DEEP LEARNING BASED SLEEP STAGE CLASSIFICATION

A Thesis submitted by

Xiaopeng Ji

For the award of

Doctor of Philosophy

2023

ABSTRACT

Sleep plays an essential role in humans' life and sleep stage classification is the first step for sleep research and sleep disorder diagnosis. This research aims to classify sleep stages automatically using deep learning methods, which can help clinicians identify sleep problems. Polysomnograms (PSGs), including electroencephalography (EEG), electromyogram (EMG), electrocardiogram (ECG), and electrooculogram (EOG), are signals collected through placing electrodes on the scalp cross different locations, which are powerful tools for sleep stages classification and sleep disorders identification. To classify sleep stages more effectively and efficiently, three models were developed in this research, namely the jumping knowledge spatial-temporal graph convolutional network (JK-STGCN) model, the 3DSleepNet model, and the MixSleepNet model. For all the three models, key features are extracted from multi-channel signals to aggregate spatial and temporal information. For the JK-STGCN model, the connections among different bio-signal channels from the identical epochs and their neighbouring epochs can be obtained through two adaptive adjacency matrices learning methods. A jumping knowledge spatial-temporal graph convolution module helps this model to extract spatial features from the graph convolutions efficiently and temporal features are extracted from its common standard convolutions to learn the transition rules among sleep stages. For the 3DSleepNet model, the intrinsic connections among different bio-signals and different frequency bands in time series and time-frequency are learned by the 3D convolutional layers, while the frequency relations are learned by the 2D convolutional layers. The partial dot-product attention layers help this model find the most important channels and frequency bands in different sleep stages. A long short-term memory unit is added to learn the transition rules among neighbouring epochs. For the MixSleepNet model, the 3D convolution branch can explore the correlations between multi-channel signals and multi-band waves in each channel in the time series, while the graph convolution branch can explore the connections between each channel and each frequency band. The experiments on ISRUC-S3 and ISRUC-S1 demonstrate that the JK-STGCN model outperforms the 1D-CNNs, 2D-CNNs, U2-Net, and other GCN models on these two subsets, while the 3D-CNN model achieves similar performances with a faster speed, and the MixSleepNet model achieves better performances based on the second expert's label with a notable improvement of sleep stage 1 precision.

CERTIFICATION OF THESIS

I, Xiaopeng Ji, declare that this Thesis entitled *Deep Learning Based Sleep Stage Classification* is not more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references, and footnotes. The thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Date: 31/8/2023

Endorsed by:

Professor Yan Li
Principal Supervisor

Professor Paul (Peng) Wen
Associate Supervisor

Student and supervisors' signatures of endorsement are held at the University.

STATEMENT OF CONTRIBUTION

Xiaopeng Ji makes the majority of contributions to the research and each of these papers

Paper 1:

Ji, X., Li, Y., Wen, P., 2022. Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 1464–1472. https://doi.org/10.1109/TNSRE.2022.3176004.

Xiaopeng Ji contributed 70% to this paper.

Corresponding author: Xiaopeng Ji

Paper 2:

Ji, X, Li, Y, Wen, P., 2023. 3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning. IEEE Trans Neural Syst Rehabil Eng. 2023 Aug 28;PP. doi: 10.1109/TNSRE.2023.3309542. Epub ahead of print. PMID: 37639413.

Xiaopeng Ji contributed 70% to this paper.

Corresponding author: Xiaopeng Ji

Paper 3:

Ji, X., Li, Y., Wen, P., Acharya, R., Barua, P. 2023. MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model. Comput. Methods Programs Biomed. 107992. https://doi.org/10.1016/j.cmpb.2023.107992.

Xiaopeng Ji contributed 70% to this paper.

Corresponding author: Xiaopeng Ji

ACKNOWLEDGEMENTS

It took me twenty-seven years of study to finally make this possible, but I didn't realize it until I finished all the writing parts. Twenty-seven years is a long time for a human's life, but it is just a moment in the history of humans. From Confucius and Socrates to Alan Turing and Longping Yuan, no one lives forever, but their knowledge and ideas do. I gave it my all and ended up accomplishing only a meager contribution to our world. Even so, I will do my best to meet new challenges in the future. It would be a great honor if in the future I or others could build on my achievements! The honor belongs not only to me but to all those who have helped me.

I would first express my heartfelt gratitude to my principal supervisor, Professor Yan Li, who gave me the opportunity to start my Ph.D. program. Your unwavering support, insightful feedback, and encouragement have been instrumental in shaping the direction of this study.

I am thankful to my associate supervisor, Professor Paul (Peng) Wen. Your expertise and guidance have been instrumental in broadening my perspective and refining my research skills.

I extend my appreciation to Dr. Tianning Li. The opportunity of teaching machine learning courses has not only given me great teaching experience but has also provided me with a subsistence income. I would also like to thank Dr. Bo Song. Your feedback on my candidate confirmation has improved the quality of the content and writing, resulting in a smooth process overall.

I want to express my deep appreciation to all my friends, who have been with me during my Ph.D. journey, especially Wei Pei and his wife, Jiaxin Li. Your understanding, warmth, patients, and assistance made me feel at home. I will always remember this precious memory of sharing the house. Thanks to Liqiang Wang, who first shared his house. Your car and tableware have provided me with much convenience in my daily life. Thanks to James (Zebing Hua) and Ashely (Hao Sun). I enjoyed the time with you two after our courses. Thanks to Dr. Yi Huang and his wife Hui Liang. You brought us so many joys and your experience about Australian life helped me a lot. Thanks to MingKan Shen. Your advice on writing and publication helped me to improve my writing skills and enriched my publishing knowledge. Thanks

to my former roommates, Danny, Zoe (Xinyu Zhang), and Zhihui Geng. The memory of playing video games and board games will always be my treasure.

I would like to acknowledge the financial support provided by the University of Southern Queensland through the UniSQ International Fees Research Scholarship Their funding has facilitated the successful completion of this research.

I express my deepest gratitude to my parents (Jincheng Ji, Ping Yang, and my stepmother Qing Zhang). Their constant encouragement and belief in my abilities have motivated me to pursue my dreams and overcome challenges with determination. I am forever grateful for your love, sacrifices, and the endless opportunities you have provided.

To my wife, Pei Li. Thank you for being my partner in every sense of the word. Together, we have weathered storms and celebrated victories, and I am excited to continue our journey side by side.

Finally, I'd like to end with an old Chinese poem to express my feelings: ‘If the spring breeze could show tenderness towards the flower's feelings, could it grant me youth again?’ (春风若有怜花意，可否许我再少年？).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

1D-CNN ..... one-dimensional convolutional neural network
2D-CNN ..... two-dimensional convolutional neural network
3D-CNN ..... three-dimensional convolutional neural network
AASM ..... American Academy of Sleep Medicine
AdaBoost ..... Adaptive boosting
AFR ..... Adaptive feature recalibration
ANN ..... Artificial neural network
AR ..... Auto-regressive
Bagging ..... Bootstrap aggregating
BERT ..... Bidirectional encoder representations from transformers
BiLSTM ..... Bidirectional long short-term memory
BP ..... Back-propagation
BPNN ..... Back-propagation neural network
CNN ..... Convolutional neural network
CWT ..... Continuous wavelet transform
DBN ..... Deep belief network
Densenet ..... Dense convolutional neural network
DSVM ..... Dendrogram-support vector machine
DT ..... Decision tree
DTCWT ..... Dual tree complex wavelet transform
DWT ..... Discrete wavelet transform
ECG ..... Electrocardiogram
EEG ..... Electroencephalography
EMD ..... Empirical mode decomposition
EMG ..... Electromyogram
EOG ..... Electrooculogram
FFT ..... Fast Fourier transform
FT ..... Fourier transform
GCN ..... Graph convolutional network
GNN ..... Graph neural network
GRL ..... Gradient reversal layer
HHT ..... Hilbert-Huang transform

IIR ..... Infinite impulse response

IMF ..... Intrinsic mode functions

JK-STGCN Jumping knowledge based spatial-temporal graph convolutional network

LDA ..... Linear discriminate analysis

LS-SVM ..... Least square support vector machine

LSTM ..... Long short-term memory

MD ..... Mean degree

MHA ..... Multi-head attention

MILBP ..... Multiple channels information local binary pattern

MLP ..... Multilayer perceptrons

MSTGCN Multi-view spatial-temporal graph convolutional network

MTSA ..... Multitaper spectral analysis

NLP ..... Natural language processing

NREM ..... Non-rapid eye movement

OAA-SVM ..... One-against-all support vector machine

PSD ..... Power spectral density

PSG ..... Polysomnogram

RCNN ..... Recurrent-convolutional neural networks

REM ..... Rapid eye movement stage

RF ..... Random forest

RNN ..... Recurrent neural network

SE ..... Squeeze-and-excitation

SNet ..... Spectrogram Net

STFT ..... Short term Fourier transform

ST-GCN ..... Spatial-temporal graph convolution network

SVM ..... Support vector machine

TCE ..... Temporal context encoder

VG ..... Visibility graph

ViT ..... Vision transformer

WT ..... Wavelet transform

XGBoost ..... eXtreme gradient boosting

CHAPTER 1: INTRODUCTION

1.1. Background

Sleep takes up around one-third of human lives, and this process helps human beings to rejuvenate and relieve fatigue, while insufficient sleep impairs our cognitive performance and affects our daily life psychologically and physically (Weber and Dan, 2016). According to the American National Highway Traffic Safety Administration, falling asleep during driving leads to at least 100,000 traffic accidents annually (Garces Correa and Laciar Leber, 2010). It is also reported that there are approximately 33% population has insomnia problems (Ohayon, 2002). Other sleep disorders, like apnea and circadian rhythm sleep disorders, also afflict millions of people (Ivanenko and Gururaj, 2009). Therefore, sleep stages analysis is necessary for doctors to diagnose sleep disorders and to make scientific recommendations for treatment.

Sleep stage classification has been studied for decades and one of the most powerful tools is polysomnograms (PSGs). PSGs, including electroencephalography (EEG), electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (ECG), are bio-signals conventionally collected through invasive or non-invasive methods. The invasive signals are recorded from intracranially implanted electrodes, while the non-invasive method obtains bio-signals through electrodes attached to the scalp surface (Ball et al., 2009). In recent years, wearable bio-signal collection devices have been developed with a fast progress, which enables facilitating a diverse range of experiments outside traditional laboratory settings (Casson, 2019).

PSGs-based classification algorithms can be divided into multi-channel-based methods and single-channel-based methods broadly, according to the number of channels being input. The multi-channel-based methods have wide implementations in many research fields related to brain activities like measuring the depth of anesthesia (Nguyen-Ky et al., 2013, 2011), motor imagery classification (Hou et al., 2020; Liu and Yang, 2021; Siuly et al., 2013; Zhao et al., 2019), seizure prediction (Covert et al., 2019; Tsiouris et al., 2018; Wang et al., 2021), sleep stage classification (Yuan et al., 2023), etc. Among all PSGs, EEG signals are the most important ones but other types of PSGs, like EOG, and EMG also make great contributions to improve performance. For example, EOGs may help classifiers to improve the performance of the REM stage identification (Wang et al., 2023). Although multi-channel PSGs have

demonstrated their necessity in these tasks, many researchers attempt to use single-channel bio-signal to analyse brain activities, like drowsiness detection (B and Chinara, 2021), emotion recognition (Taran and Bajaj, 2019) and sleep stage identification (Rahman et al., 2018). Compared with multi-channel data, single-channel bio-signal requires fewer computing resources and storage. Figure 1.1 shows an example of single-channel EEG signals in five different sleep stages.

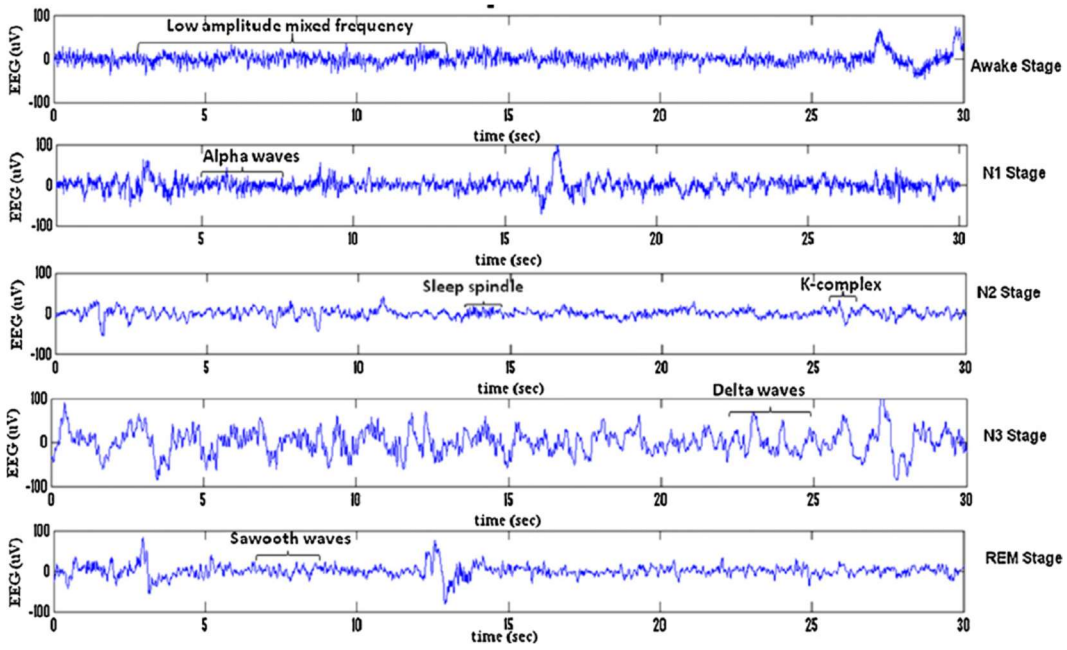

Figure 1.1 An example of EEG signals in five different sleep stages (Fraiwan et al., 2012).

The details of sleep stages were first introduced by Loomis (Loomis et al., 1937, 1936) in the mid of 1930s, when sleep was divided into various stages based on EEG. After then, the sleep stages were studied deeper and deeper. The rapid eye movement stage (REM) was introduced by Aserinsky and Kleitman in 1953 (Aserinsky and Kleitman, 1953), and is related to dreaming. Sleeps were also divided into two main categories: REM and non-Rapid eye movement (NREM) and these two sleep types happen in alternating cycles during sleep. Within an adult's 6-8 hours of sleep, this cycle will alternate 4-6 times, with each cycle lasting approximately 90-110 minutes. However, the length of lasting time and the number of cycles will be different on different subjects, and are affected by age, mental health, and so on (Roebuck et al., 2013). In 1968, the golden criteria named R&K rules, were proposed by Rechtschaffen & Kales (Kales and Rechtschaffen, 1968). According to R&K rules, the NREM stage was divided into four further stages, including stage I, stage II, stage III, and stage IV. Stage I is a transition stage of the brain from alpha waves (8-13 Hz) to theta waves

(4-7 Hz). Sleep spindles ranging from 11-16 Hz and K-complexes appear in stage II. These two sleep stages are also known as light sleep. Stage III and stage IV are deep sleep, and they are characterized by slow waves. R&K rules were reviewed by the American Academy of Sleep Medicine (AASM) in 2004, and the AASM standard was published in 2007. According to these new criteria, stage III and stage IV are merged into stage III, and other bio-signals, like respiratory, cardiac, and movement events are also used for sleep stage classification tasks.

1.2. Methods of sleep stage classification

Traditionally, sleep stage classification requires experts to complete all processes manually. Since bio-signal collection requires a professional to place electrodes in correct places on human bodies, patients must sleep in a study centre or a hospital with some sensors attached to their heads. Otherwise, the quality of the collected signals is very low. After data collection, specialists are required to read PSGs and classify each epoch into five or six stages according to R&K rules or AASM standards. Therefore, the visual inspection process is very expensive and limited by the number of available experts (Yildirim et al., 2019). Moreover, the sleep stages identification works heavily depends on experts' experience, which means that the agreement of this subject work among several specialists can be very low (Norman et al., 2000). To solve the problems above, a possible way is to develop high-performance automatic sleep stage classification systems to help doctors classify sleep stages and diagnose sleep disorders.

Machine learning methods have demonstrated their ability in many prediction tasks, such as object detection (Redmon et al., 2016), image recognition (He et al., 2016; Qassim et al., 2018; Simonyan and Zisserman, 2015), natural language processing (Devlin et al., 2019; Vaswani et al., 2017; Zhou et al., 2021), etc. As a result, many researchers turn to machine learning methods to improve both classification performance and efficiency (Chambon et al., 2018; Pei et al., 2022; Phan et al., 2019). Automatic sleep stage classification algorithms can be divided into two main categories, namely, traditional machine learning methods and deep learning classifiers. In terms of shallow classifiers, many acceptable results have been reported in the sleep stages identification fields. For example, support vector machines (Alickovic and Subasi, 2018; Koley and Dey, 2012; Şen et al., 2014; Zhu et al., 2014a),

random forests (Fraiwan et al., 2012; Memar and Faradji, 2018), and complex networks (Diykh et al., 2020; Diykh and Li, 2016), are the most widely used classifiers for sleep identification tasks. However, an obvious drawback of these methods is that feature engineering and feature selection are inevitable, which means that the classification results are heavily dependent on researchers' prior knowledge and understanding of data. As a result, the performance is limited by selected features and classifiers themselves.

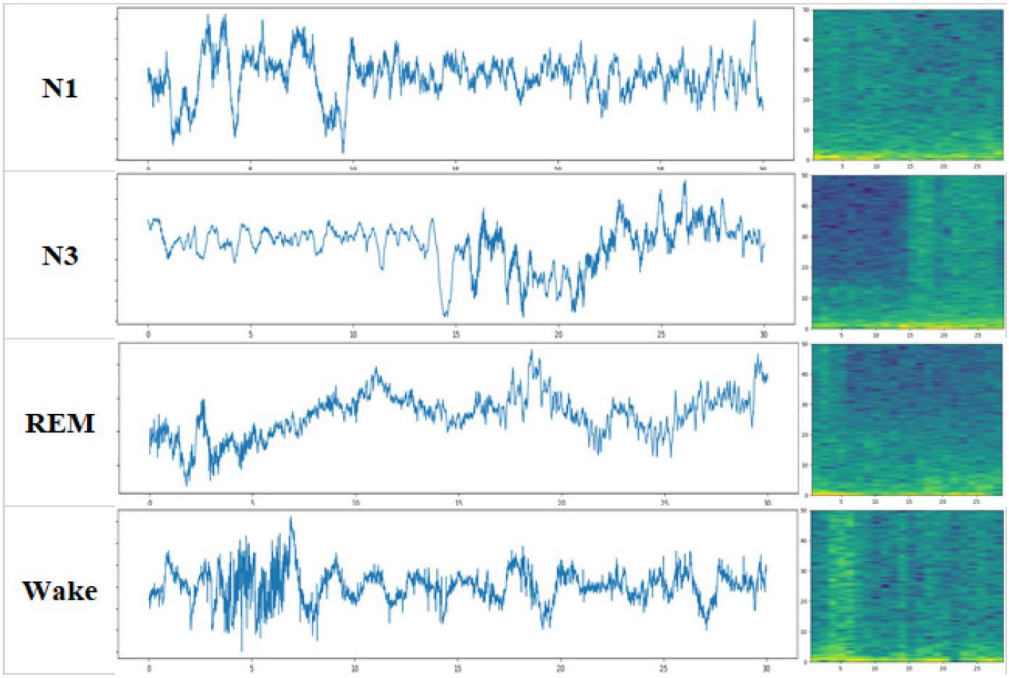

Figure 1.2 1D and 2D representation of EEG signals in four-states (Li et al., 2022).

Unlike shallow machine learning algorithms, deep learning methods allow to input raw data and extract high-level features automatically (Yildirim et al., 2019; Zhu et al., 2020). Convolutional neural networks (CNNs) are classic deep learning models and have achieved extraordinary success in the sleep stage classification field (Goshtasbi et al., 2022; Zhang and Wu, 2017). According to the representation of PSGs in models, CNNs can be simply categorized into 1D-CNN (Eldele et al., 2021), 2D-CNN (Kuo et al., 2021; Li et al., 2022). 1D-CNN models normally focus on raw signals and aggregate temporal information within each epoch (Sors et al., 2018), while 2D-CNNs normally work on 2D spectrogram representation of PSGs (Fang et al., 2023; Li et al., 2022). Figure 1.2 shows 1D and 2D representation of EEG signals in four sleep stages. However, CNNs fail to learn the correlation between neighbouring epochs, namely, transition rules, which play an important role in sleep staging. Recurrent neural networks (RNNs) have great robustness in processing time series

(Funahashi and Nakamura, 1993; Hüsken and Stagge, 2003; Michielli et al., 2019). The added long short-term memory layer (LSTM) to CNNs can help models learn transition rules among neighbouring epochs, which improves classification results notably (Supratak et al., 2017; Supratak and Guo, 2020). Other forms of deep learning models, like deep belief networks (DBN) (Yulita et al., 2017), U-Net (Perslev et al., 2021, 2019), and transformer (Phan et al., 2022), etc also demonstrate their performance.

However, an inevitable shortcoming of all models above is that the intrinsic relationship among brain regions is easily ignored during the exploration of brain activities. Structural connections and functional connections are two main types of connections in the brain, where structural connection patterns are indeed major constraints for the dynamics of cortical circuits and systems, which are captured by functional connectivity (Dimitriadis et al., 2009). The activation of specific brain regions is closely linked to different sleep stages (Killgore et al., 2023). For instance, the REM stage has been associated with the activation of the pons, thalamus, limbic areas, and temporo-occipital cortices, and the deactivation of prefrontal areas, in line with theories of REM sleep generation and dreaming properties (Dang-Vu et al., 2010). GCNs are designed to solve the problem of exploring brain connections in sleep stage classification, whose adaptive graphs are seen as the functional connections in sleep stages for performance enhancement. However, the computational complexity limits their application in practice (Jia et al., 2020b, 2021a).

Currently, the challenges of automated sleep stage identification methods are various. For traditional machine learning, the main limitation is the shallow understanding of data, which indicates that more feature extraction methods or new features need to be explored deeply. In terms of deep learning methods, 1D-CNNs and 2D-CNNs based approaches commonly fail to investigate the connections between brain activities during sleep, whose gaps can be addressed by GCNs. However, the computational complexity of GCN based methods limits its capacity to aggregate temporal information.

1.3. Research objectives

This research aims to develop new models with less computational resources and higher performance for sleep stage classification. Five objectives will be focused on in this project:

1. A novel brain graph learning method is designed to investigate the connections among different brain regions for graph convolutional operations and a jumping-knowledge-based GCN architecture is built for the sleep stage classification purpose.
2. A partial dot-product attention mechanism is designed to find the most important information among each epoch for a 3D-CNN model to identify sleep stages.
3. A model combining the GCN and 3D-CNN model is designed to classify the sleep stages. The GCN branch aims to extract spatial-spectral features, while the 3D-CNN branch aims to extract spatial-temporal features.
4. All the models are evaluated and compared with the state-of-the-art on ISRUC-S3 and ISRUC-S1 datasets with the evaluation metrics of accuracy, precision, recall, F1-score, and Cohen's Kappa. The calculation speed is tested and compared with other existing models as well.

1.4. Contributions

This research includes three key published research articles that all focused on sleep stage classification tasks based on bio-signals. The major contributions of each paper were summarised as follows:

Paper 1: "Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification".

- A novel adaptive graph learning method is designed to aggregate the temporal functional relationship among different bio-signal channels from neighbouring epochs for the localized spatial graph convolution.
- A novel jumping knowledge spatial-temporal graph convolutional module is proposed to capture the localized spatial correlations and temporal features directly.

- Sleep stage classification experiments are conducted on the ISRUC-S3 and ISRUC-S1 (https://sleeptight.isr.uc.pt/) to test the performance of the JK-STGCN model on healthy subjects and sleep-disordered cases. The experimental results demonstrate that the proposed model achieves a competitive overall performance compared to existing baselines. The experimental results of sleep stage classification on healthy and unhealthy mixed cases indicate that the JK-STGCN model achieves the best performance to classify sleep stages for both healthy and unhealthy cases when the unhealthy samples take around 60% in the training set.
- Ablation experiments are also carried out on the ISRUC-S3 dataset to explore the effects of different modules on the sleep stage classification performance and the experimental results show that the JK-STGCN model has the best performance when there is a jumping knowledge spatial-temporal graph convolutional module.

Paper 2: “3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning”.

- A 3D-CNN and 2D-CNN mixed deep learning model named 3DSleepNet is proposed to classify sleep stages automatically. 3D convolutional operations are used to extract spatial-temporal features and spatial-spectral-temporal features from temporal inputs and temporal-frequency inputs, respectively. 2D convolutional operations are also utilized in the proposed model to extract spatial-spectral features from frequency inputs.
- A novel partial dot-product attention mechanism is designed for 3D convolutional operations to efficiently capture the most relevant information. A spatial-spectral attention mechanism is designed for 2D convolutional operations to capture the most relevant spatial-spectral information.
- To evaluate the classification performance on healthy and unhealthy subjects, the classification experiments were performed on ISRUC-S3 and 50 random subjects from ISRUC-S1 (https://sleeptight.isr.uc.pt/). The accuracy, F1-score, and Cohen’s kappa on ISRUC-S3 are 0.832, 0.814, and 0.783, respectively, which indicates that the proposed model achieves a state-of-the-art performance. The overall accuracy, F1-score, and Cohen

kappa on ISRUC-S1 (the datasets with sleep-disorder patients) achieved 0.820, 0.797, and 0.768, respectively, which also demonstrates its generality on unhealthy subjects. The training speed experiments on ISRUC-S3 show that the proposed model outperforms other GCN models and U2-Net architecture models in terms of the model training time.

- The impact of the ratio of unhealthy and healthy subjects in the training set is explored using a set of mixed training data from ISRUC-S1 (unhealthy datasets) and ISRUC-S3 (healthy datasets). The experimental results show that the classification performance on unhealthy patients achieved the best when the training set consists of 100% abnormal patients.
- Incremental experiments are conducted on the ISRUC-S3 dataset to explore the effects of different model variants. The experimental results show that the proposed 3DSleepNet model achieves its best performance when the attention layers and a long short-term memory layer (LSTM) are added with all three input branches.

Paper 3: “3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning”.

- A GCN and 3D-CNN combined deep learning model is proposed for the automatic sleep stage classification task. The differential entropy, a frequency domain feature, is extracted and fed into the graph convolution branch to explore the correlation between frequency bands and channels in the spatial dimension. Additionally, the time domain feature is extracted from down-sampled time series and fed into the 3D convolution branch to investigate the correlation between frequency bands and channels in the temporal dimension.
- Classification experiments were conducted on two datasets, namely, ISRUC-S3 and 50 random selected subjects from ISRUC-S1 (https://sleeptight.isr.uc.pt/) to evaluate the classification performances. The obtained results indicate that the proposed model achieves a state-of-the-art performance when the first expert's labels are used, with an accuracy, F1-score, and Cohen's kappa of 0.830, 0.821, and 0.782, on ISRUC-S3, respectively; and 0.813, 0.787, and 0.757, on ISRUC-S1,

respectively. On the other hand, based on the second expert's labels, the proposed model achieves an accuracy, F1-score, and Cohen's kappa of 0.837, 0.820, and 0.789 on ISRUC-S3, and 0.829, 0.791, and 0.775 on 50 randomly selected subjects from ISRUC-S1, which are outperformed all the compared models.

- To further explore the contribution of each module from the proposed model, incremental experiments were performed on the ISRUC-S3 dataset. The experimental results indicate that when the graph convolutional branch and 3D convolutional branch are added, the model outperformed any other variations. Furthermore, when partial-dot attention layers are added to the 3D convolutional branch, the proposed model can achieve the highest performance.

1.5. Presentation of the thesis

The thesis consists of six chapters as follows:

Chapter 1 introduces the importance of sleep stage classification, the most widely used automatic sleep staging methods, research objectives, and the outline of this thesis.

Chapter 2 provides a comprehensive literature review on feature extraction methods, shallow classifiers, and deep learning algorithms.

Chapter 3 consists of a published journal paper, namely, ‘Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification’, which is a GCN model for sleep identification.

Chapter 4 is an accepted journal paper, namely, ‘3DSleepNet: A Multi-Channel Bio-Signal Based Sleep stage classification Method Using Deep Learning’. We propose a 3D-CNN model for the same purpose.

Chapter 5 is a submitted journal paper, namely, ‘MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model’.

Chapter 6 summarises the conclusions of this study. The potential improvements and developments are also discussed in this chapter.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the most widely used pre-process methods and features in the sleep stages identification tasks. Both shallow classifiers and deep learning algorithms with high classification performance are reviewed as well.

2.1. Pre-process

Figure 2.1 represents the general processing steps of sleep stage classification, where pre-process is the first step for this purpose. Bio-signals are highly irregular, nonlinear, and non-stationary signals (Acharya et al., 2015). The quality of collected signals can be easily affected by collection devices, body movements, faulty electrodes, etc. One way to decrease noises or physiological artifacts is to apply precautions to avoid unnecessary motion, but it cannot work well if patients fail to follow experts' instruction (Jiang et al., 2019). Many artifact removal algorithms and noise filtering methods have been proposed for pre-processing step.

In terms of artifact removal methods, one way is to utilize reference channels to eliminate the artifactual signals another way is to decompose signals into other domains (Jiang et al., 2019). The most used decomposition methods include Regression (Al-Nuaimi et al., 2018), Blind Source Separation (Sweeney et al., 2012), Wavelet Transform algorithm (James and Hesse, 2005), etc.

