IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2023, accepted 6 July 2023, date of publication 11 July 2023, date of current version 17 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3294410

RESEARCH ARTICLE

4s-SleepGCN: Four-Stream Graph Convolutional Networks for Sleep Stage Classification

MENGLEI LI[1], (Graduate Student Member, IEEE), HONGBO CHEN[1], (Graduate Student Member, IEEE), YONG LIU[2], AND QIANGFU ZHAO[2], (Senior Member, IEEE)

[1]Graduate School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan
[2]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan

Corresponding author: Menglei Li (d8231104@u-aizu.ac.jp)

ABSTRACT Sleep staging serves as a critical basis for assessing sleep quality and diagnosing sleep disorders in clinical practice. Most existing methods rely solely on a single channel for sleep staging, thereby neglecting the complementary nature of multimodal electrophysiological signal characteristics. In contrast, the current multi-stream sleep staging network primarily utilizes electrooculogram (EOG) and electroencephalogram (EEG) signals as inputs and efficiently fuses the extracted multimodal features. However, the importance of motion information in electrophysiological signals is rarely investigated, which could improve the classification performance. Moreover, recent sleep staging models have been plagued by issues of overparameterization and suboptimal classification accuracy. Moreover, EOG and EEG are non-Euclidean graph-structured data that can be effectively handled by graph convolutional networks. To address the aforementioned issues, we propose an efficient graph-based multi-stream model named 4s-SleepGCN, which combines biological signal features to classify sleep stages. In each single-stream model, the positional relationship of the modal sequences is incorporated into the proposed model to enhance the feature representation for sleep staging. On this basis, graph convolutions are utilized to capture spatial features, while multi-scale temporal convolutions are employed to model temporal dynamics and extract more discriminative contextual temporal features. The EEG signal, EOG signal, and corresponding motion information are separately fed into the single-stream model comprising our 4s-SleepGCN. Experimental results show that the proposed 4s-SleepGCN achieves the highest accuracy compared to state-of-the-art methods in both the Sleep-EDF-39 dataset (92.3%) and Sleep-EDF-153 dataset (85.5%). Additionally, we conduct numerous experiments on two representative datasets that demonstrate the validity of the motion modalities in sleep stage classification. Also, the proposed single-stream network shows higher accuracy (89.2% and 89.8%) in classification while requiring 33% fewer parameters. Our proposed 4s-SleepGCN model serves as a powerful tool to assist sleep experts in assessing sleep quality and diagnosing sleep-related diseases.

INDEX TERMS Sleep staging, graph convolutional networks, multi-stream networks, multimodal electrophysiological signal, motion information.

I. INTRODUCTION

Human beings generally sleep for approximately one-third of their lives. It is undeniable that high-quality sleep can help protect the mental and physical health of an individual [1]. Over the past few decades, sleep disorders [2] such as insomnia, sleep apnea, sleep-disordered breathing, and circadian rhythm sleep disorders have affected millions of people worldwide and can be considered a growing epidemic [3]. Sleep disorders exhibit different incidences and characteristics across various sleep stages. As a result, a large portion of researchers [4], [5], [6], [7] employ sleep stage scoring to objectively evaluate sleep quality and effectively assist in the prevention and diagnosis of sleep disorders. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey.

monitoring and analyzing sleep based on the observed different stages of sleep throughout the night is highly desirable.

As of date, the analysis of polysomnography (PSG) is considered a representative criterion for sleep stage scoring [8]. PSG is a collection of various biological signals, including the electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG), which are recorded using attached electrodes and various sensors placed on different parts of the body [9]. In particular, the consecutive 30-second sleep epochs from the PSG recordings are utilized for performing sleep stage classification epoch-by-epoch. EEG can be regarded as a cost-effective, typical, and scientifically-proven solution for monitoring and recording electrical activity during sleep. Moreover, EOGs and EMGs are also considered essential signals for sleep analysis [10]. Therefore, sleep staging is mainly scored manually by human experts according to the biological signals of overnight PSG recordings [11]. Rechtschaffen and Kales (R&K) [12] have proposed the only widely accepted standard for delineating six sleep stages, which include wakefulness (*W*), rapid eye movement (*REM*), and non-REM (*NREM*). Within *NREM*, there are further categorizations into four sleep stages (S_1, S_2, S_3, and S_4). Nevertheless, this variability in sleep stage classification evaluation has led to a need for standardization. To address this, the American Academy of Sleep Medicine (AASM) [13] has revised and updated guidelines, based on the R&K sleep stage criteria, establishing them as a definitive reference for PSG assessment. Specifically, a major change in the AASM manual is the integration of NREM stages S_3 and S_4 into a single deep sleep stage called N_3. The AASM manual is widely utilized for scoring sleep stages, encompassing both manual and automatic classification. Manual sleep stage classification is a labor-intensive, time-consuming, tedious, and error-prone task [14]. In comparison, the reliable and high-accuracy approach of automatically classifying sleep stages deserves significant attention in sleep research.

Over the past decade, there has been an influx of automatic sleep stage classification methods in relevant research. These methods can not only effectively identify sleep stages but also provide a basis for the diagnosis and prevention of sleep-related disorders. In the previous conventional methods, hand-engineered features in the time domain, frequency domain, and time-frequency domain are usually used for automatic sleep stage scoring. For example, consideration of feature extraction in the time-frequency domain is raised by Tsinalis et al. [15] who reach the accuracy of 78.9% in sleep stage classification. In addition, great progress in classifying sleep stages has been made in machine learning-based methods, e.g., Support Vector Machine (SVM) [16] and Random Forest (RF) [17]. However, machine learning-based methods often exhibit unsatisfactory performance, and more significantly, these methods typically rely on prior knowledge of sleep analysis. As the elapse of time, deep learning techniques have become extremely mainstream in sleep stage classification, which can be categorized into Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs). From our own perspective, the latest deep learning-based methods for sleep stage classification are based on two concepts: the single-channel EEG-based method and the multi-modal physiological signals-based method.

A. SINGLE-CHANNEL EEG-BASED METHODS

Given the growing trend in the application of deep learning, recent studies have been focusing on the task of sleep stage classification on EGG signals to achieve outstanding performance, which can be roughly divided into three main approaches, namely recurrent neural networks (RNNs), convolutional neural networks (CNNs), and graph convolutional networks (GCNs). RNNs [18] are considered to be able to model the long-term contextual dependencies of temporal sequences in EEG signals. More recently, specific RNN-based methods [19], [20] that learn sequential features from EEG signals have achieved success in automatic sleep staging. In addition, the Long Short-Term Memory (LSTM), a representative structure of RNN, has demonstrated great effectiveness and is utilized in IITnet [21] to learn the transition rules among sleep stages. However, due to the long-term dependence of the data on RNNs, the problem of gradient disappearance or explosion is extremely prone to occur, leading to instability in training the model. In contrast, CNNs have better parallelizability and have the ability to directly extract sleep stage transition features from texture images encoded from sleep stage sequences. The CNN-based method proposed by Tsinalis et al. [22] demonstrates the ability to reliably score sleep stages using a single-channel EEG signal. Sors et al. [23] employ CNNs to extract appropriate features directly from raw EEG. Fang et al. [24] design a novel adaptive-boosting-based dual-stream network framework to extract different modalities features of single-channel EEG signals for sleep staging. In addition, a novel CNN framework based on single-channel EEG signals, called SleepEEGNet [25], has been proposed for sleep stage evaluation using extracted time-invariant features. However, most CNN-based methods struggle to capture temporal dependencies from EEG signals. To address this issue, several integrated systems (i.e., DeepSleepNet [26] and TinySleepNet [27]) have been proposed, which combine CNN and RNN to simultaneously extract features in the spatial and temporal domains, resulting in accurate models for sleep stage discrimination. Considering that EEG electrodes are distributed in a non-Euclidean space, CNNs and RNNs are limited in that the grid data are used as model input and the connection between spatial correlations between electrodes is ignored. GCNs [28] have been shown to be powerful in modeling the topological relationship of EEG electrodes. In this regard, the spatiotemporal graph convolutional network (STGCN) [29], as one of the most advanced extensions of GCN-based models, has exhibited outstanding performance in sleep staging. A quintessential example should be cited that the

TABLE 1. Representative EEG and EOG characteristics during different sleep stages.

Sleep stages	EEG-characteristics	EOG-characteristics
REM	Low-amplitude, mixed-frequency EEG activity without K complexes or sleep spindles. (Resembles eyes open wake epoch)	Rapid eye movements.
N_1	Low-amplitude, predominantly 4-7 Hz, mixed EGG activity.	Slow, rolling eye movements.
N_2	Sleep spindles: a train of distinct 11-16 Hz waves (most frequently 12-14 Hz) with a duration between 0.5 and 2 seconds. K complex: negative, well-delineated, sharp waveforms immediately followed by a high-voltage slow wave, with a total duration of at least 0.5 seconds.	Either slow eye movements or absence of slow eye movements.
N_3	Delta waves of high amplitude (greater than 75μV) and low frequency (0.5-2 Hz).	None
Wake	Eye-close wakefulness: sinusoidal alpha rhythm (8-13 Hz activity). Eye-open wakefulness: Beta wave(highest frequency and lowest amplitude).	Eye-close wakefulness: slow-rolling eye movements. Eye-open wakefulness: rapid eye movements.

