EXPLOITING INTERACTIVITY AND HETEROGENEITY FOR SLEEP STAGE CLASSIFICATION VIA HETEROGENEOUS GRAPH NEURAL NETWORK

Ziyu Jia[1, 4], Youfang Lin[1], Yuhan Zhou[1], Xiyang Cai[2], Peng Zheng[1], Qiang Li[3], Jing Wang[1]

[1]School of Computer and Information Technology, Beijing Jiaotong University, China
[2]Samueli School of Engineering, University of California, Los Angeles, USA
[3]Department of Computer Science, Eidgenössische Technische Hochschule Zürich, Switzerland
[4]School of Computing, National University of Singapore, Singapore

ABSTRACT

Sleep stage classification based on physiological time-series is essential for sleep quality evaluation and the diagnosis of sleep disorders in clinical practice. Existing machine learning studies have achieved adequate results in sleep stage classification. However, those methods neglect the significance of simultaneously capturing the interactivity and heterogeneity of physiological signals. In this paper, we propose a novel Sleep Heterogeneous Graph Neural Network (SleepHGNN) to employ these essential features. The SleepHGNN is a deep graph network consisting Heterogeneous Graph Transformer layers, which are composed of Heterogeneous Message Passing module for capturing the heterogeneity and Target-Specific Aggregation module for capturing the interactivity of physiological signals. The experiments conducted on two benchmark datasets show that the SleepHGNN outperforms the state-of-the-art models on the sleep stage classification task for both healthy subjects and subjects with sleep disorders.

***Index Terms*—** Sleep Stage Classification, Heterogeneous Graph Neural Network, Multimodal Physiological Signal

1. INTRODUCTION

Sleep quality is closely related to both our physical and mental health. According to the sleep standard, to assess sleep quality, sleep experts use polysomnogram (PSG) to classify sleep stages. PSG is a set of multimodal signals, i.e., a set of physiological signals that records the activity of different human organs, including electroencephalogram (EEG), electromyogram (EMG), electrocardiogram (ECG), and electrooculogram (EOG), etc. American Academy of Sleep Medicine (AASM) [1] standard is a generally applied sleep staging standard, which combines the characteristics of multimodal signals together to distinguish the five sleep stages: a wake stage (W), three non-rapid eye movement stages (N1, N2, N3), and a rapid eye movement stage (REM). For instance, in addition to typical EEG activities, the low amplitude of EMG and the rapid oscillation of EOG are essential criteria for classifying the REM stage. Therefore, the employment of multimodal signals is vital for high-accuracy sleep stage classification. However, challenges still exist in fully exploiting the typical natures of multimodal signals.

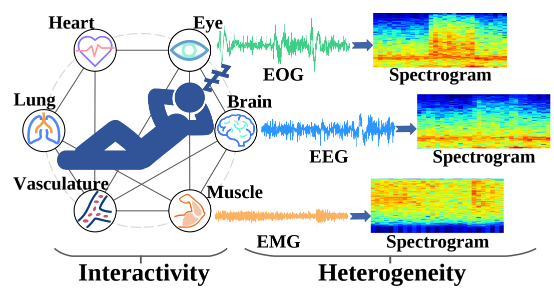

Fig. 1. The interactivity and the heterogeneity of multimodal signals.

The interactivity and heterogeneity of multimodal signals are both important natures for accurate sleep stage classification. As shown in Fig. 1, the **interactivity** among human body systems demonstrates that human organs constantly interact with each other during sleep [2]. Specifically, the strength of interactivity among human body systems continuously alters with the transition of sleep stages. For sleep stage classification tasks, the interactivity of organs is reflected by that of the corresponding multimodal signals. The **heterogeneity** of multimodal signals in our work refers to the difference among the signals from different modalities. As illustrated in Fig. 1, the morphology and amplitudes of multimodal raw signals and their spectrogram are significantly distinct, i.e., each modality has its unique feature distribution.

Although the interactivity and heterogeneity of multimodal signals are both essential, prior studies only attempted to capture one of them. Specifically, Some existing methods tried to capture the interactivity among different channels in PSG via Convulutional Neural Networks (CNNs) [3], Recurrent Neural Networks (RNNs) [4], and Graph Convolutional Networks (GCNs) [5]. The other studies endeavored to exploit the heterogeneity of different modalities by modality-dependent features extractors [6] and handcraft features [7].

