ENHANCING CONTEXTUAL ENCODING WITH STAGE-CONFUSION AND STAGE-TRANSITION ESTIMATION FOR EEG-BASED SLEEP STAGING

Jaeun Phyo[1], Wonjun Ko[1], Eunjin Jeon[1], and Heung-Il Suk[1, 2, †]

[1]*Department of Brain and Cognitive Engineering, Korea University*
[2]*Department of Artificial Intelligence, Korea University*
Anam-ro 145, Seoul 02841, Republic of Korea

ABSTRACT

Sleep staging is essential for sleep assessment and plays a vital role as one of the health indicators. It is challenging to correctly classify stage-transitioning epochs of sleep electroencephalography (EEG) because of their mixed signals of stages. To this end, recent studies exploited and devised various deep learning architectures. However, those are still suffering from confusing two or more stages, especially in stage-transitioning epochs. In this work, we propose a novel network architecture that takes advantage of two auxiliary classification tasks and exploits their outputs to adapt feature representations, thus effectively discriminating confusing stages. Specifically, one auxiliary task is an *epoch-level stage classification* to produce confidence scores about stages. The other is a *stage-transition detection* to learn inter-epoch relations. Using inferred information about stage-confusion at an epoch level and stage-transition across neighboring epochs helps learn more concrete representations for stage identification. We demonstrated and analyzed the validity of our proposed method over two publicly available datasets, achieving promising performances.

***Index Terms*—** sleep staging; electroencephalography; deep learning; sequence-to-sequence;

1. INTRODUCTION

Sleep staging is an important factor in measuring sleep quality and diagnosing sleep-related diseases [1]. The brain goes through series of different stages during sleep that can be specified by distinct waveform, amplitude, and dominant frequency range. Typically, those sleep stages can be categorized into five stages accroding to the American Academy of Sleep Medicine (AASM) manual: Wake (W), Rapid Eye Movements (REM), Non-REM1 (N1), Non-REM2 (N2), and Non-REM3 (N3) [2]. For sleep staging, sleep experts identify stages using brain monitoring methods such as electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG), which is manual, labor-intensive, and subjective [3].

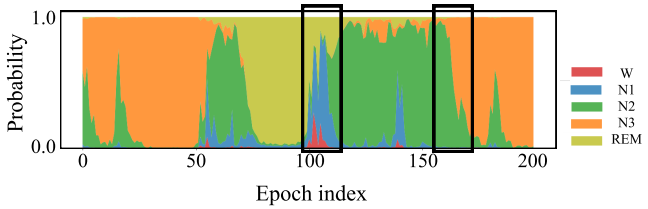

Fig. 1. Posterior probability distribution over sleep stages. The solid rectangles denote transitioning epochs, which have multipule stage properties.

To this end, many pioneering studies have developed machine learning or deep learning-based automatic sleep staging methods [4–6]. In particular, thanks to its methodological impacts, recent state-of-the-art deep networks for sleep staging have adopted an architectural form that contains a *feature extractor* and a *context encoder* jointly [5, 7, 8]. In this form, the feature extractor learns spectro-temporal features in the intra-epoch of an input signal, and the context encoder aims to represent inter-epoch relations [5, 7, 8].

Commonly, it is more challenging to classify transitioning epochs compared to non-transitioning ones correctly [5, 8, 9]. In Fig.1, transitioning epochs in solid rectangles may contain mixed signals of multiple stages, which cause a classifier hard to predict confidently. To tackle this issue, we hypothesize that if the context encoder knows which stages are confusing and exploits such information to adapt representations, it can help classify stages more accurately. In this regard, we propose to take advantage of two auxiliary tasks that allow inferring the information about which stages are confusing and whether stage-transition occurs. The inferred information is then used via an attention mechanism for a context encoder to update representations adaptively and make a better decision.