Sleep stage classification focuses on some specific frequency bands, namely, delta (0.5-2 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (13-22 Hz), and gamma (30-50 Hz), so that noises need to be removed by filters (Sen et al., 2023). Notch filters aim to remove the noise generated by alternating current from a power supply, which is a

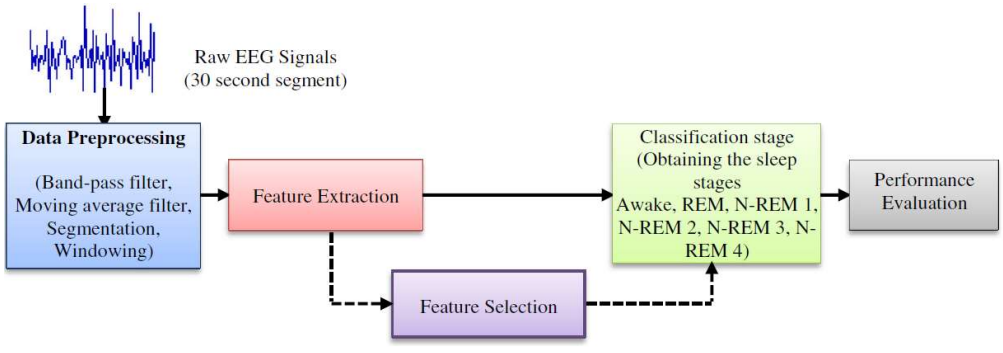

Figure 2.1 General processing steps of sleep stage classification, including the evaluation (Şen et al., 2014).

major noise (Malghan and Hota, 2020). High pass filtering, low pass filtering and band pass filtering are also used to remove undesired frequencies to increase signal quality.

Apart from artifact reduction and noise filter, there are some other common operations that can be performed in the pre-processing stage, like normalization, calibration, detrending, and equalization (Motamedi-Fakhr et al., 2014).

2.2. Feature extraction

Feature extraction is an important process in machine learning methods. Considering the raw data of automatic sleep stage classification are time series, extracted features not only can decrease the computational complexity but also can improve the classification performance. Numerous techniques have been applied to extract features from multiple types of domains, including temporal features, time-frequency features, frequency features, and nonlinear features/complexity measures.

2.2.1. Temporal features

The sampling rate of PSGs is normally over 100Hz, which means that an epoch contains at least 3000 data points. Due to the large size of the data, the correlations among neighbouring data points are overlooked, while some representative features are extracted to represent the overall temporal characteristics of each epoch.

2.2.1.1. Standard statistics

Standard statistics are the simplest and most frequently used features. It is believed that the statistical distribution of the same stage is similar in the time series. As a result, statistical features can represent the overall characteristics of the temporal data (Motamedi-Fakhr et al., 2014). The most widely used statistical features are mean value (Yu et al., 2012), standard deviation/Variance (Aboalayon and Faezipour, 2014; Hassan and Bhuiyan, 2016; Hassan and Hassan Bhuiyan, 2016), skewness (Zoubek et al., 2007), kurtosis (Hassan et al., 2015), median (Vural and Yildiz, 2010), maximum, and minimum (Garcés Correa et al., 2014).

2.2.1.2. Zero crossing

The zero crossing feature is the number of zero crossings or the number of sign changes in an epoch (Şen et al., 2014). This feature is determined by the central frequency of a dominant band (Carrozzi et al., 2004). The high zero crossing value represents that the data is dominated by a high-frequency band, while low-frequency bands give a low zero crossing value (Motamedi-Fakhr et al., 2014).

2.2.1.3. Hjorth parameters

The Hjorth parameters were proposed by Hjorth in 1970 (Hjorth, 1970). There are three parameters derived from PSGs, namely, *Activity*, *Mobility*, and *Complexity,* which represent amplitude, time scale, and complexity, respectively. *Activity* equals the variance of the raw signal. *Mobility* is a measure of the signal mean frequency (Ansari-Asl et al., 2007). *Complexity* is obtained from the variance of the raw data and its first and second derivatives (Khalighi et al., 2013).

2.2.1.4. Shannon entropy

Shannon entropy was first introduced by Shannon in 1948 (Shannon, 1948). This entropy can be defined as a measure of information of the entire epoch (Fraiwan et al., 2011).

2.2.1.5. Renyi entropy

Renyi entropy (Rényi, 1961) can be seen as a superclass of Shannon entropy, where a parameter α is added to increase or decrease its sensitivity towards the shape of probability distributions (Chen et al., 2015).

2.2.2. Frequency features

Frequency features or spectral features are extracted to explore the characteristics of signals in the frequency domain. Frequency domain features facilitate classifiers to discover the frequency band waves that are unique in each

stage. Therefore, spectral features are important factors to improve classification performance for both visual scoring and automatic classifiers.

The non-parametric methods are normally based on the Fourier transform (FT), or Fast Fourier transform (FFT) and they do not depend on an explicit model of the analysed signals. Power spectral density (PSD) is one of the most frequently used spectral features (Welch, 1967), and they are often obtained from some special transforms, like FT or Welch method (Radha et al., 2014). Absolute and relative spectral powers(Šušmáková and Krakovská, 2008), relative spectral powers (Zoubek et al., 2007), and spectral entropy (Koley and Dey, 2012), etc. have been in the sleep pattern recognition task as well.

The parametric methods may give accurate spectral results with little size data if data adheres to a specific model. Consequently, the spectral estimation problem transforms into the task of estimating the parameters of the model (Motamedi-Fakhr et al., 2014). Auto-regressive (AR) model is the most popular signal modelling method, where each sample of a given signal is considered a prediction of the previous weighted samples of that signal (Kayikcioglu et al., 2015).

2.2.3. Time-frequency features

Unlike temporal features or frequency features, which only focus on the time domain or frequency domain respectively, the time-frequency analysis can decompose data into time and frequency.

2.2.3.1. Short time Fourier transform

Short-time Fourier transform or Short-term Fourier transform (STFT) is one of the simplest time-frequency analyses (Motamedi-Fakhr et al., 2014). Temporal data are divided into several segments with the same length, and a windowed Fourier transform is calculated on each segment, after then the window slides along the time axis (Wu et al., 2022). An inevitable problem is that the resolution of time and the resolution of frequency are in conflict, which means that the higher time resolution leads to a lower frequency resolution. A larger time segment size gives higher frequency resolution, but time resolution is reduced as a result.

In sleep identification tasks, the STFT is utilized to obtain the spectrogram images, which means that a one-dimensional signal is mapped to a two-dimensional matrix. Extracted spectrogram images are analysed using multiple channels information local binary pattern (MILBP) approach and fed to ensemble classifiers after feature selection (Abdulla et al., 2023).

2.2.3.2. Wavelet transform

Wavelet transform (WT) decomposes a signal into a set of wavelet functions, which are scaled and shifted versions of a base wavelet. By using different versions of wavelet functions, the high-frequency components can be obtained by short-duration functions and low frequencies can be obtained by long duration (Vatankhah et al., 2010). The wavelet transform can be broadly categorized into the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). Due to their advantages in non-stationary signal decomposition, both CWT (Fraiwan et al., 2012, 2011) and DWT (Al-Salman et al., 2023; See and Liang, 2011) have been applied in sleep stage classification.

2.2.3.3. Empirical mode decomposition

Empirical mode decomposition (EMD) (Huang et al., 1998) aims to represent the raw signal by decomposing it into a finite sum of intrinsic mode functions (IMF) or modes (Hassan and Bhuiyan, 2016). The EMD consists of two components, one is a series of independent intrinsic signals or modes and the other one is the residue of the signal (Fraiwan et al., 2011). The obtained IMFs normally cannot be input as features. As a result, to extract time-frequency features Hilbert-Huang transform (HHT) (Fraiwan et al., 2011; Li et al., 2009) is applied to the obtained IMFs.

2.2.4. Non-linear features

Since PSGs are non-linear signals, it is impossible to fully characterize them by linear stochastic models (Fell et al., 1996). Therefore, many non-linear features are explored to provide complementary information for specific waveforms (Motamedi-Fakhr et al., 2014). The most widely used non-linear features include fractal dimension

(Koley and Dey, 2012; Krakovská and Mezeiová, 2011; Radha et al., 2014; Sekkal et al., 2022), approximate entropy (Chen et al., 2015), sample entropy (See and Liang, 2011), Hurst exponent (Ghimatgar et al., 2019; Memar and Faradji, 2018), Permutation entropy (Liu et al., 2021), and multiscale entropy (Rodríguez-Sotelo et al., 2014), etc.

2.3. Machine learning methods

Machine learning algorithms are the most widely used automatic sleep stage classification methods and machine learning algorithms can be categorized into shallow classifiers and deep learning methods. Shallow machine learning methods normally cannot deal with raw data directly, which means that manual feature extraction is required. Although the performance of these shallow classifiers is limited by the extracted features and classifiers themselves, this exploration still provides a more efficient method for automatic sleep stage classification.

2.3.1. Support vector machine

Support vector machines (SVMs) or support vector networks (Cortes and Vapnik, 1995) were developed by AT & T Bell Laboratories for two-group classification problems with a variety of kernel functions (Boser et al., 1992). Due to its robust prediction, this machine learning algorithm has been one of the most popular shallow machine learning methods, which had numerous applications in multiple fields, like image classification (Chaganti et al., 2020; Chandra and Bedi, 2021), handwriting recognition (Darmatasia and Fanany, 2017; Hamdan and Sathish, 2021), etc. SVMs have been studied in bio signal-related fields for decades and many results have been reported, like sleep spindle detection (Al-Salman et al., 2019), alcoholic EEG signals analysis (Zhu et al., 2014b), epilepsy detection (Li et al., 2016; Omidvar et al., 2021a; Shiao et al., 2017), etc.

In terms of sleep stage classification, some very high results based on SVM classification algorithms have been reported. See and Liang (See and Liang, 2011), extract sample entropy, infinite impulse response (IIR), wavelet multi-resolution analysis, and PSD after filtering a single channel signal using a 5-*th* order Butterworth bandpass and feed these 13 features into an SVM classifier for the final classification, which achieves 96.2% accuracy on a Wake-REM classification task. Aboalayon et al.

filter the input EEG signal into five frequency bands, namely, delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (> 30 Hz), after which, energy features, temporal features and entropy features are extracted from these frequency bands and fed into a SVM for a two-state sleep classification (Aboalayon et al., 2014). Based on similar features and the SVM classifier, Aboalayon and Faezipour test the performance on the Sleep-EDF database (http://www.physionet.org/physiobank/database/sleep-edfx/) and achieves 92% accuracy (Aboalayon and Faezipour, 2014). Some visibility graph features (VG), horizontal visibility graph features, mean degrees (MDs), etc, are also applied in the SVM algorithm for sleep stage classification (Zhu et al., 2014a).

Many modified versions of SVMs have also been reported in sleep identification. A SVM based hierarchical classifier is proposed by Huang et al. (Huang et al., 2013). In this study, the classification system is divided into five layers, where the first layer is used to classify the wake state and layers 2-4 with SVMs using different feature sets aim to classify the stages of wake, light sleep, deep sleep, and REM. A least square support vector machine (LS-SVM) is a famous and popular modified version of the traditional SVM algorithm, and it has been used in EEG-related classification or regression analysis. In a study by (Diykh et al., 2020), statistic features are obtained from original EEG signals and transferred into weighted undirected networks. Then a set of structural and spectral attributes are pulled out and fed into the LS-SVM for the final classification. Al-Salman et al. design a novel technique based on probability distribution features from a single channel and input these features into a LS-SVM to get the identified results (Al-Salman et al., 2023).

2.3.2. Decision trees

Decision trees (DTs) are tree-like models, where the root node is the start point of a tree and the leaf nodes are the last yield. The C4.5 decision tree model is the most popular modified version for classification purposes (Quinlan, 1986). In EEG analysis, DTs have been applied to sleep spindles detection (Duman et al., 2009), seizure detection (Polat and Güneş, 2007), sleep apneas classification (Rohan and Kumari, 2021), and sleep stage identification (Gunnarsdottir et al., 2018), etc.

Beyond decision trees, the combination of a SVM and a DT can be a better option for some specific application data, like those collected at the DyCog Lab of the

Lyon Neuroscience Research Center (Eichenlaub et al., 2014). Lajnef et al. design a decision-tree-based support vector machine approach named Dendrogram-SVM (DSVM) for sleep stage classification (Lajnef et al., 2015). In that study, time domain features, frequency domain features, linear features, and non-linear features are extracted from the filtered EEG signals and input into the DSVM. The DSVM model is a decision tree-based model, where each binary classification node is replaced by a binary SVM. Compared with linear discriminate analysis (LDA) and one-against-all SVM (OAA-SVM), the DSVM achieve the best performance on the data mentioned above.s

2.3.3. Neural networks

Artificial neural networks (ANNs or called Neural networks (NN) for short) are one of the most famous and most widely used classifiers of machine learning methods. Perceptrons were first designed by Rosenblatt in 1958, which consist of a retina, a single layer of input functions, and a single output and can be seen as an abstraction of nerve cells (Rosenblatt, 1958). Even though a perceptron can solve linear problems easily, it is still impossible to solve many pattern recognition problems with a single layer, such as the XOR logical function (Marvin and Seymour, 1969). As a result, multilayer perceptrons (MLPs) are designed to address this issue, and the back-propagation (BP) algorithm is later developed to improve their performance (Minsky and Papert, 1988; Rumelhart et al., 1986).

ANNs have been applied in EEG analysis for decades, like depression identification (Mohan et al., 2016; Puthankattil and Joseph, 2012), autism diagnosis (Djemal et al., 2017), epileptic seizures detection (Choubey and Pandey, 2021; Kumar et al., 2014; Omidvar et al., 2021b), depth of Anesthesia monitoring (Gu et al., 2019), etc. In terms of sleep stage estimation, many studies have been reported. Tagluk et al. design a multi-channel based three-layer back-propagation neural network (BPNN), where an EEG channel, two EOG channels, and an EMG channel are used to extract features and fed into their model for sleep stage classification (Tagluk et al., 2010). Peker extract features of complex values from a dual tree complex wavelet transform (DTCWT) and these features are classified through a complex-valued neural network (Peker, 2016).

However, an inevitable disadvantage of ANNs is that correlations among neighbouring epochs are easily ignored, which are important factors in classification performance. Recurrent neural networks (RNNs) are developed to address this problem. RNNs not only can provide feedback connections but also use internal memory to process data in time series, which may help to remember the correlations among neighbouring epochs. Hsu et al. design a three-layer RNN for sleep stage classification, where the context layer works to learn the transition rules in the short term (Hsu et al., 2013).

A long short-term memory (LSTM) layer can be seen as a special RNN unit, which can solve the vanishing gradient problem and provide a longer memory than the original RNN. The LSTM layer can also be added to original neural networks to improve the prediction capabilities (Dong et al., 2017).

2.3.4. Ensemble classifiers

Ensemble classification algorithms aim to train a set of weak classifiers for classification instead of just one classifier and the prediction results are decided by the voting results of all weak classifiers, which are hoped to provide better performance than a single one. Individual classifiers in an ensemble are known as base classifiers. If the base classifiers are all of the same kind (e.g. decision trees) the ensemble is known as homogeneous. Otherwise, it is known as heterogeneous. (Bramer, 2013).

In sleep stage classification, almost all ensemble classifiers have been evaluated and normally have a better performance than single shallow classifiers. Hassan et al. design a decision tree based bootstrap aggregating (Bagging) classifier, where ten statistical and spectral features in total are extracted and selected from a single EEG channel. The training sets consist of several subsets, whose data are randomly selected from the entire set, and several decision trees are trained using all the subsets for the final results (Hassan et al., 2015). Unlike the Bagging algorithm combining all weak classifiers with the same weights, the results of boosting are decided by the combination of the weighted outcomes from all the weak classifiers, where the weights are determined by the performance in each iteration (Freund and Schapire, 1997). Hassan and Bhuiyan also develop an Adaptive Boosting (AdaBoost) algorithm in 2015 (Hassan and Bhuiyan, 2015), and the classification accuracy of 2-class, 3-class, 4-class, 5-class, and 6-class increase by 3.24%, 3.41%, 3.99%, 2.99%,

and 1.32%, respectively. Abdulla et al. use a genetic algorithm to select the optimal weights for ensemble classifiers to further improve the classification performance (Abdulla et al., 2023). Other ensemble methods, like boosting (Hassan and Bhuiyan, 2017), eXtreme Gradient Boosting (XGBoost) (Liu et al., 2021), and random forests (RF) (Memar and Faradji, 2018), etc. also give acceptable results.

2.4. Deep learning methods

In recent 10 years, deep learning methods have become the most popular algorithms in almost all research fields, like image classification (Dimitrovski et al., 2023; Kaur et al., 2023), medical image analysis (Gupta and Bajaj, 2023; Jiang et al., 2023; Mohammed Alqahtani, 2023; Narayan et al., 2023), natural language process (Mehrish et al., 2023; Weng et al., 2023), etc.

The structure of deep learning models also improve a lot in the development of deep learning. The basic deep learning networks are well-known as AlexNet, which is developed by Krizhevsky et al., where two-branch convolutional neural networks (CNNs) are designed to solve the image classification problem. After that many researchers joined the development of CNNs. For example, He et al. designed ResNet in 2016, which used residual learning to ensure that a deeper CNN structure would not reduce the classification performance (He et al., 2016). Other models like graph neural networks (GNNs) (Scarselli et al., 2009a, 2009b), graph convolutional networks (GCNs) (Kipf and Welling, 2017), U-Nets (Ronneberger et al., 2015), three-dimensional convolutional networks (3D-CNN) (Ji et al., 2013), transformer (Vaswani et al., 2017), etc, are further developed for different purposes.

2.4.1. Convolutional neural networks

Convolutional neural networks are the simplest deep learning methods, which were first introduced by Krizhevsky et al. in 2012 (Krizhevsky et al., 2012). In this research, a picture is split into two parts, which are input into a branch of convolution to extract features, and two fully-connected layers are added for final classification. CNNs are further developed into one-dimensional convolutional neural networks (1D-CNNs), two-dimensional convolutional neural networks (2D-CNNs), and three-

dimensional convolutional neural networks (3D-CNNs) for different purposes, where 1D-CNNs may help to aggregate temporal information in the time dimension (Yubo et al., 2022; Zhao et al., 2017), 2D-CNNs can aggregate spatial information of images to classify images (Kausar et al., 2018; Li et al., 2014), and 3D-CNNs can aggregate temporal and spatial information parallelly for video analysis (Diba et al., 2017; Hou et al., 2019; Qiu et al., 2017).

Compared with traditional machine learning methods, CNNs can extract high-level features automatically without any prior knowledge and have the capability to improve classification performances. As a result, many researchers turn to using CNNs to analyse EEG signals.

2.4.1.1. One-dimensional CNN

1D-CNNs are the simplest CNN models among all CNN types, where they can aggregate temporal information by convolutional operations through temporal dimension and this characteristic allows them to learn the correlations among neighbouring data points (Phan et al., 2018).

Tsinalis et al. design an end-to-end model with convolutional layers, pooling layers, and fully-connected layers for sleep stage classification (Tsinalis et al., 2016). In this research, single-channel EEG data are input into a one-dimensional convolutional layer with 20 filters and then subsampled by a pooling layer. The subsampled features are stacked by a ‘stacking’ layer, which means that extracted features are transformed from a 1D representation into a 2D representation. After 2D convolutional operations, two fully-connected layers and a softmax layer are utilized for final classification. To solve the imbalanced class issue, the preceding two and succeeding two epochs are also input as a single, continuous signal. The generalization is also tested by a 20-fold cross-validation. The experimental results show that this end-to-end CNN model can achieve comparable results as the state-of-the-art shallow classifiers.

Pure one-dimensional CNNs only focus on the correlations among neighbouring data points, rather than the correlations among neighbouring epochs. As a result, an obvious drawback of basic CNNs is that the transition rules among sleep stages can be easily ignored. An easy way to solve this problem is to add an LSTM layer or a bidirectional long short-term memory (BiLSTM) layer to learn and remember

these rules. DeepSleepNet is a very famous CNN-based BiLSTM model in the sleep stage classification field (Supratak et al., 2017). In this research, a single EEG channel-based end-to-end BiLSTM model is designed with a representation learning part and a sequence residual learning part. The representation learning part consists of two CNN branches, where the branch with a large filter size aims to extract frequency information and the one with a small filter size aims to capture temporal information. After this representation learning, a residual learning framework with two BiLSTM layers is added to learn transition rules. The representation learning part is pre-trained by class-balance data at first and the whole model is fine-tuned on a sequential dataset. It is believed that this two-step procedure can solve the problem of the imbalanced class issue. According to the experimental results, the BiLSTM-based CNN model can achieve similar classification performance compared to the state-of-the-art shallow classifiers on both the MASS and Sleep-EDF datasets. The DeepSleepNet model is further developed by a model named TinySleepNet to reduce the complexity of the architecture (Supratak and Guo, 2020). The TinySleepNet model consists of two parts, namely, a representation learning part and a sequence learning part, where the representation learning part uses only one CNN branch to extract features and the sequence learning part replaces two BiLSTM layers by a

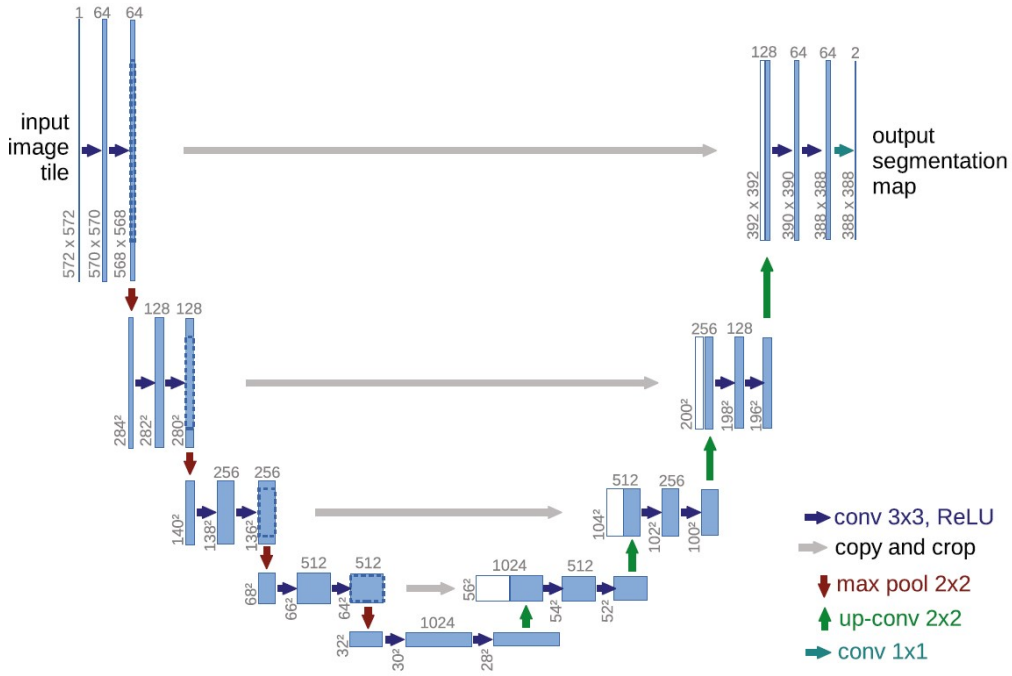

Figure 2.2 The architecture of original U-Net model for biomedical image segmentation (Ronneberger et al., 2015).

unidirectional LSTM layer. The pre-training step is replaced by data augmentation to ensure that the imbalanced data will not reduce the classification performance.

As shown in Figure 2.2, the name of U-Net model is from its U-shaped network structure. The U-Net model is first developed for biomedical image segmentation by Ronneberger et al. (Ronneberger et al., 2015) and has been developed at a very fast pace in this field (Çiçek et al., 2016; Oktay et al., 2018; Siddique et al., 2021). Motivated by its success in biomedical image segmentation, many researchers turned to using this architecture to deal with bio-signals, especially those related to sleep stage classification tasks. Unlike the original U-Net, which uses 2D-CNN to extract features, U-Net architecture-based models in the sleep staging field are normally 1D-CNNs. The explanation of this phenomenon may be that the data size of a 30s-length epoch is quite huge and the complex architecture of the U-Net limits its applications on such big data size. The U-Time model is a one-dimensional convolutional operation-based U-Net model for sleep stage classification (Perslev et al., 2019). U-Time consists of three modules, namely, an encoder submodule, a decoder submodule, and a segment classifier. The encoder submodule consists of four convolution blocks and all convolutional operations keep the input dimensionality through zero-paddings and followed by a batch normalization layer and a max-pooling layer. Two additional convolutional layers with a batch normalization are applied to the outputs of all four encoder blocks. The decoder submodule consists of four transposed-convolution blocks (Long et al., 2015) and each convolution block consists of an up-sampling layer followed by a convolution layer and a normalization step. The resulting features are concatenated and further computed by two convolutional layers with normalization. Finally, the outputs of the decoder submodule are input into the segment classifier, which predicts the classification results based on sample-wise scores. Based on the U-Time model, another U-Net model named U-Sleep is later developed for sleep stage identification (Perslev et al., 2021). The U-Sleep model also consists of three submodules, namely, an encoder submodule, a decoder submodule, and a segment classifier, which play the same roles as they are in the U-Time model with a different architecture. U^2-Net, a two-level nested U-structure model, which is designed for salient object detection (Qin et al., 2020), is also designed and applied to sleep stage classification tasks (Jia et al., 2021b).

Transformer is a sequence-to-sequence model developed for the natural language processing (NLP) purposes, where a scaled dot-product attention

mechanism and a multi-head attention mechanism are designed to replace the RNN layer and reduce the computational complexity (Vaswani et al., 2017). Due to the success of the attention mechanism in NLP, more attention-based models have been developed, like the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), Vision Transformer (ViT) (Dosovitskiy et al., 2021), Swin Transformer (Liu et al., 2021), and Informer (Zhou et al., 2021), etc. Attention mechanisms have also been used in sleep stage identification and many one-dimensional CNNs also add attention layers to improve the classification performance . Eldele designs an attention-based model named AttnSleep, which consists of three submodules, namely, the feature extraction submodule, the temporal context encoder submodule, and the classification submodule (Eldele et al., 2021). The feature extraction submodule consists of two branches, where the branch with a small kernel can extract high-frequency features, and the low-frequency features are extracted by the wide kernel convolutions. The extracted features are further processed by an adaptive feature recalibration (AFR) block for performance improvements. The temporal context encoder (TCE) submodule aims to capture temporal dependencies in the input features, where a multi-head attention (MHA) layer is utilized for this purpose. DynamicSleepNet is a multi-channel based 1D-CNN model, where both channel attention mechanisms and spatial attention mechanisms are used to select the most helpful features for classification and enhance the most important parts of each feature, respectively (Wang et al., 2023).

2.4.1.2. Two-dimensional CNN

Compared with one-dimensional CNNs, two-dimensional CNNs are more flexible, which means that some more features and architectures can be selected to build effective models.

In 1D-CNNs, the convolutional operation only performs along temporal dimension and this characteristic allows models to explore the relationships in time series, while the correlations among virtual channels (Chambon et al., 2018) are easily ignored. However, 2D-CNNs can aggregate information from both the temporal dimension and the virtual channel dimension. As a result, 2D-CNNs have more advantages to explore brain activities. Chambon et al. design a multi-channel end-to-end 2D-CNN model for sleep stage classification, where multi-channel EEG and EOG signals are processed jointly due to their comparable magnitudes, and EMG signals are processed in a parallel pipeline (Chambon et al., 2018). The classification results

demonstrate that this 2D-CNN model overperforms 1D-CNN models (Supratak et al., 2017; Tsinalis et al., 2016), and the experiments on increasing sensors (channels) also show that additional modalities give them a significant boost of performance.

Two-dimensional manual features can also be input to 2D-CNNs for sleep stage identification tasks. The simplest 2D-CNN architecture for this purpose can be a model performing the convolutional operation on a time-frequency image (TFI) obtained by using the smoothed short-time Fourier transform (Xie et al., 2017). A model named Spectrogram Net (SNet) is also designed based on the ensemble 2D-CNN method and continuous wavelet transform (CWT) features (Kuo et al., 2021). In this study, the squeeze-and-excitation (SE) block (Hu et al., 2019) is further combined with an inception module (Szegedy et al., 2015) and a residual learning module (He et al., 2016) to build an Inception-Residual-SE block, where 11 Inception-Residual-SE blocks are used in total to extract abstract features from the input spectrograms. The ensemble technique, smoothing rules, and data augmentation are also utilized to further improve the prediction capabilities.

To learn the correlations among neighbouring sleep epochs, RNN layers can also be added to 2D-CNN models for sleep stage classification. A SLEEPNET is a Recurrent-Convolutional Neural Networks (RCNN) model, which exploits spectrograms as inputs and learns transition rules through five RNN layers (Biswal et al., 2017). In this study, each 30-second epoch is segmented into 29 sub-epochs that are 2 seconds long with a 1-second overlap and, multitaper spectral analysis (MTSA) is applied to every sub-epoch for feature extraction. The extracted features are fed into 2D-CNN layers for “spatial” feature extraction, which are passed to RNN layers to further learn temporal dependency for final classification. An EEGSNet model extracts spectrograms and exploits the BiLSTMs for sleep identification tasks as well (Li et al., 2022).

