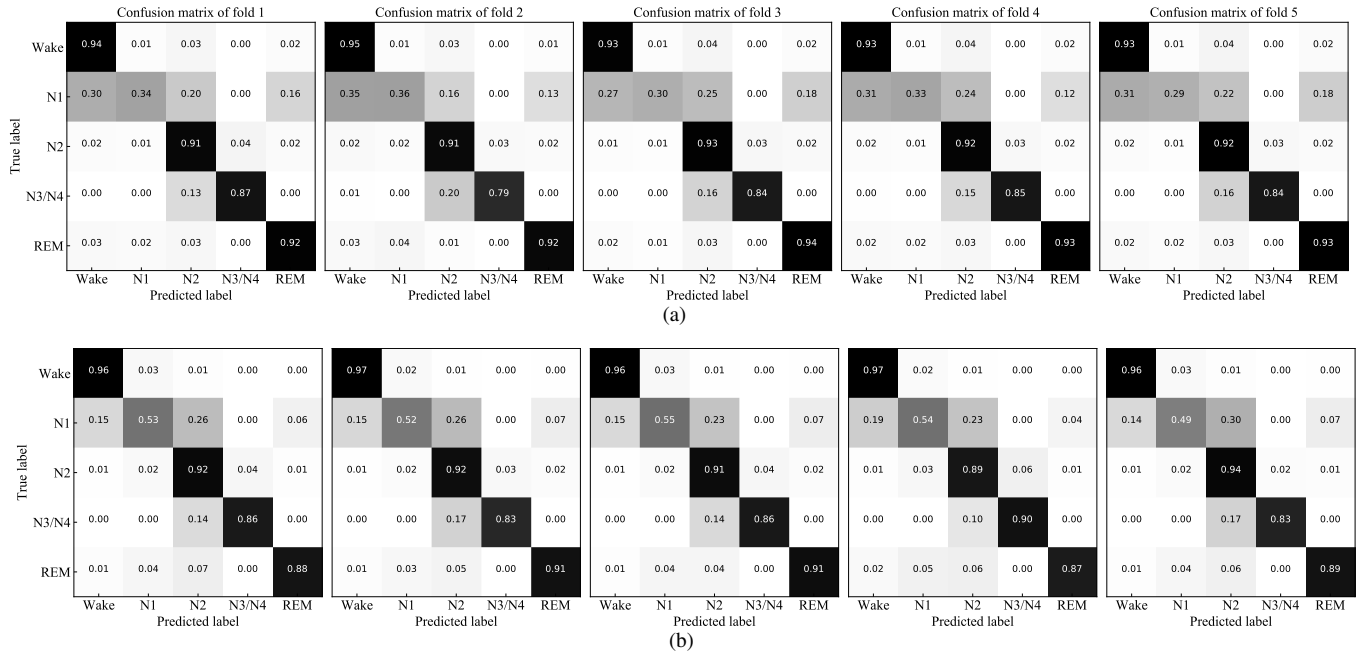

Fig. 3. The normalized confusion matrices of each fold cross-validation from two datasets. (a) SHHS dataset and (b) Sleep-EDF dataset.

0.999 respectively. Moreover, ReduceLROnPlateau of Callback in Keras is implemented to reduce the lr. Specifically, when the model monitors the validation accuracy showing no improvement within 3 epochs, the lr would drop to half of it. The minimum lr is set to 10^{-7}. To find out appropriate batch size of mini-batch, size of 32, 64, 128, and 256 are evaluated, we select 64 as the size of mini-batch finally. The categorical cross entropy is chosen as the loss function of the model which is always used for classifying multi-class tasks. The model converges to the optimal solution within 40 iterations, hence the number of iteration is set to 40.

We use the 20% of whole data set as the test set and the remaining 90s epochs as the train set. Furthermore, we use the 5-fold cross-validation (80% train set for training, 20% train set for validation) scheme to train and evaluate our model for both datasets. It is important to notice that the epochs from test set are not involved in optimizing hyper-parameters of the model, thus the test set can be completely independent. We argue that the training and evaluation process should not include any epochs from the test set. It has been reported non-independent test set would result in an improvement of the performance [13], [31]. In addition, only 90s epochs from the SHHS dataset are used to determine the hyper-parameters of the proposed model. Once achieving optimal hyper-parameters, they would be used in all experiments. To be specific, when the model is applied to another dataset, there would be no need to modify the architecture and hyper-parameters of the model except for the input length which should adapt to the f_s of EEG from different datasets.