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government under Grant 2017-0-00451 (Development of BCI based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning) and Grant 2019-0-00079 (Department of Artificial Intelligence, Korea University). †: corresponding author

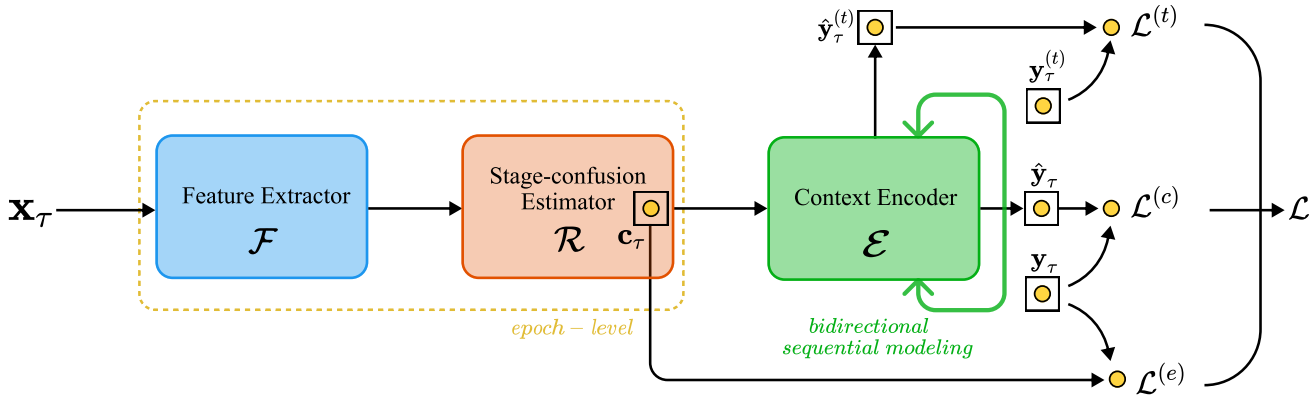

Fig. 2. Overall architecture of the proposed model for sleep staging. Our model consists of an epoch-level *feature extractor* $\mathcal{F}$, a *stage-confusion estimator* $\mathcal{R}$, and a *context encoder* $\mathcal{E}$. The proposed model exploits two auxiliary tasks to utilize their outputs to adapt feature representations, therefore effectively discriminates confusing stages. In detail, class probabilities $\mathbf{c}_\tau$ in the stage-confusion estimator module infer which stages are confusing. Meanwhile, a predicted stage-transition vector $\mathbf{y}_\tau^{(t)}$ in the context encoder gives explicit information about stage-transition across neighboring epochs. The total objective function $\mathcal{L}$ consists of three loss functions: $\mathcal{L}^{(e)}$ for the epoch-level stage classification, $\mathcal{L}^{(t)}$ for the stage-transition detection, and $\mathcal{L}^{(c)}$ for the final sleep stage prediction $\hat{\mathbf{y}}_\tau$.

2. RELATED WORK

There have been many approaches to learn contextual dependencies from neighboring epochs in sleep staging methods. Modeling contextual information is significant since sleep stages are labeled by sleep scorers considering transition rules [2]. Supratak *et al.* [5] and Phan *et al.* [9] independently employed Recurrent Neural Network (RNN) to model the contextual dependencies of neighboring epochs. Qu *et al.* [8] utilized multi-head attention for contextual encoding. Although those methods extract contextual information from neighboring epochs, they ignore which feasible stages are confused at an epoch, thus leading to limited modeling of contextual information. Unlike previous works, our proposed framework allows the context encoder to learn and adapt to a circumstance of stage-confusion and stage-transition, inferred from specially designed modules with two auxiliary tasks.

Attention can be used as guidance to bias the portion of available features to the most informative components of an input. For instance, squeeze and excitation network [10] exploited a channel-wise attention mechanism by adaptively calibrating the channel features. They parameterized the gating mechanism by forming a bottleneck with two densely connected layers, where the network could learn the attentive features. Compared to the previous method, we let a bottleneck layer intend to learn the probability of each sleep stage of an epoch as a pre-classification phase. Instead of learning the channel-wise dependencies, our model learns possible stages within an epoch through pre-classification, which leads to effective modeling of contextual dependencies.