Since bio-signals can be represented as ‘images’ (such as PSGs), a natural idea is to analyse PSGs by image processing methods. Kanwal et al. design a deep and dense 2D-CNN model based on this idea (Kanwal et al., 2019). According to this research, multi-channel signals are concatenated to generate two-dimensional images with red, green and blue (RGB) channels, after which, high bit-depth frequency domain features are obtained by computing the two dimensional FFT for dynamic range

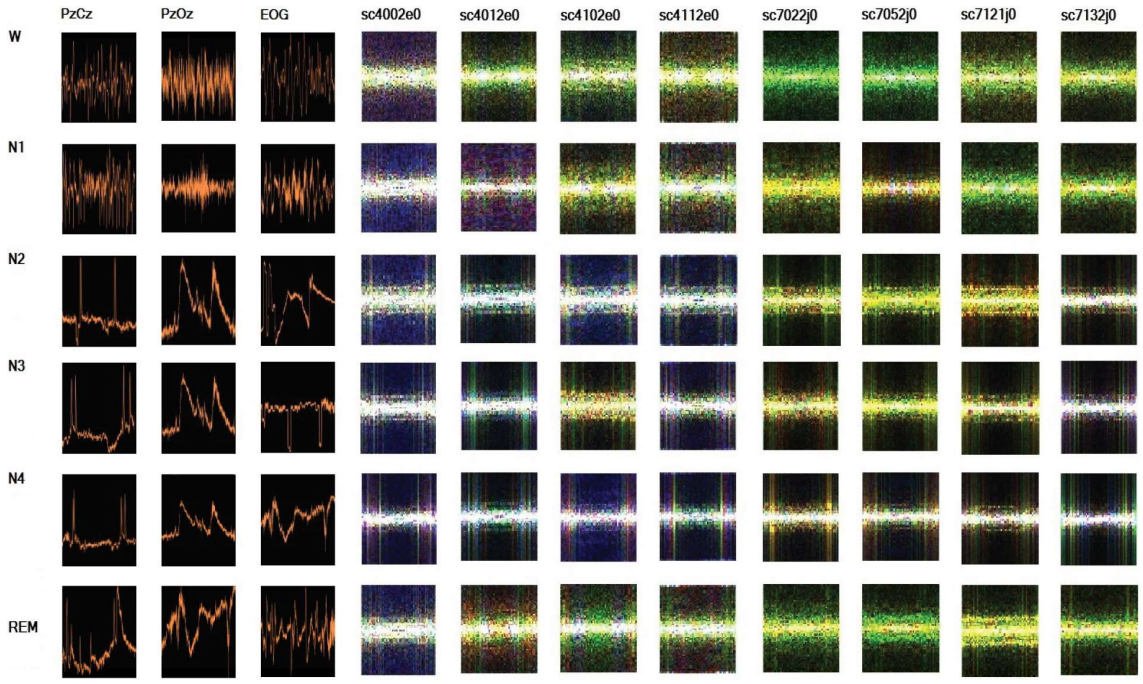

Figure 2.3 The 2D image representations and Fourier feature representations of signals (Kanwal et al., 2019).

increase. Dense convolutional neural networks (Densenet) (Huang et al., 2018), consisting of a single block of 13 layers, including four convolutional layers, are selected as the final classifier. Figure 2.3 shows the image representations of the raw signals in the first three columns, and the remaining columns are Fourier feature representations obtained by the concatenating all the three channels.

Attention mechanisms have also been applied to 2D-CNN models in sleep stage classification fields. Fang et al. design a dual-stream deep neural network with adaptive boosting techniques, where a multi-scale attention layer is added to the two-dimensional branch, the modified ResNet50 branch, to be more specific (Fang et al., 2023).

2.4.1.3. Three-dimensional CNN

Three-dimensional convolutional neural networks (3D-CNNs) are widely used in EEG analysis, like emotion recognition (Jia et al., 2020a; Salama et al., 2018; Wang et al., 2018; Zhao et al., 2020), motor imagery classification (Liu and Yang, 2021; Zhao et al., 2019), and epileptic seizure prediction (Wang et al., 2021), etc. Compared with 1D-CNNs and 2D-CNNs, 3D-CNNs have more advantages in brain activity exploring, which requires uncovering the correlations among brain regions, and frequency bands

in the temporal dimension. However, little research about sleep stage classification has been reported to use 3D-CNNs. To the best of our knowledge, there is only one research using 3D-CNNs for five-sleep state prediction (Duan et al., 2021) except ours, which is introduced in detail in Chapter 4.

In the published study, the 3D-CNN branch aims to learn the high-level representation of the time-frequency features of multi-channel signals, while an LSTM branch processes the 2D time-frequency features generated from 30-second EOG data. Both two branches are followed by a flatten layer for concatenation, which are input into a deep belief network (DBN) for classification (Duan et al., 2021).

2.4.2. Graph convolutional networks

In recent years, graph neural networks (GNNs) have become popular in many fields. Unlike traditional neural networks or CNNs which require fixed structural inputs, GNNs are designed to classify non-Euclidean inputs, and they have been used in several areas like calculating molecular fingerprints (Duvenaud et al., 2015), text classification (Huang et al., 2019; Malekzadeh et al., 2021), and 3D object detection (Shi and Rajkumar, 2020), etc. A graph convolutional network (GCN) is inherited from GNN and CNN since it combines the properties of these two networks with non-Euclidean inputs and convolutional calculation. GCNs are widely used in social network-related analysis (Liu et al., 2022; Wu et al., 2021), image classification (Cheng et al., 2022; Mou et al., 2020; Wan et al., 2020), event extraction (Nguyen and Grishman, 2018), and disease classification (Rhee et al., 2018), etc. Since the inputs of GCNs are non-Euclidean structures, it is more advantageous in representing brain connections and their activities. Therefore, many EEG-related analyses turn to GCN models, like epilepsy classification (Chen et al., 2020), emotion recognition (Gao et al., 2022; Qiu et al., 2023), and motor imagery recognition (Hou et al., 2021), etc.

In terms of sleep stage classification using GCN algorithms (Zhao et al., 2023), GraphSleepNet is the first GCN model (Jia et al., 2020b). In this research, EEG, EOG, ECG, and EMG signals are filtered by bandpass filters with frequency bands of 0.30-100 Hz, 0.10-100 Hz, 0.10-100 Hz, and 10-100 Hz, respectively. As shown in Figure 2.4, for each channel, differential entropy (DE) features are extracted from nine-crossed frequency bands: 0.5-4 Hz, 2-6 Hz, 4-8 Hz, 6-11 Hz, 8-14 Hz, 11-22 Hz, 14-31 Hz, 22-40 Hz, 31-50 Hz. The extracted features are utilized for two purposes, the

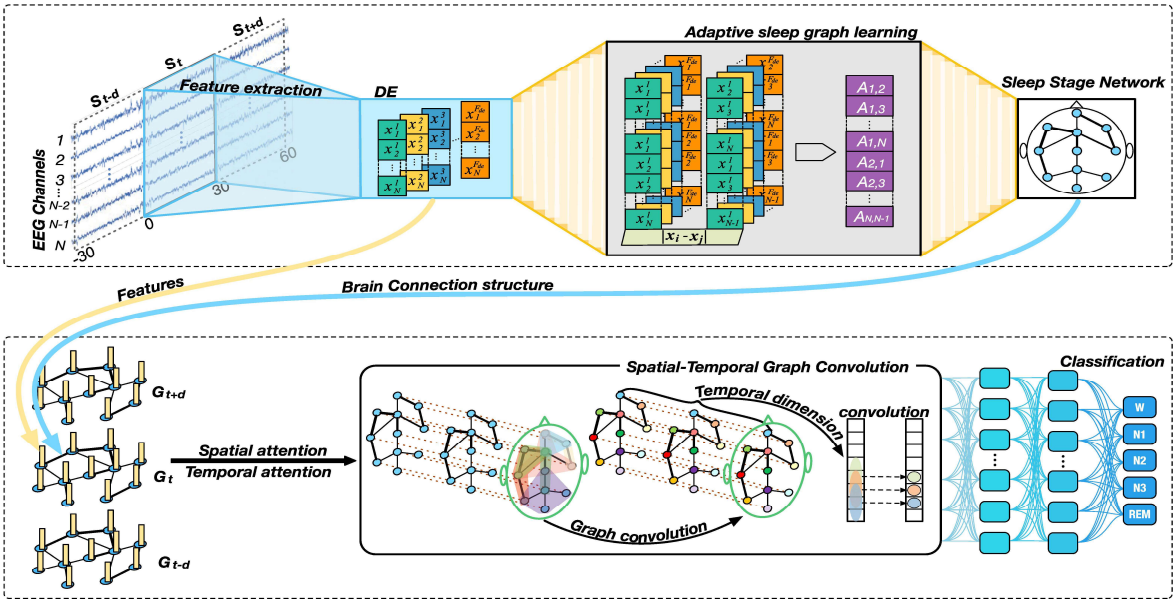

Figure 2.4 The overall architecture of the GraphSleepNet model (Jia et al., 2020b).

first one is to generate an adaptive adjacency matrix (brain connection) by learning, and the second one is to train the spatial-temporal graph convolution networks (ST-GCN) as the inputs. In the spatial-temporal convolution block, the spatial features are extracted by a graph convolution layer and the temporal features are extracted by standard 2D convolutional layers. The temporal attention mechanism and spatial attention mechanism are added to capture valuable spatial-temporal information. After the spatial-temporal convolution block, two fully-connected layers are followed for classification. This ST-GCN model is further developed by a GCN model, namely, the multi-view spatial-temporal graph convolutional networks (MSTGCN) (Jia et al., 2021a). The MSTGCN model consists of two branches, which use two views of brain networks (connection). The brain networks are built based on functional connectivity and spatial distance, where the former represents the collaboration of different brain regions in space and the latter represents the physical distance proximity of the brain regions. The architecture of GCN blocks in the GraphSleepNet model is reused in both two branches in the MSTGCN model, where only one adaptive graph learning module is replaced by the spatial distance. An adversarial domain generalization method and a special layer called Gradient Reversal Layer (GRL) are implemented to enhance the robustness. Motivated by the GraphSleepNet and the idea of ResNet, we also built a GCN model called jumping knowledge based spatial-temporal graph convolutional networks (JK-STGCN), where a new adaptive graph learning algorithm is designed and a jump knowledge (Xu et al., 2018) module is implemented to enhance predictive

capacity (Ji et al., 2022). Detailed information on this model can be found in Chapter 3.

2.5. Datasets

In this study, two datasets are employed to evaluate the classification performance, namely, ISRUC (Khalighi et al., 2016) and Sleep-EDF-153 (Goldberger et al., 2000).

ISRUC dataset consists of three subsets, namely, ISRUC-S1, ISRUC-S2, and ISRUC-S3. These subsets comprise 100 subjects (55 male and 45 female), 8 subjects (6 male and 2 female), and 10 subjects (9 male and 1 female), respectively. For each subset, six EEG channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, and O2-A1), two EOG channels (LOC-A2 and ROC-A1), three EMG channels (Chin EMG, left leg movements and right leg movements), and one ECG channel are collected according to the 10-20 international standard system. These recordings are segmented into epochs with 30s and each sleep epoch is labelled by two well-trained experts according to the AASM standard. The sampling rate of these PSGs is 200 Hz and they are filtered by notch filters with 50 Hz to eliminate electrical noises. EEGs and EOGs are filtered by Butterworth filters with a range of 0.3 Hz - 35 Hz, while EMGs are filtered by Butterworth filters with a range of 10 Hz - 70 Hz.

The Sleep-EDF-153 dataset consists of 78 healthy subjects (34 males and 44 females, aged 25-101). Each subject has two full days of PSGs, except subject 13, 36, and 52, whose one night's data is lost, and this leads to the total number of recordings decreasing to 153. Each recording encompasses two EEG channels (Fpz-Cz and Pz-Oz), one EOG channel (horizontal), and one EMG channel (submental chin EMG). The sampling rate of EEGs and EOG is 100 Hz, while the submental-EMG signal was electronically high-pass filtered, rectified and low-pass filtered after which the resulting EMG envelope expressed in uV rms (root-mean-square) was sampled at 1Hz (Goldberger et al., 2000). Each sleep epoch is labelled according to R&K rules.

2.6. Research achievements and outcomes

The major research achievements and outcomes in the literature review of this thesis are summarized in Table 2.1, Figure 2.1, and Table 2.2.

Table 2.1 Performance of sleep stage classification methods on ISRUC-S3 subset.

References	Methods	Overall performance			Per class F1 score (F1)				
		ACC	F1	Kappa	W	N1	N2	N3	REM
(Alickovic and Subasi, 2018)	SVM	0.733	0.721	0.657	0.868	0.523	0.699	0.786	0.731
(Memar and Faradji, 2018)	RF	0.729	0.708	0.648	0.858	0.473	0.704	0.809	0.699
(Dong et al., 2017)	MLP+LSTM	0.779	0.758	0.713	0.860	0.469	0.760	0.875	0.828
(Hassan and Subasi, 2017)	Bagging	0.740	0.706	0.662	0.847	0.465	0.751	0.843	0.625
(Hassan and Bhuiyan, 2017)	Boosting	0.714	0.681	0.625	0.821	0.458	0.725	0.825	0.576
(Liu et al., 2021)	XGBoost	0.749	0.722	0.676	0.866	0.481	0.751	0.848	0.666
(Supratak et al., 2017)	CNN+BiLSTM	0.788	0.779	0.730	0.887	0.602	0.746	0.858	0.802
(Supratak and Guo, 2020)	CNN+LSTM	0.753	0.737	0.682	0.809	0.533	0.758	0.851	0.734
(Jia et al., 2021b)	U^2-Net	0.807	0.791	0.751	0.867	0.581	0.808	0.895	0.805
(Jia et al., 2020b)	STGCN	0.799	0.787	0.741	0.878	0.574	0.776	0.864	0.841
(Jia et al., 2021a)	MSTGCN	0.821	0.808	0.769	0.894	0.596	0.806	0.890	0.856
(Zhao et al., 2023)	STDP-GCN	0.826	0.810	-	0.835	0.629	0.831	0.860	0.906
(Wang et al., 2023)	CNN	0.829	0.812	0.791	0.899	0.624	0.807	0.907	0.846
(Yubo et al., 2022)	CNN	0.819	0.806	0.768	0.888	0.596	0.820	0.870	0.869
My Research (Paper 1)	JK-STGCN	0.831	0.814	0.782	0.900	0.598	0.826	0.901	0.845
My Research (Paper 2)	3D-CNN	0.832	0.814	0.783	0.896	0.596	0.832	0.909	0.838
My Research (Paper 3)	GCN+3DCNN	0.830	0.821	0.782	0.899	0.625	0.819	0.899	0.830

'ACC'=Accuracy, F1=F1-Score.

Table 2.1 shows that the overall classification performance of all our models has achieved higher results than the existing models on the ISRUC-S3 subset. However, the execution efficiency is an important factor to be considered in the deployment. The training time of our JK-STGCN model and 3D-CNN model are also compared.

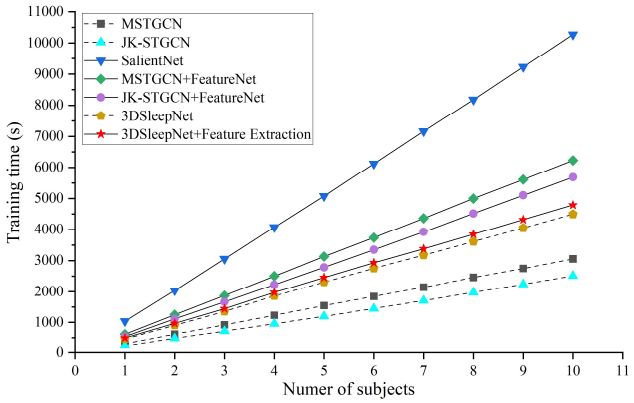

Figure 2.5 Training time comparison among high classification performance on the ISRUC-S3 subset.

Figure 2.5 shows the execution efficiency of the JK-STGCN model and 3D-CNN model, which was further improved compared with those high classification

performance models, like the MSTGCN model and the U^2-Net model (SalientSleepNet).

TABLE 2.2 Performance of sleep stage classification methods on the ISRUC-S3 subset (Expert 2).

References	Methods	Overall performance			Per class F1 score (F1)				
		ACC	F1	Kappa	W	N1	N2	N3	REM
(Alickovic and Subasi, 2018)	SVM	0.714	0.666	0.624	0.824	0.428	0.724	0.815	0.569
(Memar and Faradji, 2018)	RF	0.709	0.693	0.623	0.837	0.475	0.681	0.762	0.708
(Supratak et al., 2017)	CNN+BiLSTM	0.724	0.693	0.645	0.842	0.422	0.759	0.853	0.590
(Supratak and Guo, 2020)	CNN+LSTM	0.778	0.760	0.714	0.792	0.519	0.798	0.891	0.799
(Jia et al., 2021b)	U2-Net	0.810	0.792	0.756	0.895	0.566	0.802	0.896	0.802
(Jia et al., 2021a)	MSTGCN	0.831	0.813	0.781	0.893	0.585	0.821	0.891	0.876
My Research (Paper 1)	JK-STGCN	0.833	0.819	0.784	0.897	0.617	0.824	0.896	0.859
My Research (Paper 3)	GCN+3DCNN	0.838	0.820	0.790	0.891	0.620	0.833	0.904	0.853

'ACC'=Accuracy, F1=F1-Score.

According to Table 2.2, the classification performance of the MixSleepNet model (paper 3) was further evaluated based on the second expert's score on the ISRUC-S3 subset, which achieves the best results among all the existing state-of-the-art methods.

CHAPTER 3: Paper 1 - Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification

3.1. Overview of Paper 1

The details of Paper 1 are given below:

- Paper title: “Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification.”
- Paper length: 9 pages
- Journal: IEEE Transactions on Neural Systems and Rehabilitation Engineering
  - Rank: Q1 (Biomedical Engineering)
  - Impact factor: 4.528 (2021-2022)
  - Cite Score: 8.8 (2022)
  - SJR: 1.260 (2022)
  - SNIP: 1.675 (2022)
- DOI: https://doi.org/10.1109/TNSRE.2022.3176004
- First author: Xiaopeng Ji
- Corresponding author: Xiaopeng Ji

HDR thesis author’s declaration

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

TABLE 3.1: Authorship contributions of Paper 1

Conception and design of study	Xiaopeng Ji, Yan Li, Peng Wen
Analysis and interpretation of data	Xiaopeng Ji
Drafting the manuscript	Xiaopeng Ji

Revising the manuscript critically for important intellectual content	Xiaopeng Ji, Yan Li, Peng Wen
Approval of the version of the manuscript to be published	Xiaopeng Ji, Yan Li, Peng Wen

3.2. Paper file

Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification

Xiaopeng Ji, Yan Li, and Peng Wen

***Abstract*— A novel jumping knowledge spatial-temporal graph convolutional network (JK-STGCN) is proposed in this paper to classify sleep stages. Based on this method, different types of multi-channel bio-signals, including electroencephalography (EEG), electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (ECG) are utilized to classify sleep stages, after extracting features by a standard convolutional neural network (CNN) named FeatureNet. Intrinsic connections among different bio-signal channels from the identical epoch and neighboring epochs can be obtained through two adaptive adjacency matrices learning methods. A jumping knowledge spatial-temporal graph convolution module helps the JK-STGCN model to extract spatial features from the graph convolutions efficiently and temporal features are extracted from its common standard convolutions to learn the transition rules among sleep stages. Experimental results on the ISRUC-S3 dataset showed that the overall accuracy achieved 0.831 and the F1-score and Cohen kappa reached 0.814 and 0.782, respectively, which are the competitive classification performance with the state-of-the-art baselines. Further experiments on the ISRUC-S3 dataset are also conducted to evaluate the execution efficiency of the JK-STGCN model. The training time on 10 subjects is 2621s and the testing time on 50 subjects is 6.8s, which indicates its highest calculation speed compared with the existing high-performance graph convolutional networks and U-Net architecture algorithms. Experimental results on the ISRUC-S1 dataset also demonstrate its generality, whose accuracy, F1-score, and Cohen kappa achieve 0.820, 0.798, and 0.767 respectively.**

***Index Terms*— Deep learning, graph convolutional networks, sleep stage classification.**

I. Introduction

SLEEP plays an important role in human life. Sleep disorders, like insomnias, apnea, and circadian rhythm sleep disorders affect our daily life psychologically and physically. Disturbed sleep patterns lead to sleeplessness at night, which affects our mental status and results in poor mental functions [1]. Poor sleep quality also raises risks of cardiovascular diseases and strokes [2]. Bio-signals, including electroencephalography (EEG), electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (ECG), collected through electrodes placed in different locations in humans, such as the brain, chest, and face, called polysomnograms (PSGs), are powerful tools to help experts and researchers to diagnose sleep disorders [3]. These PSGs are segmented into epochs, which are classified into sleep stages by experienced experts according to the sleep staging criteria such as the Rechtschaffen and Kales sleep staging rules (R&K rules) [4] and American Academy of Sleep Medicine (AASM) standards [5]. Although the PSG-based sleep stage classification is a powerful tool for experts to analyze sleep quality and diagnose sleep disorders, this visual inspection-based manual sleep scoring is a tedious and time-consuming task for trained specialists [6].

To identify sleep stages efficiently, many automatic sleep stage classification methods have been reported. Traditional machine learning methods have given reasonably high sleep stage classification performance in past decades. Inputs of traditional machine learning algorithms are usually extracted from the time-domain [7], [8], frequency-domain [9], [10], or time-frequency domain [11], [12], which requires a lot of prior knowledge [8], [13]. For example, a preprocessing phase is required to eliminate cognitive noise and interference among channels. Often principal component analysis is a typical data reduction technique to seek undesired linear correlation among variables [14]. Due to this limitation, the performance of those algorithms heavily depends on feature engineering and feature selections. Compared to traditional machine learning algorithms, deep learning methods can extract higher-level features from original inputs and output classification results directly. Convolutional neural networks (CNNs) have demonstrated their advanced performance in sleep stage classification [15], [16], while other popular deep learning algorithms like recurrent neural networks (RNNs) [17], [18] and deep belief networks [19] have achieved reliable results as well.

CNNs have the capacity of extracting high-level features from raw data, which allows researchers to input raw data directly instead of hand-crafted features. However, these methods require Euclidean inputs and ignore connections among

Manuscript received January 8, 2022; revised May 2, 2022; accepted May 10, 2022. Date of publication May 18, 2022; date of current version June 3, 2022. The work of Xiaopeng Ji was supported by the University of Southern Queensland International Tuition Fee Scholarship. *(Corresponding author: Xiaopeng Ji.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Xiaopeng Ji and Yan Li are with the School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: xiaopeng.ji@usq.edu.au; yan.li@usq.edu.au).

Peng Wen is with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: paul.wen@usq.edu.au).

Digital Object Identifier 10.1109/TNSRE.2022.3176004

brain regions. Considering the limited understanding of the intrinsic relationship among different channels in different sleep stages, graph-based methods are more advantageous in representing brain connections and their activities. Compared to CNNs, graph convolution networks (GCNs) [20], [21] have the capacity to extract spatial features efficiently on topological data structures, which would provide a potential way to explore the relationship among multiple bio-signal channels during the sleep stage classification.

According to the R&K rules and AASM standards, the transition pattern between neighboring sleep stages is also an essential factor to be considered when sleep stages are identified. However, most of those deep learning algorithms are focused on model development while little attention is paid to the transition mechanism during the sleep process.

To tackle the above challenges, a jumping knowledge spatial-temporal graph convolutional network (JK-STGCN) is proposed to identify sleep stages automatically in this study. With two adaptive graph learning layers and a jumping knowledge graph convolution structure, the JK-STGCN not only learns functional connections among brain regions at each epoch and aggregates temporal functional connections from neighboring epochs but also extracts spatial and temporal features from inputs. The main contributions of this paper are summarized as follows:

- A novel adaptive graph learning method is designed to aggregate the temporal functional relationship among different bio-signal channels from neighboring epochs for the localized spatial graph convolution.
- A novel jumping knowledge spatial-temporal graph convolutional module is proposed to capture the localized spatial correlations and temporal features directly.
- Sleep stage classification experiments are conducted on the ISRUC-S3 and ISRUC-S1 (https://sleeptight.isr.uc.pt/) to test the performance of the JK-STGCN model on healthy subjects and sleep-disordered cases. The experimental results demonstrate that the proposed model achieves the competitive overall performance compared to existing baselines. The experimental results of sleep stage classification on healthy-unhealthy mixed cases indicate that the JK-STGCN model achieves the best performance to classify sleep stages of both healthy and unhealthy cases when the unhealthy samples take around 60% of occupancy.
- Ablation experiments are also carried out on the ISRUC-S3 dataset to explore the effects of different modules on the sleep stage classification performance and the experimental results show that the JK-STGCN model has the best performance when there is a jumping knowledge spatial-temporal graph convolutional module with no attention mechanism.

II. Related Work

A. Sleep Stage Classification

Traditional machine learning classification algorithms, such as support vector machines [22], [23] and random forest [24], [25], have been used for decades in bio-signal analysis, and many studies have reported their high performance in sleep scoring. However, these algorithms require prior knowledge about signal characteristics and feature engineering. That means that the performance may be severely limited by researchers' understanding of data. Due to the fact that deep learning has brought significant breakthroughs in many research areas, such as image processing [26], [27] and natural language processing [28], more and more researchers apply deep learning to sleep stage classification [29], [30].

Unlike traditional machine learning methods, deep learning algorithms, such as CNNs and RNNs [31], have the capacity to extract abstract and high-level features from raw data directly, which allows researchers to use the raw data instead of hand-picked features. Sors *et al.* [32] proposed a 14-layer CNN to extract features from the original single EEG channel inputs. A two-step training CNN model named DeepSleepNet [15] extracts time-invariant features and bidirectional-long short-term memory to learn transition rules among EEG segments. The combination of a long short-term memory unit and a deep belief network [19] has also been applied to identify sleep stages. The U-Net is a very complex architecture with a multi-scale extraction module, which also demonstrates its performance in sleep stage classification [33], [34].

Although these algorithms can extract spatial features and temporal features manually or automatically, they still failed to explore the functional connections among different brain regions during sleep stage classification.

B. Graph Convolutional Networks

Recently, visibility graphs have been utilized in the bio-signal analysis [35] and the sleep stage classification area [36]. Experimental results indicate that graph features make many contributions to improving the classification accuracies. Combining the graph construction and convolutional operation, also known as GCN, has become popular in many fields, like calculating molecular fingerprints [37], text classification [38], neural machine translation [39], etc. Motivated by its success, many researchers have turned to these non-Euclidean input neural networks in the bio-signal processing area, including motor imagery recognition [40], emotion recognition [41], [42], and epileptic seizure detection [43]. However, for sleep scoring, only a few GCN models have been reported. GraphSleepNet [44] is a spatial-temporal graph convolutional network with a spatial attention layer and a temporal attention layer [45], which inputs differential entropy features [46] extracted from multi-channel bio-signals into a learnable adjacency matrix to calculate the graph convolution and classify sleep stages. Jia *et al.* [47] designed a multi-view spatial-temporal graph convolutional network (MSTGCN) and applied the spatial-temporal graph neural network with domain generalization for sleep stage classification. Although the existing GCN models were claimed to be able to solve the problem of obtaining dynamical functional connections among different brain regions and achieved some higher classification accuracies than traditional methods, they fail to aggregate temporal information from neighboring epochs.

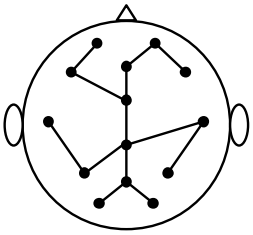

Fig. 1. An example of sleep graph mapped from electrodes at time t.

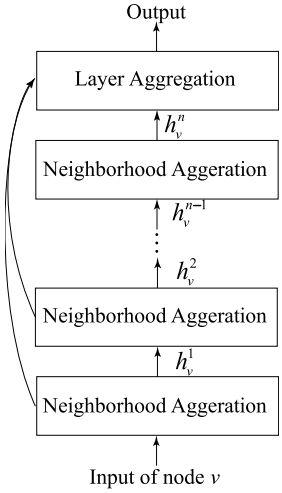

Fig. 2. The structure of a n-layer jumping knowledge network.

III. Preliminaries

In this study, a sleep graph is defined as an undirected graph $\mathcal{G} = (V, E, A)$, where V denotes the set of nodes with the number of $|V| = N$; E denotes the set of edges connecting these nodes; $A \in \mathbb{R}^{N\times N}$ denotes an adjacency matrix of $\mathcal{G}$. At epoch t, attached electrodes will be mapped to a graph as shown in Fig. 1. The connections (edges) between nodes are controlled by a learnable adjacency matrix A.

The raw signal sequences containing L samples are defined as $S = (s_1, s_2, \ldots, s_L) \in \mathbb{R}^{L\times N\times T_s}$, where N denotes the number of channels, T_s denotes sample data points. For each sleep epoch $s_i \in S (i \in \{1, 2, \ldots, L\})$, features are extracted from a CNN named FeatureNet [47] and a N-channel feature matrix of the i-th epoch is defined as $X_i = (x_1^i, x_2^i, \ldots, x_N^i)^T \in \mathbb{R}^{N\times F}$, where $x_n^i \in \mathbb{R}^F (n \in \{1, 2, \ldots, N\})$ denotes features extracted from channel n at epoch i.

The jumping knowledge spatial-temporal graph convolution module is a combination of spatial graph convolution and temporal convolution based on a JK-Net structure [48]. It aggregates both neighborhoods at each independent layer and neighborhoods from previous layers, which increases the size of the influence distribution. As Fig. 2 shows, for each independent node in a graph, the last layer can select from all of those intermediate representations to adapt an effective neighborhood size for each node as needed, and this can lead to a desired adaptivity.

IV. Jumping Knowledge Spatial-Temporal GCN

Fig. 3 illustrates the architecture of the proposed model. There are three key components in this model: 1) Two adaptive graph learning layers are designed to construct adjacency matrixes for the two graph convolutional layers. 2) Based on the JK-Net [48], graph convolutional layers with residual connections are utilized to capture the localized spatial features from neighboring nodes at the same epoch and to aggregate information from different layers. 3) A jumping knowledge spatial-temporal graph convolution module is designed to extract both spatial features and temporal features.

A. Adaptive Graph Learning

Motivated by their high performance of adaptive graph learning methods in the studies of [44], [47], two different graph learning layers are utilized in this study for the localized spatial graph convolution operation.

1) Function-Based Adaptive Graph Learning: As proposed in [44], this connection A_{mn} between node n and node m in an adaptive graph is defined by a non-negative function:

$$\begin{aligned} A_{mn} &= g(x_m, x_n) \\ &= \frac{\exp(\mathrm{ReLU}(\boldsymbol{\omega}^T |x_m - x_n|))}{\sum_{n=1}^{N} \exp(\mathrm{ReLU}(\boldsymbol{\omega}^T |x_m - x_n|))} \end{aligned} \tag{1}$$

where x_m and x_n are the nodes of the adaptive graph, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_F)^T \in \mathbb{R}^{F\times 1}$ is a learnable parameter set. The activation function ReLU guarantees that A_{mn} is non-negative. The softmax operation normalizes each row of A. Weight vector $\boldsymbol{\omega}$ is updated by minimizing the following loss function,

$$\mathcal{L}_{\text{graph\_learning}} = \sum_{m,n=1}^{N} \|x_m - x_n\|_2^2 A_{mn} + \lambda \|A_F\|^2 \tag{2}$$

where $\lambda \geq 0$ is a regularization parameter.