To tackle this challenge, we propose the Sleep Heterogeneous Graph Neural Network (SleepHGNN) based on the heterogeneous graphs constructed from the PSG recordings. The heterogeneity and interactivity among the multimodal signals are modeled by the nodes and edges in the graph, respectively. SleepHGNN comprises several Heterogeneous Graph Transformer (HGT) layers, a graph readout layer, and a classifier. Our main contributions can be summarized as follows. We propose the HGT layers for sleep stage classification, which consist of two modules: Heterogeneous Messaging Passing for heterogeneity capturing, and Target-Specific Aggregation for interactivity capturing. Besides, experimental results demonstrate that the proposed SleepHGNN outperforms the state-of-the-art models on the benchmark datasets.

2. RELATED WORK

Many researchers tried to utilize multimodal signals for sleep staging [8, 9]. These studies can be divided into two perspectives: the interactivity capturing and the heterogeneity capturing of multimodal signals.

For the interactivity-capturing work, researchers only apply a single feature extractor to capture features from multimodal signals. For example, a CNN-based deep model [10] leverages 2D CNN to capture the interaction between EEG and EOG signals. SeqSleepNet [4] extracts time-frequency images from EEG, EOG, and EMG signals separately. A fusion network [11] is proposed to leverage the strength of two diverse sequence-to-sequence networks. To better encode the topology of the interaction among EEG signals, GraphSleepNet [5] models the signals as graphs and adopts GCNs to capture the interactivity among them.

For the heterogeneity-capturing work, researchers treat multimodal signals separately and combine their features at last. A CNN-based network [6] processes EEG/EOG signals and EMG signals in independent pipelines, considering the difference of statistical and spectral properties between EMG and the other two modalities. A hierarchical neural network [7] joins the handcraft and network-trained features, where the handcraft features are obtained from each modality in different ways.

However, these approaches fail to capture the interactivity and the heterogeneity simultaneously, which is significant for extracting valuable features from multimodal signals. Therefore, we apply the heterogeneous graph to model PSG signals to capture both of them.

3. SLEEP HETEROGENEOUS GRAPH NEURAL NETWORK

The overall architecture of SleepHGNN is illustrated in Fig. 2. It is made up of a feature extractor composed of HGT layers, a graph readout layer based on global pooling, and a simple classifier based on fully connected layers. The key points of SleepHGNN are the construction of the sleep heterogeneous graphs to model the interactivity and heterogeneity of the multimodal signals, and the employment of Heterogeneous Graph Transformer layers to simultaneously capture the interactivity and the heterogeneity of PSG recordings.

3.1. Construction of Sleep Heterogeneous Graph

In our study, all the PSG recordings are decomposed into 30-second sleep epochs, which are then represented as heterogeneous graphs. As illustrated in Fig. 2, the nodes in the graphs signify the PSG channels, and the edges in the graphs are constructed according to the correlations between each PSG channel measured by the mutual information (MI) [12, 13]. Specifically, each modality corresponds to a unique node type in the graphs, represented by different node colors in this figure. Therefore, the heterogeneity of multimodal signals is manifested by the heterogeneity of the graphs.

3.2. Heterogeneous Graph Transformer

To capture the heterogeneity and interactivity of multimodal signals simultaneously, we adopt the Heterogeneous Graph Transformer (HGT) [14] to extract features from sleep heterogeneous graphs. It comprises two modules: Heterogeneous Messaging Passing to capture heterogeneity and Target-Specific Aggregation to capture interactivity.

3.2.1. Heterogeneous Message Passing

Treating each node and edge type separately, Heterogeneous Message Passing considers the heterogeneity of nodes and edges. It contains two parts: Mutual Attention and Message Extracting.

Mutual Attention. Mutual Attention is to calculate the attention vector for each target node, the elements of which are the normalized similarity scores between each target node and its neighboring source nodes.

Firstly, the key vector $K(src)$ and the query vector $Q(tgt)$ are calculated as follows:

$$K(src) = \text{K-Lin}_{\varphi(src)}\Big(G_s^{(l-1)}[src]\Big), \tag{1}$$

$$Q(tgt) = \text{Q-Lin}_{\varphi(tgt)}\Big(G_s^{(l-1)}[tgt]\Big), \tag{2}$$

where src and tgt denote $\varphi(src)$-type source node and $\varphi(tgt)$-type target node, respectively. Note that $\text{K-Lin}_{\varphi(src)}$ and $\text{Q-Lin}_{\varphi(tgt)}$ are both node-type-dependent linear projections. Thereby, the heterogeneity of the nodes is captured.