GraphSleepNet [30] utilizes spatial graph convolutions in conjunction with interleaving temporal convolutions to effectively capture the transition rules among different sleep stages. Furthermore, Jia et al. [31] have developed a novel deep graph neural network named MSTGCN to extract time-varying spatial and temporal features from multi-channel brain signals, using the spatial topological information between brain regions to distinguish different sleep stages. However, these methods overlook the significance of spatiotemporal relations in sleep staging. To address this limitation, Li et al. [32] propose a combination of dynamic and static STGCN, incorporating inter-temporal attention blocks. This approach effectively captures long-range dependencies among different EEG signals and achieves superior performance in sleep stage classification. Despite achieving better performance, single-channel EEG-based methods are frequently limited by the fact that a single fixed physiological signal is suboptimal for distinguishing specific sleep stages.

B. MULTI-MODAL PHYSIOLOGICAL SIGNALS-BASED METHODS

The multi-modal fusion strategy aims to integrate diverse media types, capturing complementary information and thereby enhancing the performance and robustness of learning [33], [34]. Sleep staging is a complex dynamic process, where different sleep stages are classified based on physiological signals that exhibit varying frequencies and amplitudes at different time periods. Table 1 shows representative EEG and EOG characteristics during different sleep stages, based on information from existing studies [35], [36]. In the N_2 and N_3 stages, the EOG waves exhibit a similar pattern, whereas EEG, as a unimodal physiological signal, provides valuable and specific characteristic information, enabling better classification. In contrast, when classifying the *REM* and N_1 stages, the EEG signal, which lacks some key features, may lead to misclassification. Therefore, the effective identification of different sleep stages requires the integration of different physiological signals. In order to harness the complementary potential of PSG signals, researchers have turned to utilizing multi-modal signals to enhance sleep staging models. For instance, a variation of CNN [37] demonstrates that using multi-channel data achieves better performance compared to single-channel data. Dong et al. [38] apply a combination of DNN and RNN to extract salient features from EEG and EOG signals. Additionally, Andreotti et al. [39] highlight the advantages of incorporating multi-modal PSG signals for sleep staging. And the SeqSleepNet [40] achieves an overall classification accuracy of 87.1% based on multi-channel signals by relying solely on a hierarchical RNN. In a similar vein, Chambon et al. [41] use a spatiotemporal CNN model to capture modality-specific information from all multivariate and multi-modal PSG signals. Phan et al. [42] employ a multi-task CNN combining joint classification and prediction framework to identify sleep stages. These methods primarily focus on extracting the features from different PSG signals individually and combining them by concatenation. However, this is not sufficient to model complex relationships between multimodal signals. As a result, recent works have emerged that fully fuse multimodal feature information to showcase the distinct contributions of each modality in identifying specific sleep stages, such as SalientSleepNet [43] and SleepPrintNet [44]. Moreover, Jia et al. [45] design a squeeze-and-excitation network to model the heterogeneity between different modalities. In the latest research, MMASleepNet [46] introduces an effective feature fusion module to capture the relationships among different modalities. MaskSleepNet [47] effectively combines CNN with an attention mechanism to capture feature information from different PSG signals, leading to a classification accuracy of up to 85.0% on the Sleep-EDF-153 dataset. However, these methods fail to consider coherent features of the PSG signals, such as the speed at which different PSG signals change from frame to frame. Essentially, comprehensive spatial-temporal dependencies may be ineffectively captured.

After a thorough review of previous studies, we have identified three main limitations that need to be addressed. Firstly, the majority of existing multichannel-based methods only consider the captured features from the EEG and EOG signals and ignore the signal motion stream, which is not able to obtain more comprehensive features. Secondly, current multistream models for sleep staging are typically overparameterized to extract discriminative features from signal

sequences, resulting in high model complexity and limiting the development of multichannel-based sleep staging. Thirdly, in current GCN-based approaches to sleep staging, there is a lack of adequate exploration of the semantic information of signal sequences and long-range spatiotemporal dependencies are not well captured. To address the aforementioned limitations, we propose a novel graph-based multi-stream fusion model called 4s-SleepGCN for automatic sleep staging. Our proposed model simultaneously fuses the features of EEG signals, EOG signals, EEG motion, and EOG motion within a unified GCN framework. Our proposed model provides a better balance between performance and parameter scale than some state-of-the-art models, achieving the highest overall performance on two standard datasets.

Overall, the main contributions to this work can be summarized as follows:

- To the best of our knowledge, we are the first to utilize a multi-stream fusion strategy to facilitate the fusion of EEG signals, EOG signals, and the corresponding motion stream, which significantly outperforms the state-of-the-art methods on two benchmark datasets for sleep staging. Furthermore, the motion modality is shown to be a beneficial addition to sleep staging.
- In each single-stream model, we utilize the position embedding method along with spatial-temporal convolutions to model spatial-temporal relationships effectively and classify sleep stages.
- We propose a lightweight, single-stream solid baseline that is more potent than most previous methods. We hope that the solid baseline will be helpful for the study of automatic sleep staging.
- On the Sleep-EDF-39 and Sleep-EDF-153 datasets, our proposed 4s-SleepGCN outperforms both single-stream and two-stream models. The experimental results underscore the importance of multiple information. Our proposed model addresses the current deficiencies of multi-modal learning in sleep staging, paving the way for multi-modal learning in sleep stage classification.

The remainder of this paper is organized as follows: Section II elaborates on the proposed 4s-SleepGCN and explains its components in detail. Next, the dataset used and the experimental settings are described in Section III. Meanwhile, Section III verifies the effectiveness and advantages of the proposed model using two publicly available datasets. In Section IV, we discuss our proposed approach formally. Finally, Section V concludes this work and offers insights into future research directions.

II. METHODOLOGY

In this article, we propose a multi-stream framework to fuse the spatial information of two different PSG signals (i.e., EEGs and EOGs) and the motion information of their sequences to obtain a powerful sleep staging model. Accordingly, in this section, we present the architecture and components of our proposed network in detail. The proposed network consists of four functional modules: encoder, position embedding, graph convolutional network module, and temporal modeling module. Finally, a multi-stream feature extraction strategy is introduced to promote the sleep stage classification task.

A. NETWORK ARCHITECTURE

Inspired by the success of the two-stream framework and graph convolution [48], we design a graph-based multi-stream network to classify sleep stages from different perspectives. In Fig. 1(a), the PSG data is preprocessed to obtain EEG sequence, EOG sequence, EEG motion, and EOG motion information. Subsequently, the four data are respectively fed into the SleepGCN network to obtain the softmax scores. As described in [49], the weighted average method has been successfully applied in the field of fusing classification results and can further improve the classification results. Therefore, The prediction of sleep stage classification is calculated by the weighted summation method of the four softmax scores. Fig. 1(b) illustrates the architecture of the SleepGCN. Among them, the input signal sequence is composed of T frames, and the sleep information contained in each frame is composed of the number of electrodes (N) and the number of channels (C_0) with dimensions $C_0 \times N$, which can be represented as an input tensor with the shape of $C_0 \times T \times N$, where C_0 is equal to 3. Then we use two fully connected layers to encode the position to a dimension of 64 (C_1) and then merge it with the position of the same dimension to obtain the new input for 128. The GCN module is adopted to capture long-range spatial dependencies. In order to mitigate the prevalent issue of over-smoothing encountered in most GCN-based models, which has been documented in previous studies [50], we employ ReLU activation functions for each GCN block of our proposed model. By applying activation functions after each GCN block, our classification network can effectively capture complex patterns in the PSG graph data and preserve the expressive power of the node representations, enhancing the model's ability to perform accurate sleep graph classification. The temporal modeling module uses different dilation convolutions to effectively aggregate contextual information. The Global Average Pooling (GAP) layer is introduced to aggregate spatio-temporal features and pool feature maps of distinct samples to a similar size of $1 \times 1 \times 512$. Finally, the softmax layer is used to obtain probabilities for the sleep stage. Each module is presented separately in the following subsections.

B. ENCODER

Since sleep staging based on PSG data can be formulated as a graph modeling problem, the raw PSG sequence of sleep staging can be represented as an undirected graph $G = (V, E)$ with N electrodes and T frames, including a node set $V = \{V_1, V_2, \cdots, V_N\}$ of electrodes N and E is the edge set representing the connection between the electrodes captured by an adjacency matrix $A \in \{0, 1\}^{N \times N}$. A denotes the relationship