2) Temporal-Information-Based Graph Learning: A function-based adaptive graph learning method can learn the intrinsic connections among different bio-signal channels at one epoch. However, it fails to aggregate functional connections from neighboring epochs, which means that, for each node, the temporal influences from its neighboring nodes of previous epochs and coming epochs are ignored. A temporal-information-based graph learning method considers the intrinsic connections from both temporal and spatial view. A $2d+1$ time steps temporal-information-based adaptive graph is defined as

$$A_T = avg(X \cdot W) \tag{3}$$

where $X = (x_{t-d}, \ldots, x_t, \ldots, x_{t+d}) \in \mathbb{R}^{(2d+1)\times N\times F}$ is a feature set. $W = (w_{t-d}, \ldots, w_t, \ldots, w_{t+d}) \in \mathbb{R}^{(2d+1)\times F\times N}$ is a learnable parameter set. The avg function calculates the mean values of $2d + 1$ adjacency matrixes from time step $t - d$ to $t + d$, which helps to aggregate connections from $2d + 1$ neighboring epochs. The loss of this temporal-information-based graph learning will be considered during calculating the overall loss which is defined as in equation (4):

$$\mathcal{L}_{\text{loss}} = \mathcal{L}_{\text{cross\_entropy}} + \mathcal{L}_{\text{graph\_learning}} + \beta \|A_T\|^2 \tag{4}$$

where β denotes the strength of L2 regularization for temporal-information-based adjacency matrix A_T, and $\mathcal{L}_{\text{graph\_learning}}$ is the loss of the function-based adaptive graph learning as

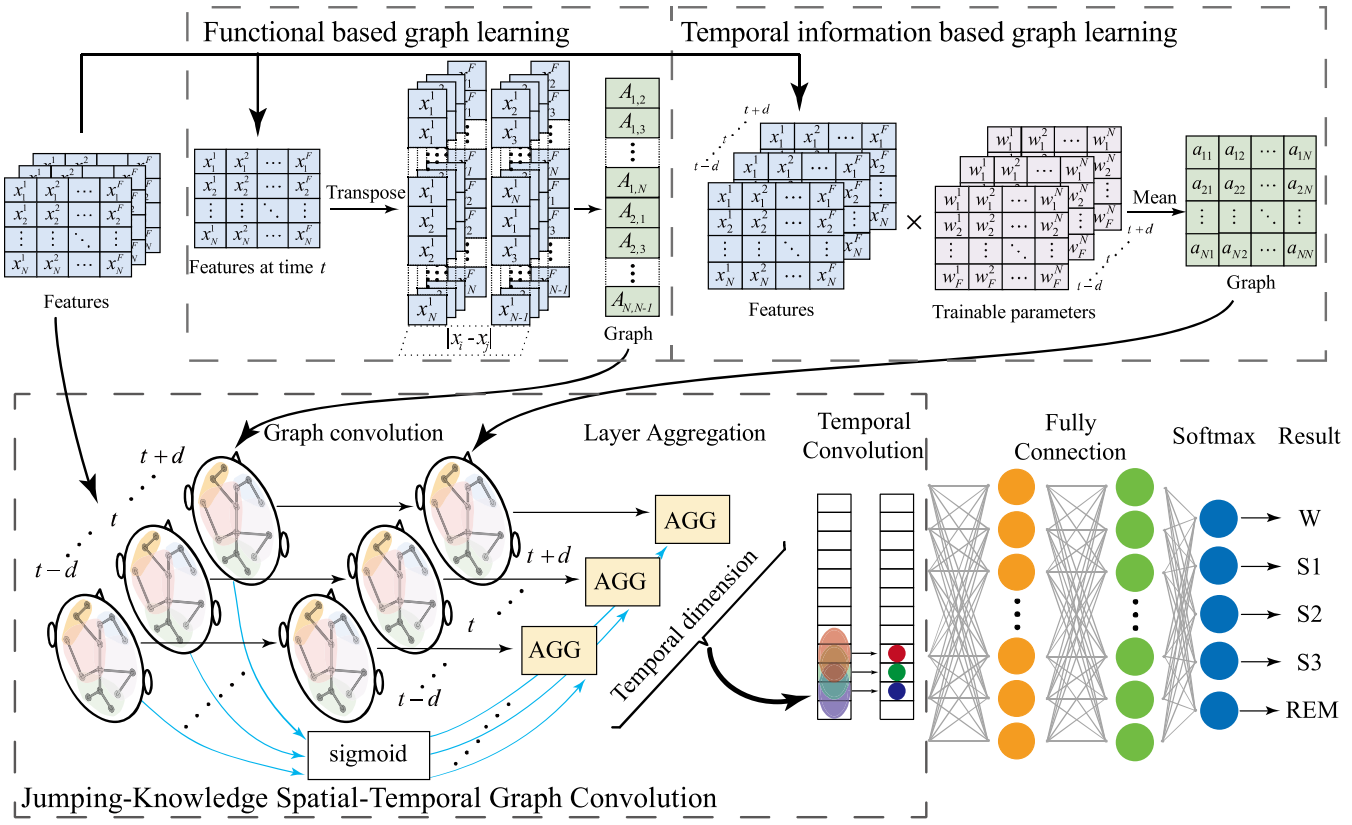

Fig. 3. The structure of the JK-STGCN model. The features are used to generate two adaptive graphs for the jumping knowledge spatial-temporal graph convolution module. The features of $t+2d$ time steps are utilized for the spatial features extraction. The temporal features are extracted by the 2D standard convolution.

defined in the equation (4). $\mathcal{L}_{\text{cross\_entropy}}$ denotes the original loss function as defined in equation (5):

$$\mathcal{L}_{\text{cross\_entropy}} = -\frac{1}{L}\sum_{i=1}^{L}\sum_{r=1}^{R} y_{i,r}\log \hat{y}_{i,r} \tag{5}$$

where L denotes the number of samples, R denotes the number of classes. y is the true label and $\hat{y}$ is the predicted value.

B. Jumping Knowledge Spatial-Temporal Graph Convolution

The jumping knowledge spatial-temporal graph convolution module is a combination of spatial graph convolution and temporal convolution based on the JK-Net structure as mentioned previously, and the spatial graph convolution has the ability to capture spatial features from neighboring graph nodes at the same epoch and the temporal convolution exploits temporal dependencies from nearby epochs.

1) Spatial Graph Convolution: In this study, a GCN is utilized from the perspective of spectral graph theory, and the $K-1$ order Chebyshev polynomials is adopted to reduce the computational complexity

The Laplacian matrix is defined as [42]:

$$L = D - A \tag{6}$$

where A is an adjacency matrix learned based on equation (1) or equation (3), and $D \in \mathbb{R}^{N\times N}$ denotes the diagonal degree matrix of A.

The graph convolution on input x is defined as [49]:

$$g_\theta *_G x = g_\theta(L)x = \sum_{k=0}^{K-1}\theta_k T_k(\tilde{L})x \tag{7}$$

where g_θ denotes the convolution kernel, $*_G$ is the graph convolutional operation, $\theta \in \mathbb{R}^K$ is a vector of polynomial coefficients. $\tilde{L} = 2/\lambda_{\max}L - I_N$, where $\lambda_{\max}$ denotes the Laplacian matrix's maximum eigenvalue, I_N denotes the unit matrix. The $K-1$ order Chebyshev polynomials is recursively defined as:

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \tag{8}$$

where $T_0(x) = 1$, $T_1(x) = x$.

2) Jumping Knowledge Graph Convolution: Based on the JK-Net, a jumping knowledge module is used to extract the spatial information from each node and to aggerate features from different layers. This aggregating layer can be formulated as:

$$\begin{aligned}\text{AGG}^h = {} & \text{ReLu}(g_\theta *_G \chi^{(l-1)}) \\ & + \text{sigmoid}(\text{ReLu}(g_{\theta'} *_{G'} \chi^{(l-2)}))\end{aligned} \tag{9}$$

where g_θ and $g_{\theta'}$ are different convolution kernels defined by equation (7), $*G$ and $*G'$ are graph convolution based on two adaptive graphs learned through equation (1) and equation (3), $\chi^{(l-1)}$ and $\chi^{(l-2)}$ are inputs of graph convolution layer $l-1$ and graph convolution layer $l-2$, ReLu and

TABLE I
NUMBERS OF EPOCHS FOR EACH SLEEP STAGE FROM ISRUC-S1 AND ISRUC-S3 DATASETS

	W	N1	N2	N3	REM	Total
ISRUC-S3	1651	1215	2609	2014	1060	8549
ISRUC-S1	20098	11062	27511	17251	11265	87187

sigmoid are activation functions as defined below:

$$\mathrm{ReLu}(x) = \max(0, x) \tag{10}$$

$$\mathrm{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

3) Temporal Convolution: Common standard 2D convolution layers are utilized to extract temporal features after a spatial graph convolution layer. Based on the combination of sufficient extracted localized spatial features and aggregated localized spatial features at each epoch, the temporal convolution learns the transition rules from the neighboring epochs of the current sleep stages. The temporal convolution of the l-th 2D convolution layer is defined as:

$$\chi^{(l)} = \sigma\left(\Phi * \left(\sigma\left(\mathrm{AGG}^{(l-1)}\right)\right)\right) \in \mathbb{R}^{N \times C_l \times T_l} \tag{12}$$

where σ is the activation function, Φ denotes the convolution kernel, $*$ is the standard convolution operation, and AGG denotes the output of the aggregate layer defined as the equation (9), C_l is the number of channels, and T_l is the l-th layer's temporal dimension.

V. EXPERIMENTS

A. Datasets Used and Experiment Setting

In this study, experiments are conducted on two subsets of the ISRUC-Sleep database [50]: 1) Both the ISRUC-S3 and the ISRUC-S1 data are utilized to evaluate the classification performance of the proposed model. The ISRUC-S3 subgroup contains 10 healthy adults (9 males and 1 female, aged from 30 to 58). The ISRUC-S1 subgroup contains 100 adults with evidence of having sleep disorders (55 males and 45 females, aged from 20 to 85). Each recording from these two subgroups contains 2 EOG channels (LOC-A2 and ROC-A1), 6 EEG channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, and O2-A1), 3 EMG channels (Chin EMG, left leg movements and right leg movements), and 1 ECG channel, and all signals were sampled at 200Hz. The PSG was segmented into 30-second-length epochs and annotated by two experts according to the AASM standards. 2) The ISRUC-S1 data is also used to test the generality of the proposed method. The distribution of sleep stages is shown in TABLE I.

The inputs to the proposed model are extracted from a standard CNN named FeatureNet. This feature extractor aims to extract high-level features from the raw input feature matrix, which means that 3000 original data points from each channel at each epoch will be transferred into a 256-dimension feature vector. 2 EOG channels, 6 EEG channels, 1 EMG channel (Chin EMG), and 1 ECG channel are fed into the CNN to extract features. After that, these extracted features are fed into the proposed model for classifying sleep stages. Detailed hyper-parameters are shown in TABLE II, where the parameter 'neighboring epoch size' means the number of neighboring epochs to aggregate temporal functional connections among brain regions, and the parameter 'Order of Chebyshev polynomials' is set to 9 to aggerate the spatial information from all nine neighboring channels at each epoch.

TABLE II
HYPER-PARAMETERS OF JK-STGCN

Hyperparameter	Value
Neighboring epoch size	5
Layer number of functional graph learning	1
Layer number of temporal graph learning	1
Layer number of graph convolution	2
Temporal convolution kernels	10
Order of Chebyshev polynomials	9
Regularization parameter of graph learning	0.0005
Dropout probability	0.5
Number of training epochs	80
Batch size	64
Learning rate	0.0001
Optimizer	Adam

To evaluate the classification performance of the proposed method, we compare it with traditional machine learning methods, Euclidean-inputs deep learning algorithms like CNNs, RNNs, U-Nets, and existing GCN models on the ISRUC-S3 subgroup and further evaluation experiments for deep learning methods are carried out on the ISRUC-S1 subset. For a fair comparison with the MSTGCN model proposed in [47], we use the same features extracted from the FeatureNet to test the performance, due to the fact that the inputs of both the proposed model and MSTGCN are extracted from the FeatureNet, which means the performance of these two methods may be influenced by the CNN. Moreover, the code is uploaded on Github (https://github.com/XiaopengJi-USQ/JK-STGCN).

The evaluation measures accuracy (ACC), Cohen's kappa (κ), precision (PR), recall (RE) and F1-score (F1) are defined as below:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}\% \tag{13}$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{14}$$

where p_0 is the overall accuracy of the model and p_e is the hypothetical probability of chance agreement.

$$precision = \frac{TP}{TP + FP}\% \tag{15}$$

$$F1 = \frac{2 \times RE \times precision}{RE + precision} \tag{16}$$

All these experiments are conducted in a computer with an Intel I9-10900K CPU, 64 GB Memory and a Nvidia 2080ti GPU.

B. Comparison With the State-of-the-Art Methods

The details of the performance comparison with these baselines on the ISRUC-S3 subgroup data are presented in TABLE III.

The performance of traditional machine learning algorithms heavily depends on researchers' prior knowledge and feature engineering, which means both the spatial features and temporal features cannot be extracted effectively. As a result, their

TABLE III
COMPARISON BETWEEN JK-STGCN AND OTHER DEEP LEARNING METHODS ON ISRUC-S1 SUBGROUP

	Method	Overall Metrics			Per-class F1-score (F1)				
		ACC	F1	κ	W	N1	N2	N3	REM
Alickovic et al. [22]	SVM	0.733	0.721	0.657	0.868	0.523	0.699	0.786	0.731
Memar et al. [25]	RF	0.729	0.708	0.648	0.858	0.473	0.704	0.809	0.699
Supratak et al. [15]	CNN+BiLSTM	0.788	0.779	0.730	0.887	**0.602**	0.746	0.858	0.808
Supratak & YiKe [18]	CNN+RNN	0.746	0.736	0.672	0.797	0.540	0.741	0.833	0.768
Jia et al. [34]	U^2-Net	0.799	0.786	0.742	0.860	0.589	0.793	0.886	0.802
Perslev et al. [33]	U-Net	0.770	0.764	-	0.900	0.550	0.780	0.740	0.850
Jia et al. [44]	STGCN	0.799	0.787	0.741	0.878	0.574	0.776	0.864	0.841
Jia et al. [47]	MSTGCN	0.821	0.808	0.769	0.894	0.596	0.806	0.890	**0.856**
proposed model	JK-STGCN	**0.831**	**0.814**	**0.782**	**0.900**	0.598	**0.826**	**0.901**	0.845

* W=wake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement.

performance cannot be as high as those by deep learning methods.

In terms of deep learning algorithms, both CNNs and RNNs have the ability to extract spatial features or temporal features from original data effectively. However, they ignore the importance of potential connections (relationships) among different channels, which also limits their performances. Because of their special architecture, the U-Net model and U^2-Net model perform as well as the traditional CNNs and RNNs on the ISRUC-S3 dataset. However, their complex architecture and large size training set requirements limit their application.

Although the GraphSleepNet model and the MSTGCN model consider the functional connection among different brain regions and reach higher performance than most Euclidean-inputs deep learning methods like CNNs, RNNs, and the U-Net model, they fail to consider the effects of neighboring nodes from neighboring epochs.

For the classification results, the JK-STGCN can identify most of the corresponding stages. The stage of Wake, N2, and N3 achieve the highest performance among all the algorithms. The reason that stage N1 has a lower classification is because N1 is a transitional stage between the Wake and N2 stages, which means its characteristic is not as clear as the deep sleep stages. From the classification results in TABLE IV, we can find that the JK-STGCN model can classify most Wake stage and most deep sleep stages successfully. The non-symmetric confusion matrix indicates that these misclassifications are caused by the imbalanced class data.

The classification performance of deep learning algorithms can be affected by the dataset size. To further evaluate the classification performance of non-Euclidean inputs models and Euclidean inputs models on large dataset size, the classification experiments were also conducted on the ISRUC-S1 subgroup. 50 subjects are randomly selected from the ISRUC-S1 subset for 25-fold cross validation. The results in TABLE V demonstrate that the JK-STGCN model has more reliable performance compared with other models.

The execution time of a model reflects the complexity of its architecture and its efficiency. Under the same computer setting and similar classification accuracy, the shorter time it takes, the higher efficiency the model has. Models with the top three classification performances on both the ISRUC-S1 and ISRUC-S3 are selected to compare their training time and testing time. Considering the different structures of each model, several training parameters, such as the features extracted from the FeatureNet, the training epochs, batch size, and others are set the same for the MSTGCN and the JK-STGCN. However, the architecture of the U^2-Net model is much more complex than these two GCN methods. As a result, all training parameters of the U^2-Net are set the same as those in [34]. As Fig. 4 illustrates, both the training time and the training time plus the feature extraction time of the proposed model are much lower than those by the MSTGCN and the U^2-Net model. The main reason is that the parameter size of MSTGCN or the U^2-Net is much larger than that in JK-STGCN, which means the two models are much more complex than the JK-STGCN model. Fig. 5 illustrates the testing time of the JK-STGCN, the MSTGCN and the U^2-Net model on 50 subjects. The testing time of the proposed method is a little shorter than the MSTGCN's testing time, which are both around 7 seconds, whereas the U^2-Net takes about 37 seconds to complete the same prediction task. The faster prediction speed and smaller storage space requirement of the JK-STGCN model make it possible to deploy this algorithm to some edge artificial intelligence devices, like the smartphone and the smartwatch.

TABLE IV
CONFUSION MATRIX OBTAINED FROM 10-FOLD VALIDATION ON ISRUC-S3 DATASET

	Predicted					Per-class Metrics	
	W	N1	N2	N3	REM	PR	RE
W	1499	106	30	8	8	0.891	0.908
N1	143	656	273	3	140	0.670	0.540
N2	30	142	2258	138	41	0.790	0.865
N3	3	1	234	1776	0	0.921	0.882
REM	7	74	62	3	914	0.829	0.862

C. Model Analysis

Experiments above demonstrate that the JK-STGCN model has the capacity to classify sleep stages on both healthy subjects and unhealthy cases. The results in TABLE III and

TABLE V
COMPARISON BETWEEN JK-STGCN AND OTHER DEEP LEARNING METHODS ON ISRUC-S1 SUBGROUP

	Method	Overall Metrics			Per-class F1-score (F1)				
		ACC	F1	κ	W	N1	N2	N3	REM
Supratak et al. [15]	CNN+BiLSTM	0.717	0.691	0.638	0.823	0.466	0.738	0.809	0.621
Supratak & YiKe [18]	CNN+RNN	0.778	0.758	0.714	0.883	0.532	0.764	0.848	0.763
Jia et al. [34]	U^2-Net	0.815	**0.801**	0.762	**0.899**	**0.570**	0.800	0.878	0.857
Perslev et al. [33]	U-Net	0.770	0.770	-	0.890	0.520	0.790	0.770	**0.880**
Jia et al. [44]	STGCN	0.786	0.754	0.723	0.884	0.437	0.775	0.838	0.835
Jia et al. [47]	MSTGCN	0.804	0.785	0.748	0.887	0.545	0.791	0.872	0.832
proposed model	JK-STGCN	**0.820**	0.798	**0.767**	0.895	0.550	**0.811**	**0.883**	0.850

* W=wake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement.

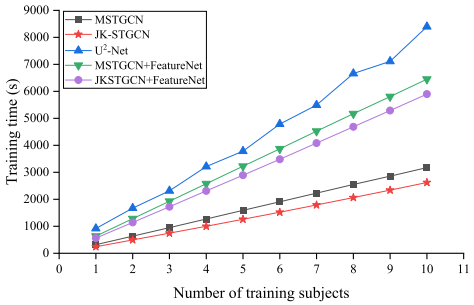

Fig. 4. Training time for the U^2-Net, MSTGCN and JK-STGCN based on *k*-fold training.

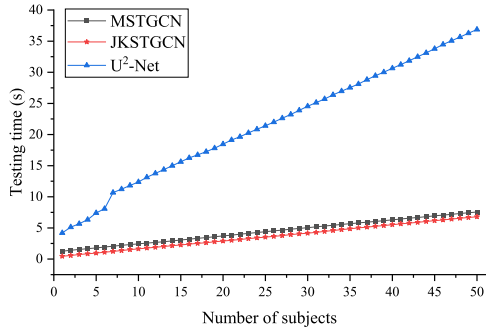

Fig. 5. Testing time for the U^2-Net, MSTGCN and JK-STGCN based on 50 subjects.

TABLE V imply that the classification performance may be affected by two factors, one being the proportion of unhealthy subjects in the training set and the other being the size of the training set. To further evaluate the effects of these two factors on the classification performance, two more experiments are conducted on the ISRUC-S1 subset which contains 100 patients with sleep disorders like sleep apnea obstructive syndrome, periodic limb movements of sleep, etc.

1) Effects of the Proportion of the Unhealthy Subjects in Training Set: The testing set contains 10 patients which are randomly selected from the ISRUC-S1 subset. The initial training set contains ten healthy subjects selected from the ISRUC-S3. The proportion of the unhealthy subjects in the training set is changed by removing one healthy subject from the training set randomly and adding a new random unhealthy subject from the rest subjects in ISRUC-S1 and this operation repeats until all healthy subjects are removed.

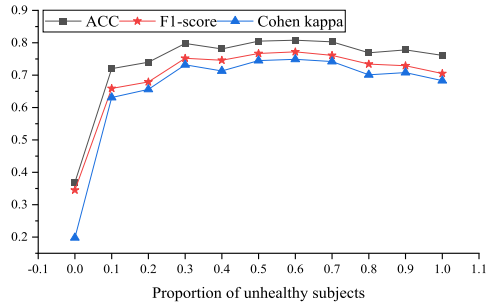

Fig. 6. The trend of accuracy, F1-score and Cohen kappa with different proportion of unhealthy subjects in training set.

Fig. 6 illustrates the changing trend of accuracy, F1-score, and Cohen kappa of the experiment above. The classification accuracy of the JK-STGCN model is much lower in this disordered sleep stage classification task when all the training data is from the healthy subjects. However, the classification accuracies rise rapidly when there are 10% unhealthy subjects. Then the performance improves slowly as the proportion of unhealthy subjects in the training set increases. The JK-STGCN achieves the best performance to classify disordered sleep stages when the unhealthy subjects reach 60%. After that, the performance reduces slightly as the proportion of unhealthy data increases. It is believed that the classification performance improves as the unhealthy data increase at first mainly because the JK-STGCN model starts to learn and recognize the features in abnormal bio-signals and this leads to the improvement of abnormal bio-signals classification. However, the JK-STGCN model starts to misclassify the normal bio-signals, when abnormal bio-signals and abnormal transition ratio reach a high level, resulting in a performance reduction. Even though the classification accuracies are heavily affected by the ratio of the healthy subjects to unhealthy patients, this negative effect may be eliminated by increasing the training set size.

2) Effects of the Size of the Training Set: The ISRUC-S1 subset is randomly divided into four disjoint subgroups, and each subgroup contains 10, 20, 30, and 40 patients respectively. One-subject validation is carried out on each subgroup to validate the influence of the train set size on the JK-STGCN model. As shown in Fig. 7, the classification performance keeps rising with the training set size increases. It is believed that the JK-STGCN model has the capacity to learn and

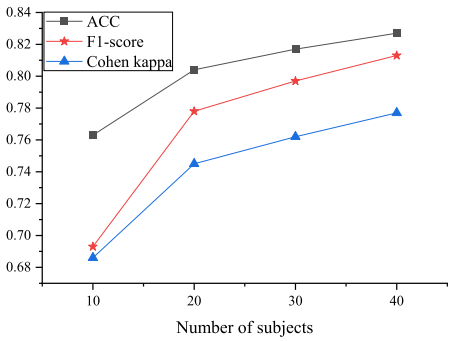

Fig. 7. The performance of JK-STGCN on four disjoint groups from ISRUC-S1.

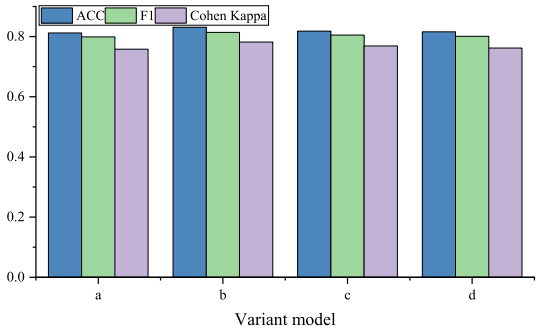

Fig. 8. Comparison of the designed variant models.

recognize both the normal data and the abnormal data if there are sufficient training data.

3) Ablation Experiment: To explore the effects of each module used in the proposed model, four variant models are designed and evaluated using the ISRUC-S3 database. The details are described below:

1. *variant a (basic model)*: The basic model is an independent adaptive graph learning STGCN model without the jumping knowledge spatial-temporal graph convolutional module.

2. *variant b (+ jumping knowledge spatial-temporal graph convolutional module)*: The jumping knowledge spatial-temporal graph convolutional module is added to the basic model to form a JK-STGCN model.

3. *variant c (+spatial attention)*: A spatial attention layer is added to the JK-STGCN to indicate the importance of different channels.

4. *variant d (+spatial attention and temporal attention)*: A spatial attention layer and a temporal attention layer are both added to the JK-STGCN to learn the importance of different channels and different sleep epochs.

As Fig. 8 illustrates, the basic model has the lowest performance among all these variant models. The main reason may be that the parameter size is too small to learn such complex spatial-temporal features, even the adaptive graph learning algorithm provides the optimal connections among channels.

The performance improves when the jumping knowledge spatial-temporal graph convolutional module is added to the basic STGCN model. According to [48], GCN models can achieve the best accuracies when there are two graph convolutional layers with residual connections. The classification results also demonstrate that the two-layer GCN model with residual connections may extract sufficient spatial features, which are more important than global information. The classification accuracies decrease when the attention mechanisms are added. The reason is that the attention mechanisms pay more attention to the important channels and sleep EEG segment sequences, which means some unimportant factors that are related to channels and epochs are ignored, resulting in an inefficient information extraction for sleep stage classification.

VI. Conclusion

In this paper, a JK-STGCN model is proposed to classify sleep stages. The JK-STGCN model contains two adaptive graph learning layers that explore intrinsic connections and relationships among multi-channel bio-signals during sleep stage classification. A jumping knowledge spatial-temporal graph convolution module is designed to extract spatial features and temporal features, which helps the model learn transitional rules among epochs. The experimental results on the ISRUC-S3 subset show that the overall accuracy, the F1-score, and Cohen kappa reached 0.831, 0.814, and 0.782, respectively, which is much better in performance compared to those Euclidean-input deep learning methods and the existing STGCN methods. The experimental results on the ISRUC-S1 subset demonstrate its high performance of sleep stage classification on unhealthy subjects, compared to other deep learning baselines. In addition, extensive experiments are carried out to evaluate the training time and testing time among the top three models. The fastest training speed and prediction speed imply that the proposed model has the ability to be deployed on edge artificial intelligence devices. Moreover, the effects of the distribution of the datasets on the classification performance are explored. The results indicate that the proposed model has reliable robustness to classify both normal data and abnormal data when there is sufficient training data. The ablation experiment is also conducted to find the most important module of the proposed model. Even though the JK-STGCN demonstrates its high performance on sleep stage classification, there is still some space to improve. One drawback is that the GCN model is a multi-channel-based classification algorithm, which means the storage space of the dataset it requires is larger than single-channel-based classification algorithms. One solution is to use the connections among frequency bands instead of the connections among channels and this change allows GCN to classify sleep stages by using a single channel bio-signal, which can decrease the storage space and accelerate the training speed and testing speed. Another improvement that may be considered in the future is the jumping-knowledge module. In the proposed model, the jumping-knowledge operation only happens in each epoch, rather than happens among neighboring epochs. It is believed that the neighboring-epoch-crossed jumping operation would help the aggregate layers to aggregate both spatial and temporal information from the graph convolutional layers, and it would also help the standard temporal convolutional layers to learn the transition rules effectively.

REFERENCES

[1] 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.

[2] H. J. Cho, H. Lavretsky, R. Olmstead, M. J. Levin, M. N. Oxman, and M. R. Irwin, "Sleep disturbance and depression recurrence in community-dwelling older adults: A prospective study," *Amer. J. Psychiatry*, vol. 165, no. 12, pp. 1543–1550, Dec. 2008.

[3] N. A. Siuly, Y. Li, and P. Wen, "Identification of motor imagery tasks through CC-LR algorithm in brain computer interface," *Int. J. Bioinf. Res. Appl.*, vol. 9, no. 2, p. 156, 2013.

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

[5] R. B. Berry *et al.*, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology Tech. Specifications, Darien, Illinois, Amer. Acad. Sleep Med.*, vol. 176, pp. 2012, 2012.

[6] A. Roebuck *et al.*, "A review of signals used in sleep analysis," *Physiol. Meas.*, vol. 35, no. 1, pp. R1–R57, Jan. 2014.

[7] M. Diykh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016.

[8] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017.

[9] L. Zoubek, S. Lesecq, F. Chapotot, S. Charbonnier, and A. G. C. Buguet, "Feature selection for sleep/wake stages classification using data driven methods," *Biomed. Signal Process. Control*, vol. 2, no. 3, pp. 171–179, Jul. 2007.

[10] A. Stochholm, K. Mikkelsen, and P. Kidmose, "Automatic sleep stage classification using ear-EEG," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 4751–4754.

[11] 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.

[12] W. Al-Salman, Y. Li, and P. Wen, "Detection of EEG K-complexes using fractal dimension of time frequency images technique coupled with undirected graph features," *Frontiers Neuroinform.*, vol. 13, p. 45, Jun. 2019.

[13] M. Diykh, Y. Li, and S. Abdulla, "EEG sleep stages identification based on weighted undirected complex networks," *Comput. Methods Programs Biomed.*, vol. 184, no. 6, Feb. 2020, Art. no. 105116.

[14] M. T. Sadiq, X. Yu, and Z. Yuan, "Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 114031.

[15] 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.

[16] L. Wei, Y. Lin, J. Wang, and Y. Ma, "Time-frequency convolutional neural network for automatic sleep stage classification based on single-channel EEG," in *Proc. IEEE 29th Int. Conf. Tools With Artif. Intell. (ICTAI)*, Nov. 2017, pp. 88–95.

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

[18] 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) Conjunction With 43rd Annu. Conf. Can. Med. Biol. Eng. Soc.*, Jul. 2020, pp. 641–644.

[19] I. N. Yulita, M. I. Fanany, and A. M. Arymuthy, "Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification," *Proc. Comput. Sci.*, vol. 116, pp. 530–538, Jan. 2017.