Graphic card Nvidia Tesla P100 with 16 Gbytes memory is used for model training. The implementation is written in Keras [41] with the Tensorflow backend [42].

IV. Experimental Results

TABLE III
Mean Confusion Matrix of 5-Fold Cross-validation on Raw Single-channel EEG C4/A1 from the SHHS Dataset

	Predicted					Per-class Metrics			Overall Metrics	
	Wake	N1	N2	N3/N4	REM	$PR(\%)$	$RE(\%)$	$F1(\%)$	$ACC(\%)$	$K(\%)$
Wake	4447	62	176	5	77	**91.5**	**93.3**	**92.4**		
N1	192	202	134	0	96	47.4	32.4	38.5		
N2	142	100	7583	271	171	89.9	91.7	90.1	89.2	84.8
N3/N4	9	0	449	2350	1	89.5	83.7	86.5		
REM	69	62	91	1	2763	88.9	92.5	90.7		

TABLE IV
Mean Confusion Matrix of 5-Fold Cross-validation on Raw Single-channel EEG Fpz-Cz from the Sleep-EDF Dataset

	Predicted					Per-class Metrics			Overall Metrics	
	Wake	N1	N2	N3/N4	REM	$PR(\%)$	$RE(\%)$	$F1(\%)$	$ACC(\%)$	$K(\%)$
Wake	10478	270	80	5	41	**95.7**	**96.4**	**96.0**		
N1	356	1186	588	6	142	64.6	52.1	57.7		
N2	63	246	8368	335	121	87.8	91.6	89.7	89.4	85.0
N3/N4	4	2	308	1831	1	84.0	85.3	84.7		
REM	46	132	189	2	2929	90.6	88.8	89.7		

A. Performance Metrics

We evaluate the model performance using accuracy (ACC), precision (PR), recall (RE), F1 score ($F1$), and Cohen's kappa coefficient (K). ACC is the proportion of correct predictions made by the model to the total predications. PR calculates the ratio of correctly predicted positives to all positives. RE means the fraction between true positives and all predications in the actual class. $F1$ represents the weighted average of PR and RE. K measures the agreement between