Secondly, the similarity scores between the key vectors and query vectors are measured by:

$$\text{Score}(src, e, tgt) = \frac{K(src)W_{\psi(e)}^{\text{ATT}}Q(tgt)^{\text{T}}}{\sqrt{d}}, \tag{3}$$

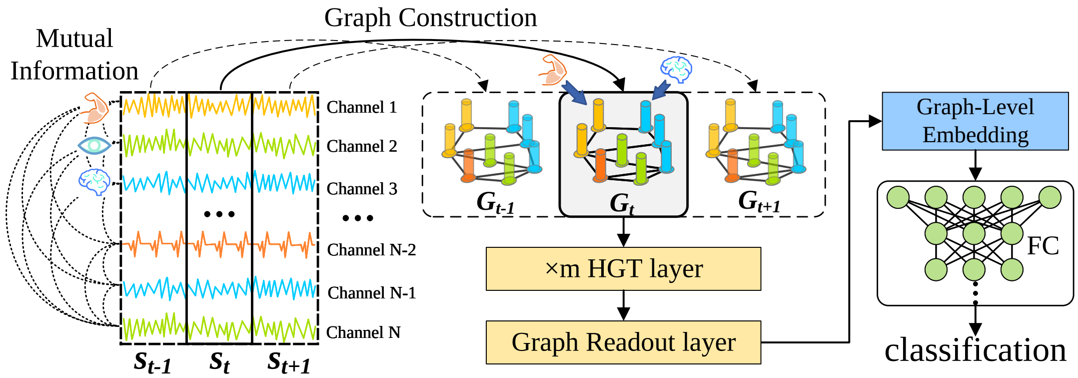

Fig. 2. The overall architecture of SleepHGNN. Firstly, we construct the sleep heterogeneous graphs. Then, the interactivity and heterogeneity of them are captured by the HGT layers. Finally, the graph-level embeddings are calculated and fully connected layers (FC) are employed for classification.

where e denotes the edge with edge type $\psi(e)$ between node pair (src, tgt). In contrast to Transformer, HGT applies weight matrix $W^{\text{ATT}}_{\psi(e)}$ to obtain the edge-dependent scores, and the heterogeneity of the edges is captured.

Finally, the attention vectors $\mathbf{Att}_{\text{HGT}}(src, e, tgt)$ are calculated by softmax normalization.

Message Extracting. Parallel to Mutual Attention, the Message Extracting part, which is to extract the raw messages from source nodes, is executed as:

$$\mathbf{Msg}_{\text{HGT}}(src, e) = \text{V-Lin}_{\varphi(src)}\left(G_s^{(l-1)}[src]\right) W^{\text{MSG}}_{\psi(e)}. \quad (4)$$

Similar to Mutual Attention, $\varphi(src)$-type source nodes are projected into the message vectors by $\text{V-Lin}_{\varphi(src)}$ to capture the heterogeneity of nodes. Then, $W^{\text{MSG}}_{\psi(e)}$ is employed to capture the heterogeneity of edges.

Then, a weighted raw message according to the attention vectors will be passed forward:

$$\mathbf{Msg}'_{\text{HGT}}[src] = \mathbf{Att}_{\text{HGT}}(src, e, tgt) \cdot \mathbf{Msg}_{\text{HGT}}(src, e). \quad (5)$$

3.2.2. *Target-Specific Aggregation*

Target-Specific Aggregation considers the interactivity. After the Heterogeneous Message Passing, we additively aggregate the messages from the neighboring source nodes to each target node, which captures the interactivity of the multimodal signals: $\widetilde{G}_s^{(l)}[tgt] = \sum_{\forall src \in N(tgt)} \mathbf{Msg}'_{\text{HGT}}[src]$.

Finally, we map the $\varphi(tgt)$ target nodes back to their initial feature distribution by $\text{A-Lin}_{\varphi(tgt)}$, then apply the nonlinear activation and the weighted residual connection as follows:

$$\begin{aligned} G_s^{(l)}[tgt] = (1-\alpha) \cdot \sigma\Big(\text{A-Lin}_{\varphi(tgt)}\big(\widetilde{G}_s^{(l)}[tgt]\big)\Big) \\ + \alpha \cdot G_s^{(l-1)}[tgt], \end{aligned} \quad (6)$$

where σ denotes the non-linear activation, and α is a learnable parameter, serving as the weight of the residual connection.

4. EXPERIMENTS

4.1. Datasets and Preprocessing

We evaluate the performance of the models on two benchmark datasets ISRUC-1 and ISRUC-3 [15]. ISRUC-1 contains 87,187 samples from PSG recordings of 100 subjects with sleep disorders. The subjects' age ranges from 20 to 85 years old, with an average of 51. Most of the subjects have detected sleep apnea events. The cases could be under medication, but all were in the position to breathe without equipment assistance. ISRUC-3 contains 8,589 samples from PSG recordings of 10 healthy subjects. The age ranges from 30 to 58 years old, with an average of 40.