3. METHODS

Given a sleep sample $\mathbf{X}$ that includes N single-channel raw EEG epochs $\left[\mathbf{x}_\tau \in \mathbb{R}^T\right]_{\tau=1}^N$, where T denotes the number of time points in an epoch, the task of sleep staging is to assign a stage label to each epoch. This work focuses on learning feature representations of the EEG signals by jointly exploiting the stage-confusion inferred at an epoch-level and the stage-transition over the neighboring epochs for contextual information. Our proposed network consists of three main modules: epoch-level feature extractor $\mathcal{F}$, stage-confusion estimator $\mathcal{R}$, and inter-epoch context encoder $\mathcal{E}$ as schematized in Fig.2.

3.1. Epoch-level Feature Extractor

We use the multi-scale neural network [11] to obtain an F-dimensional feature representation $\mathbf{f}_\tau$ for an epoch EEG $\mathbf{x}_\tau$. To be specific, the epoch-level feature extraction module $\mathcal{F}$ exploits two multi-scale 1D convolutional layer paths. Each path has a spectral convolution, followed by temporal convolutions. We employ different kernel sizes for temporal convolutions to learn intra-epoch dynamics [5, 11]. Thus, each path in this module learns different ranges of spectro-temporal properties [5]. Moreover, intermediate activations of each temporal convolution are gathered to represent multi-scale spectro-temporal information [11]. Finally, features extracted from two paths are concatenated and *global average pooled* [12].

3.2. Stage-confusion Estimator

The stage-confusion estimator module $\mathcal{R}$ estimates confusing stages by an auxiliary classifier at an epoch level and reflects that information to the epoch-level feature representation via an attention mechanism. Particularly, given the epoch-level feature representation $\mathbf{f}_\tau$ from the preceding step, it first computes the class probabilities $\mathbf{c}_\tau$ with a logistic regression function as follows: $\mathbf{c}_\tau = \mathrm{softmax}(\mathbf{f}_\tau \mathbf{W}_q + \mathbf{b}_q)$, where $\mathbf{W}_q \in \mathbb{R}^{F \times C}$ and $\mathbf{b}_q \in \mathbb{R}^C$ are the learnable parameters of weighting coefficients and bias, respectively. The individual class probability denotes the confidence of the respective class membership. Meanwhile, the class probabilities jointly carry the information that which classes are confusing. In this sense, we use the class probabilities $\mathbf{c}_\tau$ as a latent source of information to update the epoch-level feature representation with an attention mechanism. Specifically, we obtain an attention vector $\mathbf{a}_\tau$ from the class probabilities by applying a series of linear and non-linear operations as follows: $\mathbf{a}_\tau = \mathrm{sigmoid}(\mathbf{W}_r \mathbf{c}_\tau + \mathbf{b}_r)$, where $\mathbf{W}_r \in \mathbb{R}^{C \times F}$ and $\mathbf{b}_r \in \mathbb{R}^F$ are the tunable parameters. Finally, the epoch-level feature representation is updated according to the attention vector via Hadamard multiplication, i.e., $\tilde{\mathbf{f}}_\tau = \mathbf{a}_\tau \otimes \mathbf{f}_\tau$. It is noteworthy that the attention-guided representation $\tilde{\mathbf{f}}_\tau$ entails both the signal-level features and the stage-confusing information jointly.