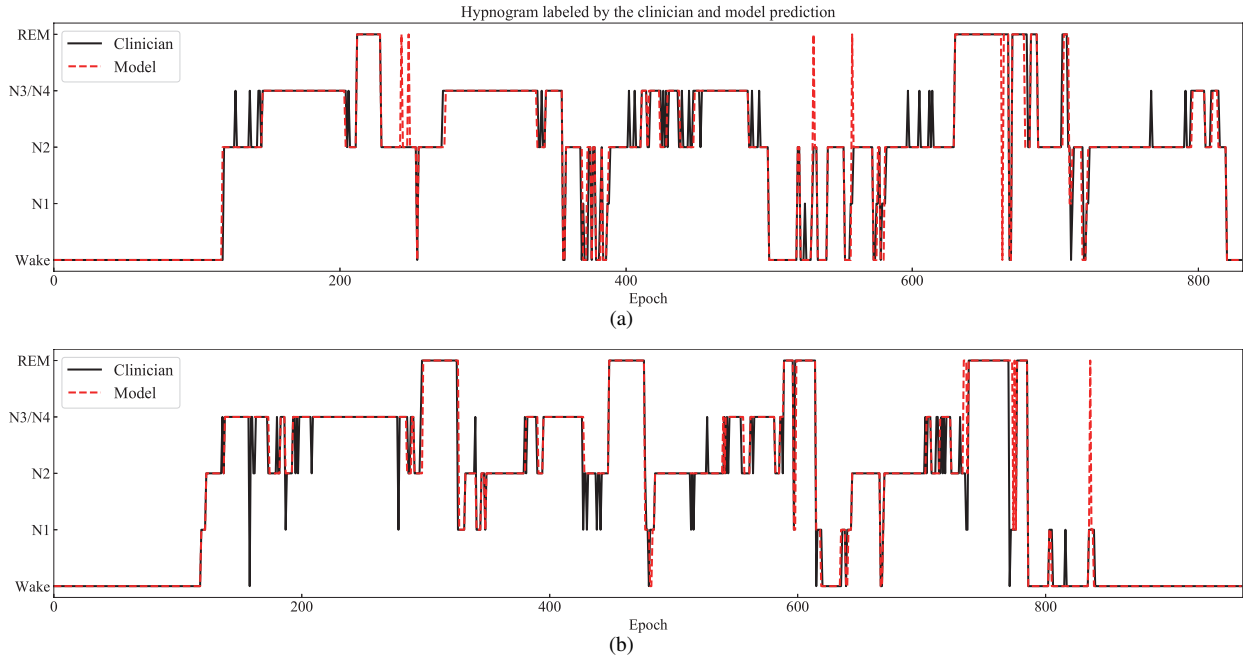

Fig. 4. The comparison between hypnogram labeled by the clinician and the model's prediction of two subjects. The solid black line is the ground truth, the dotted red line donates the hypnogram labeled by the prediction of the proposed model. (a) SHHS dataset and (b) Sleep-EDF dataset.

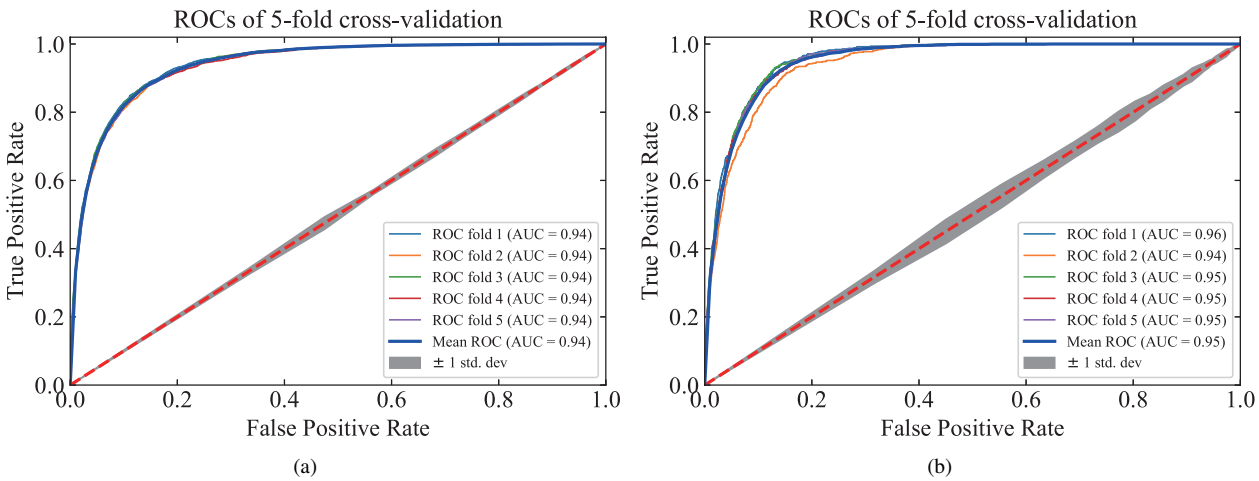

Fig. 5. The ROCs together with corresponding AUCs from 5-fold cross-validation. (a) SHHS dataset and (b) Sleep-EDF dataset.

true labels and predicted labels. A large value of K can indicate good performance of the model. They are calculated as follows:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}. \tag{3}$$

$$PR = \frac{TP}{TP + FP}. \tag{4}$$

$$RE = \frac{TP}{TP + FN}. \tag{5}$$

$$F1 = 2 \cdot \frac{RE \cdot PR}{RE + PR}. \tag{6}$$

$$K = \frac{\frac{\sum_{i=1}^{n} x_{ii}}{N} - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}{1 - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}. \tag{7}$$

where TP, TN, FN and FP donate the true positives, true negatives, false negatives and false positives, respectively. N is the number of 90s epochs of the test set, n represents the number of classes. In this work, n equals 5, x_{ii} $(1 \leq i \leq 5)$ represents the diagonal value of the confusion matrix.

To show the performance of each fold cross-validation from the SHHS and Sleep-EDF datasets, we present the normalized confusion matrices (CM) in Fig. 3. For each fold cross-validation, the test set is the same and independent from model training and validation. Firstly, we use single-channel EEG C4/A1 (90s epochs) from the SHHS dataset to tune the hyper-parameters. Once getting the best performance, the hyper-parameters and model architecture are fixed for all experiments. Table III provides the mean CM of 5-fold cross-validation from the SHHS dataset, we can see that the overall accuracy and K are respectively 89.2% and 84.8%.