Each PSG recording from these two datasets includes 6 EEG channels, 2 EOG channels, and 2 EMG channels. The sampling rate of all signals is 200 Hz. Because the collected PSG recordings have a lower signal-to-noise ratio (SNR), filters are applied on some channels to eliminate noise and unnecessary background signals. It is worth noting that all PSG recordings are preprocessed with a notch filter to eliminate 50 Hz electrical noise. In addition, the bandpass Butterworth filters of 0.3-35 Hz and 10-70 Hz are adopted to process EEG/EOG and EMG signals, respectively.

4.2. Experiment Settings

To evaluate the performance of the proposed method, we compare our SleepHGNN with the widely used baselines. The same experiment and data settings are used for fair comparison. Specifically, the 10-fold cross-validation is applied to evaluate the performance. We adopt the subject-independent strategy for cross-validation. Explicitly, each fold's training samples and testing samples should come from different subjects. For example, in ISRUC-1, we utilize samples from 90 subjects (about 78,500 samples) as the training set, and the rest as the testing set for each fold.

4.3. Overall Comparison

The comparison of the sleep stage classification performance between our model and other benchmark methods on ISRUC-1 and ISRUC-3 is demonstrated in Table 1. Overall, SleepHGNN outperforms other baselines on both ISRUC-1 and ISRUC-3.

As shown in Table 1, deep learning networks DeepSleepNet and MNN can learn discriminative features for sleep staging from the signals. In contrast, traditional machine learning methods, such as RF and SVM, are not capable of capturing useful features, leading to poor results. However, MMCNN captures the heterogeneity of multimodal signals, while SeqSleepNet and GraphSleepNet capture the interactivity. None of these networks can simultaneously process both the essential natures of multimodal signals mentioned above. To take one step further, we construct a sequence of sleep heterogeneous graphs to model the PSG recordings, which enables our model to capture the interactivity and heterogeneity of multimodal signals simultaneously. Therefore, SleepHGNN achieves better accuracy and F1-score than any other state-of-the-art methods. More importantly, SleepHGNN is capable of clinical implementation since its accuracy for sleep staging is similar to or even better than human experts, whose accuracy is about 80% in clinical diagnosis according to a study involving more than 2,500 sleep experts [15], most with three or more years of experience.

Method	ISRUC-1		ISRUC-3	
	Accuracy(%)	F1-score(%)	Accuracy(%)	F1-score(%)
RF [16]	68.08	63.66	61.87	55.96
SVM [17]	63.61	56.05	68.22	66.53
DeepSleepNet [18]	73.92	72.17	75.83	73.23
MNN [3]	71.02	66.30	73.28	69.90
MMCNN [6]	76.90	73.60	77.07	75.33
SeqSleepNet [4]	76.96	68.29	77.45	75.80
GraphSleepNet [5]	76.27	72.72	78.83	72.68
SleepHGNN	**79.80***	**76.64***	**80.56***	**78.65***

* indicates the significant differences between our model and other models ($p < 0.05$).

Table 1. The performance comparison between our model and other state-of-the-art models of the subject-independent experiments on ISRUC-1 and ISRUC-3.

4.4. Model Analysis and Discussion

Ablation Study. To verify the effectiveness of simultaneously capturing the interactivity and heterogeneity of multimodal signals, we conduct an ablation study on the HGT layers in SleepHGNN. As demonstrated in Table 2, we compare the performance of SleepHGNN and its counterpart where the HGT layers are replaced with homogeneous GCNs [19]. Experimental results show that HGT layers of SleepHGNN contribute to better sleep stage classification accuracy by capturing both the interactivity and the heterogeneity of multimodal signals.

Model Setting	Accuracy(%)	F1-score(%)
SleepHGNN	80.56	78.65
w/o HGT	78.91	75.46

Table 2. The result of the ablation study on ISRUC-3.

The connection of PSG signals in sleep heterogeneous graphs under different sleep stages. To demonstrate the interactivity of the PSG signals, we visualize their connection by the edges in the sleep heterogeneous graphs, as shown in Fig. 3. The figure indicates that the interactive connection between the signals varies with the sleep stages. Specifically, in the wake stage W, the connection among different signals are relatively more complicated. By contrast, the connection during the light sleep stages such as N1 becomes sparser. Further, there are even fewer connections in the sleep heterogeneous graphs during the deep sleep stage N3. This phenomenon is consistent with the existing studies [20], where the brain in the light sleep stages is believed to be more active than that in the deep sleep stages.