[20] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," 2018, *arXiv:1812.08434*.

[21] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

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

[23] Y. L. Siuly and W. Peng, "Classification of EEG signals using sampling techniques and least square support vector machines," in *Rough Sets and Knowledge Technology*. Berlin, Germany: Springer-Verlag, 2009.

[24] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012.

[25] 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.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

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

[28] A. W. Yu, H. Lee, and Q. V. Le, "Learning to skim text," 2017, *arXiv:1704.06877*.

[29] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Comput. Biol. Med.*, vol. 106, pp. 71–81, Mar. 2019.

[30] H. Sun *et al.*, "Sleep staging from electrocardiography and respiration with deep learning," *Sleep*, vol. 43, no. 7, p. zsz306, Jul. 2020.

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

[32] 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.

[33] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. J. Jennum, and C. Igel, "U-sleep: Resilient high-frequency sleep staging," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–12, Dec. 2021.

[34] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," 2021, *arXiv:2105.13864*.

[35] G. Zhu, Y. Li, P. Wen, and S. Wang, "Analysis of alcoholic EEG signals based on horizontal visibility graph entropy," *Brain Inform.*, vol. 1, nos. 1–4, pp. 19–25, 2014.

[36] G. Zhu, Y. Li, and P. P. Wen, "An efficient visibility graph similarity algorithm and its application on sleep stages classification," in *Proc. Int. Conf. Brain Inform.*, 2012, pp. 185–195.

[37] D. Duvenaud *et al.*, "Convolutional networks on graphs for learning molecular fingerprints," 2015, *arXiv:1509.09292*.

[38] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," 2017, *arXiv:1706.02216*.

[39] A. Vaswani *et al.*, "Attention is all you need," 2017, *arXiv:1706.03762*.

[40] Y. Hou *et al.*, "Deep feature mining via attention-based BiLSTM-GCN for human motor imagery recognition," 2020, *arXiv:2005.00777*.

[41] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.

[42] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, early access, May 11, 2020, doi: 10.1109/TAFFC.2020.2994159.

[43] I. C. Covert *et al.*, "Temporal graph convolutional networks for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2019, pp. 160–180.

[44] Z. Jia *et al.*, "GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2020, pp. 1324–1330.

[45] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 922–929.

[46] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 1–7.

[47] Z. Jia *et al.*, "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.

[48] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-I. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," 2018, *arXiv:1806.03536*.

[49] M. Simonovsky and N. Komodakis, "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3693–3702.

[50] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-sleep: A comprehensive public dataset for sleep researchers," *Comput. Methods Programs Biomed.*, vol. 124, pp. 180–192, Feb. 2016.

CHAPTER 4: Paper 2 - 3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning

4.1. Overview of Paper 2

The details of Paper 2 are given below:

- Paper title: “3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning.”
- Paper length: 11 pages
- Journal: IEEE Transactions on Neural Systems and Rehabilitation Engineering
  - Rank: Q1 (Biomedical Engineering)
  - Impact factor: 4.528 (2021-2022)
  - Cite Score: 8.8 (2022)
  - SJR: 1.260 (2022)
  - SNIP: 1.675 (2022)
- DOI: https://10.1109/TNSRE.2023.3309542
- First author: Xiaopeng Ji
- Corresponding author: Xiaopeng Ji

HDR thesis author’s declaration

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

TABLE 4.1: Authorship contributions of Paper 2

Conception and design of study	Xiaopeng Ji, Yan Li, Peng Wen
Analysis and interpretation of data	Xiaopeng Ji
Drafting the manuscript	Xiaopeng Ji

Revising the manuscript critically for important intellectual content	Xiaopeng Ji, Yan Li, Peng Wen
Approval of the version of the manuscript to be published	Xiaopeng Ji, Yan Li, Peng Wen

4.2. Paper file

3DSleepNet: A Multi-Channel Bio-Signal Based Sleep Stages Classification Method Using Deep Learning

Xiaopeng Ji, Yan Li, and Peng Wen

***Abstract*— A novel multi-channel-based 3D convolutional neural network (3D-CNN) is proposed in this paper to classify sleep stages. Time domain features, frequency domain features, and time-frequency domain features are extracted from electroencephalography (EEG), electromyogram (EMG), and electrooculogram (EOG) channels and fed into the 3D-CNN model to classify sleep stages. Intrinsic connections among different bio-signals and different frequency bands in time series and time-frequency are learned by 3D convolutional layers, while the frequency relations are learned by 2D convolutional layers. Partial dot-product attention layers help this model find the most important channels and frequency bands in different sleep stages. A long short-term memory unit is added to learn the transition rules among neighboring epochs. Classification experiments were conducted using both ISRUC-S3 datasets and ISRUC-S1, sleep-disorder datasets. The experimental results showed that the overall accuracy achieved 0.832 and the F1-score and Cohen's kappa reached 0.814 and 0.783, respectively, on ISRUC-S3, which are a competitive classification performance with the state-of-the-art baselines. The overall accuracy, F1-score, and Cohen's kappa on ISRUC-S1 achieved 0.820, 0.797, and 0.768, respectively, which also demonstrate its generality on unhealthy subjects. Further experiments were conducted on ISRUC-S3 subset to evaluate its training time. The training time on 10 subjects from ISRUC-S3 with 8549 epochs is 4493s, which indicates its highest calculation speed compared with the existing high-performance graph convolutional networks and U^2−Net architecture algorithms.**

***Index Terms*— Deep learning, 3D convolutional networks, sleep stages classification.**

I. Introduction

SLEEP disorders, including insomnia, apnea, and circadian rhythm sleep disorders, are widespread in most populations. Sleep stages classification is the first step for sleep research and sleep disorder diagnosis. Polysomnograms (PSGs) are physiological signals, which are collected and analyzed by experts to explore brain activities during humans action [1], like measuring the depth of anesthesia [2], [3], identifying the motor imagery [4], seizure prediction [5], etc. Rechtschaffen and Kales sleep staging rules (R&K rules) [6] and American Academy of Sleep Medicine (AASM) standards [7] are golden criteria for experts to evaluate sleep quality through observing their bio-signals. The R&K rules divide sleep into six stages, namely awake (W), rapid eye movement (REM), and four non-REM stages (N1, N2, N3, and N4), but the AASM standards merged N3 and N4 into N3, called slow wave sleep (SWS). The exhaustive manual classification work is not only time-consuming and labor-consuming but also subjective depending on trained experts.

Various prior machine learning methods have been attempted to classify sleep stages automatically and efficiently. Traditional machine learning methods, including random forests, and support vector machines, have been investigated in sleep stages classification for decades, and many other classifiers also demonstrated their high performance in this task. Despite their success in sleep stages classification, shallow learning algorithms normally extract features according to experts' knowledge, which means that the classification effectiveness is limited by feature engineering and feature selections. Even though experts have extracted features from time domain [8], [9], frequency domain [10], [11], and time-frequency domain [12], [13], it is still a huge challenge to find new and effective features to improve the classification performance of traditional classifiers. Deep learning algorithms, including convolutional neural networks (CNNs) [14], recurrent neural networks (RNNs) [15], [16], deep belief networks (DBNs) [17], and graph convolutional networks (GCN) [18] have been proposed for this shortcoming.

Based on different data representations, the inputs of CNNs are normally one-dimensional signals [14] or two-dimensional signals [19], and the convolutional operations are with one-dimension and two-dimension data, respectively. These two types of models aggregate temporal information in time series, while they ignore the intrinsic connections among different bio-signals. GCN models have thus been proposed for the shortcoming. Unlike CNNs and RNNs that require Euclidean data, the inputs to GCNs are non-Euclidean structures, which means that the functional connections among

Manuscript received 22 December 2022; revised 23 May 2023; accepted 17 August 2023. Date of publication 28 August 2023; date of current version 7 September 2023. *(Corresponding author: Xiaopeng Ji.)*

This work involved human subjects from publicly available databases. The authors confirm that all human subject research procedures and protocols are exempted from review board approval.

Xiaopeng Ji and Yan Li are with the School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: xiaopeng.ji@usq.edu.au; yan.li@usq.edu.au).

Peng Wen is with the School of Engineering, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: paul.wen@usq.edu.au).

Digital Object Identifier 10.1109/TNSRE.2023.3309542

different channels and spatial correlations can be explored by GCN models automatically. However, due to their low computing efficiency, it would still have a long way to eventually apply GCNs for clinical diagnosis. Compared with CNNs and GCNs, 3D-CNNs not only can extract temporal features from raw data but also have the capacity to aggregate spatial information with less computing complexity. Transitional rules may also impact the final classification results, which means that they need to be considered during sleep stages identification to improve classification performance. However, few algorithms pay much attention to it, and sometimes this important factor is even ignored.

To tackle the challenges above, a 3D-CNN model based on a backbone from the spatial-spectral-temporal based attention 3D dense network [20] is proposed for automatic sleep staging. To explore relations among signals and frequency bands deeply, temporal features, frequency features, and temporal-frequency features are extracted and fed into this proposed model.

The main contributions of this paper are summarized as follows:

- A 3D-CNN and 2D-CNN mixed deep learning model named 3DSleepNet is proposed to classify sleep stages automatically. 3D convolutional operations are used to extract spatial-temporal features and spatial-spectral-temporal features from temporal inputs and temporal-frequency inputs, respectively. 2D convolutional operations are also utilized in the proposed model to extract spatial-spectral features from frequency inputs.
- A novel partial dot-product attention mechanism is designed for 3D convolutional operations to efficiently capture the most relevant information. A spatial-spectral attention mechanism is designed for 2D convolutional operations to capture the most relevant spatial-spectral information.
- To evaluate the classification performance on healthy and unhealthy subjects, the classification experiments were performed on ISRUC-S3 and 50 random subjects from ISRUC-S1 (https://sleeptight.isr.uc.pt/). The accuracy, F1-score, and Cohen's kappa on ISRUC-S3 are 0.832, 0.814, and 0.783, respectively, which indicates that the proposed model achieves a state-of-the-art performance. The overall accuracy, F1-score, and Cohen kappa on ISRUC-S1 (the datasets with sleep-disorder patients) achieved 0.820, 0.797, and 0.768, respectively, which also demonstrates its generality on unhealthy subjects. The training speed experiments on ISRUC-S3 show that the proposed model outperforms other GCN models and U^2-Net architecture models in terms of the model training time.
- The impact of the ratio of unhealthy and healthy subjects in the training set is explored using a set of mixed training data from ISRUC-S1 (unhealthy datasets) and ISRUC-S3 (healthy datasets). The experimental results show that the classification performance on unhealthy patients achieved the best when the training set consists of 100% abnormal patients.
- Incremental experiments are conducted on the ISRUC-S3 dataset to explore the effects of different model variants. The experimental results show that the proposed 3DSleepNet model achieves its best performance when the attention layers and a long short-term memory layer (LSTM) are added with all three input branches.

This paper is organized as follows: Section II introduces related works about sleep stages classification and 3D convolutional neural networks. In Section III, the details of the proposed models are illustrated. Experimental data, experimental setting, experimental results, and discussions are presented in Section IV. Finally, conclusions are drawn in Section V.

II. RELATED WORK

A. Sleep Stages Classification

Automatic sleep stages classification algorithms have been studied for decades and many shallow machine learning classifiers have been reported in sleep scoring. Support vector machine (SVM) is one of the most widely used classifiers in many classification tasks and has demonstrated its high performance in identifying sleep stages. Zhu et al. [21] extracted graph domain features through a visibility graph similarity method to perform a five-state classification based on single-channel EEG. Naive Bayes [22], random forest (RF) [23], [24], complex networks [25], [26], and ensemble learning-based classifiers [27], [28], [29] also gave acceptable classification results. However, all these research methods are based on hand-crafted features, the classification performance and efficiency depend on feature engineering and researchers' understanding of data.

Compared to other shallow machine learning algorithms, deep learning methods not only allow experts to input original/raw data but also can extract higher-level features. Motivated by their breakthroughs in many fields, such as image recognition [30], [31] and natural language processing [32], many researchers have applied deep learning algorithms to process bio-signals. DeepSleepNet [14] is a two-step training CNN model with an BiLSTM layer for sleep stages classification. A representation learning module helps this model to capture both temporal information and frequency information through two CNNs with small and large filter sizes for the first layers. A sequence residual learning module is used to learn stage transition rules. Ji et al. [18] proposed a multi-channel-based graph convolutional network to perform five-stage classification, which achieved the best performance on ISRUC-S3 and ISRUC-S1. U-Nets and U^2-Nets are novel complex models with a multi-scale extraction module, which also gave acceptable results in sleep scoring tasks [33], [34], [35].

B. 3D Convolutional Neural Networks

2D-CNNs are normally utilized to recognize images [36], [37], [38] through 2D convolutional operations, which means that spatial features can be extracted efficiently. Compared with 2D-CNNs, 3D-CNNs can capture spatial features and temporal features simultaneously. Motivated by this characteristic, many researchers turn to build 3D-CNNs from time series data, such as human action recognition [39] and pose estimation [40]. In the bio-signal analysis area, some 3D-CNN

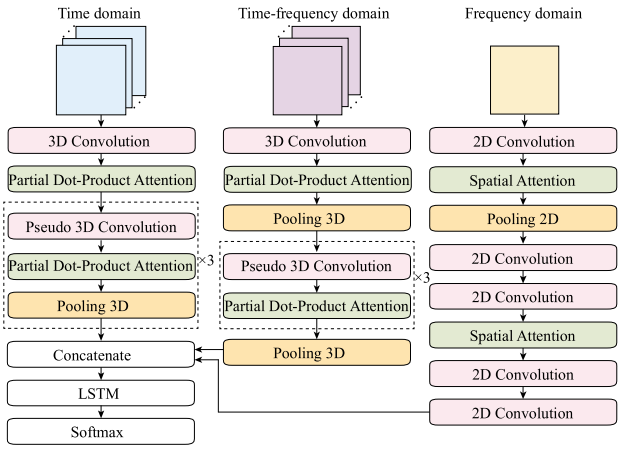

Fig. 1. The overall architecture of the 3DsleepNet. Time domain features, time-frequency domain features and frequency domain features are inputted into this model after feature extraction. Time domain features are down-sampled from raw/original signals. Time-frequency domain features come from short-term differential entropies and the frequency domain features come from power spectral densities.

models are also reported in emotion recognition [41], [42], [43], epileptic seizure prediction [44], and motor imagery analysis tasks [45], but there is still little study to apply 3D-CNNs for classifying sleep stages.

III. Methodology

Fig. 1 shows the overall architecture of the proposed model in this paper. The proposed model consists of one spatial-temporal stream, one spatial-spectral-temporal stream, and one spatial-spectral stream. The inputs of the spatial-temporal stream and spatial-spectral-temporal stream are 3D representations of raw/original signals in the time domain and time-frequency domain. The inputs of the spatial-spectral stream are 2D representations of raw signals in the frequency domain. In terms of the spatial-temporal stream and spatial-spectral-temporal stream, a partial dot-product attention mechanism is designed to help the proposed model to pay more attention to valuable information in the time series of each frequency band from each channel. A long-short term memory layer is added to learn transition rules among neighboring epochs for classification.

There are four key components in this proposed 3DSleepNet model: 1) A three-stream 3D-CNN model is designed for automatic sleep staging through time space, time-frequency space, and frequency space after extracting features from multi-channel bio-signals. 2) For each 3D-CNN stream, partial dot-product attention layers are added to help the proposed model to focus on more valuable information. 3) Pseudo 3D-CNN modules are used to decrease computing complexity. 4) A LSTM layer is added to learn transitional rules among neighboring epochs.

A. Feature Extraction and 3D Representation

Fig. 2 shows the procedure of feature extraction, where time domain features and time-frequency domain features are spatial-temporal and spatial-spectral-temporal 3D representations of bio-signals, respectively, and frequency domain features are 2D representations. The bio-signal of Nchannels are defined as $S = (s_1, s_2, \ldots, s_N) \in \mathbb{R}^{N\times L}$, where $s_i \in \mathbb{R}^L (i \in \{1, 2, \ldots, N\})$ is a channel of EEG, ECG or EMG signal with L data points. Before feature extraction, all channels to be used will be filtered by M bandpass filters and the filtered signals are defined as $S' = (s'_1, s'_2, \ldots, s'_N) \in \mathbb{R}^{N\times M\times L}$, where the $i-$th channel with M frequency band waves of L data points is represented by $s'_i \in \mathbb{R}^{M\times L} (i \in \{1, 2, \ldots, N\})$. In sleep stages classification tasks, $L-$length bio-signals are usually segmented into epochs of 30 seconds (s), and the filtered multi-channel signals in each epoch can be represented as $E = (e_1, e_2, \ldots, e_N) \in \mathbb{R}^{N\times M\times T}$, where e_i is the $i-$th channel with M frequency band waves of T data points in that epoch. To extract temporal features, the filtered signals E are down-sampled, and the length of signals changes from T to τ. The temporal features can be represented as $\chi_t = (x^1, x^2, \ldots, x^\tau) \in \mathbb{R}^{N\times M\times \tau}$, each $x^j \in \mathbb{R}^{N\times M} (j \in \{1, 2, \ldots, \tau\})$ is a 2D feature map of time step j. In terms of 3D representations of time-frequency features, the short-term differential entropy is calculated based on the filtered signals E. The spatial-spectral-temporal 3D representation of bio-signals is defined as $\chi_{tf} = (\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^{\hat{\tau}}) \in \mathbb{R}^{N\times M\times \hat{\tau}}$, where $\hat{x}^k \in \mathbb{R}^{N\times M} (k \in \{1, 2, \ldots, \hat{\tau}\})$ is a 2D differential entropy feature map of time step k. The 2D representation of frequency domain features are defined by the power spectral density of M frequency band waves from all Nchannels of that epoch.

B. 3D Convolution

The 3D convolution is achieved by convolving a 3D kernel to the cube formed by stacking multiple temporal-contiguous 2D feature maps together [39]. As a result, comparing with 1D temporal convolution and 2D spatial convolution, 3D temporal-spatial convolution is more advantageous in both representing brain connections and their activities. The convolutional value at (x, y, z) on the $j-$th feature map in the $i-$th layer is given by [39]:

$$v_{ij}^{xyz} = \sigma\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \quad (1)$$

where P_i, Q_i and R_i are the size of the 3D kernels along the three dimensions. w_{ijm}^{pqr} is the $(p, q, r)-$th value of the kernel connected to the $m-$th feature map in the previous layer, and σ is an activation function. Fig. 3 shows an example of the 3D convolution on the extracted 3D features. The kernel size along the time dimension in the current layer is 3, which means that each convolutional value is decided by the 3D kernel and 3 contiguous 2D feature maps. This characteristics of aggregating spatial information and temporal information help the proposed model to capture valuable features comprehensively.

C. Pseudo-3D Convolution

Even though 3D convolutions are more advantageous in exploring temporal features and spatial features, the requirements of a large amount of computing resources limit its

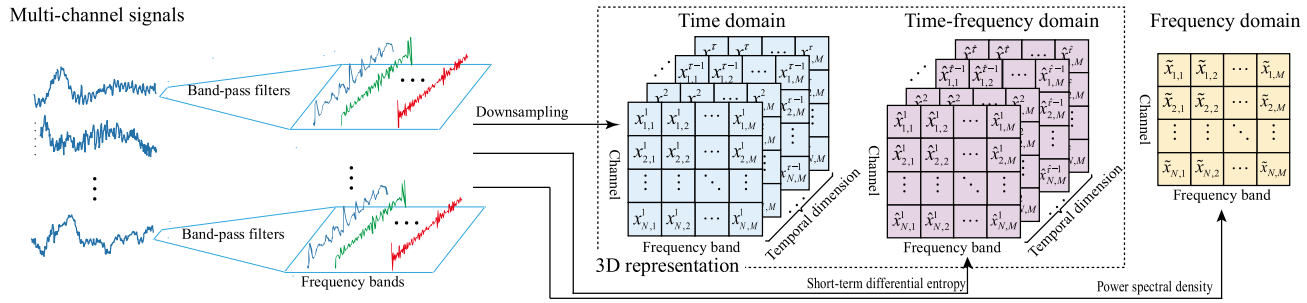

Fig. 2. Each channel is filtered by m filters to band waves. 3D temporal features are extracted through down-sampling from filtered band waves of n channel bio-signals. 3D temporal-frequency features are extracted by calculating the short-term differential entropy of each band waves from each channel. The frequency domain features are extracted by calculating the power spectral density of m band waves of n channels.

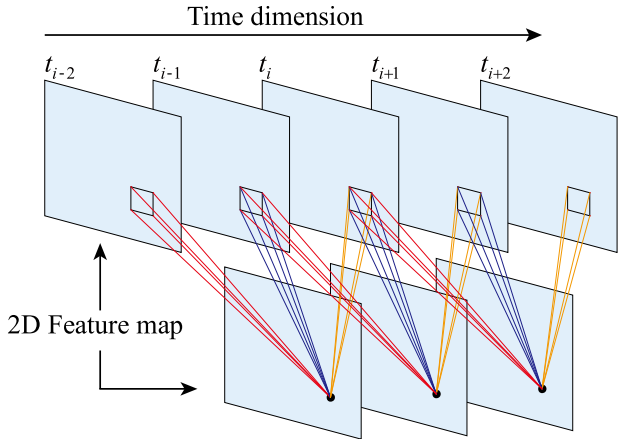

Fig. 3. An example of 3D convolution on the extracted 3D features. The kernel size is three in the time dimension.

applications. To tackle this problem, Pseudo-3D convolutions [46] are adopted in the proposed model to reduce the computational complexity. The kernel of the standard 3D convolutions are (P, Q, R), where P and Q can be seen as the kernel size of 2D spatial convolutions and R is the kernel size along the time dimension. In Pseudo-3D convolution, the kernel (P, Q, R) are decoupled into $P \times Q \times 1$ and $1 \times 1 \times R$, where $P \times Q \times 1$ represents convolutional filters equivalent to 2D CNN on spatial domain and $1 \times 1 \times R$ convolutional filters like 1D CNN tailored to the temporal domain. Hence, the output of a Pseudo-3D convolution module l can be defined as:

$$\chi^l = \Phi^{1\times1\times R}(\Phi^{P\times Q\times1}(\chi^{l-1})) \tag{2}$$

where χ^{l-1} is the output of $l-1$th layer, $\Phi^{P\times Q\times1}$ and $\Phi^{1\times1\times R}$ denote the 2D convolution on spatial domain with a kernel of $P \times Q$ and 1D convolution on temporal domain with a kernel of R, respectively.

D. Partial Dot-Product Attention

The attention mechanism is often utilized to automatically extract the most relevant information. In the proposed model, a simple but effective partial dot-product attention is designed to quantify the importance of input features, where higher weights are assigned to the most relevant information and lower weights are assigned to the less relevant information. For a given input $\chi \in \mathbb{R}^{N\times M\times T}$, the partial dot-product attention is computed as:

$$Att = \chi \otimes \sigma((\chi \cdot M_1) \cdot M_2 + b) \tag{3}$$

where $M_1 \in \mathbb{R}^{T\times M}$, $M_2 \in \mathbb{R}^{M\times T}$, $b \in \mathbb{R}^{N\times M\times T}$ are learnable parameters, $\cdot$ denotes dot-product, $\otimes$ refers to the point-wise multiplication, and σ is a softmax function.

IV. Experiments

A. Experimental Data and Experimental Settings

ISRUC-Sleep [47] is an open-source database, which consists of three subsets, namely, ISRUC-S1, ISRUC-S2, and ISRUC-S3. All signals were collected according to the international 10–20 standard electrode placement. The detailed information of subjects and the distribution of sleep stages are listed in TABLE I. Each recording consists of 19 channels, including, EOG, EEG, EMG, ECG, snore, body position, etc. The sampling rate of EOG, EEG, and EMG are 200Hz.

Classification experiments are conducted on ISRUC-S1 and ISRUC-S3 to evaluate the classification performance for both healthy cases and sleep-disorder patients and the evaluation measures, including accuracy (ACC), precision (PR), recall (RE), F1-score (F1), and Cohen's kappa (κ) are defined as below:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}\% \tag{4}$$

$$precision = \frac{TP}{TP + FP}\% \tag{5}$$

$$recall = \frac{TP}{TP + FN}\% \tag{6}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives and false negatives, respectively.

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \tag{7}$$

where $precision$ and $recall$ are defined as in equation (5) and equation (6).

$$\kappa = 1 - \frac{1 - Accuracy}{1 - p_e} \tag{8}$$

where $Accuracy$ is defined as question (4), and p_e is the hypothetical probability of chance agreement calculated:

$$p_e = \frac{1}{N^2}\sum_k n_{k1}n_{k2} \tag{9}$$

TABLE I
THE COMPREHENSIVE INFORMATION OF ISRUC DATASETS

Subset	Subject Number	Age	Sex ratio (Male: Female)	Sleep health conditions	Distribution of each sleep stage					
					Wake	N1	N2	N3	REM	Total
ISRUC-S1	100	51±16	55:45	Unhealthy	20098	11062	27511	17251	11265	87187
ISRUC-S2	8	46.8±18.8	6:2	Unhealthy	2282	2211	5042	2609	2063	14207
ISRUC-S3	10	40±10	9:1	Healthy	1651	1215	2609	2014	1060	8549

TABLE II
DETAILED INFORMATION OF CHANNELS USED IN EXPERIMENTS

Signal type	Label	Description
EOG	LOC-A2	Left eyes movements.
	ROC-A1	Right eyes movements.
EEG	F3-A2	A1 and A2 are placed in
	C3-A2	the left and right
	O1-A2	ear-lobes.
	F4-A1	
	C4-A1	
	O2-A1	
Chin EMG	X1	Placed between the chin and the lower lip.

where N is the total number of samples, and k is the number of categories, and n_{ki} is the number of times rater i predicted category k.

Incremental experiments and executing efficiency experiments are also carried out on ISRUC-S3 to explore the contribution of each module in the proposed model and computing complexity, respectively. Moreover, the ISRUC-S1 data are also used to test the generality of the proposed method on unhealthy patients. The code will be uploaded on Github (https://github.com/XiaopengJi-USQ/3DSleepNet) once the paper is published.

All these experiments are carried out on a computer with an Intel I9-12900KF CPU, 128 GB memory, and an Nvidia 3090 GPU.

B. Preprocessing and Feature Extraction

All PSGs were pre-processed by the data provider: 1) A notch filter was used to eliminate the 50 Hz electrical noise; 2) For EEG and EOG data, a bandpass Butterworth filter was utilized to obtain waves from 0.3 Hz to 35 Hz. The EMG data were filtered by a lower cutoff frequency of 10 Hz and a higher cutoff frequency of 70 Hz. 3) The last 30 epochs of each subject were removed due to noise.

The detailed information of channels used to extract features in all experiments is listed in TABLE II

In terms of time domain feature extraction, the features are obtained through down-sampling from 200 Hz to 10 Hz for each selected channel. According to our experiments, the excessive down-sampling in time series has little negative effects on the classification results, but the training time is substantially reduced. To extract time-frequency domain features, the differential entropy (STDE) of 9 crossed frequency bands: 0.5-4 Hz, 2-6 Hz, 4-8 Hz, 6-11 Hz, 8-14 Hz, 11-22 Hz, 14-31 Hz, 22-40 Hz and 31-49 Hz are calculated for each channel. These time-frequency features are obtained from filtered bio-signals with a 200 Hz sampling rate and the window size of the short-term differential entropy is set to 3s, which means that there are 10 feature maps in each epoch. The frequency domain features are obtained by computing the power spectral density of each epoch with 200 Hz sampling.

C. Comparison With the State-of-the-Art Methods

We compare the proposed model with traditional machine learning methods, including SVM [48], Random Forest [24], and Multilayer Perceptron neural network [16] on both ISRUC-S1 and ISRUC-S3 datasets. Ensemble classifiers, like Bootstrap aggregating (Bagging) [27], boosting [28], eXtreme Gradient Boosting (XGBoost) [29] are also evaluated on these two subsets. Multi-types deep learning algorithms with different architectures, such as CNNs [14], [49], U^2-Net [35], and GCNs [18], [50], [51] are included for comparisons purpose as well. For a fair comparison, all models were reproduced based on our hardware computational environments, except the JK-STGCN model, which we have all results in our previous work [18] and subject-independent cross-validation strategy were adopted on both of the subsets.

TABLE III shows the comparison results on ISRUC-S3 and random selected 50 unhealthy subjects from ISRUC-S1 subsets. 10-fold cross-validation and 25-fold cross-validation are carried out on ISRUC-S3 and 50 unhealthy subjects from ISRUC-S1, respectively, to test the classification performance on healthy subjects and unhealthy cases.

The classification performances of the SVM model and the RF model are the lowest among all these methodologies, even more features are extracted and selected by researchers. There are two reasons leading to this phenomenon. On the one hand, the inefficient extracted features cannot represent the original data comprehensively. On the other hand, the classification results are limited by the performance of the classifier itself, which are improved by ensemble methods.