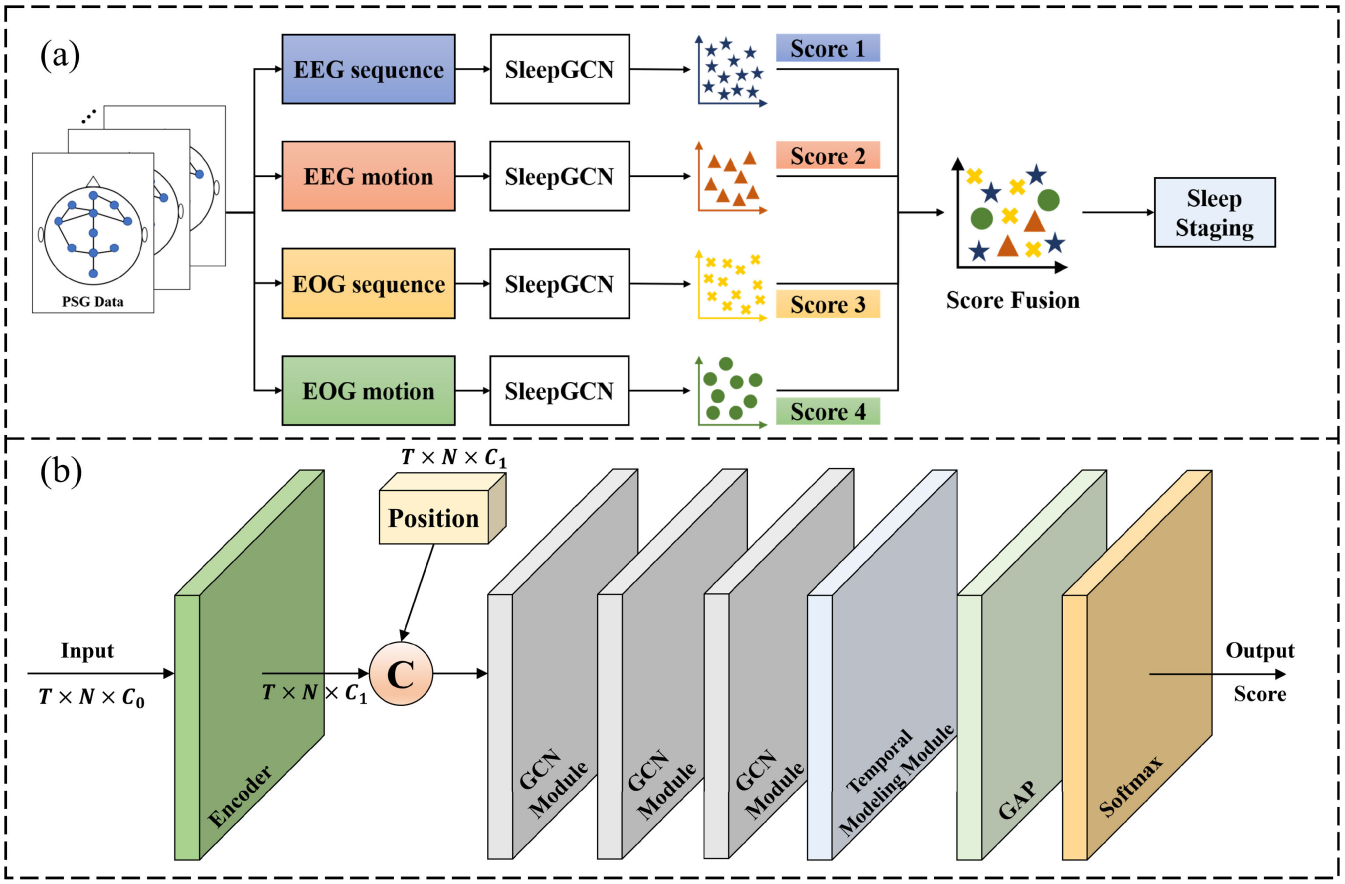


FIGURE 1. The proposed network architecture for sleep staging. (a) Illustration of the overall architecture of the multi-stream fusion sleep staging network (4s-SleepGCN). The scores of four streams are fused to predict the final sleep staging; (b) Overview of the SleepGCN. The network consists of an encoder, a position embedding, three graph convolutional network (GCN) modules, a temporal modeling module, and a global average pooling (GAP) layer. C_0, T, and N denote the number of channels, sequences, and electrodes of input data, respectively. Using the encoder gives a dimension of 64 (C_1), which is the same dimension as the position. For the GCN module, the output channel numbers are 128, 256, and 384. The temporal modeling module is used to extract the temporal feature from PSG sequences. The final output channel becomes 512 through the GAP layer. We also use the batch normalization and activation function for each block of the model.

between the electrodes, where initially $A_{i,j} = 1$ if there is a functional connection between electrodes i and j, and 0 otherwise. The PSG signal sequence can provide the coordinates of each electrode in graph convolutional networks, which can be described as $X \in \mathbb{R}^{T\times N\times C}$. Therein, N denotes the total number of electrodes in a frame, T is the number of frames in the raw signal sequence, and C represents the coordinates of all electrodes in the entire frame sequence. We denote all electrode features as a feature set X, which can be represented as a matrix:

$$X = \left\{X_{n,t} \in \mathbb{R}^{C_0} | n \in N; t \in T\right\} \tag{1}$$

where the electrode of type $n = \{1, 2, \cdots N\}$ at time $t = \{1, 2, \cdots T\}$ generates the dimensional feature vector $X_{n,t}$. Our goal is to employ the encoder including two fully connected (FC) layers to encode the original position information into a high-dimensional space, which can be described as follows:

$$X' = ReLU\left(g\left(ReLU\left(\tilde{g}\cdot X + k_1\right) + k_2\right)\right) \in \mathbb{R}^{C_1} \tag{2}$$

where $g \in \mathbb{R}^{C\times C_1}$ and $\tilde{g} \in \mathbb{R}^{C_1\times C_1}$ denotes weight matrices. k_1 and k_2 are the bias vectors. We use the ReLU function as the activation function. In this work, the higher order information by encoding instead of the original position is used as input to improve the ability of personalized expression.

C. POSITION EMBEDDING

Position embedding is a widely employed technique for capturing location information within sequences. It has shown successful applications across various domains, with particular effectiveness in natural language processing. Since EEGs and EOGs are time-series data, the sequential relationship between frames affects the meaning of the entire signal. Considering only the coordinate information of the electrodes and the graph structure of the biosignals, it is difficult for the model to capture the sequential relationships between different time steps in the signal, which may result in suboptimal classification performance. Therefore, the absence of the position relationships of sequences could weaken the classification ability of sleep stage models. To address this

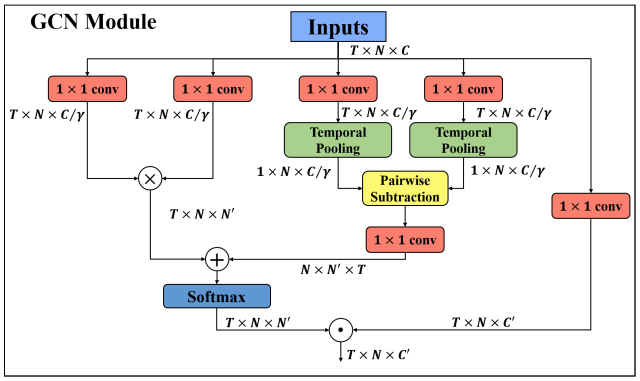


FIGURE 2. The architecture of the GCN module. The input feature map is used as the input signal with dimension $T \times N \times C$, where T, N, and C are the number of frames, electrodes, and channels, respectively. We set the reduction rate γ to 8 in our work to extract compact representations. $\otimes$ denotes matrix multiplication operation, $\oplus$ denotes the elementwise summation, and $\odot$ denotes element-wise multiplication.

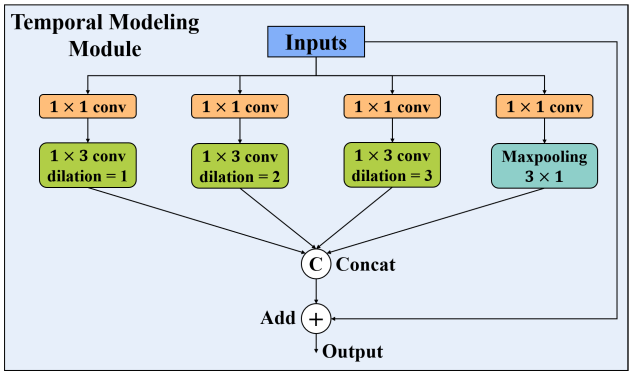


FIGURE 3. The architecture of temporal modeling module. In order to lower the computational costs due to the extra branches, we fix kernel sizes at 1 × 3 and use different dilation rates for larger receptive fields. Meanwhile, the 3 × 1 max-pooling layer is used to capture the most salient feature.

issue, position embedding is applied in our model to incorporate positional information in the model input, which can better capture the sequential relationships between different time steps in the sleep signals, leading to improved sleep stage classification performance. Inspired by the previous works [51], [52], two one-hot vectors are applied to characterize the position relations of electrodes and frames. In frame sequences $T = \{T_1, T_2, \cdots, T_w\}$, the w^{th} frame T_w is denoted by a one-point vector, where the w^{th} dimension is set to one and the others are zero. As for the same operation of the frame sequences, we proceed to obtain a one-hot vector as T_w for the electrode sequences. Similar to the encoding of the inputs according to Eq. 2, the embedding representation in the electrode and frame sequences can be expressed as $N'_w \in \mathbb{R}^{C_1}$ and $T'_w \in \mathbb{R}^{C_1}$, respectively. Subsequently, the embedding vectors in the frame- and electrode- dimensions are fused and concatenated with the original features X'. Finally, the output feature maps $X'' \in \mathbb{R}^{2C_1 \times N \times T}$ can be obtained by the concatenation operation $\frown$, as given in Eq. 3. Notably, we use the original position as the residual embedding to make the position encoding information explicitly.