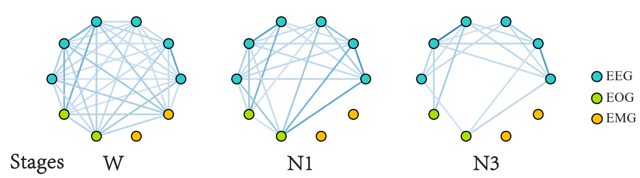

Fig. 3. The connection of sleep heterogeneous graphs under different sleep stages.

5. CONCLUSION

In this paper, we propose a novel deep model SleepHGNN for sleep stage classification. The Heterogeneous Graph Transformer is applied for capturing the interactivity and heterogeneity of the multimodal signals. To the best of our knowledge, this is the first attempt to leverage heterogeneous graph neural networks for sleep stage classification. The experiments on the benchmark datasets prove that SleepHGNN outperforms the state-of-the-art models on the sleep stage classification task for both healthy subjects and subjects with sleep disorders. Since the SleepHGNN is a universal framework for the graph-level classification task based on the heterogeneous graph, we will generalize model to other domains like protein classification and molecular graph classification including anticancer chemical compound classification in the future.

6. REFERENCES

[1] Richard B Berry, Rita Brooks, Charlene E Gamaldo, et al., "The aasm manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, vol. 176, pp. 2012, 2012.

[2] Ronny P Bartsch, Kang KL Liu, Amir Bashan, and Plamen Ch Ivanov, "Network physiology: how organ systems dynamically interact," *PloS one*, vol. 10, no. 11, pp. e0142143, 2015.

[3] Hao Dong, Akara Supratak, Wei Pan, et al., "Mixed neural network approach for temporal sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 324–333, 2017.

[4] Huy Phan, Fernando Andreotti, Navin Cooray, et al., "Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 400–410, 2019.

[5] Ziyu Jia, Youfang Lin, Jing Wang, et al., "Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *IJCAI*, 2020, pp. 1324–1330.

[6] Stanislas Chambon, Mathieu N Galtier, Pierrick J Arnal, et al., "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2018.

[7] Chenglu Sun, Chen Chen, Wei Li, et al., "A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1351–1366, 2019.

[8] Roy R Lederman, Ronen Talmon, Hau-tieng Wu, Yu-Lun Lo, and Ronald R Coifman, "Alternating diffusion for common manifold learning with application to sleep stage assessment," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5758–5762.

[9] Mathias Perslev, Michael Jensen, Sune Darkner, et al., "U-time: A fully convolutional network for time series segmentation applied to sleep staging," in *Advances in Neural Information Processing Systems 32*, pp. 4415–4426. Curran Associates, Inc., 2019.

[10] Michael Sokolovsky, Francisco Guerrero, Sarun Paisarnsrisomsuk, et al., "Deep learning for automated feature discovery and classification of sleep stages," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.

[11] Huy Phan, Oliver Y Chén, Philipp Koch, et al., "Fusion of end-to-end deep learning models for sequence-to-sequence sleep staging," in *EMBC*. IEEE, 2019, pp. 1829–1833.

[12] Seung-Hyun Jin, Peter Lin, Sungyoung Auh, and Mark Hallett, "Abnormal functional connectivity in focal hand dystonia: mutual information analysis in eeg," *Movement Disorders*, vol. 26, no. 7, pp. 1274–1281, 2011.

[13] Sun Hee Na, Seung-Hyun Jin, Soo Yong Kim, and Byung-Joo Ham, "Eeg in schizophrenic patients: mutual information analysis," *Clinical Neurophysiology*, vol. 113, no. 12, pp. 1954–1960, 2002.

[14] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun, "Heterogeneous graph transformer," in *Proceedings of The Web Conference 2020*, 2020, pp. 2704–2710.

[15] Sirvan Khalighi, Teresa Sousa, José Moutinho Santos, and Urbano Nunes, "Isruc-sleep: a comprehensive public dataset for sleep researchers," *Computer methods and programs in biomedicine*, vol. 124, pp. 180–192, 2016.

[16] Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] Johan AK Suykens and Joos Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[18] Akara Supratak, Hao Dong, Chao Wu, and Yike Guo, "Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[20] Victor I Spoormaker, Manuel S Schröter, Pablo M Gleiser, et al., "Development of a large-scale functional brain network during human non-rapid eye movement sleep," *Journal of Neuroscience*, vol. 30, no. 34, pp. 11379–11387, 2010.