The proposed model shows the best ability to detect the Wake stage with the PR of 91.5%. By contrast, the performance of stage N1 classification is the worst which is consistent with the results of existing works. To be specific, there are 32.4% of N1 90s epochs being recognized correctly. In addition, 30.8% of N1 samples are misclassified as Wake, 21.5% as N2 and 15.3% as REM. Stages N2, N3/N4 and REM have similar classification results in terms of the PR corresponding to 89.9%, 89.5% and 88.9% respectively.

To demonstrate the generalization capability of the proposed architecture, we conduct the 5-fold cross-validation using the same model determined by the SHHS dataset (i.e., without any hyper-parameters modification except for the input length) on the Sleep-EDF dataset. As can be seen from Table I, the distribution of the numbers of five stages is a bit different. Stage Wake has the biggest proportion and the number of N3/N4 is the smallest in Sleep-EDF dataset, whereas the smallest percentage is stage N1 in the SHHS dataset. Besides, the EEG channel used in two datasets is also distinct, C4/A1 for the SHHS dataset and Fpz-Cz for the Sleep-EDF dataset. It is worthy to note that despite the EEG channel and the size of the input length ($90 \times f_s$, 1) are quite different, the proposed model can obtain promising performance on two different datasets by comparing Table III and Table IV.

We further reveal the hypnogram comparison labeled by experts and the model's prediction for one subject of SHHS and Sleep-EDF datasets in Fig. 4.

B. ROC and AUC

A receiver operating characteristic (ROC) curve is originally used as a performance measurement for the binary classification problem at different thresholds settings. The ROC curve is plotted with the true positive rate (TPR) against the false positive rate (FPR) where TPR is on the y-axis and FPR is on the x-axis. In our multi-class classification task, we treat the true label as positive, others as false for each 90s epoch in the test set. Furthermore, we calculate the area under the curve (AUC) of ROC which represents the degree or measure of separability. The AUC can indicate how much the model is capable of distinguishing between classes, higher the AUC, the better the model is. We show in Fig. 5 the ROCs and AUCs from 5-fold cross-validation. Moreover, mean ROC and corresponding AUC are obtained from the ROC of each fold cross-validation, the gray area represents the standard deviation (STD). As can be seen from Fig. 5, for both datasets, the ROC of each fold cross-validation as well as mean ROC can achieve a large AUC which nears to 1, showing a good measure of separability. More specifically, the mean AUC is 0.94 and 0.95 from SHHS and Sleep-EDF datasets, respectively.

C. Performance comparison

We make a comparison between the proposed model with some existing works using the same datasets in terms of the ACC and K in Table V and Table VI. Table V provides detailed comparison information about recent studies and the proposed method on the SHHS dataset. It shows that the proposed framework can achieve higher ACC and K using raw single-channel C4/A1 EEG compared to approaches using hand-crafted features as input [12] or multi-channel PSG data [27], [30] or the single-channel EEG [32], [33]. Similarly, Table VI demonstrates that the proposed model outperforms state-of-the-art methods on the Sleep-EDF dataset. Some studies [4], [13], [43] extract features manually or multi-channel signals are used as input [6], [29] or some methods adopt single-channel EEG [31], [34], [35]. Considering Table V and Table VI, the proposed framework can achieve promising performance on SHHS and Sleep-EDF datasets.

V. Discussion and Conclusion

In this paper, we propose an end-to-end framework with CNNs, namely SCNet, which combines the feature learning ability and classification capacity. The proposed model is applied to classify sleep stages automatically from raw single-channel EEG without using any hand-engineered features and any other preprocessing (e.g., signal filtering and resample implementation). There are two main advantages that we train and evaluate the model with raw single-channel EEG. Comparing with those methods with hand-crafted features [4], [12], [43], where extracting hand-engineered features is conducted with priori knowledge and not in a data-driven way, and it is time-consuming for the researchers. Moreover, the selection of types and number of features would result in different model performance, there is no gold standard about the extraction of hand-crafted features. The second advantage is that it is much easier and more comfortable to record single-channel EEG data compared to the multi-channel scheme [27], [29], [31] either at the hospital or home. Moreover, multi-channel PSG data used as input can increase the computational cost. Considering practical applications, the use of raw single-channel EEG can simplify the measurement scheme and reduce the related cost.