$$X'' = \left(N'_w + T'_w\right) \frown X' \tag{3}$$

D. GRAPH CONVOLUTIONAL NETWORK MODULE

Indeed, capturing long-range dependencies of PSG sequence data is crucial for sleep staging. Inspired by the idea of semantics-guided neural network [51] and non-local block [53], we adopt the GCN module (see Fig. 2) to extract correlations between electrodes, thereby capturing rich features of sleep stages from PSG data to achieve sleep staging. More specifically, the similarity between the electrodes in the feature space is used to construct the sleep graph. The long-range weight can be modeled by the pairwise similarity between every two electrodes a^{th} and b^{th} in the same frame T, which is defined as follows:

$$f\left(a, b\right) = \varphi\left(X''_a\right)^T \lambda\left(X''_b\right) \tag{4}$$

where φ and λ represent two transformations of the original features. Since the long-range transformed feature $f\left(a, b\right)$ characterizes only the long-range spatiotemporal relationship of the electrode pair, we use the following form to define the relationship between shared bias on the channel dimension:

$$\mathbb{B}\left(a, b\right) = \delta\left(\alpha\left(\mathbb{TP}\left(X''_a\right)\right) - \beta\left(\mathbb{TP}\left(X''_b\right)\right)\right) \tag{5}$$

therein, the function of temporal pooling $\mathbb{TP}$ is to aggregate temporal features, whereas in our work we use mean pooling. The $\delta \in \mathbb{R}^{T \times C/8}$, $\alpha \in \mathbb{R}^{C \times C/8}$ and $\beta \in \mathbb{R}^{C \times C/8}$ are three linear embedding functions implemented by the 1×1 convolutional layer. The distances along the channel dimension $\mathbb{B}\left(\cdots\right) \in \mathbb{R}^{N \times N \times T}$ uses the nonlinear transformations to model the topological relationship on the channels. Furthermore, we use the bias for attention score calculation to update the weighting information. We update the weights using an overall attention score that is the sum of the two component weights, thus the updated weights can be formulated as follows:

$$Output_G = X'' \odot \left(\sigma\left(f\left(a, b\right) + \mathbb{B}\left(a, b\right)\right)\right) \tag{6}$$

where $\odot$ is the element-wise multiplication. σ is the softmax activation function. X'' and $Output_G$ denote input and output feature maps.

E. TEMPORAL MODELING MODULE

The duration of the different sleep stages varies. Therefore, temporal modeling is also essential for sleep staging. Current methods [54], [55] still use temporal convolutions with a single fixed scale to perform temporal modeling. The feature information obtained from distant frames is very limited, and the long-range temporal dependence is not well captured, which affects the accuracy of sleep stages. It is not optimal to use temporal convolutions with a fixed kernel size

to deal with the problem of sleep staging. Consequently, the multi-scale temporal features extracted by convolution kernels with different scales are fused to better model the temporal topological features. The difference from the previous method is that we use four parallel temporal convolution branches to achieve temporal modeling, as shown in Fig. 3. In each branch, we introduce a bottleneck architecture [56] that uses 1×1 convolution to reduce the computational cost and thus speed up the training and model inference. In addition, the first three branches of the model utilize temporal convolutions with a kernel size of 1×3, employing different dilations [57] to analyze short-term and long-term temporal dependencies, thus obtaining multi-scale temporal receptive fields. In the final branch, a 3×1 max-pooling layer is utilized to extract the most important features. Finally, we use a concatenation strategy to fuse the features. In conclusion, the temporal modeling module is proposed to extract richer temporal features from the physiological signal sequences, which can be used to capture the temporal dependencies between sleep stages and distinguish the different duration dynamics.

F. MULTI-STREAM FUSION

In this work, we utilize multi-stream fusion strategies to model the first-order information (EEG and EOG signals) and the corresponding motion information for sleep staging. In the Sleep-EDF dataset, the sequence of electrode motion information can be obtained by calculating the difference of the same electrode between two consecutive frames, typically in terms of the differences in the coordinates of EEG and EOG electrodes. The position of the electrode of the human brain can be defined as $V_{g,t}\{g \in N, t \in T\}$, where N and T denote the number of electrodes and the number of frames of the signal sequences, respectively. The g represents the electrode in the frame t. As for the motion information, the position difference of the same electrode in two consecutive frames can be calculated to obtain a sequence of electrode motion information, namely the displacement information. This displacement information can then be used as an additional input feature to help the model learn dynamic features. The sequence of electrode motion information $\mathbb{M}$ for electrode g in frame t is obtained by subtracting the position of the electrode in the next frame $t+1$ from its position in the current frame t, which can be expressed as follows:

$$\mathbb{M} = V_{g,t+1} - V_{g,t} \tag{7}$$

Therein, $V_{g,t+1}$ is the position of the electrode g in frame $t+1$. $\mathbb{M}$ is a vector representing the motion of the electrode between the two frames. Finally, the EEG, EOG, and corresponding motion information are fed into four streams and fused to classify different sleep stages.

III. EXPERIMENTAL RESULTS

In this section, the effectiveness of the proposed approach is evaluated using two publicly available datasets. The first subsection provides a comprehensive description of the Sleep-EDF-39 and Sleep-EDF-153 datasets, along with the experimental setups employed in this study. Subsequently, the metrics utilized to evaluate the performance of the sleep stage model are explained. Finally, we present the performance results of our proposed model and discuss its effectiveness in comparison to other state-of-the-art models.

TABLE 2. Details of the number of sleep stages in the sleep-EDF-39 and sleep-EDF-153 datasets.

Dataset	W	N_1	N_2	N_3	R	Total
Sleep-EDF-39	7927	2804	17799	5703	7717	41950
Sleep-EDF-153	65951	21522	69132	13039	25835	195479

A. DATASET AND EXPERIMENTAL SETTINGS

1) SLEEP-EDF-39 AND SLEEP-EDF-153 DATASETS

The Sleep-EDF-39 and Sleep-EDF-153 datasets are two versions of the Sleep-EDF dataset [58]. The Sleep-EDF-153 dataset is an expanded version of the Sleep-EDF-39 dataset. The two publicly available datasets are commonly utilized in sleep staging research and are sourced from the PhysioBank. The participants are enrolled in the sleep cassette (SC) and sleep telemetry (ST) studies. In our experiment, we adopt the PSG sleep recordings from SC. They record the PSGs of healthy Caucasians without any sleep-related medications. Each subject records PSG recordings during two subsequent day-night periods, which include two scalp-EEG, horizontal EOG, chin EMG, and event markers. Therein, EEG is sampled from F_{pz}-C_z and P_z-O_z electrde locations. In our experiments, the F_{pz}-C_z EEG is used as the input EEG signal. All EEG and EOG are acquired at a sampling rate of 100 Hz. The sleep-EDF-39 dataset contains data files for 20 male and female subjects (age 28.7 ± 2.9). The number of participants in the Sleep-EDF-153 data set is 78, ranging in age from 25 to 101 years. Consistent with some baseline approaches [26], [43], the Sleep-EDF-39 and Sleep-EDF-153 datasets in our experiment contain 41950 and 195479 sleep epochs, respectively, as shown in Table 2. Moreover, in two datasets, each 30-s recording is manually classified into eight stages (*wake*, S_1, S_2, S_3, S_4, *REM*, movement, and unknown) according to the R&K standard [12]. In the latest AASM manual [13], movement and unknown stages are excluded and the S_3 and S_4 stages are combined into one signal stage N_3. Therefore, sleep stages in the datasets consist of W (Wake), N_1 (S_1), N_2 (S_2), N_3 ($S_3 + S_4$) and R (*REM*).

2) EXPERIMENTAL SETTING

In our experiment, the proposed model is implemented on the Pytorch platform with an RTX3060 GPU card. The network is trained with a batch size of 64. The Adam optimizer as an optimization strategy is used to train the model for 120 epochs. The learning rate is set to 10^{-3} and is decayed by 10 at the 30^{th}, 60^{th}, and 90^{th} epochs, respectively. In our work, we set the weights of the EEG stream, the EOG stream, and the corresponding motion stream to 0.6, 0.6, 0.4,