Compared to shallow classifiers, 1D-dimension convolution models, including TinySleepNet [49] and DeepSleepNet [14], not only can input original data without feature engineering but also improve the classification performance. There are two reasons to explain this improvement. CNNs can extract high-level features and aggregate temporal information in continuous time series, which means that the correlation between previous and next data points can be learned comprehensively. As a result, the classification accuracy increases. Moreover, LSTM layers are added to learn transition rules, which help these models to further improve classification performance. However, the 1D-dimension convolution only can focus on one-dimension data, the spatial connections among

TABLE III
COMPARISON AMONG 3DSLEEPNET AND OTHER METHODS ON ISRUC S3 AND 50 RANDOM SUBJECTS FROM ISRUC S1 SUBGROUP

Subset		Method	Overall Metrics			Per class F1 score (F1)				
			ACC	F1	κ	W	N1	N2	N3	REM
ISRUC-S3	Alickovic et al.	SVM	0.714	0.672	0.626	0.824	0.428	0.724	0.815	0.569
	Memar et al.	RF	0.702	0.685	0.616	0.838	0.470	0.671	0.763	0.684
	Dong et al.	MLP+LSTM	0.751	0.708	0.675	0.852	0.378	0.758	0.872	0.682
	Hassan & Subasi	Bagging	0.740	0.706	0.662	0.847	0.465	0.751	0.843	0.625
	Hassan & Bhuiyan	Boosting	0.714	0.681	0.625	0.821	0.458	0.725	0.825	0.576
	Liu et al.	XGBoost	0.749	0.722	0.676	0.866	0.481	0.751	0.848	0.666
	Supratak et al.	CNN+BiLSTM	0.719	0.696	0.643	0.831	0.463	0.742	0.851	0.595
	Supratak & Guo	CNN+LSTM	0.753	0.737	0.682	0.809	0.533	0.758	0.851	0.734
	Jia et al.	U^2-Net	0.807	0.791	0.751	0.867	0.581	0.808	0.895	0.805
	Jia et al.	STGCN	0.786	0.770	0.724	0.864	0.540	0.782	0.869	0.793
	Jia et al.	MSTGCN	0.818	0.803	0.765	0.898	0.581	0.808	0.880	**0.848**
	Ji et al.	JK-STGCN	0.831	0.814	0.782	**0.900**	**0.598**	0.826	0.901	0.845
	This study	3D-CNN	**0.832**	**0.814**	**0.783**	0.896	0.596	**0.832**	**0.909**	0.838
ISRUC-S1	Alickovic et al	SVM	0.684	0.608	0.583	0.793	0.242	0.708	0.808	0.490
	Memar et al.	RF	0.699	0.649	0.607	0.841	0.307	0.705	0.750	0.640
	Dong et al.	MLP+LSTM	0.703	0.648	0.614	0.807	0.301	0.724	0.817	0.591
	Hassan & Subasi	Bagging	0.693	0.621	0.595	0.813	0.248	0.715	0.798	0.529
	Hassan & Bhuiyan	Boosting	0.663	0.614	0.555	0.789	0.325	0.687	0.766	0.504
	Liu et al.	XGBoost	0.736	0.691	0.657	0.866	0.372	0.742	0.835	0.638
	Supratak et al.	CNN+BiLSTM	0.730	0.691	0.654	0.850	0.385	0.739	0.830	0.648
	Supratak & Guo	CNN+LSTM	0.764	0.745	0.695	0.846	0.548	0.729	0.830	0.794
	Jia et al.	U^2-Net	0.816	**0.800**	0.764	0.903	**0.577**	0.801	**0.886**	0.832
	Jia et al.	STGCN	0.780	0.751	0.715	0.889	0.463	0.763	0.825	0.813
	Jia et al.	MSTGCN	0.808	0.787	0.752	0.885	0.539	0.799	0.876	0.838
	Ji et al.	JK-STGCN	0.820	0.798	0.767	0.895	0.550	**0.811**	0.883	0.850
	This study	3D-CNN	**0.820**	0.797	**0.768**	**0.908**	0.534	0.808	0.880	**0.855**

* W=awake, N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement, REM= rapid eye movement.

different brain regions are ignored consequently, which finally limits their performance. Even though pure 1D-dimension convolution models cannot reach a very high classification accuracy, U^2-Net-based models using 1D-dimension convolution layers still improve the performance slightly through their complex architecture and exponentially higher computational resources. GCN classifiers including a GraphSleepNet classifier, an MSTGCN classifier, and a JK-STGCN classifier, can extract spatial features and temporal features efficiently and this characteristic ensures their high performance in sleep stages classification tasks. Compared with the algorithms above, the 3DSleepNet model not only requires fewer crafted features than the traditional machine learning methods but also can capture spatial-temporal information more effectively than CNN, U^2-Net, and GCN models. The classification results on ISRUC-S3 and random subjects from ISRUC-S1 demonstrate that the 3DSleepNet model not only can classify sleep stages with high accuracy for healthy subjects but also have good generality on unhealthy cases.

Fig. 4 shows the confusion matrix of classifications results obtained from all compared models on ISRUC-S3. For each model, the performance of classifying stages of Wake, N2, N3, and REM have higher results than N1 stage, but some minor samples are still misclassified into other classes. Compared with multi-channel-based models, single-channel-based classifiers have lower REM accuracy. One explanation is that EOG channels help to classify REM stages. As a result, algorithms using single EEG channels without EOG signals fail to classify REM stages correctly. The N1 stage has the lowest classification results, and some samples are incorrectly classified into Wake, N2, and REM. Due to the fact that slow eye movements also make great contribution in classifying N1, all multi-channel-based models have better results for N1 than single-channel-based classifiers. TABLE IV shows the comparison results among several deep learning methods [14], [34], [35], [49], [52], [53], [54] on other public datasets [55], [56], [57]. Compared with these algorithms, the accuracy of the proposed model on ISRUC-S3 is 0.832 which stays similar level, but N1 stages and N3 stages of the 3DSleepNet model outperforms other models.

In practice, when the classification performance of the model is similar, the one taking a shorter training time will have the advantage of capturing the market. According to TABLE III, the 3DSleepNet model and the JK-STGCN model can achieve the top 2 classification accuracy on both ISRUC-S3 and ISRUC-S1, where the MSTGCN model and

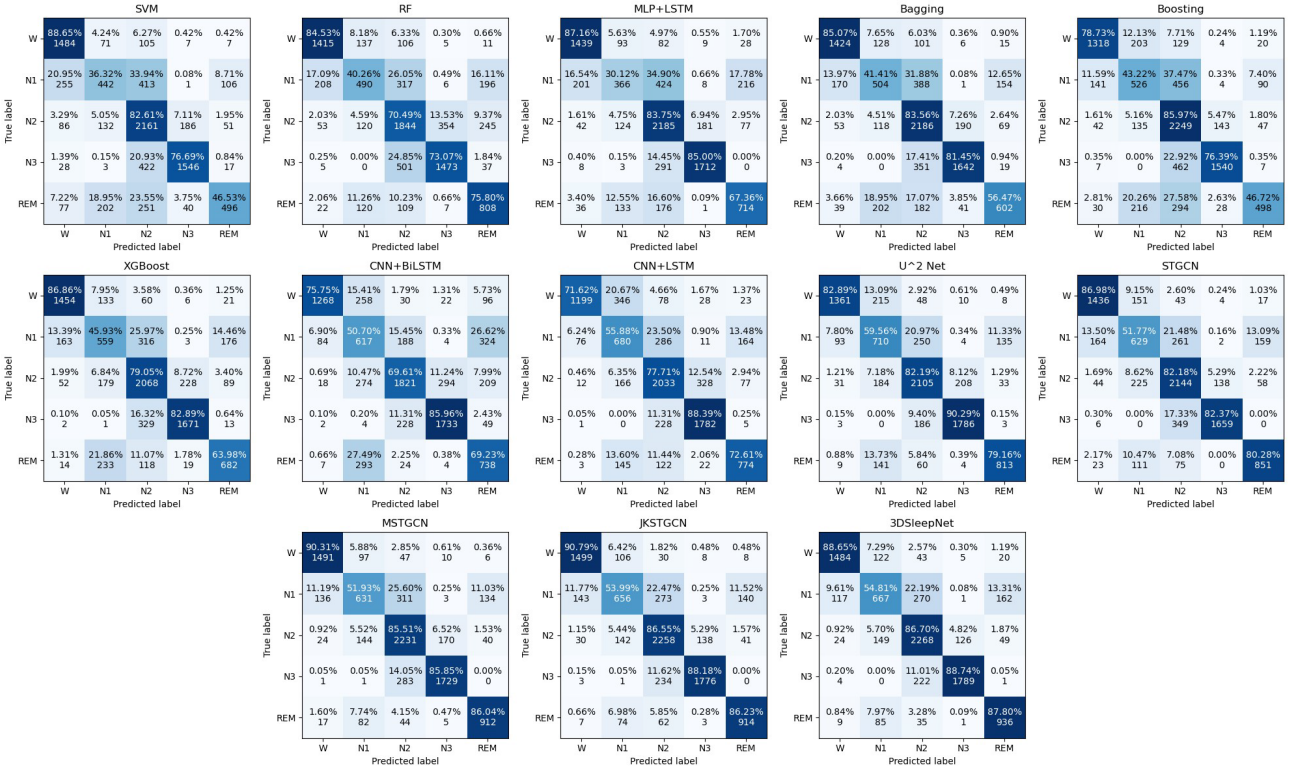

Fig. 4. Confusion matrix from all the compared models on ISRUC-S3 subset.

TABLE IV
COMPARISON RESULTS AMONG SEVERAL DEEP LEARNING METHODS BASED ON OTHER PUBLIC DATASETS

Methods	datasets	Overall Metrics			Per class F1 score (F1)				
		ACC	F1	κ	W	N1	N2	N3	REM
DeepSleepNet	Sleep-EDF-v1	0.820	0.769	0.760	0.847	0.466	0.859	0.848	0.824
SleepEEGNet	Sleep-EDF-v1	0.843	0.797	0.790	0.892	0.522	0.868	0.851	0.850
TinySleepNet	Sleep-EDF-v1	0.854	0.805	0.800	0.901	0.514	0.885	0.883	0.843
SeqSleepNet	Sleep-EDF-153	0.838	0.782	-	0.928	0.489	0.854	0.786	0.851
SleepUtime	Sleep-EDF-153	-	0.76	-	0.920	0.510	0.840	0.750	0.800
SalientSleepNet	Sleep-EDF-153	0.841	0.795	-	0.933	0.542	0.858	0.783	0.858
CNN+GRU	SHHS1-700	0.832	-	0.760	0.897	0.311	0.850	0.781	0.808
This study	ISRUC-S3	0.832	0.814	0.783	0.896	0.596	0.832	0.909	0.838

* W=awake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement.

the SalientSleepNet model also perform very well on ISRUC-S3 or ISRUC-S1 subsets. Therefore, these four models are selected to compare the training efficiency. Considering the architecture and parameters of all these four models are totally different, in order to fairly compare execution efficiency, parameters including batch size, learning epochs, etc. are set the same as those in [18], [35], and [51]. Fig. 5 shows the training time comparison among the top 3 models on ISRUC-S3 or ISRUC-S1. The MSTGCN and JK-STGCN models take the least time to train the classification, but their feature extraction steps take more time, which leads to a higher overall training time compared with the proposed model. Due to its complex architecture, the SalientNet model takes the longest time to complete the training tasks.

D. Model Analysis

According to TABLE III, the proposed 3DsleepNet model can have a good capacity to identify sleep stages for healthy and unhealthy subjects/patients. But it also implies that there are still two important factors, namely the proportion of unhealthy patients and the size of training sets, which can affect the generality on unhealthy patients. To explore the influence of these two factors, two more experiments are carried out on the ISRUC-S3 and ISRUC-S1 datasets.

1) Influence of the Ratio of the Unhealthy Patients in the Training Set: For a fair comparison of the influence of the ratio of unhealthy patients in the training set, all comparison experiments are conducted on a fixed testing set, which consists of ten random subjects from the ISRUC-S1 dataset. The training set initially consists of 10 healthy subjects from the ISRUC-S3 dataset, and the ratio of unhealthy patients is changed by replacing a random healthy subject in it with a random data sample from the remaining 90 patients of the ISRUC-S1 dataset.

It can be seen from Fig. 6 that the classification performance is very low at the beginning when all data in the

training set are healthy samples. The accuracy, F1-score and Cohen's kappa only achieve 0.53, 0.47, and 0.38, respectively. However, the classification accuracy increases noticeably and achieves 0.73 when there are 10% unhealthy samples. The F1-score and Cohen's kappa results are also improved a lot and achieve 0.68, and 0.64, respectively. After a steady rising, the classification performance trends keep improving with a slow growth rate, even though there are two decreasing points. Finally, when all data samples are from sleep-disorder cases, the classification performance reaches the highest, with the accuracy, F1-score, and Cohen's kappa are 0.78, 0.73, and 0.69, respectively. Since the distribution of sleep stages from ISRUC-S1 and ISRUC-S3 is totally different, it is reasonable and acceptable that the performance is very low at the beginning, when the model learns normal bio-signals and transitional rules from healthy subjects and has little knowledge about abnormal features. However, it starts to recognize the abnormal features or patterns when an unhealthy subject is added to the training set, and this leads to the improvement of the classification accuracy. The trend of the slow growth of accuracy keeps until all training data are from unhealthy patients.

Under the same size of the training set, the classification results with unhealthy patients are lower than those from healthy subjects. This phenomenon is also caused by the different distribution of sleep stages from ISRUC-S1 and ISRUC-S3, which means that the transitional rules and signal characteristics are quantitatively and complexly higher than those of healthy cases. As a result, models cannot learn features comprehensively with a small dataset, which leads to lower classification performance.

2) Influence of Training Data Size: To explore whether an increased dataset size may increase the classification performance on sleep-disorder cases, an extra experiment is carried out.

The ISRUC-S1 dataset is divided into four disjoint subgroups, and each subgroup contains 10, 20, 30, and 40 patients, respectively. For each subgroup, the Leave-One-Out cross-validation is conducted to validate the influence of the train data size on the proposed model.

Fig. 7 shows the experimental results on the effects of the train set. The classification accuracy, F1-score, and Cohen's kappa on unhealthy patients are 0.79, 0.75, and 0.73, respectively, which are the lowest. However, the classification measurements are improved when the size of the training set increases and finally achieves 0.82, 0.81, and 0.76, respectively.

This experiment also demonstrates that a big training data set size can weaken the negative impacts on classification results from abnormal bio-signals and abnormal transitional rules.

3) Incremental Comparison Experiments: To further investigate the influence of each module used in the proposed model, eight variant models are designed and evaluated using the ISRUC-S3 database. The details are described below:

1. *variant a (basic model)*: The basic model is a one-branch 3D-CNN model of a temporal stream without any attention layer or LSTM layer.

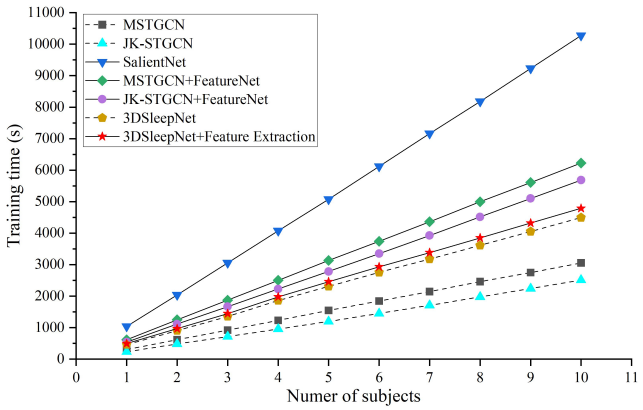

Fig. 5. Training time for the top 3 models on ISRUC-S3 or ISRUC-S1 based on 10-fold training.

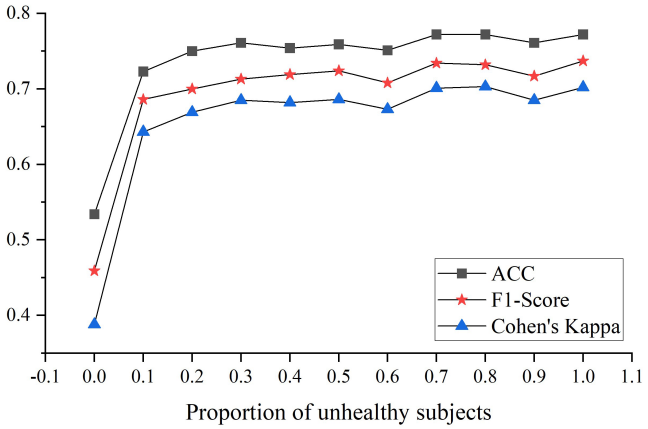

Fig. 6. The trend of classification results with different ratio of unhealthy patients in the training data set.

2. *variant b (basic model + temporal-frequency stream):* A temporal-frequency stream is added to construct a two-stream 3D-CNN model without any attention layer or LSTM layer.

3. *variant c (variant b + frequency stream):* A frequency stream is added to *variant b* to construct a three-stream 3D-CNN model without any attention layer or LSTM layer

4. *variant d (variant c + partial dot-product attention):* A partial dot-product attention is added to quantify the importance of input features.

5. *variant e (variant d + LSTM):* A LSTM layer is added to learn transitional rules among neighboring epochs.

6. *variant f (variant c + spatial-temporal attention + LSTM):* The partial dot-product attention layer is replaced by a spatial-temporal attention layer to indicate the importance of different channels and different sleep epochs. An LSTM layer is added to learn transitional rules among neighboring epochs.

7. *variant g (variant c + self-attention + LSTM):* The partial dot-product attention layer is replaced by a self-attention layer to indicate the interdependence within input features.

8. *variant h (using EEG channels only + the complete model e):* Features are extracted from six EEG signals to test the importance of the complementary channels to the overall performance.

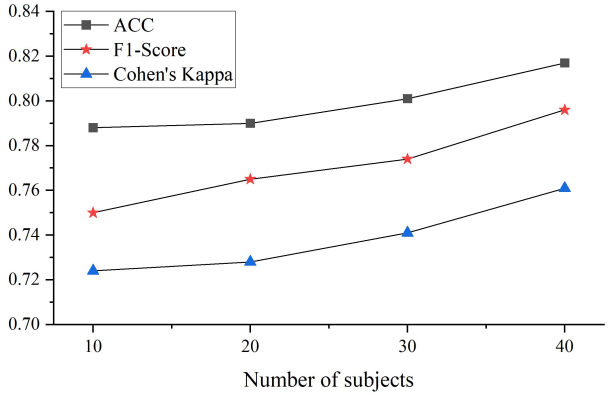

Fig. 7. The classification results of 3DSleepNet on four disjoint groups from ISRUC-S1.

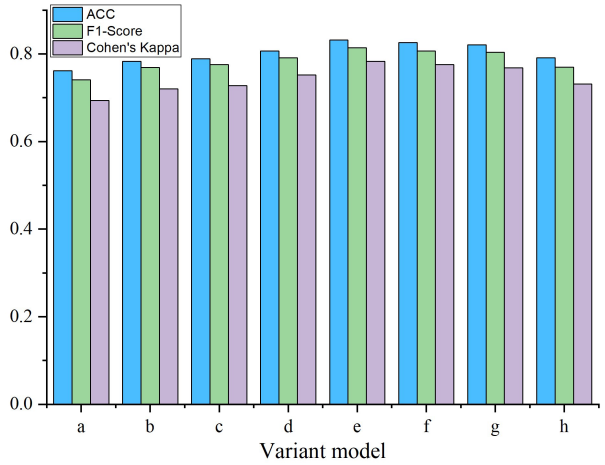

Fig. 8. Comparison of the designed variant models.

Fig. 8 illustrates the classification performance of eight variant models on ISRUC-S3. The basic model which inputs the 3D temporal features has the lowest performance, but all measurements increase when time-frequency features and frequency features are fed into the 3D-CNN model. The reason for this improvement is that the 3D-CNN model only learns the correlation among signals and frequency bands in time series but lacks the knowledge of time-frequency and frequency of signals which plays an important role in sleep stages identification. As a result, the classification accuracy is not very high at the beginning but increases with the added time-frequency features and frequency features. The performance further improves when partial dot-product attention layers and an LSTM layer are added. In terms of the partial dot-product attention layers, they indicate the most important frequency bands in each channel at each epoch and this helps to capture the most relevant information. The LSTM layer helps the proposed model learn the transitional rules among neighboring epochs, which also improves the classification performance. The spatial-temporal attention mechanism shrinks the input features through one dimension which loses more information than the partial dot-product attention mechanism and this leads to the ineffective quantifying of the importance of input features. The self-attention pays more attention role to the correlation among inner elements but the excessive down-sampling in time series weakens the connection among inner elements, which makes it underperform in this sleep stages classification task. To further evaluate the contribution of different channels in the classification performance, the 3DSleepNet model with six EEG channels inputs is tested. Compared to the complete model with all channels, the accuracy, F1-score, and Cohen's kappa of the model only using EEG signals, achieve 0.791, 0.770 and 0.731, respectively, which decrease heavily from 0.832, 0.814 and 0.783, respectively. Lacking EMG and EOG inputs leads to the result that the classifier fails to classify REM and N1 correctly, and these misclassifications also decline the performance of other stages.

V. Conclusion

In this study, a 3D-CNN based sleep stages classification model named as 3DSleepNet is proposed. The 3DSleepNet consists of two 3D-CNN streams and one 2D-CNN stream. The inputs of 3D-CNN streams are time domain features, and time-frequency domain features, and the inputs of the 2D-CNN stream are frequency domain features. For 3D-CNN branches, Pseudo-3D convolution layers are utilized to decrease the computing complexity and partial dot-product attention layers are designed to help the proposed model pay attention to valuable information. After the fusion layer of three streams, an LSTM layer is used to learn the transitional rules among neighboring epochs. Compared with the best results reported by other models, the 3D-CNN model also can achieve competitive performance on both healthy and unhealthy datasets with less computational demand. Based on the classification results, two more factors that may impact the performance are further explored. The experimental results indicate that the poor classification performance on unhealthy cases, which are caused by abnormal bio-signals, can be improved limitedly by increasing the proportion of sleep-disorder patients in the training set or increasing the number of the training data. An incremental experiment is also conducted to identify the contributions of each model variant and several different attention layers are tested to find the best one. A limitation of the proposed model is that the multi-channel-based algorithm requires large storage and memory for computation. However with modern computer hardware technology, this shouldn't be a main problem for its applications and deployment on edge artificial intelligence devices. In the future, we will improve and explore a new 3D representation of a single-channel signal, such as one EEG channel, so that the storage requirements and computing complexity can be further decreased.

References

[1] S. Siuly and Y. Li, "Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach," *Neural Comput. Appl.*, vol. 26, no. 4, pp. 799–811, May 2015, doi: 10.1007/s00521-014-1753-3.

[2] T. Nguyen-Ky, P. Wen, and Y. Li, "Consciousness and depth of anesthesia assessment based on Bayesian analysis of EEG signals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1488–1498, Jun. 2013, doi: 10.1109/TBME.2012.2236649.

[3] T. Nguyen-Ky, P. Wen, Y. Li, and R. Gray, "Measuring and reflecting depth of anesthesia using wavelet and power spectral density," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 4, pp. 630–639, Jul. 2011, doi: 10.1109/TITB.2011.2155081.

[4] N. A. Siuly, Y. Li, and P. Wen, "Identification of motor imagery tasks through CC-LR algorithm in brain computer interface," *Int. J. Bioinf. Res. Appl.*, vol. 9, no. 2, p. 156, 2013, doi: 10.1504/IJBRA.2013.052447.

[5] M. Li, W. Chen, and T. Zhang, "Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM," *Biocybern. Biomed. Eng.*, vol. 36, no. 4, pp. 708–718, 2016, doi: 10.1016/j.bbe.2016.07.004.

[6] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Arch. Gen. Psychiatry*, vol. 20, no. 2, pp. 246–247, Feb. 1969, doi: 10.1001/archpsyc.1969.01740140118016.

[7] 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, Amer. Acad. Sleep Med., Darien, IL, USA, 2012, vol. 176.

[8] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017, doi: 10.1007/s00521-017-2919-6.

[9] M. Diykh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016, doi: 10.1109/TNSRE.2016.2552539.

[10] A. Stochholm, K. Mikkelsen, and P. Kidmose, "Automatic sleep stage classification using ear-EEG," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 4751–4754.

[11] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods," *Biomed. Signal Process. Control*, vol. 2, no. 3, pp. 171–179, Jul. 2007, doi: 10.1016/j.bspc.2007.05.005.

[12] 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, doi: 10.1007/s10439-015-1444-y.

[13] W. Al-Salman, Y. Li, and P. Wen, "Detection of EEG K-complexes using fractal dimension of time frequency images technique coupled with undirected graph features," *Frontiers Neuroinform.*, vol. 13, p. 45, Jun. 2019.

[14] 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, doi: 10.1109/TNSRE.2017.2721116.

[15] E. Bresch, U. Großekathöfer, and G. Garcia-Molina, "Recurrent deep neural networks for real-time sleep stage classification from single channel EEG," *Frontiers Comput. Neurosci.*, vol. 12, p. 85, Oct. 2018, doi: 10.3389/fncom.2018.00085.

[16] 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, doi: 10.1109/TNSRE.2017.2733220.

[17] I. N. Yulita, M. I. Fanany, and A. M. Arymuthy, "Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification," *Proc. Comput. Sci.*, vol. 116, pp. 530–538, Jan. 2017, doi: 10.1016/j.procs.2017.10.042.

[18] X. Ji, Y. Li, and P. Wen, "Jumping knowledge based spatial–temporal graph convolutional networks for automatic sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1464–1472, 2022, doi: 10.1109/TNSRE.2022.3176004.

[19] 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, doi: 10.1109/TNSRE.2018.2813138.

[20] Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou, and J. Wang, "SST-EmotionNet: Spatial–spectral–temporal based attention 3D dense network for EEG emotion recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2909–2917.

[21] 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 Inform.*, vol. 18, no. 6, pp. 1813–1821, Nov. 2014, doi: 10.1109/JBHI.2014.2303991.

[22] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016, doi: 10.1016/j.bspc.2015.09.002.

[23] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, Oct. 2012, doi: 10.1016/j.cmpb.2011.11.005.

[24] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 84–95, Jan. 2018, doi: 10.1109/TNSRE.2017.2776149.

[25] M. Diykh, Y. Li, and S. Abdulla, "EEG sleep stages identification based on weighted undirected complex networks," *Comput. Methods Programs Biomed.*, vol. 184, Feb. 2020, Art. no. 105116, doi: 10.1016/j.cmpb.2019.105116.

[26] M. Diykh and Y. Li, "Complex networks approach for EEG signal sleep stages classification," *Expert Syst. Appl.*, vol. 63, pp. 241–248, Nov. 2016, doi: 10.1016/j.eswa.2016.07.004.

[27] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017, doi: 10.1016/j.knosys.2017.05.005.

[28] A. R. Hassan and M. I. H. Bhuiyan, "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting," *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.015.

[29] C. Liu et al., "Automatic sleep staging with a single-channel EEG based on ensemble empirical mode decomposition," *Phys. A, Stat. Mech. Appl.*, vol. 567, Apr. 2021, Art. no. 125685, doi: 10.1016/j.physa.2020.125685.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[32] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 1–11. Accessed: May 19, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[33] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. J. Jennum, and C. Igel, "U-sleep: Resilient high-frequency sleep staging," *NPJ Digit. Med.*, vol. 4, no. 1, Apr. 2021, Art. no. 72, doi: 10.1038/s41746-021-00440-5.

[34] 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," Oct. 2019, *arXiv:1910.11162*.

[35] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," May 2021, *arXiv:2105.13864*.

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2014, *arXiv:1409.1556*. Accessed: Jul. 20, 2022.

[37] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 169–175, doi: 10.1109/CCWC.2018.8301729.

[38] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *Proc. 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, Dec. 2018, pp. 278–282, doi: 10.1109/ICSCCC.2018.8703316.

[39] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013, doi: 10.1109/TPAMI.2012.59.

[40] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "Real-time 3D hand pose estimation with 3D convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 956–970, Apr. 2019, doi: 10.1109/TPAMI.2018.2827052.

[41] Y. Wang, Z. Huang, B. McCane, and P. Neo, "EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489715.

[42] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, 2018.

[43] Y. Zhao, J. Yang, J. Lin, D. Yu, and X. Cao, "A 3D convolutional neural network for emotion recognition based on EEG signals," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–6.

[44] Z. Wang, J. Yang, and M. Sawan, "A novel multi-scale dilated 3D CNN for epileptic seizure prediction," in *Proc. IEEE 3rd Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Jun. 2021, pp. 1–4, doi: 10.1109/AICAS51828.2021.9458571.

[45] T. Liu and D. Yang, "A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding," *Brain Sci.*, vol. 11, no. 2, p. 197, Feb. 2021, doi: 10.3390/brainsci11020197.

[46] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5534–5542, doi: 10.1109/ICCV.2017.590.

[47] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-sleep: A comprehensive public dataset for sleep researchers," *Comput. Methods Programs Biomed.*, vol. 124, pp. 180–192, Feb. 2016, doi: 10.1016/j.cmpb.2015.10.013.

[48] 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, doi: 10.1109/TIM.2018.2799059.

[49] 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, doi: 10.1109/EMBC44109.2020.9176741.

[50] Z. Jia et al., "GraphSleepNet: Adaptive spatial–temporal graph convolutional networks for sleep stage classification," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 1324–1330.

[51] Z. Jia et al., "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.

[52] 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, doi: 10.1371/journal.pone.0216456.

[53] 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, doi: 10.1109/TNSRE.2019.2896659.

[54] X. Shao and C. S. Kim, "A hybrid deep learning scheme for multi-channel sleep stage classification," *Comput., Mater. Continua*, vol. 71, no. 1, pp. 889–905, 2022, doi: 10.32604/cmc.2022.021830.

[55] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.

[56] S. F. Quan et al., "The sleep heart health study: Design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, Dec. 1997.

[57] G.-Q. Zhang et al., "The national sleep research resource: Towards a sleep data commons," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, Oct. 2018, doi: 10.1093/jamia/ocy064.

CHAPTER 5: Paper 3 - MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model

5.1. Overview of Paper 3

The details of Paper 3 are given below:

- Paper title: “MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model.”
- Paper length: 13 pages
- Journal: Computer Methods and Programs in Biomedicine
  - Rank: Q1 (Computer science application)
  - Impact score: 6.1 (2021-2022)
  - Cite Score: 10.10 (2022)
  - SJR: 1.12 (2022)
- First author: Xiaopeng Ji
- Corresponding author: Xiaopeng Ji

HDR thesis author’s declaration

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

TABLE 5.1: Authorship contributions of Paper 3

Conception and design of study	Xiaopeng Ji, Yan Li, Peng Wen
Analysis and interpretation of data	Xiaopeng Ji
Drafting the manuscript	Xiaopeng Ji
Revising the manuscript critically for important intellectual content	Xiaopeng Ji, Yan Li, Peng Wen, Rajendra Acharya, Prabal Barua
Approval of the version of the manuscript to be published	Xiaopeng Ji, Yan Li, Peng Wen, Prabal Barua, U Rajendra Acharya

5.2. Paper file

MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model

Xiaopeng Ji, Yan Li, Peng Wen, Prabal Barua, U Rajendra Acharya

ABSTRACT

Background and Objective: Sleep staging is an essential step for sleep disorder diagnosis, which is time-intensive and laborious for experts to perform this work manually. Automatic sleep stage classification methods not only alleviate experts from these demanding tasks but also enhance the accuracy and efficiency of the classification process.