3.3. Inter-epoch Context Encoder

The context encoder module $\mathcal{E}$ embeds the inter-epoch relations from $\mathbf{F}_{1:N} = [\mathbf{f}_1 \cdots \mathbf{f}_N]$, $\tilde{\mathbf{F}}_{1:N} = [\tilde{\mathbf{f}}_1 \cdots \tilde{\mathbf{f}}_N]$ and predicts the sleep stages over all N epochs, i.e., $\mathcal{E}(\mathbf{F}_{1:N}, \tilde{\mathbf{F}}_{1:N}) = \hat{\mathbf{Y}}_{1:N}$, where $\hat{\mathbf{Y}}_{1:N} = [\hat{\mathbf{y}}_1 \cdots \hat{\mathbf{y}}_N]$ denotes the predicted labels. In this module, we additionally define an auxiliary task of stage-transition detection, with which the inter-epoch relations embedding can be better estimated. Basically, given a sequence of ground-truth stage labels in the training set, we can obtain the stage-transition labels, i.e., '*transition*' (1) and '*no-transition*' (0), as byproducts. In particular, we define a transition label for a τ-th epoch as follows:

$$\mathbf{y}_\tau^{(t)} = \begin{cases} 1 & \text{if } \mathbf{y}_\tau \neq \mathbf{y}_{\tau-1} \text{ or } \mathbf{y}_\tau \neq \mathbf{y}_{\tau+1} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Armed with the additional label information of stage-transition, we formulate the inter-epoch context encoder training in multi-task learning. That is, the module is trained to optimally predict the sleep stages and the occurrence of stage transition at an epoch compared to the neighboring epochs.

We use a bidirectional Long Short-Term Memory [13] for the inter-epoch relations embedding and the sequence-to-sequence classification. In regards to the transition detection and the sleep stage classification, the hidden state $\mathbf{h}_\tau$ at each epoch is fed into the respective classifiers to make detections.

3.4. Training Procedure

To train model parameters of three modules in our framework, we jointly optimize three learning tasks, two auxiliary tasks and one main task: early epoch-level stage classification, sequence-level transition detection, and sequence-level stage classification. First, for the early epoch-level stage classification task, we define a class-weighted cross-entropy (WCE) function defined as follows:

$$\mathcal{L}^{(e)} = \mathrm{WCE}(\mathbf{y}_\tau, \mathbf{c}_\tau) = -w_c \sum_\tau \mathbf{y}_\tau \cdot \log \mathbf{c}_\tau \tag{2}$$

where, $\mathbf{y}_\tau$ is the ground truth stage of the τ-th epoch and w_c denotes the inverse proportion of c-class samples in the training set.

The objective function for the stage-transition detection task is defined as $\mathcal{L}^{(t)} = \mathrm{WCE}(\mathbf{y}_\tau^{(t)}, \hat{\mathbf{y}}_\tau^{(t)})$, where $\hat{\mathbf{y}}_\tau^{(t)}$ is stage-transition prediction.

Finally, the total loss function $\mathcal{L}$ is computed as $\mathcal{L} = \mathcal{L}^{(c)} + \lambda^{(e)}\mathcal{L}^{(e)} + \lambda^{(t)}\mathcal{L}^{(t)}$ where $\mathcal{L}^{(c)} = \mathrm{WCE}(\mathbf{y}_\tau, \hat{\mathbf{y}}_\tau)$ where $\hat{\mathbf{y}}_\tau$, $\lambda^{(e)}$ and $\lambda^{(t)}$ are predicted sleep stage and coefficient constants multiplied to each loss term.

4. EXPERIMENTS

4.1. Datasets

We used two public EEG datasets for sleep staging to evaluate our proposed method. The Sleep-EDF [14] contains the polysomnography (PSG) recordings from 20 healthy subjects. This dataset is a widely-used dataset for sleep analysis, which consists of horizontal EOG, Fpz-Cz and Pz-Oz EEG at a sampling rate of 100 Hz along with an EMG. According to other baselines [5, 15], we adopted Fpz-Cz channels of recordings. The stages N3 and N4 were merged as N3, i.e., $C = 5$, and continuous wake epochs longer than 30 minutes outside the sleep period were also ignored in our experiments. We further evaluated MASS-SS3 [16] dataset, which contains the PSG recordings from 62 healthy subjects (28 male and 34 female). Similar to existing studies [5, 8], we used F4-LER channel among 27 channels for fair comparison. All EEG signals were scaled to median 0 and inter quantile range 1, band-pass filtered within 0.5 to 49 Hz, and downsampled to 100 Hz.