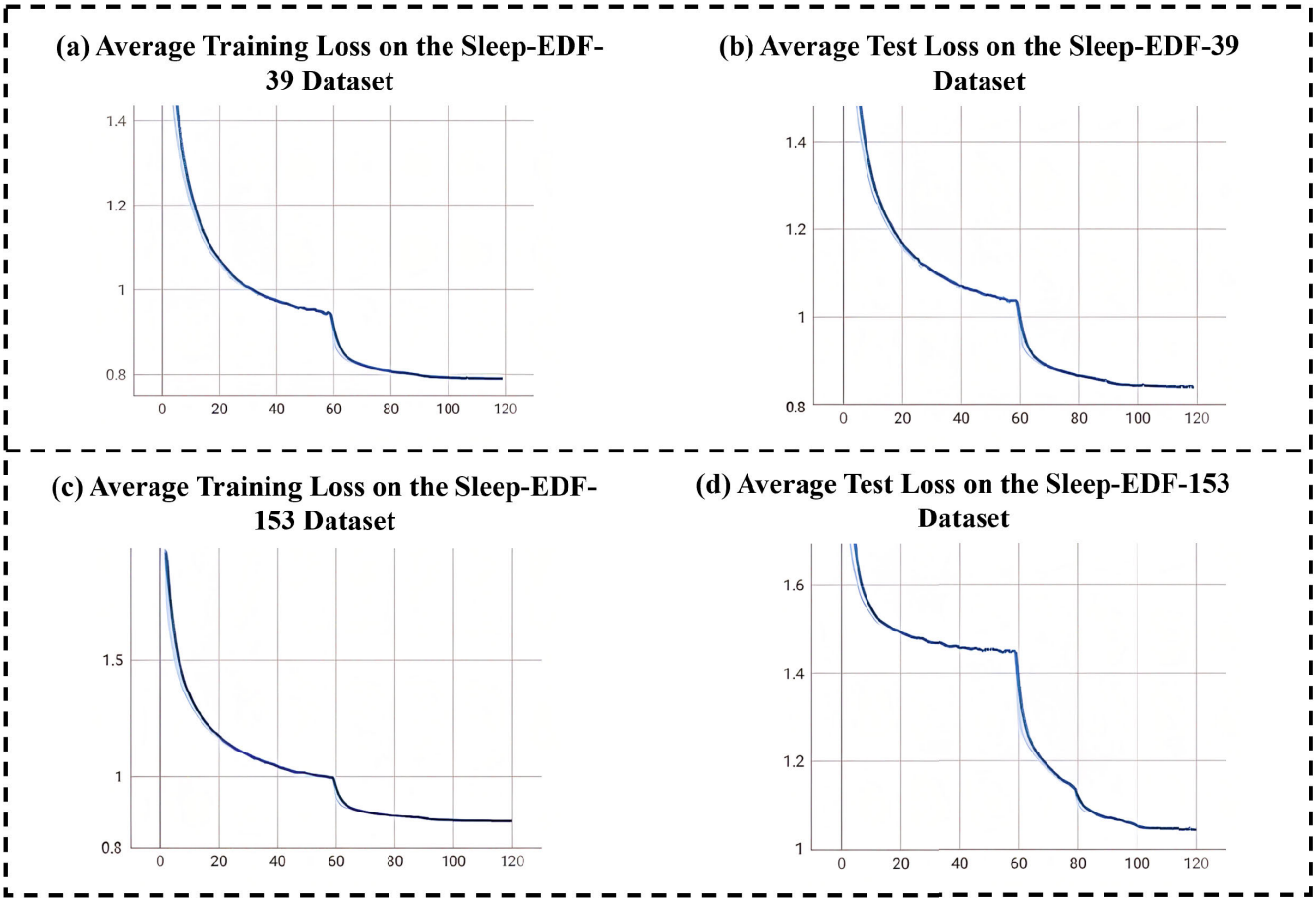


FIGURE 4. Training and test loss vs. a number of epochs of the proposed model. The horizontal axes and the vertical axes represent epochs and the value of the loss function, respectively. The sub-figure(a) and sub-figure(b) show the training loss and test loss on the Sleep-EDF-39 dataset. The sub-figure(c) and sub-figure(d) show the proposed model loss for training and testing on the Sleep-EDF-153 dataset.

and 0.4 for weighted fusion like other multi-stream GCN methods. To improve the generalization performance and reliability of our proposed model and reduce the risk of overfitting, we implement dropout and label smoothing [59] during the training process. Specifically, in our experimental setup, we set the dropout rate to 0.2 and employ label smoothing for better calibrated classification networks with a smoothing factor of 0.1. In addition, we use K-fold cross-validation to evaluate the performance of our sleep staging model. We follow a rigorous evaluation methodology, using a 20-fold cross-validation scheme with K set at 20 to ensure a fair comparison with baseline models. For this purpose, subjects in the sleep-EDF-39 and sleep-EDF-153 datasets are divided into 20 groups. Accordingly, experimental results for 20-fold cross-validation are obtained. Eventually, we calculate the average of the results of all 20 test samples as the final experimental results of our model, which provide reliable performance metrics for assessing the performance of the network. Moreover, we use the TensorBoard to monitor the training progress to evaluate the performance of our proposed model on two public datasets. As shown in Fig. 4, we observe that the training loss gradually decreases and stabilizes over iterations. This trend indicates that our model is effectively learning patterns and features from the training data.

B. EVALUATION METRICS

To provide a comprehensive evaluation of the performance of the sleep staging model, we introduce several metrics including accuracy, macro-precision, macro-recall, macro-averaged F1 score, and Cohen's Kappa coefficient. The overall accuracy (ACC), macro-precision (P_{macro}), macro-recall (R_{macro}), macro-averaged F1 score ($MF1$), and Cohen's Kappa coefficient(κ) are defined as follows:

$$ACC = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP+TN}{TP+FP+FN+TN}\right)_i \tag{8}$$

$$P_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FP}\right)_i \tag{9}$$

$$R_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FN}\right)_i \tag{10}$$

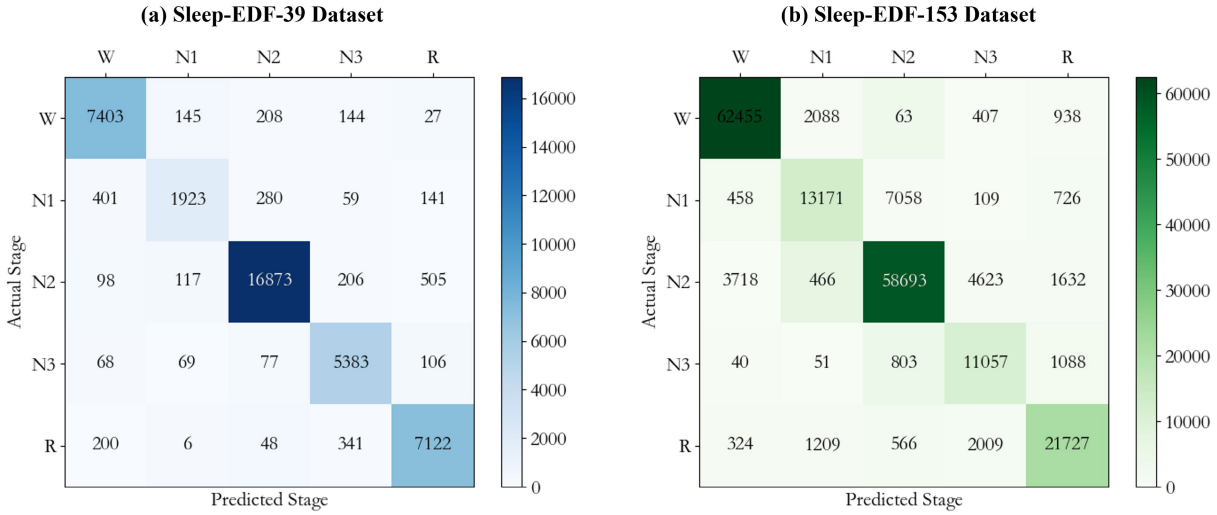


FIGURE 5. Visualization of the experimental confusion matrix obtained from 20-fold validation. We employ the Sleep-EDF-39 and Sleep-EDF-153 datasets to obtain two confusion matrices. The sub-figure(a) and sub-figure(b) show the confusion matrix for the Sleep-EDF-39 dataset and the Sleep-EDF-153 dataset, respectively.

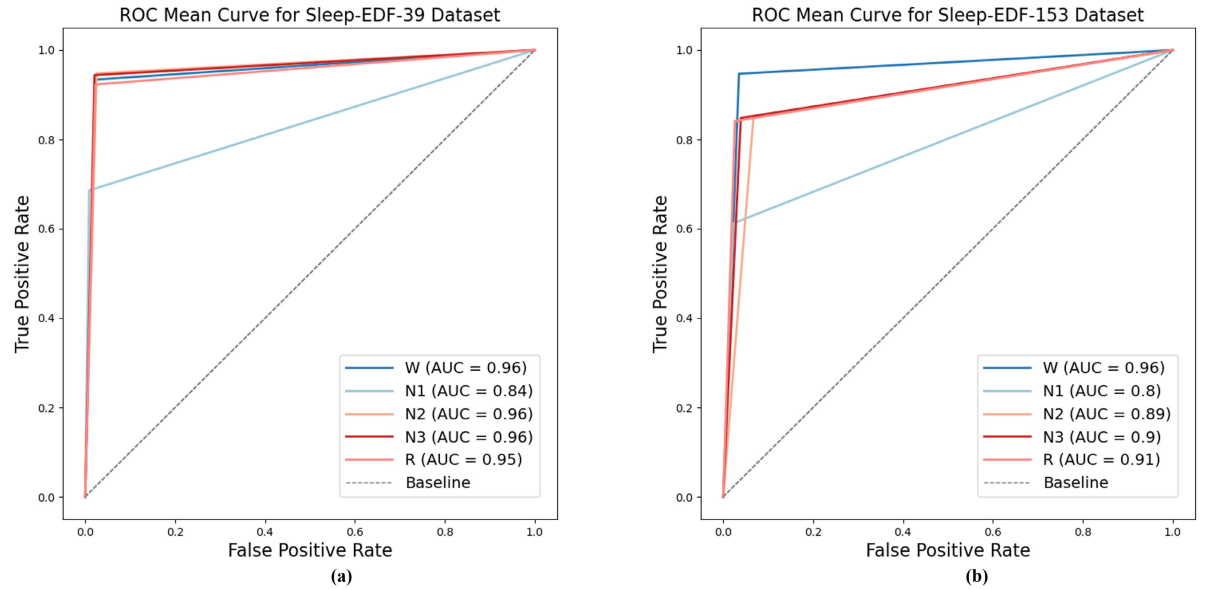


FIGURE 6. The mean ROC curve and AUC values for different sleep stages based on 20-fold cross-validation. The ROC mean curves in sub-figure(a) and sub-figure(b) respectively use the Sleep-EDF-39 and Sleep-EDF-153 datasets as the testing datasets. The AUC values for the five sleep stages are included in the legend.

$$MF1 = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{2\times TP}{2\times TP + FN + FP}\right)_i \tag{11}$$

$$\kappa = \frac{ACC - p_e}{1 - p_e} \tag{12}$$

where TP, FP, FN, and TN stand for the true positives, false positives, true negatives, and false negatives, respectively. K represents the total number of epochs used in the cross-validation, which is defined as 20 in this work. p_e denotes the hypothetical probability of chance agreement.