Methods: A novel multi-channel biosignal-based model constructed by the combination of a 3D convolutional operation and a graph convolutional operation is proposed for the automated sleep stages using various physiological signals. Both the 3D convolution and graph convolution can aggregate information from neighboring brain areas, which helps to learn intrinsic connections from the biosignals. Electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG) and electrocardiogram (ECG) signals are employed to extract time domain and frequency domain features. Subsequently, these signals are input to the 3D convolutional and graph convolutional branches, respectively. The 3D convolution branch can explore the correlations between multi-channel signals and multi-band waves in each channel in the time series, while the graph convolution branch can explore the connections between each channel and each frequency band. In this work, we have developed the proposed multi-channel convolution combined sleep stage classification model (MixSleepNet) using ISRUC datasets (Subgroup 3 and 50 random samples from Subgroup 1).

Results: Based on the first expert's label, our generated MixSleepNet yielded an accuracy, F1-score and Cohen kappa scores of 0.830, 0.821 and 0.782, respectively for ISRUC-S3. It obtained accuracy, F1-score and Cohen kappa scores of 0.812, 0.786, and 0.756, respectively for the ISRUC-S1 dataset. In accordance with the evaluations conducted by the second expert, the comprehensive accuracies, F1-scores, and Cohen kappa coefficients for the ISRUC-S3 and ISRUC-S1 datasets are determined to be 0.837, 0.820, 0.789, and 0.829, 0.791, 0.775, respectively.

Conclusion: The results of the performance metrics by the proposed method are much better than those from all the compared models. Additional experiments were carried out on the ISRUC-S3 sub-dataset to evaluate the contributions of each module towards the classification performance.

Keywords: 3D convolutional networks, graph convolutional networks, sleep stage classification

1. INTRODUCTION

High-quality sleep helps human beings to rejuvenate and relieve fatigue, while low-quality sleep affects their physical and mental health [1]. Most populations suffer from sleep disorders, including insomnia, obstructive sleep apnea, and disruptions in circadian rhythm synchronization [2]. Sleep scoring is a critical approach to identifying problems related to sleep rhythm disruption [3]. To classify sleep stages, experts collect biosignals by placing electrodes in different locations on human heads and analyzing those signals. Electroencephalography (EEG), electromyography (EMG), and electrooculography (EOG) constitute the principal polysomnographic (PSG) methodologies for the examination of cerebral activities [4]. According to the American Academy of Sleep Medicine (AASM) standards [5], the process of visual inspection necessitates experts to categorize every sleep epoch into five distinct stages, specifically, wakefulness (W), rapid eye movement (REM), and three non-rapid eye movement (NREM) stages denoted as N1, N2, and N3 [6]–[8]. Although manual sleep scoring can help the experts to analyze and diagnose sleep-related problems effectively, the high workload of sleep stages identification and subjectivity of experts limits their clinical applications. Therefore, semi-automatic and automatic sleep staging systems have increasingly drawn attention [9]–[11].

Many researchers have employed machine learning techniques to analyze PSGs [12]–[15]. Shallow classifiers, including decision trees [16], support vector machines [17]–[19], and random forests [20], etc. have demonstrated satisfactory classification performance. However, one of the drawbacks of these shallow classifiers is that the feature extraction process is required, whose classification performance depends on the selected features and feature selection algorithms. Although researchers have tried to extract features from multiple perspectives [21], compared with deep learning methods, there is still a lot of space to improve. Deep

Manuscript received on January 1, 2023; revised January 1, 2023; accepted on January 1, 2023. Date of publication January 1, 2023; date of current version January 1, 2023. Corresponding author: Xiaopeng Ji.

Xiaopeng Ji, Yan Li and Rajendra Acharya are with the School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: xiaopeng.ji@usq.edu.au; yan.li@usq.edu.au; rajendra.acharya@usq.edu.au).

Peng Wen is with the School of Engineering, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: paul.wen@usq.edu.au).

Prabal Barua is with the Cogninet Brain Team, Sydney, NSW 2010, Australia (e-mail: prabal.barua@usq.edu.au)

learning models possess the capability to directly extract intricate features from raw data, leading to significant advancements in sleep stage classification [22]–[24]. Convolutional neural networks (CNN) [25], [26], including 1D-CNN [27]–[29], 2D-CNN [30], [31] have shown promising results in sleep stage classification tasks. Even though CNNs can learn features from time series in each epoch, the correlations among neighboring epochs can be easily ignored. To address this limitation, recurrent neural networks (RNNs) [32], particularly long short-term memory (LSTM) networks, have been introduced. [33]–[35]. The LSTM module has both long-term and short-term memories of inputs, which means that the transition rules of epochs can be learned comprehensively, while the intrinsic connection among different channels remains unexplored. Unlike CNNs or RNNs, the inputs of graph convolutional networks (GCNs) are non-Euclidean structures, and this characteristic allows GCNs to learn the functional connections and spatial connections among brain areas [36]–[38]. Even though GCNs can extract spatial features effectively, their high computing complexity limits their applications in times series, whose density of data points is higher than other tasks. As a result, all existing GCN models in sleep stage classifications use frequency features or extracted high-level features from CNNs as inputs. In contrast, 3D convolutional neural networks (3D-CNNs) can extract intricate spatial features and temporal features simultaneously, allowing them to investigate the correlations of channels in the temporal dimension, while the correlations of the frequency domain are easily ignored.

1.1 Related work

Automatic sleep staging has advanced rapidly in recent years and the classification algorithms can be divided into *three* categories, namely, rule-based algorithms, traditional machine learning models and deep learning methods. The rule-based algorithms classify sleep epochs into five or six categories according to the AASM standards or the R&K rules, respectively. Experts focus on the characteristics of waves, like amplitude ranges, wavelet power spectrum ranges and spectral coefficient ratio and establish some rules to classify sleep stages [39]. However, because of the insufficient extracted features, the classification accuracy is quite low and unacceptable. Compared with rule-based algorithms, machine learning methods have significantly improved classification performances. Support vector machines (SVMs) stand out as one of the extensively utilized classifiers across various classification tasks, exhibiting notable efficacy in the identification of sleep stages. For instance, Zhu et al [40], extracted features from the graph domain using a visibility graph similarity technique. This approach was employed to achieve a five-state classification based on single-channel EEG data. Ignacio et al [17] used multitapers with a convolution method to extract time-frequency features from two EEG channels for the general features and muscle movement features, extracted from two EEG channels and two EOG channels as the supplementary features. The extracted features were fed into a support vector machine classifier with a quadratic equation for the final classification. Other forms of shallow machine learning approaches, like Naive Bayes [41], random forest (RF) [20], [42], complex networks [43], and ensemble learning-based classifiers [44], have similarly demonstrated credible classification outcomes. Although experts have tried to find more comprehensive features from the time domain [45], [46], frequency domain [47]–[50], time-frequency domain [51]–[54], or even the graph domain, there is still a lot of space to improve the classification performance.

Due to their achievements in many fields, such as image recognition and natural language processing, several deep learning algorithms have been reported in biosignals processing. DeepSleepNet [33] is a CNN model with two branches, where a larger filter captures the frequency information and a smaller filter captures the temporal information. The two-step training model is first trained with a pre-training on shuffled balanced data and then it is fine-tuned with imbalanced data for the final classification. In the training process, a Bi-LSTM layer is employed to acquire an understanding of the transition patterns among adjacent epochs. Akara and Yike [34] designed a more efficient CNN model named TinySleepNet based on the DeepSleepNet model. In the representation learning part, raw signals are fed into a single branch of several convolutional layers instead of two branches as in the DeepSleepNet, and the Bi-LSTM is replaced by a LSTM to learn the transition rules. Compared with CNNs, GCNs are more advantageous in representing brain connections and their activities. Jia et al [36] designed a GCN model named GraphSleepNet for sleep stage classification tasks. This model incorporates both a spatial attention layer and a temporal attention layer, facilitating the capture of significant spatial information from each channel and crucial temporal information from adjacent epochs. The multi-view spatial-temporal graph convolutional networks [37] improve the performance of GraphSleepNet through domain generalization. The jumping knowledge-based spatial-temporal graph convolutional networks [38] further improve the classification accuracy and execution efficiency through the jumping knowledge module. However, a limitation of GCNs is that they take a lot of computational resources than CNNs. Therefore, the existing GCNs do not work on temporal data directly, where a feature extraction step is required to transfer temporal data to the frequency domain or extract temporal data first. The U^2-Net model proposed in [55] incorporates a multi-scale extraction module and demonstrates satisfactory performance in sleep stage classification tasks.

1.2 Contributions

In this study, we proposed a combined model that utilizes graph convolutional networks (GCN) and 3D-CNN to address the challenges of automatic sleep stage classification. Unlike pure GCNs or 3D-CNN models, the proposed model captures not only spatial features but also frequency features and temporal features from the GCN and 3D-CNN branches, respectively.

The principal contributions of this paper are delineated as follows:

- A novel deep learning model that integrates GCN and 3D-CNN is introduced for the task of automated sleep stage classification. The differential entropy, a frequency domain feature, is extracted for the graph convolution branch to explore

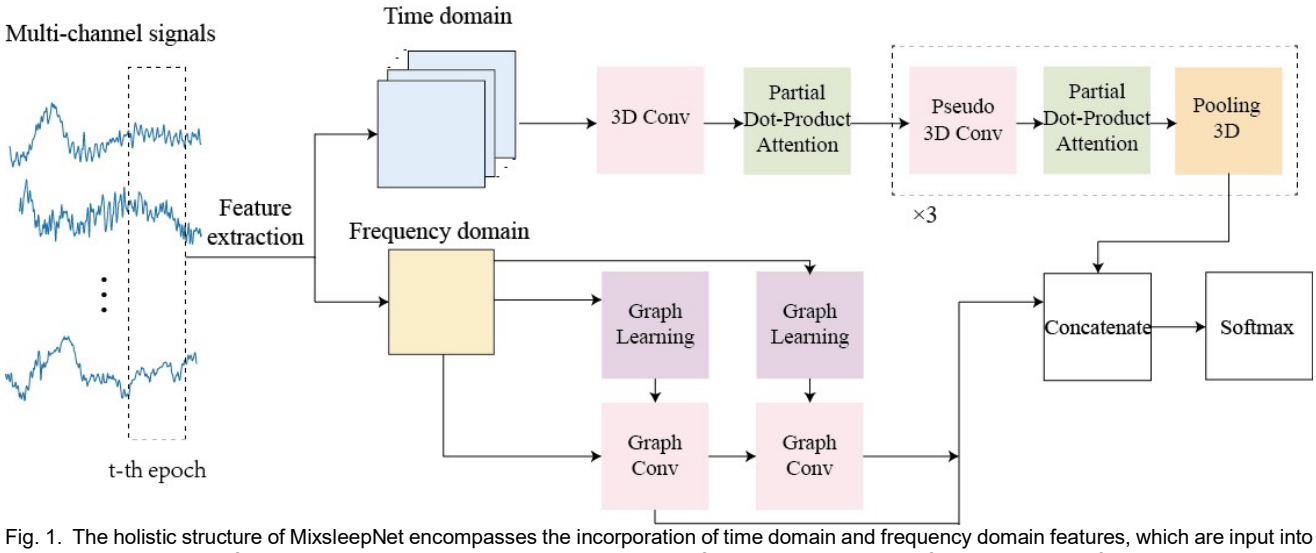

Fig. 1. The holistic structure of MixsleepNet encompasses the incorporation of time domain and frequency domain features, which are input into the model. Time domain features are acquired through the down-sampling of the initial signals, while frequency domain features are derived via differential entropy calculations.

the correlation between frequency bands and channels in the spatial dimension. Additionally, the time domain feature is extracted from down-sampled time series and fed into the 3D convolution branch to investigate the correlation between frequency bands and channels in the temporal dimension.

- Classification experiments were conducted on two datasets, namely, ISRUC-S3 and 50 randomly selected subjects from ISRUC-S1 (https://sleeptight.isr.uc.pt/) to evaluate the classification performances. The obtained results indicate that the proposed model attains a cutting-edge performance when the first expert's labels are used, with an accuracy, F1-score, and Cohen's kappa of 0.830, 0.821, and 0.782, on ISRUC-S3, respectively; and 0.813, 0.787, and 0.757, on ISRUC-S1, respectively. On the other hand, based on the second expert's labels, the proposed model achieves an accuracy, F1-score, and Cohen's kappa of 0.837, 0.820, and 0.789 on ISRUC-S3, and 0.829, 0.791, and 0.775 on 50 randomly selected subjects from ISRUC-S1, which are outperformed all the compared models [18], [33], [34], [36]–[38], [42], [55].
- To delve deeper into the individual contributions of each module within the proposed model, a series of incremental experiments were executed using the ISRUC-S3 dataset. The experimental results indicate that when the graph convolutional branch and 3D convolutional branch are added, the model outperformed other variations. Furthermore, the incorporation of partial-dot attention layers into the 3D convolutional branch leads to the attainment of the highest performance by the proposed model.

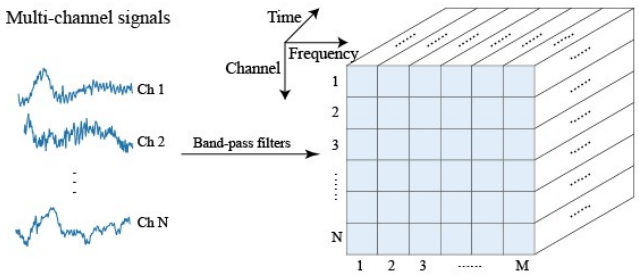

Fig. 2. For each of the N channels, M filters are applied to extract pertinent band waves. Subsequently, 3D temporal features are extracted from the filtered band waves by means of down-sampling.

The subsequent sections of this paper are structured as follows: In Section II, an exposition of shallow classifiers and deep learning models employed in sleep stage classification tasks is presented. Section III outlines the comprehensive architecture of the proposed model. The dataset adopted for this study is introduced in Section IV, which also encompasses the delineation of the experimental setup, resultant findings, and model analysis. Ultimately, Section V encapsulates the drawn conclusions.

2. Methods

Fig. 1 depicts the overall structure of the proposed MixSleepNet model. The MixSleepNet model consists of two branches one is a graph convolutional module with the inputs of frequency domain features and another branch is a 3D-CNN convolution with the inputs of time domain features. The graph convolutional module aims to uncover the relations among different channels in the frequency domain, whereas the 3D-CNN module aims to capture the effects of different signals in the time series.

2.1 Feature extraction

Inputs originating from the frequency domain are represented in 2D, while inputs stemming from the time domain take the form of spatial-temporal 3D representations of biosignals. Fig. 2 illustrates the generation of the spatial-temporal 3D representation of multi-channel bio-signals. Original bio-signals of N channels can be denoted by $S = (s_1, s_2, \ldots, s_N) \in \mathbb{R}^{N \times L}$, where $s_i \in \mathbb{R}^L (i \in \{1, 2, \ldots, N\})$ is the i-th channel with L data points in total. For each channel, M bandpass filters are used to filter one channel signal into M frequency band waves. As a result, N channel signals filtered by M filters with L data points are defined as $S' = (s'_1, s'_2, \ldots, s'_N) \in \mathbb{R}^{N \times M \times L}$, where

$s'_i \in \mathbb{R}^{M\times L}(i \in \{1,2,\ldots,N\})$ the i-th filtered channel. An epoch containing N channel signals can be defined as $E=(e_1,e_2,\ldots,e_N)\in\mathbb{R}^{N\times M\times T}$. The e_i is the i-th channel consists of M frequency bands of T data points in that epoch. Temporal features are extracted on an epoch-by-epoch basis, which are down-sampled from E. The new length of down-sampled signals is denoted by τ. Therefore, the 3D representation of the t-th epoch with τ data points can be $\chi_t=(x^1,x^2,\ldots,x^\tau)\in\mathbb{R}^{N\times M\times\tau}$. Frequency features are also extracted from E, where differential entropy is calculated for each frequency band in each channel [36].

2.2 3D convolution

The process of 3D convolution is executed by convolving a 3D kernel with the cube formed through the stacking of numerous temporally contiguous 2D feature maps. Compared with 1D-CNN and 2D-CNN, which only focus on temporal information or multi-dimensional temporal information, 3D-CNN can capture brain connections and their activities by simultaneously aggregating spatial information and multi-dimensional temporal information. Let P_i, Q_i and R_i be the size of the 3D kernels along the three dimensions and (x,y,z) be the position of convolutional to be calculated on the j-th feature map in the i-th layer. The convolutional value v_{ij}^{xyz} can be calculated by [56]:

$$v_{ij}^{xyz}=\sigma\left(b_{ij}+\sum_m\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1}w_{ijm}^{pqr}v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \tag{1}$$

where w_{ijm}^{pqr} is the (p,q,r)-th value of the kernel connected to the m-th feature map in the previous layer, and σ is an activation function.

2.3 Pseudo-3D convolution

Pseudo-3D convolutional operation [57] takes inspiration from Residual Networks and aims to reduce the computational complexity by splitting a 3D convolution operation into two separate convolution operations. Therefore, a standard 3D convolutional kernel defined by (P,Q,R) is decoupled into a 2D convolutional kernel with the size of $P\times Q\times 1$ and a 1D convolutional kernel with the size of $1\times 1\times R$. Let $\Phi^{P\times Q\times 1}$ and $\Phi^{1\times 1\times R}$ be a 2D spatial convolution and a 1D temporal convolution, respectively. The output of the l-th Pseudo-3D convolution layer is obtained by:

$$O^l=\Phi^{1\times 1\times R}(\Phi^{P\times Q\times 1}(O^{l-1})) \tag{2}$$

where O^{l-1} is the input of the l-th layer.

2.4 Partial Dot-Product Attention

Motivated by the success of the attention mechanism in time series problems, a straightforward yet impactful attention layer named partial dot-product attention is devised to capture the most important information from temporal inputs. Let $\chi\in\mathbb{R}^{N\times M\times T}$ be the input of the attention layer, the partial dot-product attention is defined as

$$Att=\chi\otimes\sigma((\chi\cdot M_1)\cdot M_2+b) \tag{3}$$

where $M_1\in\mathbb{R}^{T\times M}$, $M_2\in\mathbb{R}^{M\times T}$, $b\in\mathbb{R}^{N\times M\times T}$ are learnable parameters, the sign $\cdot$ denotes the inner product, $\otimes$ refers to the element-wise multiplication, and σ is a softmax function.

2.5 Adaptive Graph Learning

Adaptive graph reflects the dynamic connections and activities among brain areas, which makes a great contribution to improving the classification performance in sleep staging. Within the graph convolutional branch, two distinct adaptive graph learning approaches are employed to create dynamic adjacency matrices for the JK-Net module. The first adaptive graph learning is based on brain functions [36]. Let E_{ij} be the edge between electrodes, i and j, in a brain graph and it can be obtained by:

$$\begin{aligned}E_{ij}&=g(x_i,x_j)\\&=\frac{\exp(\mathrm{ReLU}(\boldsymbol{\omega}^T|x_i-x_j|))}{\sum_{j=1}^N\exp(\mathrm{ReLU}(\boldsymbol{\omega}^T|x_i-x_j|))}\end{aligned} \tag{4}$$

where x_i and x_j are extracted features from channel i and channel j and the activation function ReLu keeps E_{ij} non-negative. The learnable parameters $\boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_F)^T\in\mathbb{R}^{F\times 1}$ are iteratively updated by minimizing the ensuing loss function:

$$\mathcal{L}_{\text{adaptive\_graph}}=\sum_{i,j=1}^N\left\|x_i-x_j\right\|_2^2A_{ij}+\lambda\|A\|_F^2 \tag{5}$$

where $\lambda\geq 0$ is a regularization parameter. The other graph learning is obtained through simplifying temporal-information-based graph learning in [38], and it can be obtained by:

$$A=X\cdot W \tag{6}$$

where $X=(x_1,x_2,\ldots,x_N)\in\mathbb{R}^{N\times F}$ is the feature matrix with N channels and $x_j\in\mathbb{R}^F(j\in\{1,2,\ldots,N\})$ is the features extracted from channel j. W is a learnable parameter set, which is updated by minimizing the overall loss function:

$$\mathcal{L}_{\text{overall\_loss}}=\mathcal{L}_{\text{cross\_entropy}}+\mathcal{L}_{\text{adaptive\_graph}}+\beta\|A\|^2 \tag{7}$$

where the parameter β represents the strength of L2 regularization applied to the adjacency matrix A. $\mathcal{L}_{\text{adaptive\_graph}}$ denotes the loss function defined in equation (5) and $\mathcal{L}_{\text{cross\_entropy}}$ is the cross-entropy loss function.

2.6 Jumping Knowledge Graph Convolution

The JK-Net [38] was motivated by the ResNet model [58], which adds residual modules to enhance the overall performance. The output of the l-th JK-Net layer can be obtained by:

$$\mathrm{O}^l=\sigma(\mathrm{G}_\theta(\chi^{(l-1)}))+\sigma'(G'_{\theta'}(\chi^{(l-2)})) \tag{8}$$

where G_θ and $G'_{\theta'}$ signify the results of graph convolutional operations with kernels θ applied to the output of the $l-1$-th layer and θ' applied to the output of the $l-2$-th layer, respectively. The adaptive graphs employed in these graph convolutional operations stem from equations (4) and (6) correspondingly.

TABLE I
DETAILED CHANNEL INFORMATION OF ISURC.

Signal type	Label	ButterWorth
EOG	LOC-A2	0.3-35 Hz
	ROC-A1	
EEG	F3-A2	0.3-35 Hz
	C3-A2	
	O1-A2	
	F4-A1	
	C4-A1	
	O2-A1	
Chin EMG	X1	10-70Hz
ECG	X2	-

TABLE II
HYPERPARAMETERS USED IN OUT MIXSLEEPNET MODEL

Parameter	Value
Temporal input dim	(10, 9, 300)
Frequency input dim	(10, 9)
Pseudo 3D Conv.kernel size	(3, 3, 1) & (1, 1, 3)
3D Conv.kernel size	(3, 3, 3)
Filter size	25, 50, 100, 200
3D pooling kernel size	(2, 2, 2)
Filter size	25, 50, 100, 200
Layers of graph convolution	2
Order of Chebyshev polynomials	7
Number of training epochs	30
Batch size	16
Optimizer	Adam
Learn rate	0.0001
Dropout	0.5

3 Experiments and results

3.1 Materials and experimental settings

The evaluation experiments are carried out on ISRUC-S1 and ISRUC-S3, which are subsets from the ISRUC-Sleep database [59]. The complete database encompasses three distinct subsets: ISRUC-S1, ISRUC-S2, and ISRUC-S3. These subsets comprise 100 subjects (55 male and 45 female), 8 subjects (6 male and 2 female), and 10 subjects (9 male and 1 female), respectively. All polysomnograms (PSGs), encompassing signals such as EOG, EEG, EMG, ECG, snore, and body position, were acquired through non-invasive means, adhering to the international 10-20 standard electrode placement. The collected data were pre-processed by the data provider, which means that all EOG, EEG, EMG, and ECG signals underwent filtration with a 50 Hz notch filter to eradicate electrical noise and 30 epochs at the tail of each channel were removed to reduce noise. Furthermore, different bandpass butter filters were applied to clear useless noise in channels based on the types of bio-signals. The detailed cutoff of frequency and description of the used channels are listed in TABLE I.

In this study, we assessed the classification performance by calculating several evaluation metrics, including *Accuracy* (ACC), $F1$ -score (F1), Cohen's kappa (κ) and confusion matrix based on the values of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The specific parameter configurations of the proposed model for all datasets can be found in TABLE II. All experiments were conducted on a computer equipped with an Intel I9-12900KF CPU, 128 GB of memory, and an Nvidia 3090 GPU. The software environment was tailored to meet the specifications of each compared model. The source code will

TABLE III
OVERALL RESULTS AMONG MIXSLEEPNET AND OTHER METHODS ON ISRUC-S3 AND 50 RANDOM SUBJECTS FROM SISRUC-S1 SUBGROUP (EXPERT 1)

Subset	Model	Parameter	Overall Metrics			Per-class F1-score (F1)				
			ACC	F1	κ	W	N1	N2	N3	REM
ISRUC-S3	SVM [18]	<0.1 M	0.714	0.672	0.626	0.824	0.428	0.724	0.815	0.569
	RF [42]	<0.1 M	0.702	0.685	0.616	0.838	0.470	0.671	0.763	0.684
	DeepSleepNet [33]	21 M	0.719	0.696	0.643	0.831	0.463	0.742	0.851	0.595
	TinySleepNet [34]	1.3 M	0.753	0.737	0.682	0.809	0.533	0.758	0.851	0.734
	SalientSleepNet [55]	0.9 M	0.807	0.791	0.751	0.867	0.581	0.808	0.895	0.805
	GraphSleepNet [36]	-	0.786	0.770	0.724	0.864	0.540	0.782	0.869	0.793
	MSTGCN [37]	0.4 +1.5 M	0.818	0.803	0.765	0.898	0.581	0.808	0.880	0.848
	JK-STGCN [38]	-	**0.831**	0.814	0.782	**0.900**	0.598	**0.826**	**0.901**	0.845
	This work	**2.4 M**	0.830	**0.821**	**0.782**	0.899	**0.625**	0.819	0.899	**0.860**
ISRUC-S1	SVM [18]	<0.1 M	0.684	0.608	0.583	0.793	0.242	0.708	0.808	0.490
	RF [42]	<0.1 M	0.699	0.649	0.607	0.841	0.307	0.705	0.750	0.640
	DeepSleepNet [33]	21 M	0.730	0.691	0.654	0.850	0.385	0.739	0.830	0.648
	TinySleepNet [34]	1.3 M	0.764	0.745	0.695	0.846	0.548	0.729	0.830	0.794
	SalientSleepNet [55]	0.9 M	0.816	**0.800**	0.764	0.903	0.577	0.801	0.886	0.832
	GraphSleepNet [36]	-	0.780	0.751	0.715	0.889	0.463	0.763	0.825	0.813
	MSTGCN [37]	0.4 + 1.5 M	0.808	0.787	0.752	0.885	0.539	0.799	0.876	0.838
	JK-STGCN [38]	-	**0.820**	0.798	**0.767**	0.895	0.550	0.811	0.883	0.850
	This work	**2.4 M**	0.813	0.787	0.757	0.908	0.512	0.799	0.871	0.844

* W=awake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement. PD = patient dependent.

TABLE IV
OVEARLL RESULTS AMONG MIXSLEEPNET AND OTHER METHODS ON ISRUC-S3 AND 50 RANDOM SUBJECTS FROM SISRUC-S1 SUBGROUP (EXPERT 2)

Subset	Model	Overall Metrics			Per-class F1-score (F1)				
		ACC	F1	κ	W	N1	N2	N3	REM
ISRUC-S3	SVM [18]	0.714	0.666	0.624	0.820	0.375	0.739	0.819	0.578
	RF [42]	0.709	0.693	0.623	0.837	0.475	0.681	0.762	0.708
	DeepSleepNet [33]	0.724	0.693	0.645	0.842	0.422	0.759	0.853	0.590
	TinySleepNet [34]	0.778	0.760	0.714	0.792	0.519	0.798	0.891	0.799
	SalientSleepNet [55]	0.810	0.792	0.756	0.895	0.566	0.802	0.896	0.802
	GraphSleepNet [36]	0.809	0.795	0.754	0.881	0.565	0.797	0.876	0.853
	MSTGCN [37]	0.831	0.813	0.781	0.893	0.585	0.821	0.891	**0.876**
	JK-STGCN [38]	0.833	0.819	0.784	**0.897**	0.617	0.824	0.896	0.859
	This work	**0.838**	**0.820**	**0.790**	0.891	**0.620**	**0.833**	**0.904**	0.853
ISRUC-S1	SVM [18]	0.693	0.583	0.585	0.789	0.08	0.719	0.809	0.515
	RF [42]	0.702	0.651	0.610	0.841	0.307	0.708	0.752	0.645
	DeepSleepNet [33]	0.742	0.693	0.663	0.860	0.364	0.749	0.815	0.679
	TinySleepNet [34]	0.788	0.759	0.725	0.895	0.498	0.775	0.843	0.786
	SalientSleepNet [55]	0.811	0.783	0.755	0.900	0.515	0.802	0.867	0.830
	GraphSleepNet [36]	0.792	0.740	0.725	0.888	0.356	0.784	0.815	0.857
	MSTGCN [37]	0.825	**0.796**	0.770	0.891	0.525	0.819	0.878	0.868
	JK-STGCN [38]	0.827	0.793	0.773	0.898	**0.537**	0.819	0.869	**0.868**
	This work	**0.829**	0.791	**0.775**	**0.903**	0.482	**0.826**	**0.878**	0.864

* W=awake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement.

be made available on GitHub (https://github.com/XiaopengJi-USQ) upon publication of the paper.

3.2 Feature extraction

Since different signals have different roles in sleep classification tasks. As a result, signal selections are seriously considered. For example, EEGs are the most crucial channels to be analyzed and the correlations are influenced by brain activities, thus all the six EEG channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1 and O2-A1) are selected for feature extraction. EOGs, including left eye movements (LOC-A2) and right eye movements (ROC-A1), help to distinguish REM and non-REM. One EMG channel (Chin-EMG) and one ECG channel are also selected to help classify sleep stages.

All the 10 original signals are filtered by 9 crossed frequency bands: 0.5-4 Hz, 2-6 Hz, 4-8 Hz, 6-11 Hz, 8-14 Hz, 11-22 Hz, 14-31 Hz, 22-40 Hz and 31-49 Hz. These filtered band waves are down-sampled from 200 Hz to 10 Hz to obtain time domain features. The frequency domain features are obtained by calculating the differential entropy of the 9 mentioned frequency bands above for each channel in every epoch. This down-sampling of time series results in the loss of high-frequency components [60]. Nonetheless, the incorporation of frequency domain features offsets this limitation. Our experiments additionally reveal that this approach has minimal adverse impact on the classification outcomes, while significantly reducing the training duration.

3.3 Experimental results

Drawing on labels provided by two experts, we conducted a comparative analysis of the proposed model's performance against several baseline models featuring diverse architectures. This assessment was carried out on both the ISRUC-S3 dataset and a subset of 50 subjects chosen at random from ISRUC-S1. All experiments conducted on these two subsets employed subject-independent validation. This validation involved a 10-fold cross-validation for ISRUC-S3 and a 25-fold cross-validation for ISRUC-S1.