4.2. Training Details

We used a densely connected layer with the softmax activation function for all classifiers in the model. Also, we employed the Adam optimizer [17] at a learning rate of 10^{-3} for the loss function defined in Section 3.4. We regularized all tunable parameters by a Ridge regression ($\ell_2 = 0.001$). In the experiment, N, $\lambda^{(e)}$ and $\lambda^{(t)}$ were set as 25, 0.5 and 0.5 respectively. Further, for fair comparison, we performed 20 and 31-fold cross-validation for the Sleep-EDF and MASS

Table 1. Performance comparison of the Sleep-EDF Fpz-Cz channel.

Method	Overall results					
	MF1	W	N1	N2	N3	REM
Supratak *et al.* [5]	76.9	84.7	46.6	85.9	84.8	82.4
Perslev *et al.* [15]	78.6	87.1	51.5	86.4	84.2	83.70
Seo *et al.* [18]	77.6	87.7	43.4	87.7	84.8	82.4
Qu *et al.* [8]	79.0	**90.2**	48.3	87.8	85.6	83.0
Ours	**81.0**	89.1	**54.8**	**88.5**	**87.2**	**85.0**

Table 2. Performance comparison of the MASS F4-LER Channel.

Method	Overall results					
	MF1	W	N1	N2	N3	REM
Supratak *et al.* [5]	81.2	**87.5**	55.4	91.3	84.8	87.2
Qu *et al.* [8]	81.0	87.2	52.9	91.5	**87.0**	86.6
Ours	**81.3**	86.0	**56.9**	**91.6**	84.4	**87.8**

datasets, respectively. Our experimental code is available at: `https://github.com/ku-milab/CE_SCST`.

4.3. Results

Table 1 and Table 2 report performance comparison on Sleep-EDF and MASS between our method and other comparable sleep stage scoring methods. We only considered studies that utilized deep learning methods which learn sequence-to-sequence with raw single-channel EEGs. We reported classification performances across Macro-averaged F1 scores (MF1) and F1 scores for each sleep stage: W, N1, N2, N3, and REM. F1 score is a common metric to evaluate the performance of the models on imbalanced datasets [19] and adopted in several previous sleep staging studies [4, 5].

Table 1 shows that our model outperforms the previous studies in MF1 score on Sleep-EDF datasets. Note that the proposed network achieved significant improvements in N1 and REM. These stages are almost indistinguishable by visual inspection but show difference in the viewpoint of the contextual information [20, 21]. Based on these results, we concluded that our proposed method effectively learns the inter-epoch relations.

Further, we also observed that the proposed method showed plausible scores on MASS dataset as reported in Table 2. Although the performance is slightly lower in W, N2, and N3 stages, the proposed architecture achieved the best performance in N1 and REM.

We also visualized 150 sequences of predicted label from the stage-confusion estimator $\arg\max \mathbf{c}_\tau$ (Fig. 3(a)), the final decision $\hat{\mathbf{y}}_\tau$ (Fig. 3(b)), and the ground truth $\mathbf{y}_\tau$ (Fig. 3(c)). Note that the proposed model (Fig. 3(b)) classified confusing stages better than the comparable one (Fig. 3(a)) by exploiting inter-epoch relations.

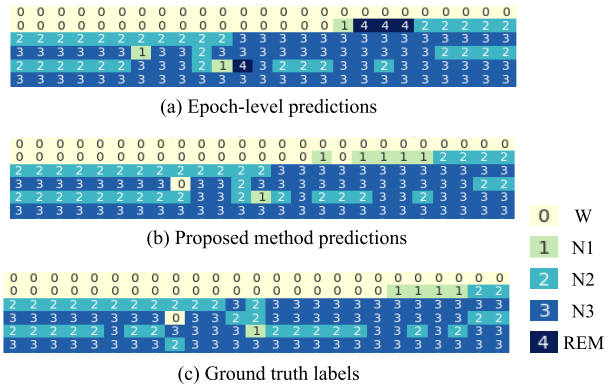

Fig. 3. Examples of predicted and ground truth labels: (a) epoch-level predictions, (b) proposed method predictions, and (c) ground truth labels.