C. EXPERIMENT RESULTS

In this subsection, the effectiveness of the proposed model is evaluated using the Sleep-EDF-39 and Sleep-EDF-153 datasets. In Fig. 5, the confusion matrices for the predicted sleep stage of each dataset are visualized, showing agreement with the expert results. Based on Eq. 8 and the confusion matrices, the overall accuracy of our model for the two datasets can be determined by calculation and is equal to 92.3% and 85.5%, respectively. For the Sleep-EDF-39 dataset, the macro-precision, macro-recall, and macro-F score are 88.7%, 90.0%, and 89.1%, respectively. Similarly, from the sub-figure(b) of Fig. 5, we obtain the macro-precision, macro-recall, and macro-F score of the Sleep-EDF-153 dataset as 81.9%, 80.4%, and 80.6%, respectively. Furthermore, we use Cohen's kappa coefficients to measure the degree of accuracy and reliability in sleep stage classification. The Cohen's kappa coefficients for

TABLE 3. Comparisons of the validation results with different input modalities on Sleep-EDF-39 and Sleep-EDF-153 datasets.

Methods	Acc. I (%)	Acc. II (%)
1s-SleepGCN (only EEG)	89.2	82.1
1s-SleepGCN (only EOG)	89.8	82.8
2s-SleepGCN	91.5	84.4
4s-SleepGCN	92.3	85.5

2s-SleepGCN represents using the EEG and EOG modalities.
4s-SleepGCN represents using EEG stream, EOG stream, EEG motion stream, and EOG motion stream.
Acc. I and Acc. II shows the overall accuracy for Sleep-EDF-39 and Sleep-EDF-153 datasets, respectively.

Sleep-EDF-39 and Sleep-EDF-153 are 0.89 and 0.80, respectively, indicating that the classification results have high consistency with the actual distribution of sleep stages, being within the standard of $0.8 \sim 1$ [60].

Moreover, to investigate the effects of the classification accuracy of different sleep stages from two publicly available datasets, the receiver operating characteristic (ROC) mean curves of different sleep stages are obtained to show the effect of the proposed sleep staging model on the final classification accuracy, as shown in Fig. 6. As expected, the ROC curves of all sleep stages, except for the N_1 stage, converge towards the upper-left corner of the graph. This convergence signifies that our model exhibits high true positive rates (TPR) and low false positive rates (FPR). This trend further demonstrates the excellent predictive performance of our model in accurately classifying different sleep stages. Nevertheless, the area under the curve (AUC) values for each sleep stage (ranging from 0.8 to 0.96) on both datasets significantly exceed the value of 0.75 in [61]. This substantial improvement in AUC values underscores the superior performance of our model, which holds high clinical value. These results indicate that our proposed model not only outperforms random classification but also demonstrates a noteworthy ability to differentiate between positive and negative instances.

To further verify the advantage of the proposed multi-stream fusion strategy in sleep staging, we test the performance using four data modalities: *single-stream model*, which uses either the EEG or EOG stream independently; *two-stream model*, which fuses the EEG and EOG modalities; *four-stream model*, which incorporates the EEG stream, the EOG stream, the EEG motion stream, and the EOG motion stream. Table 3 shows that the EOG modality performs slightly better than the EEG modality in sleep staging. The superiority of the multi-stream method over the single-stream method is evident. Compared to the two-stream model, we respectively obtain 0.8% and 1.1% improvement on two datasets with the fusion of all four streams. This suggests that the fusion of the EEG stream, the EOG stream, and the corresponding motion stream can yield better classification performance, thus becoming a better choice for sleep staging.

D. COMPARISON WITH STATE-OF-THE-ART MODELS

To evaluate the effectiveness of our proposed method, we conduct a comparison between our proposed 4s-SleepGCN model and several baseline models using the Sleep-EDF-39 and Sleep-EDF-153 datasets. The results of this comparison are presented in Table 4. In comparison to other baseline methods, our method reaches state-of-the-art (SOTA) accuracy of 92.3% and 85.5%, outperforming the baseline models by more than 1.3% and 0.2% on two public datasets.

For some traditional machine learning-based methods, e.g., SVM and RF, the inability to adequately extract various features often leads to poor results in sleep stage classification. Deep learning methods have become a predominant approach for sleep staging to achieve better performance, including those using only CNNs, only GCNs, and a mixture of CNNs and RNNs. Despite the fact that these methods perform reasonably well in the sleep stage classification, resulting in varying degrees of drawbacks. For instance, it is difficult to adjust and optimize some mixed deep-learning models with extensive parameters such as DeepSleepNet, SeqSleepNet, and TinySleepNet. Moreover, there are also methods, e.g., ResnetLSTM and SleepEEGNet, that convert physiological signals into time-frequency images, which often leads to partial information loss. This contrasts with the previous work, where our model uses a multi-information flow fusion method to capture the distinctive complementary features of the original data. Moreover, the motion information from EEG and EOG aids in further enhancing the performance of sleep staging. Therefore, our proposed 4s-SleepGCN achieves the highest accuracy compared with other baseline models.

On the Sleep-EDF-153 dataset, the classification performance of the W and N_2 stages is the best among all sleep stages. Specifically, the F1 score of the W and N_2 stages reaches 94.0% and 86.1%, respectively. Moreover, for this reason, the N_1 stage belongs to the sleep transition period [66], which can be mainly misclassified into N_2 and *REM* stages. The classification effect for the N_1 stage falls short of expectations compared to the other sleep stages, but it still achieves an optimal result compared to the other baseline methods. This is sufficient to illustrate that our model can effectively classify sleep stages in a large sample dataset. Additionally, we can observe that the F1 score of N_3 and *REM* stages is worse than that of most baseline models. The poor results attributed to the fact that $N_3 - N_2$ and *REM* $- N_2$ are also misclassified pairs. In classifying the N_3 stage, an important factor contributing to its lower classification performance is the small proportion of N_3 stage instances within the Sleep-EDF-153 dataset, representing only 6.67% of the total. The limited number of N_3 stage examples in the dataset poses a challenge for the classification model to effectively learn the specific patterns and features associated with the N_3 stage. Due to this scarcity, our proposed model may not be sufficiently familiar with the minority class, resulting in suboptimal generalization and a drop in performance in classifying the N_3 stage. However, the precision of the N_3 and *REM* stages reaches 84.8% and 84.0%, respectively. Therefore, our proposed model can to a large extent reproduce the

TABLE 4. Performance of the Sleep-EDF-39 and Sleep-EDF-153 datasets compared with baseline methods.

Methods	Sleep-EDF-39 dataset							Sleep-EDF-153 dataset						
	Overall results		F1-Score for Sleep Stag(%)					Overall results		F1-Score for Sleep Stag(%)				
	Macro-F score(%)	Accuracy(%)	*Wake*	N_1	N_2	N_3	*REM*	Macro-F score(%)	Accuracy(%)	*Wake*	N_1	N_2	N_3	*REM*
SVM [43]	63.7	76.1	71.6	13.6	85.1	76.5	71.8	57.8	71.2	80.3	13.5	79.5	57.1	58.7
RF [43]	67.6	78.1	74.9	22.5	86.3	80.8	73.3	62.4	72.7	81.6	23.2	80.6	65.8	60.8
SleepEEGNet [25]	79.7	84.3	89.2	52.2	86.8	85.1	85.0	77.0	82.8	90.3	44.6	85.7	81.6	82.9
U-time [62]	78.6	78.2	87.0	52.0	86.0	84.0	84.0	76.4	-	92.0	51.0	84.0	75.0	80.0
MultitaskCNN [42]	75.0	83.1	87.9	33.5	87.5	85.8	80.3	72.8	79.6	90.9	39.7	83.2	76.6	73.5
AttnSleep [63]	78.1	84.4	89.7	42.6	88.8	90.2	79.0	75.1	81.3	92.0	42.0	85.0	82.1	74.2
DeepSleepNet [26]	76.9	82.0	84.7	46.6	85.9	84.8	82.4	75.3	78.5	91.0	47.0	81.0	69.0	79.0
TinySleepNet [27]	80.5	85.4	90.1	51.4	88.5	88.3	84.3	78.1	83.1	92.8	51.0	85.3	81.1	80.3
SeqSleepNet [40]	79.7	86.0	91.9	47.8	87.2	85.7	86.2	78.2	83.8	92.8	48.9	85.4	78.6	85.1
ResnetLSTM [64]	73.7	82.5	86.5	28.4	87.7	89.8	76.2	71.4	78.9	90.7	34.7	83.6	80.9	67.0
MLTCN [65]	77.1	84.2	88.5	39.4	87.7	87.0	82.7	74.9	81.0	92.2	42.8	83.3	**88.3**	77.7
SleepPrintNet [44]	78.0	83.1	88.8	48.0	86.7	86.2	80.3	76.5	81.6	92.7	47.4	83.6	80.0	78.8
SalientSleepNet [43]	83.0	87.5	92.3	56.2	89.9	87.2	89.2	79.5	84.1	93.3	54.2	85.8	78.3	**85.8**
Li et al. [32]	89.0	91.0	**92.1**	**79.7**	93.2	88.2	**91.6**	**81.1**	85.3	92.9	66.6	86.0	75.2	84.6
4s-SleepGCN (ours)	**89.1**	**92.3**	92.0	75.9	**95.6**	**91.0**	91.2	80.6	**85.5**	**94.0**	**68.4**	**86.1**	70.7	83.7

The numbers in bold indicate the highest performance metrics among all approaches, while the result underlined represents the sub-optimal performance.