Comparing with the conventional deep neural network based on CNNs, where the convolutional layers with the fixed filter size are assembled in sequence. In such a case, it is not capable of capturing features representation from different scales. To address this issue, our model employs two MC blocks, which are the concatenation of several convolutional layers with four distinct filer sizes, to extract different scale features. Instead of using the traditional max-pooling layer, we adopt the M-Apooing layer to add average feature representation with maximum features simultaneously, which further improve the proposed model's ability of feature learning. The results demonstrate that, without changing the architecture, hyper-parameters of the model and the training algorithm, the proposed architecture obtains higher ACC and K compared to the existing methods. In addition, the SCNet model is quite simple and compact with a total 5×10^5 parameters compared to the methods in [13] which has 1.4×10^8 parameters and [31] in which the number of parameters of the representation learning and sequence residual learning parts has up to 6×10^5 and 2×10^7 respectively. Moreover, the proposed SCNet model can achieve the better performance with less computing resources occupied. Concerning online and realtime applications (e.g.,

TABLE V
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND PREVIOUS METHODS ON THE SHHS DATASET

Study	Method	Input channel	Preprocessing	Input type	Feature type	Subjects	$ACC(\%)$	$K(\%)$
Proposed	Deep CNN	C4-A1	No	Time series	learned	100	**89.2**	**84.8**
Sors *et al.* [32]	Deep CNN	C4-A1	No	Time series	learned	5728	87	81
Seo *et al.* [33]	CNN + LSTM	C4-A1	No	Time series	learned	5791	86.7	79.8
Varela *et al.* [27]	Deep CNN	2EEG+ 2EOG + EMG	Yes	Time series	learned	1000	78	83
Zhang *et al.* [30]	CNN + LSTM	2EEG + 2EOG + EMG	Yes	Spectrogram	learned	5793	87	82
Karimzadeh *et al.* [12]	Distributed Classifier	C4-A1	Yes	Features	hand-crafted	140	83.2	-

TABLE VI
PERFORMANCE COMPARISON BETWEEN THE PROPOSED METHOD AND PREVIOUS METHODS ON THE SLEEP-EDF DATASET

Study	Method	Input channel	Preprocessing	Input type	Feature type	Subjects	$ACC(\%)$	$K(\%)$
Proposed	Deep CNN	Fpz-Cz	No	Time series	learned	61	**89.4**	**85**
Mikkelsen *et al.* [29]	Deep CNN	Fpz-Cz + EOG	No	Time series	learned	20	84.0	-
Supratak *et al.* [31]	CNN + LSTM	Fpz-Cz	No	Time series	learned	20	82.0	76
Wei *et al.* [34]	Deep CNN	Fpz-Cz	Yes	Time series	learned	20	84.3	78
Phan *et al.* [6]	Deep CNN	Fpz-Cz + EOG	Yes	Time-frequency image	learned	20	82.3	75
Phan *et al.* [35]	1-max CNN	Fpz-Cz	Yes	Time-frequency image	learned	20	79.8	72
Phan *et al.* [36]	Attentional RNN	Fpz-Cz	Yes	Feature vectors	learned	20	79.1	70
Zhu *et al.* [4]	SVM	Pz-Oz	Yes	Features	hand-crafted	8	88.9	83
Hsu *et al.* [43]	RNN	Fpz-Cz	Yes	Features	hand-crafted	8	87.2	-
Tsinalis *et al.* [13]	Deep auto-encoder	Fpz-Cz	Yes	Features	hand-crafted	20	78	-

sleep monitoring), our model with raw single-channel EEG is more reasonable to reduce the time delay and obtain reliable results.

Furthermore, to demonstrate the generalization of the proposed architecture, different single-channel EEGs (i.e., C4/A1, Fpz-Cz) from two datasets are adopted. The length of input is not restricted to a fixed number, our model can be adapted to different length of input (125×90, 100×90) relating to the f_s of EEG efficiently, and it is not tested in other studies to our knowledge. Experimental results show that the proposed model can obtain promising performance on two datasets (SHHS: ACC-89.2%, K-84.8%; Sleep-EDF: ACC-89.4%, K-85%), which indicate the desirable generalization of the SCNet model. These properties can benefit the proposed model be more easily adaptable to clinical or wearable devices applications.

In future work, we plan to explore non-EEG signals to conduct the automatical sleep stage classification. Also, it is valuable to adopt clinic datasets that have rarely been used in previous studies.

ACKNOWLEDGMENT

This study is to memorize Prof. Tapani Ristaniemi from University of Jyväskylä for his great help to the nine authors and Prof. Tapani Ristaniemi has supervised this study very much.