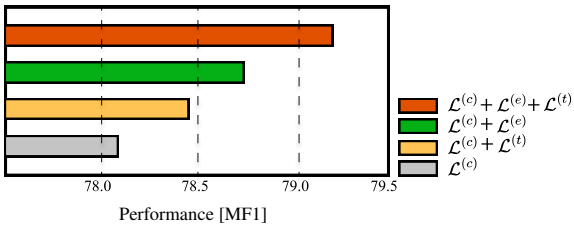

Fig. 4. Macro-averaged F1 scores of ablation cases on Sleep-EDF dataset. Our proposed method (red-colored) achieved the best performance.

4.4. Ablation Study

Note that our model introduced two auxiliary tasks, the epoch-level stage classification and the stage-transition detection. In order to further verify the efficacy of each task, we performed ablation experiments as depicted in Fig 4. The performance was best when both auxiliary tasks were used. Based on this study, we concluded that our auxiliary tasks are beneficial for well-generalization.

5. CONCLUSION

In our work, we proposed a novel deep neural network with two auxiliary tasks, namely epoch-level stage classification and stage-transition detection, to discriminate confusing stages well. Our proposed method learns inter-epoch relations effectively by introducing our newly-designed modules with two auxiliary tasks. Additionally, we observed that the proposed method achieved promising performances on sleep staging experiments on two publicly available datasets. In the meantime, there is still room for improvement. For instance, the proposed method would be better classify confusing stages with a layer-wise hierarchical classifier due to the differences and similarities of EEG signals in sleep stages. Thus, we will consider this as our future research.

6. REFERENCES

[1] J. A. Hobson and E. F. Pace-Schott, "The cognitive neuroscience of sleep: neuronal systems, consciousness and learning," *Nat. Rev. Neurosci.*, vol. 3, no. 9, pp. 679–693, 2002.

[2] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, 2012.

[3] N. A. Collop, "Portable monitoring for the diagnosis of obstructive sleep apnea," *Curr. Opin. Pulm. Med.*, vol. 14, no. 6, pp. 525–529, 2008.

[4] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomed. Signal Process. Controll*, vol. 42, pp. 107–114, 2018.

[5] A. Supratak, H. Dong, C. Wu, and Y. Guo, "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.

[6] P. An, Z. Yuan, and J. Zhao, "Unsupervised multi-subepoch feature learning and hierarchical classification for EEG-based sleep staging," *Expert Syst. Appl.*, vol. 186, p. 115759, 2021.

[7] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, and Y. Zhao, "GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, pp. 1324–1330, 2020.

[8] W. Qu, Z. Wang, H. Hong, Z. Chi, D. D. Feng, R. Grunstein, and C. Gordon, "A Residual based Attention Model for EEG based Sleep Staging," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2833–2843, 2020.

[9] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "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.

[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Comput. Vis. Patt. Recognit. (CVPR)*, pp. 7132–7141, 2018.

[11] W. Ko, E. Jeon, S. Jeong, and H.-I. Suk, "Multi-scale Neural Network for EEG Representation Learning in BCI," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 31–45, 2021.

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

[13] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[14] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[15] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pp. 4415–4426, 2019.

[16] C. O'reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research," *J. Sleep Res.*, vol. 23, no. 6, pp. 628–635, 2014.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, "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.

[19] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[20] M. Corsi-Cabrera, Z. Munoz-Torres, Y. del Río-Portilla, and M. Guevara, "Power and coherent oscillations distinguish REM sleep, stage 1 and wakefulness," *Int. J. Psychophysiol.*, vol. 60, no. 1, pp. 59–66, 2006.

[21] P. An, Z. Yuan, J. Zhao, X. Jiang, and B. Du, "An effective multi-model fusion method for EEG-based sleep stage classification," *Knowl.-Based Syst,*, vol. 219, p. 106890, 2021.