TABLE 5. Comparison of model parameters on Sleep-EDF-39 dataset.

Methods	Param.(M)	Acc.(%)
SleepEEGNet [25]	2.1	84.3
TinySleepNet [27]	1.3	85.4
U-time [64]	1.1	78.2
SalientSleepNet [43]	0.9	87.5
1s-SleepGCN (only EEG)	0.6	89.2
1s-SleepGCN (only EOG)	0.6	89.8
2s-SleepGCN	1.2	91.5
4s-SleepGCN	2.5	92.3

The Acc. denotes the accuracy for Sleep-EDF-39 dataset.
2s-SleepGCN represents using the EEG and EOG modalities.
4s-SleepGCN represents using EEG stream, EOG stream, EEG motion stream, and EOG motion stream.

sleep scoring of human experts and thus provide assistance in the diagnosis of sleep problems.

Besides, we show the comparative results in terms of accuracy and model complexity (number of parameters) with some SOTA methods to demonstrate the superiority of our model. As can be seen in Table 5, the efficiency of our model has improved compared to previous models for the Sleep-EDF-39 dataset. At first glance, our proposed 4s-SleepGCN has a larger number of parameters than SalientSleepNet. However, our method has adopted the four-stream network architecture, which consists of four backbones. In comparison, the proposed single-stream model based on the EEG or EOG modality achieves relatively great results with an accuracy of 89.2% and 89.8%, respectively. Besides, the proposed single-stream model requires only 0.6 million parameters, which reduces the number of parameters by about 0.3 million. This proves that our proposed single-stream solid baseline can be introduced as a strong and powerful baseline for sleep staging. The proposed 2s-SleepGCN and 4s-SleepGCN require about 0.3M+ and 1.6M+ more parameters compared to the SalientSleepNet, while improving the accuracy by 4% and 4.8%, respectively. We conclude that the lightweight, single-stream solid baseline constructed in this study can significantly reduce the number of model parameters while ensuring classification accuracy. In addition, the two-stream and four-stream proposals show better performance when more parameters are requested.

IV. DISCUSSION

Sleep disorders have indeed risen in striking proportion worldwide over the past 40 years [67], [68], [69]. Sleep stage classification plays a critical role in the diagnosis and treatment of sleep disorders. Automated sleep stage scoring is expected to play a leading role in the diagnosis and treatment of sleep disorders in the future. In this work, a graph-based multi-stream fusion model named 4s-SleepGCN is proposed for sleep staging. EEG, EOG, and the corresponding motion information are fused to enhance the understanding of brain activity and aid in the identification of different sleep stages. This confirms that the motion modality holds significant potential for sleep staging and contributes to improved accuracy and temporal understanding of sleep stages. The proposed EEG or EOG single-stream method with a lightweight network has demonstrated acceptable performance on benchmark datasets, making it a promising candidate for application in residential healthcare settings. In clinical medicine, there is a need to accurately classify different sleep stages and provide reliable results for specialists. The proposed multi-stream model holds the potential to assist doctors in making accurate diagnostic and treatment decisions, thereby improving patients' sleep health outcomes.

The Sleep-EDF-39 dataset and Sleep-EDF-153 dataset utilize in our study comprise practical data obtained from patients. It is important to note that these datasets are non-independent and non-identically distributed, meaning there are significant variations in the sample sizes across different sleep stages. Nevertheless, our proposed method demonstrates robustness by achieving satisfactory classi-

fication results for each sleep stage. This also underscores its effectiveness in handling the complexities inherent in real-world patient data. In addition, our proposed multi-stream model demonstrates remarkable classification performance, particularly in the N_2 stage. Abnormalities observed in N_2 sleep features have been identified as potential indicators for various sleep disorders such as sleep apnea and parasomnias. The accurate classification of the N_2 stage by our model holds significant promise in the identification, diagnosis, and intervention of sleep disorders, ultimately leading to enhanced sleep quality and overall well-being. The exceptional classification performance of our multi-stream model, particularly in the N_2 stage, highlights its potential as a valuable tool in sleep research, clinical assessments, and interventions aimed at optimizing sleep architecture. Its robust capabilities make it an asset in furthering our understanding of sleep-related phenomena and facilitating effective interventions to address sleep disorders. By leveraging the strengths of our proposed model, researchers and clinicians can make significant strides in the field of sleep medicine, ultimately improving the lives of individuals affected by sleep-related issues. Furthermore, for the Sleep-EDF-153 and Sleep-EDF-39 datasets, the ratio of the average training time per fold (approximately 4.17 and 1.36 hours, respectively) is smaller than the ratio of the respective data sizes (195k and 42k). In other words, the training time of our proposed model does not increase proportionally to the increase in data size. Therefore, our model can effectively manage the processing of larger datasets without significantly increasing the training time. This indicates that the proposed model demonstrates a certain degree of scalability. Such scalability is particularly valuable in real-world scenarios where the volume of data is substantial.

However, there is still room for improvement. First, classifying sleep stages, especially the N_1 stage, can remain challenging due to its transitional nature between wakefulness and sleep, making correct recognition a tricky task. Second, sleep staging typically relies on the subjective interpretation and classification of physiological signals by experts. Nonetheless, different experts may interpret the same data and arrive at varying conclusions. To this end, there may be variations in sleep patterns and characteristics among individuals, necessitating an individualized approach to classification. Furthermore, existing sleep staging models are typically processed offline, analyzing and capturing post-sleep data. However, for the timely detection and intervention of potential sleep issues, real-time monitoring is crucial. This is particularly significant for patients with sleep apnea, as real-time detection enables the adjustment of ventilation pressure and treatment parameters, leading to optimized treatment outcomes.

V. CONCLUSION

In this work, we propose a novel multi-stream fusion graph convolutional network called 4s-SleepGCN to efficiently classify different sleep stages by combining multi-stream biological signal features. The positional relationship of modal sequences is embedded into the sleep staging network to improve the feature characterization capability, which can better leverage the task of sleep stage classification. Besides, the proposed 4s-SleepGCN model uses graph convolution and temporal convolution to directly model spatial-temporal dependencies from the PSG graph sequences. Graph convolution can effectively extract the long-range dependencies between electrodes. Temporal convolution can learn richer temporal features and aggregate multi-scale contextual information. Furthermore, we model EEG, EOG, and the corresponding motion information in a unified multi-stream network framework for the first time, demonstrating the validity of motion modality. Experiments on the Sleep-EDF-39 and Sleep-EDF-153 datasets evaluate the feasibility and superiority of our proposed model. Our proposed 4s-SleepGCN model achieves significantly better accuracy on both of them than the current state-of-the-art model. In addition, the proposed lightweight single-stream network with only 0.6 million model parameters achieves higher accuracy and smaller network size compared to some baseline models, which provides a new perspective in the field of sleep staging and thus can be used to monitor and track sleep in a home environment. The proposed multi-stream model can be used as a powerful tool to assist sleep experts in assessing sleep quality and diagnosing sleep-related diseases. The flexibility and adaptability of our proposed model make it suitable for various applications beyond sleep stage classification, such as medical applications, healthcare monitoring, and sports analysis.

In future work, we will focus on how to better understand the complex characteristics of the N_1 stage and improve the classification accuracy of the N_1 stage. We will also intend to collaborate with hospitals or research institutes to validate the scalability and real-world applicability of our proposed method. Furthermore, we aim to enhance the generalization ability of our 4s-SleepGCN model, allowing it to be applied to a broader range of domains beyond sleep stage classification.

ACKNOWLEDGMENT

The authors would like to thank Z. Cheng and Z. Wang from The University of Aizu for their constructive comments in the process of this project study, which helped them to improve the manuscript.

REFERENCES

[1] J. M. Siegel, "Clues to the functions of mammalian sleep," *Nature*, vol. 437, no. 7063, pp. 1264–1271, Oct. 2005.
[2] P. H. Finan, P. J. Quartana, B. Remeniuk, E. L. Garland, J. L. Rhudy, M. Hand, M. R. Irwin, and M. T. Smith, "Partial sleep deprivation attenuates the positive affective system: Effects across multiple measurement modalities," *Sleep*, vol. 40, no. 1, Jan. 2017, Art. no. zsw017.
[3] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea: A population health perspective," *Amer. J. Respiratory Crit. Care Med.*, vol. 165, no. 9, pp. 1217–1239, May 2002.
[4] H. Danker-Hopfe, D. Kunz, G. Gruber, G. Klösch, J. L. Lorenzo, S. L. Himanen, B. Kemp, T. Penzel, J. Röschke, H. Dorn, A. Schlögl, E. Trenker, and G. Dorffner, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *J. Sleep Res.*, vol. 13, no. 1, pp. 63–69, Mar. 2004.

[5] 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, Mar. 2006.

[6] 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 Informat.*, vol. 18, no. 6, pp. 1813–1821, Nov. 2014.

[7] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Comput. Biol. Med.*, vol. 40, no. 7, pp. 629–634, Jul. 2010.

[8] C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement, "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients," *Sleep Med.*, vol. 2, no. 5, pp. 389–396, Sep. 2001.