REFERENCES

[1] R. Ferri *et al.*, "A quantitative statistical analysis of the submentalis muscle emg amplitude during sleep in normal controls and patients with rem sleep behavior disorder," *J. Sleep Res.*, vol. 17, no. 1, pp. 89–100, 2008.

[2] C. Kuo and S. Liang, "Automatic stage scoring of single-channel sleep eeg based on multiscale permutation entropy," in *Proc. IEEE Biomed. Circuits Syst. Conf (BioCAS).*, pp. 448–451, 2011.

[3] 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.

[4] 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.

[5] H. G. Jo *et al.*, "Genetic fuzzy classifier for sleep stage identification," *Comput. Biol. Med.*, vol. 40, no. 7, pp. 629–634, 2010.

[6] 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.

[7] 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.

[8] 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.

[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] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel eeg signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, 2012.

[11] K. Šušmáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, 2008.

[12] F. Karimzadeh *et al.*, "A distributed classification procedure for automatic sleep stage scoring based on instantaneous electroencephalogram phase and envelope features," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 362–370, 2017.

[13] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Ann. Biomed. Eng.*, vol. 44, no. 5, pp. 1587–1597, 2016.

[14] M. Sharma *et al.*, "An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank," *Comput. Biol. Med.*, vol. 98, pp. 58–75, 2018.

[15] S. Kouchaki *et al.*, "Tensor based singular spectrum analysis for automatic scoring of sleep eeg," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 1, pp. 1–9, 2014.

[16] H. Phan *et al.*, "Metric learning for automatic sleep stage classification," in *Proc. IEEE Eng. Med. Biol. Soc (EMBC).*, pp. 5025–5028, 2013.

[17] E. Alickovic and A. Subasi, "Ensemble svm method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, 2018.

[18] T. Lajnef *et al.*, "Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods.*, vol. 250, pp. 94–105, 2015.

[19] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017.

[20] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on eeg signal using k-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, 2010.

[21] L. Fraiwan *et al.*, "Automated sleep stage identification system based on time–frequency analysis of a single eeg channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012.

[22] M. Xiao *et al.*, "Sleep stages classification based on heart rate variability and random forest," *Biomed. Signal Process. Control.*, vol. 8, no. 6, pp. 624–633, 2013.

[23] T. L. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel eeg sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Med. Biol. Eng. Comput.*, vol. 55, no. 2, pp. 343–352, 2017.

[24] P. Memar and F. Faradji, "A novel multi-class eeg-based sleep stage classification system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 84–95, 2017.

[25] H. Phan *et al.*, "Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, 2019.

[26] X. Zhang *et al.*, "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Comput. Biol. Med.*, vol. 103, pp. 71–81, 2018.

[27] I. Fernández-Varela, E. Hernández-Pereira, and V. Moret-Bonillo, "A convolutional network for the classification of sleep stages," *Multidiscip. Digit. Publ. Inst. Proc.*, vol. 2, no. 18, p. 1174, 2018.

[28] S. Chambon *et al.*, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, 2018.

[29] K. Mikkelsen and M. De Vos, "Personalizing deep learning models for automatic sleep staging," *arXiv Prepr. arXiv:1801.02645.*, 2018.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] Q. Wei *et al.*, "A residual based attention model for eeg based sleep staging," *IEEE J. Biomed. Health. Inform.*, 2020.

[35] H. Phan *et al.*, "Dnn filter bank improves 1-max pooling cnn for single-channel eeg automatic sleep stage classification," in *Proc. IEEE Eng. Med. Biol. Soc (EMBC).*, pp. 453–456, 2018.

[36] H. Phan *et al.*, "Automatic sleep stage classification using single-channel eeg: Learning sequential features with attention-based recurrent neural networks," in *Proc. IEEE Eng. Med. Biol. Soc (EMBC).*, pp. 1452–1455, 2018.

[37] 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.

[38] 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.

[39] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit (CVPR).*, pp. 1–9, IEEE, 2015.

[40] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv Prepr. arXiv:1312.4400.*, 2013.

[41] F. Chollet *et al.*, "Keras: Deep learning library for theano and tensorflow," *URL: https://keras. io/k.*, vol. 7, no. 8, p. T1, 2015.

[42] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv Prepr. arXiv:1603.04467.*, 2016.

[43] Y. L. Hsu *et al.*, "Automatic sleep stage recurrent neural classifier using energy features of eeg signals," *Neurocomputing.*, vol. 104, pp. 105–114, 2013.