[9] S. A. Keenan, "An overview of polysomnography," in *Handbook of Clinical Neurophysiology*, vol. 6. 2005, pp. 33–50.

[10] E. Estrada, H. Nazeran, J. Barragan, J. R. Burk, E. A. Lucas, and K. Behbehani, "EOG and EMG: Two important switches in automatic sleep stage classification," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2006, pp. 2458–2461.

[11] A. Malhotra, M. Younes, S. T. Kuna, R. Benca, C. A. Kushida, J. Walsh, A. Hanlon, B. Staley, A. I. Pack, and G. W. Pien, "Performance of an automated polysomnography scoring system versus computer-assisted manual scoring," *Sleep*, vol. 36, no. 4, pp. 573–582, Apr. 2013.

[12] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Arch. General Psychiatry*, vol. 20, no. 2, pp. 246–247, 1969.

[13] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. V. Vaughn, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminol. Tech. Specifications, Darien, Illinois, Amer. Acad. Sleep Med.*, vol. 176, p. 2012, Oct. 2012.

[14] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep Med. Rev.*, vol. 48, Dec. 2019, Art. no. 101204.

[15] 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, May 2016.

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

[17] 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, Jan. 2018.

[18] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*.

[19] E. Bresch, U. Großekathöfer, and G. Garcia-Molina, "Recurrent deep neural networks for real-time sleep stage classification from single channel EEG," *Front. Comput. Neurosci.*, vol. 12, p. 85, Oct. 2018.

[20] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. D. Vos, "Automatic sleep stage classification using single-channel EEG: Learning sequential features with attention-based recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1452–1455.

[21] 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," 2019, *arXiv:1902.06562*.

[22] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," 2016, *arXiv:1610.01683*.

[23] 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. Control*, vol. 42, pp. 107–114, Apr. 2018.

[24] Y. Fang, Y. Xia, P. Chen, J. Zhang, and Y. Zhang, "A dual-stream deep neural network integrated with adaptive boosting for sleep staging," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104150.

[25] 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, May 2019, Art. no. e0216456.

[26] 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, Nov. 2017.

[27] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 641–644.

[28] M. Welling and T. N. Kipf, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017, pp. 1–14.

[29] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 7444–7452.

[30] 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. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 1324–1330.

[31] Z. Jia, Y. Lin, J. Wang, X. Ning, Y. He, R. Zhou, Y. Zhou, and L. H. Lehman, "Multi-view spatial–temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1977–1986, 2021.

[32] M. Li, H. Chen, and Z. Cheng, "An attention-guided spatiotemporal graph convolutional network for sleep stage classification," *Life*, vol. 12, no. 5, p. 622, Apr. 2022.

[33] W. Xia, T. Wang, Q. Gao, M. Yang, and X. Gao, "Graph embedding contrastive multi-modal representation learning for clustering," *IEEE Trans. Image Process.*, vol. 32, pp. 1170–1183, 2023.

[34] Z. Chen, L. Fu, J. Yao, W. Guo, C. Plant, and S. Wang, "Learnable graph convolutional network and feature fusion for multi-view learning," *Inf. Fusion*, vol. 95, pp. 109–119, Jul. 2023.

[35] J. D. Geyer, S. Talathi, and P. R. Carney, "Introduction to sleep and polysomnography," in *Clinical Sleep Disorders*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2009, pp. 265–266.

[36] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, M. R. Pressman, and C. Iber, "The visual scoring of sleep in adults," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 121–131, Mar. 2007.

[37] S. Paisarnsrisomsuk, M. Sokolovsky, F. Guerrero, C. Ruiz, and S. A. Alvarez, "Deep sleep: Convolutional neural networks for predictive modeling of human sleep time-signals," in *Proc. KDD Deep Learn. Day*, Aug. 2018, pp. 1–10.

[38] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 324–333, Feb. 2018.

[39] F. Andreotti, H. Phan, N. Cooray, C. Lo, M. T. M. Hu, and M. De Vos, "Multichannel sleep stage classification and transfer learning using convolutional neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 171–174.

[40] 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, Mar. 2019.

[41] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "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, Apr. 2018.

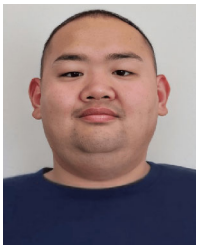
[42] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, May 2019.

[43] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," in *Proc. 30th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2021, pp. 2614–2620.

[44] Z. Jia, X. Cai, G. Zheng, J. Wang, and Y. Lin, "SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 248–257, Dec. 2020.

[45] Z. Jia, X. Cai, and Z. Jiao, "Multi-modal physiological signals based squeeze-and-excitation network with domain adversarial learning for sleep staging," *IEEE Sensors J.*, vol. 22, no. 4, pp. 3464–3471, Feb. 2022.

[46] Z. Yubo, L. Yingying, Z. Bing, Z. Lin, and L. Lei, "MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging," *Front. Neurosci.*, vol. 16, p. 1337, Aug. 2022.
[47] H. Zhu, W. Zhou, C. Fu, Y. Wu, N. Shen, F. Shu, H. Yu, W. Chen, and C. Chen, "MaskSleepNet: A cross-modality adaptation neural network for heterogeneous signals processing in sleep staging," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 5, pp. 2353–2364, May 2023.
[48] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12018–12027.
[49] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2625–2634.
[50] Z. Chen, Z. Wu, Z. Lin, S. Wang, C. Plant, and W. Guo, "AGNN: Alternating graph-regularized neural networks to alleviate over-smoothing," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 31, 2023, doi: 10.1109/TNNLS.2023.3271623.
[51] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, and N. Zheng, "Semantics-guided neural networks for efficient skeleton-based human action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1109–1118.
[52] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2019, pp. 4171–4186.
[53] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.
[54] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Actional-structural graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3590–3598.
[55] W. Peng, X. Hong, H. Chen, and G. Zhao, "Learning graph convolutional network for skeleton-based human action recognition by neural searching," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 2669–2676.
[56] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017, pp. 1–7.
[57] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–13.
[58] A. L. Goldberger, L. A. N. 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, Jun. 2000.
[59] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 558–567.
[60] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.
[61] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," *Can. J. Emergency Med.*, vol. 8, no. 1, pp. 19–20, Jan. 2006.
[62] M. Perslev, M. H. 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.*, vol. 32, 2019, pp. 1–12.
[63] E. Eldele, Z. Chen, C. Liu, M. Wu, C. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.
[64] Y. Sun, B. Wang, J. Jin, and X. Wang, "Deep convolutional network method for automatic sleep stage classification based on neurophysiological signals," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.
[65] X. Lv, J. Li, and Q. Xu, "A multilevel temporal context network for sleep stage classification," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–15, Sep. 2022.
[66] X. Chen, J. He, X. Wu, W. Yan, and W. Wei, "Sleep staging by bidirectional long short-term memory convolution neural network," *Future Gener. Comput. Syst.*, vol. 109, pp. 188–196, Aug. 2020.
[67] E. M. Wickwire, J. Geiger-Brown, S. M. Scharf, and C. L. Drake, "Shift work and shift work sleep disorder: Clinical and organizational perspectives," *Chest*, vol. 151, no. 5, pp. 1156–1172, May 2017.
[68] M. M. Ohayon, "Epidemiology of insomnia: What we know and what we still need to learn," *Sleep Med. Rev.*, vol. 6, no. 2, pp. 97–111, May 2002.
[69] D. J. Buysse, "Sleep health: Can we define it? Does it matter?" *Sleep*, vol. 37, no. 1, pp. 9–17, Jan. 2014.



MENGLEI LI (Graduate Student Member, IEEE) received the B.S. degree in computer science and technology from the Shenyang University of Technology, China, in July 2016, and the M.S. degree in computer science and engineering from The University of Aizu, Japan, in March 2021, where he is currently pursuing the Ph.D. degree with the Graduate Department of Computer and Information Systems. His research interests include graph theory, image classification, biomedical signal processing, sleep, time series data mining, machine learning, deep learning, and artificial intelligence.

HONGBO CHEN (Graduate Student Member, IEEE) was born in Tianjin, China, in 1998. He received the B.S. degree in communication engineering from Yanshan University, China, in July 2020, and the M.S. degree at The University of Aizu, Japan, in March 2022, where he is currently pursuing the Ph.D. degree. His research interests include image processing, optical communications, graph neural networks, and human action recognition.

YONG LIU was a Guest Professor with the School of Computer Science, China University of Geosciences, China, 2010. He was a Research Fellow with AIST Tsukuba Central 2, National Institute of Advanced Industrial Science and Technology, Japan, in 1999, and a Lecturer with the State Key Laboratory of Software Engineering, Wuhan University, in 1994. He is currently a Professor with The University of Aizu, Japan. His research interests include evolutionary computation and neural networks.

QIANGFU ZHAO (Senior Member, IEEE) received the Ph.D. degree from Tohoku University, Japan, in 1988. He joined the Department of Electronic Engineering, Beijing Institute of Technology, China, in 1988, first as a Postdoctoral Fellow and then an Associate Professor. Since October 1993, he has been an Associate Professor with the Department of Electronic Engineering, Tohoku University, Japan. In April 1995, he joined The University of Aizu, as an Associate Professor, where he became a tenure Full Professor, in April 1999. His research interests include image processing, pattern recognition, machine learning, and awareness computing.