Based on the first expert's labeling, the 10-fold cross-validation on ISRUC-S3 achieved an overall accuracy of 0.830, an F1-score of 0.821, and a Cohen kappa of 0.782. Additionally, experiment on ISRUC-S1 yielded an overall accuracy of 0.812, an F1-score of 0.786, and a Cohen kappa of 0.756.

When considering the second expert's labeling, the overall accuracy, F1-score, and Cohen kappa for ISRUC-S3 and ISRUC-S1 are as follows: ISRUC-S3 - 0.837, 0.820, 0.789, and ISRUC-S1 - 0.829, 0.791, 0.775, respectively

4 Discussion

4.1 Comparison with State-of-the-Art Methods

The comparison results, presented in TABLE III and TABLE IV, evince that shallow classifiers reliant on feature engineering can classify most samples into correct categories. Nevertheless, despite the extraction of hundreds of features from multiple perspectives, numerous samples remain misclassified. One plausible explanation for this phenomenon is that the constrained classification performance can be attributed to inadequacies in the extracted features.

TABLE V
CONFUSION MATRIX OF COMPARED MODELS OBTAINED FROM 10-FOLD VALIDATION ON ISRUC-S3 DATASET

Model	Labels	Prediction					Per-class Metrics	
		W	N1	N2	N3	REM	PR	RE
SVM [18]	W	1484	71	105	7	7	0.769	0.886
	N1	255	442	413	1	106	0.520	0.363
	N2	86	132	2161	186	51	0.645	0.826
	N3	28	3	422	1546	17	0.869	0.767
	REM	77	202	251	40	496	0.733	0.465
RF [42]	W	1415	137	106	5	11	0.831	0.845
	N1	208	490	317	6	196	0.565	0.403
	N2	53	120	1844	354	245	0.641	0.705
	N3	5	0	501	1473	37	0.798	0.731
	REM	22	120	109	7	808	0.623	0.758
DeepSleepNet [33]	W	1268	258	30	22	96	0.920	0.757
	N1	84	617	188	4	324	0.427	0.507
	N2	18	274	1821	294	209	0.795	0.696
	N3	2	4	228	1733	49	0.842	0.860
	REM	7	293	24	4	738	0.521	0.692
TinySleepNet [34]	W	1199	346	78	28	23	0.929	0.716
	N1	76	680	286	11	164	0.509	0.559
	N2	12	166	2033	328	77	0.740	0.777
	N3	1	0	228	1782	5	0.821	0.884
	REM	3	145	122	22	774	0.742	0.726
SalientSleepNet [55]	W	1361	215	48	10	8	0.909	0.829
	N1	93	710	250	4	135	0.568	0.596
	N2	31	184	2105	208	33	0.795	0.822
	N3	3	0	186	1786	3	0.888	0.903
	REM	9	141	60	4	813	0.820	0.792
GraphSleepNet [36]	W	1436	151	43	4	17	0.858	0.870
	N1	164	629	261	2	159	0.564	0.518
	N2	44	225	2144	138	58	0.747	0.822
	N3	6	0	349	1659	0	0.920	0.824
	REM	23	111	75	0	851	0.784	0.803
MSTGCN [37]	W	1491	97	47	10	6	0.893	0.903
	N1	136	631	311	3	134	0.661	0.519
	N2	24	144	2231	170	40	0.765	0.855
	N3	1	1	283	1729	0	0.902	0.858
	REM	17	82	44	5	912	0.835	0.860
JKSTGCN [38]	W	1499	106	30	8	8	0.891	0.908
	N1	143	656	273	3	140	0.670	0.540
	N2	30	142	2258	138	41	0.790	0.865
	N3	3	1	234	1776	0	0.921	0.882
	REM	7	74	62	3	914	0.829	0.862
This work	W	1490	143	30	8	3	0.908	0.890
	N1	118	783	212	2	102	0.608	0.643
	N2	16	245	2156	168	31	0.815	0.824
	N3	6	1	217	1791	1	0.910	0.888
	REM	11	116	31	0	908	0.869	0.852

* W=awake. N1, N2 and N3 are sleep stage 1, 2, 3, separately, and are non-rapid eye movement. REM= rapid eye movement. PR=precision and RE=recall.

Compared with traditional deep learning algorithms, deep learning methods have made significant progress in improving classification accuracy. The reasons for this improvement may vary depending on the technology used in these deep leaning models. 1D-CNNs can aggregate temporal information from neighboring data points, allowing for the exploration of correlations in the time series of an epoch at a deeper level. Therefore, CNN models can achieve superior results even when raw data are inputted. Furthermore, the use of the Bi-LSTM or LSTM layers enables the CNNs to learn transition rules among neighboring epochs, thereby further improving their performance. As a result, CNN models, such as DeepSleepNet and TinySleepNet improve in accuracy by 0.01 and 0.06 on the ISRUC-S3 subgroup. However, a notable limitation of 1D-CNNs is that they only concentrate on data in the temporal dimension, whereas the spatial dimension, specifically the interconnections among brain activity, may be easily overlooked due to the limitation of filter dimensions. Although pure 1D convolution models might not attain exceedingly high classification accuracies, models based on the U-Net architecture that employ 1D convolution layers manage to achieve slight performance enhancements. This improvement is attributed to their intricate architecture, which is composed of multiple nested U-units to detect the salient waves. The multi-scale extraction module for transition rule learning is another reason leading to this improvement. GCNs, like GraphSleepNet, MSTGCN and JK-STGCN further enhance the classification performance. One reason to explore this phenomenon is that GCNs extract spatial features from graph convolutional layers and temporal features from standard convolutional layers, allowing these models to learn brain activities and sleep transition rules. Compared to these

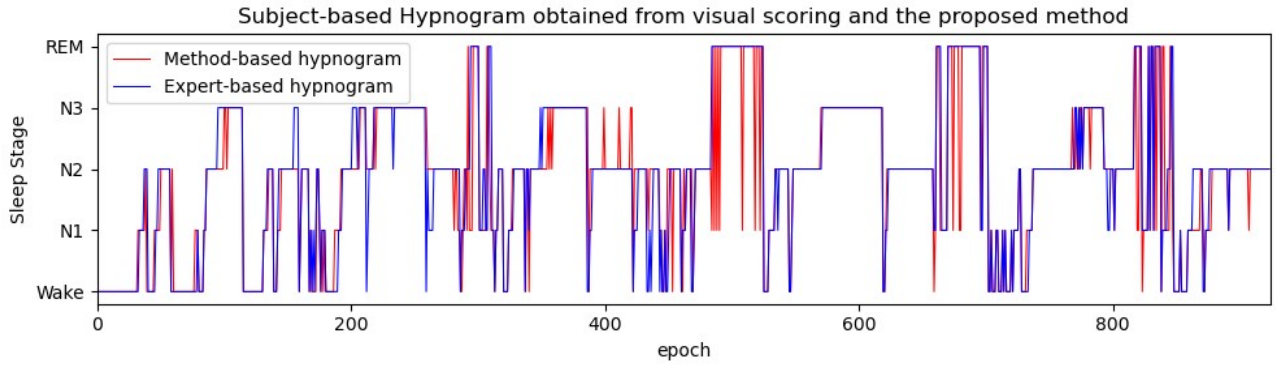

Fig. 3 Comparison between manual score annotated by expert 1 (blue) and automated scoring generated by the proposed model (red) for subject one from the ISRUC-S3 dataset.

algorithms, the MixSleepNet model outperforms all classifiers based on the second expert's labels and achieves similar performance based on the first expert's labels. The proposed model extracts spatial features from both graph convolutional layers and 3D convolutional layers, and temporal features are extracted through 3D convolutional layers. As a result, the MixSleepNet model achieves an acceptable result that is no worse than pure GCN models. Furthermore, the classification outcomes observed for both ISRUC-S3 and ISRUC-S1 underscore that the proposed model excels in classifying sleep stages with high accuracy, exhibiting strong performance across both healthy subjects and cases involving health issues.

TABLE V illustrates the confusion matrix displaying the classification results of all models evaluated on ISRUC-S3. It can be observed that the N1 stage consistently exhibits the lowest accuracy among all the various sleep stages. It is believed that stage N1 is a transitional stage between Wake and N2, which means that the N1 stage has the characteristics of both Wake and N2 and this leads to misclassification. On the other hand, all models perform exceptionally well in identifying the Wake stage, which can be explained by the distinct characteristics of brain waves during wakeful hours compared to sleep. The remaining three categories, namely, N2, N3 and REM stay at a similar level, hovering around 85%. Nonetheless, no definitive evidence exists to support the superiority of one stage over the other two in terms of classification accuracy. TABLE V also highlights an intriguing phenomenon that the precision and recall of the proposed model consistently converge to the outcomes of the more effective model categories. For example, the precision of the awake stage of all CNN-based models, including DeepSleepNet, TinySleepNet, and SalientSleepNet, achieves 0.920, 0.929, and 0.909, respectively, overperforming all GCN models, whose best result is 0.893. While the proposed model achieves 0.908, which is closer to the CNNs results than GCNs. For the N3 stage, GCNs have a better result than CNNs but the proposed model also overperforms all CNNs. It is believed that the MixSleepNet model combines the advantages of both GCNs and CNNs.

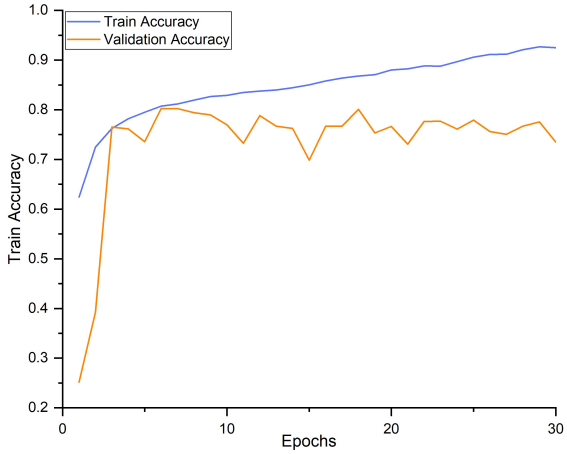

Fig. 4 An example of training curve and validation curve on the ISRUC-S3.

Fig. 3 shows an instance of a hypnogram obtained through the expert one and its corresponding generated sleep hypnogram by our method for subject one from the ISRUC-S3 subset. The comparison shows that most of the classifications are overlapped, except for those parts that have frequent transitions among stages. Fig. 4 shows an example of a training curve on the ISRUC-S3 subset, where 30 training epochs are set to avoid overfitting.

The comparison results among various machine learning methods [19], [23], [32], [50], [53], [54], [61], [62] on additional public datasets [63]–[66] are presented in TABLE VI. In contrast to these algorithms, the accuracy of the proposed model on ISRUC-S3, ISRUC-S1, and Sleep-EDF-153 dataset (153 Sleep Cassette files) [64] are 0.830, 0.813 and 0.891, respectively, which are close to the level of other methods. The proposed model has the highest classification performance for the N1 stage, N3 stage, and REM stage on ISRUC-S3, compared to other models, while the N2 and Wake stages are slightly lower than the two models. Due to the limited channels used on the Sleep-EDF-153 dataset, the classification results on this dataset indicate that the performance of the proposed model may be affected by the number of channels and the types of channels used.

4.2 Model Analysis

Several experiments are conducted on ISRUC-S3 to assess the significance of the selected channels and compare their impact on classification performance. The detailed channel selections for each experiment are listed in TABLE VII. The classification performance of each experiment is presented in TABLE VII and Fig. 5. The single channel selections are

TABLE VI
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS ON DIFFERENT DATASETS

Methods	Datasets	Cross-validation	Overall Metrics			Per class F1 score (F1)				
			ACC	F1	κ	W	N1	N2	N3	REM
RF [19]	ISRUC	Patient dependent	0.86 ± 0.02	-	-	-	-	-	-	-
Complex-valued unsupervised [23]	UCD	5-fold	0.87	-	0.8	-	-	-	-	-
CNN+RNN [32]	CAP	5-fold	0.788	0.727	0.71	0.841	0.402	0.783	0.817	0.789
Ensemble bagged trees [50]	SOF	10-fold	0.813	-	0.752	0.92	0.04	0.79	0.74	0.66
SVM [53]	SleepEDF-153	10-fold	0.917	-	-	-	-	-	-	-
CNN [61]	SleepEDF-153	10-fold	0.825	0.761	0.76	0.924	0.481	0.846	0.738	0.816
Ensemble bagged trees [62]	ISRUC-S1	10-fold	0.774	-	-	-	-	-	-	-
CNN+GRU	SHHS1-700	-	0.832	-	0.760	0.897	0.311	0.850	0.781	0.808
This work	SleepEDF-153	10-fold	0.891	0.685	0.770	0.970	0.227	0.815	0.760	0.652
This work	ISRUC-S3	10-fold	0.830	0.821	0.782	0.899	0.625	0.819	0.899	0.830
This work	ISRUC-S1 (50)	25-fold	0.813	0.787	0.757	0.908	0.512	0.799	0.871	0.813

excluded since the 3D-CNN branch will become a 2D-CNN branch and the graph in the GCN branch will become a node, which means that the architecture of the proposed model is fully changed, if a single channel data is input only.

Experiment *i*, *ii* and *iii* demonstrates that the contribution of ECG, EMG and EOG for the classification performance is different, among which, the ECG channel makes the least contribution to the performance, while the EOG channel makes the most contribution for almost all stages, especially the REM stage. The explanation for the significance of EOG is that the REM stage is an eye-movement-related stage, where EOG signals play an important role. The comparisons, *iv* vs *iii* and *ix* vs *viii* demonstrate that the EMG and ECG channels may improve the overall classification performance (ACC, >0.02), especially for the N1 stage (F1, >0.04) and the REM stage (F1, >0.06), while the comparison *ii* vs *i* shows that EMG may have more effects on these stages. An important point here is that the improvement of the REM stage is from the improvement of the N1 stage, which means that more N1 stages are classified correctly from the REM stage, leading to the improvement of REM, and this conclusion can be easily obtained by the experiment *ii*, which has higher F1 score of N1 and lower F1 score of REM than the experiment *iii*. The comparison, *vi* vs *v*, indicates that the proposed model has similar performance if pure EEGs or EOGs are used with two channels. The pure EEGs overperform the pure EOGs on the N3 stage (F1, >0.03), while pure EOGs perform better on the REM stage (F1, >0.04). However, the combination of EEG and EOG can make up for each other's shortcomings (*iii*). The performance also can be improved by adding more EEG channels (*vii* vs *v*). The improvements are from all stages except the N3 stage, while this problem can be solved by adding other channels (*ix*).

To comprehensively validate the individual impact of each module within the MixSleepNet model, five variant models are designed and tested on the ISRUC-S3 dataset. The specifics of these models are expounded upon below:

1. *variant a (simplified JK-STGCN model)*: A simplified jumping-knowledge-based graph convolutional model is selected as a variant model. Unlike the complete JK-STGCN model, which takes the output of a CNN extractor as the input, this variant model takes DE features as input and removes the convolutional layer for temporal information.

2. *variant b (pure 3D-CNN model):* A pure 3D-CNN variant is designed without any attention layers.

3. *variant c (variant b + partial dot-product attention):* The original 3D-CNN model is enhanced by the inclusion of partial dot-product attention layers.

4. *variant d (variant a +variant b):* This variant model combines the simplified JK-STGCN model and the original 3D-CNN model as two branches without any attention layers.

5. *variant e (variant a +variant c):* Partial dot-product attention layers are added to the combined model.

Fig. 6 illustrates the classification performance of the above variant models on the ISRUC-S3 dataset. The simplified JK-STGCN model has obtained the lowest performance since only input DE features are fed and the standard convolutional layers are removed, which leads to ignoring correlations among

TABLE VII
COMPARISON OF CLASSIFICATION PERFORMANCE USING DIFFERENT CHANNELS

Experiment	Channels	Overall Metrics			Per-class F1-score (F1)				
		ACC	F1	κ	W	N1	N2	N3	REM
i	1 EEG, 1 ECG	0.678	0.612	0.582	0.793	0.442	0.710	0.840	0.277
ii	1 EEG, 1 EMG	0.751	0.730	0.679	0.858	0.537	0.741	0.866	0.649
iii	1 EEG, 1 EOG	0.775	0.751	0.710	0.866	0.500	0.783	0.873	0.736
iv	1 EEG, 1 EOG, 1 EMG, 1 ECG	0.795	0.784	0.736	0.889	0.581	0.779	0.857	0.815
v	2 EEG	0.768	0.735	0.701	0.856	0.471	0.776	0.882	0.692
vi	2 EOG	0.761	0.737	0.690	0.842	0.477	0.771	0.851	0.743
vii	6 EEG	0.785	0.766	0.723	0.869	0.579	0.784	0.865	0.733
viii	6 EEG, 2 EOG	0.802	0.786	0.745	0.877	0.582	0.801	0.875	0.797
ix	6 EEG, 2 EOG, 1 EMG, 1 ECG	0.830	0.821	0.782	0.899	0.625	0.819	0.899	0.860

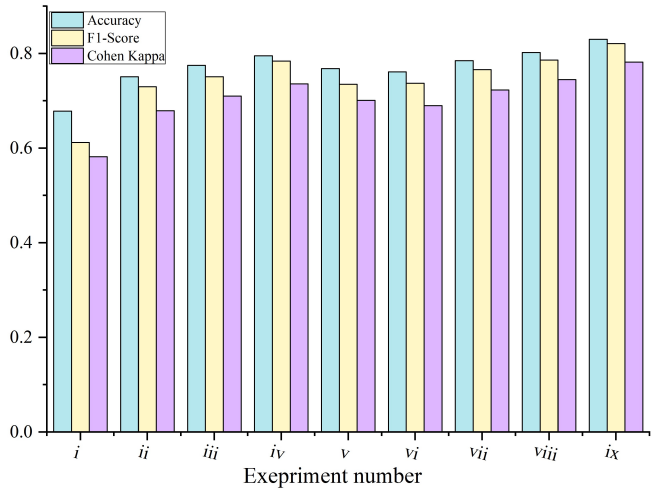

Fig. 5 Comparison of performance for different experiments.

and within epochs. As a result, the classification performance heavily decreased from the complete JK-STGCN model. The original 3D-CNN model improved significantly because both spatial information and temporal information are aggregated by 3D filters. Since attention layers have the capacity to focus on valuable information, the added attention layers on the original 3D-CNN model can further improve its performance. The combined model without attention layers overperforms all the single-branch classifiers. In the future, we intend to explore the possibility of using explainable artificial intelligence (XAI) to visualize the features responsible for various sleep stages and disorders using a huge database [67].

5 Conclusion

This study presents an automated multi-channel sleep stage classification model rooted in the fusion of 3D convolution and graph convolution techniques. Based on DE features, the simplified JK-STGCN branch with two adaptive graph learning methods explores the correlation of brain areas, while the 3D convolution branch with partial dot-product attention layers investigates brain activities in time series. Our experimentation on both ISRUC-S1 and ISRUC-S3 datasets reveals the remarkable capability of the proposed MixSleepNet model to effectively classify sleep stages for both healthy individuals and patients suffering from sleep disorders. Our proposed MixSleepNet outperformed all baselines on the second expert's labels and achieved competitive results on the first expert's labels. Incremental experiments conducted on the ISRUC-S3 dataset reveal that the combined branches with partial dot-product layers achieved the best performance. However, there is still some space to further improve our model. One of the limitations is its huge storage requirements. The feature extraction step increases the volume of the data by almost ten times, which also increases both computational resources and time. Although pseudo-3D convolution methods and K-1 order Chebyshev polynomials are adopted to address the computational complexity, this multi-channel-based classifier requires large memory and computing resources. In the future, one possible related research is to explore novel 3D representations of signals to reduce both the data size and computational resource requirements. The new implementation of the proposed model in other EEG-related fields, like Alzheimer's disease detection [13] and emotion prediction [15].

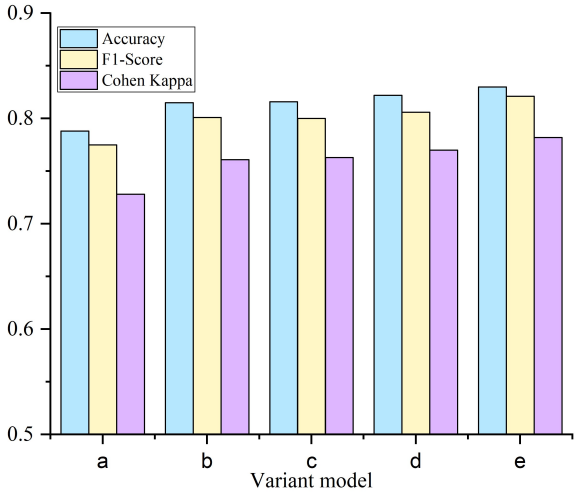

Fig. 6 Comparison of performances of different models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was based on the open-source public dataset the ISRUC (https://sleeptight.isr.uc.pt/) and Sleep-EDF-153 (https://www.physionet.org/content/sleep-edfx/1.0.0/). The patients involved in these two databases have given ethical approval.

References

[1] F. Weber and Y. Dan, "Circuit-based interrogation of sleep control," *Nature*, vol. 538, no. 7623, pp. 51–59, Oct. 2016, doi: 10.1038/nature19773.

[2] A. Ivanenko and B. R. Gururaj, "Classification and Epidemiology of Sleep Disorders," *Child Adolesc. Psychiatr. Clin.*, vol. 18, no. 4, pp. 839–848, Oct. 2009, doi: 10.1016/j.chc.2009.04.005.

[3] J. Zhang and Y. Wu, "A New Method for Automatic Sleep Stage Classification," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 5, pp. 1097–1110, 2017, doi: 10.1109/TBCAS.2017.2719631.

[4] S. Siuly and Y. Li, "Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach," *Neural Comput. Appl.*, vol. 26, no. 4, pp. 799–811, May 2015, doi: 10.1007/s00521-014-1753-3.

[5] 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. Specif. Darien Ill. Am. Acad. Sleep Med.*, vol. 176, p. 2012, 2012.

[6] K. A. I. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in

IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014, May 2014, pp. 1–6. doi: 10.1109/LISAT.2014.6845193.
[7] S. A. Keenan, "Chapter 3 An overview of polysomnography," in *Handbook of Clinical Neurophysiology*, vol. 6, C. Guilleminault, Ed., in Handbook of Clinical Neurophysiology, vol. 6. , Elsevier, 2005, pp. 33–50. doi: 10.1016/S1567-4231(09)70028-0.
[8] M. Sharma, H. Lodhi, R. Yadav, H. Elphick, and U. R. Acharya, "Computerized detection of cyclic alternating patterns of sleep: A new paradigm, future scope and challenges," *Comput. Methods Programs Biomed.*, p. 107471, 2023, doi: 10.1016/j.cmpb.2023.107471.
[9] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 7046–7059, Dec. 2013, doi: 10.1016/j.eswa.2013.06.023.
[10] J. B. Stephansen *et al.*, "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Dec. 2018, doi: 10.1038/s41467-018-07229-3.
[11] S. Roy, K. Kate, and M. Hirzel, "A semi-supervised deep learning algorithm for abnormal EEG identification," *ArXiv Prepr. ArXiv190307822*, 2019.
[12] O. Faust, H. Razaghi, R. Barika, E. J. Ciaccio, and U. R. Acharya, "A review of automated sleep stage scoring based on physiological signals for the new millennia," *Comput. Methods Programs Biomed.*, vol. 176, pp. 81–91, 2019, doi: 10.1016/j.cmpb.2019.04.032.
[13] S. K. Khare and U. R. Acharya, "Adazd-Net: Automated adaptive and explainable Alzheimer's disease detection system using EEG signals," *Knowl.-Based Syst.*, vol. 278, p. 110858, Oct. 2023, doi: 10.1016/j.knosys.2023.110858.
[14] S. K. Khare, S. March, P. D. Barua, V. M. Gadre, and U. R. Acharya, "Application of data fusion for automated detection of children with developmental and mental disorders: A systematic review of the last decade," *Inf. Fusion*, vol. 99, p. 101898, Nov. 2023, doi: 10.1016/j.inffus.2023.101898.
[15] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," *Inf. Fusion*, vol. 102, p. 102019, Feb. 2024, doi: 10.1016/j.inffus.2023.102019.
[16] S.-Y. Chang *et al.*, "An Ultra-Low-Power Dual-Mode Automatic Sleep Staging Processor Using Neural-Network-Based Decision Tree," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 66, no. 9, pp. 3504–3516, Sep. 2019, doi: 10.1109/TCSI.2019.2927839.
[17] I. A. Zapata, Y. Li, and P. Wen, "Rules-Based and SVM-Q Methods With Multitapers and Convolution for Sleep EEG Stages Classification," *IEEE Access*, vol. 10, pp. 71299–71310, Jan. 2022, doi: 10.1109/ACCESS.2022.3188286.
[18] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, 2018.
[19] M. M. Rahman, M. I. H. Bhuiyan, and A. R. Hassan, "Sleep stage classification using single-channel EOG," *Comput. Biol. Med.*, vol. 102, pp. 211–220, Nov. 2018, doi: 10.1016/j.compbiomed.2018.08.022.
[20] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012, doi: 10.1016/j.cmpb.2011.11.005.
[21] S. K. Khare and V. Bajaj, "Constrained based tunable Q wavelet transform for efficient decomposition of EEG signals," *Appl. Acoust.*, vol. 163, p. 107234, Jun. 2020, doi: 10.1016/j.apacoust.2020.107234.
[22] H. W. Loh *et al.*, "Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020)," *Appl. Sci.*, vol. 10, no. 24, p. 8963, 2020, doi: 10.3390/app10248963.
[23] J. Zhang and Y. Wu, "Complex-valued unsupervised convolutional neural networks for sleep stage classification," *Comput. Methods Programs Biomed.*, vol. 164, pp. 181–191, Oct. 2018, doi: 10.1016/j.cmpb.2018.07.015.
[24] H. W. Loh, C. P. Ooi, S. G. Dhok, M. Sharma, A. A. Bhurane, and U. R. Acharya, "Automated detection of cyclic alternating pattern and classification of sleep stages using deep neural network," *Appl. Intell.*, vol. 52, no. 3, pp. 2903–2917, Feb. 2022, doi: 10.1007/s10489-021-02597-8.
[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.
[26] S. K. Khare, V. Bajaj, S. Taran, and G. R. Sinha, "1 - Multiclass sleep stage classification using artificial intelligence based time-frequency distribution and CNN," in *Artificial Intelligence-Based Brain-Computer Interface*, V. Bajaj and G. R. Sinha, Eds., Academic Press, 2022, pp. 1–21. doi: 10.1016/B978-0-323-91197-9.00012-6.
[27] 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, p. e0216456, May 2019, doi: 10.1371/journal.pone.0216456.
[28] 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, 2018, doi: 10.1016/j.bspc.2017.12.001.
[29] Q. Cai, Z. Gao, J. An, S. Gao, and C. Grebogi, "A Graph-Temporal Fused Dual-Input Convolutional Neural Network for Detecting Sleep Stages from EEG Signals," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 68, no. 2, pp. 777–781, 2021, doi: 10.1109/TCSII.2020.3014514.
[30] S. Biswal *et al.*, "SLEEPNET: Automated Sleep Staging System via Deep Learning," arXiv, arXiv:1707.08262, Jul. 2017. doi: 10.48550/arXiv.1707.08262.
[31] 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, doi: 10.1109/TNSRE.2018.2813138.
[32] C. Zhao, J. Li, and Y. Guo, "SleepContextNet: A temporal context network for automatic sleep staging based single-channel EEG," *Comput. Methods Programs Biomed.*, vol. 220, p. 106806, Jun. 2022, doi: 10.1016/j.cmpb.2022.106806.

[33] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, 2017, doi: 10.1109/TNSRE.2017.2721116.
[34] A. Supratak and Y. Guo, "TinySleepNet: An Efficient Deep Learning Model for Sleep Stage Scoring based on Raw Single-Channel EEG," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 641–644. doi: 10.1109/EMBC44109.2020.9176741.
[35] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Comput. Biol. Med.*, vol. 106, pp. 71–81, Mar. 2019, doi: 10.1016/j.compbiomed.2019.01.013.
[36] Z. Jia *et al.*, "Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI*, in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI. 2020, pp. 1324–1330.
[37] Z. Jia *et al.*, "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.
[38] X. Ji, Y. Li, and P. Wen, "Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1464–1472, 2022, doi: 10.1109/TNSRE.2022.3176004.
[39] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *J. Neurosci. Methods*, vol. 205, no. 1, pp. 169–176, Mar. 2012, doi: 10.1016/j.jneumeth.2011.12.022.
[40] 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 Inform.*, vol. 18, no. 6, pp. 1813–1821, 2014, doi: 10.1109/JBHI.2014.2303991.
[41] Z. Zhang *et al.*, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019, doi: 10.1109/ACCESS.2019.2895133.
[42] 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, 2018, doi: 10.1109/TNSRE.2017.2776149.
[43] M. Diykh and Y. Li, "Complex networks approach for EEG signal sleep stages classification," *Expert Syst. Appl.*, vol. 63, pp. 241–248, 2016, doi: 10.1016/j.eswa.2016.07.004.
[44] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016, doi: 10.1016/j.bspc.2015.09.002.
[45] M. Diykh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, 2016.
[46] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, 2017, doi: 10.1007/s00521-017-2919-6.
[47] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature selection for sleep/wake stages classification using data driven methods," *Biomed. Signal Process. Control*, vol. 2, no. 3, pp. 171–179, 2007, doi: 10/d62rq9.
[48] A. Stochholm, K. Mikkelsen, and P. Kidmose, "Automatic sleep stage classification using ear-EEG," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2016, pp. 4751–4754.
[49] B. Fatimah, A. Singhal, and P. Singh, "A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning," *Comput. Biol. Med.*, vol. 148, p. 105877, Sep. 2022, doi: 10.1016/j.compbiomed.2022.105877.
[50] B. J. Dakhale *et al.*, "An automatic sleep-scoring system in elderly women with osteoporosis fractures using frequency localized finite orthogonal quadrature Fejer Korovkin kernels," *Med. Eng. Phys.*, vol. 112, p. 103956, Feb. 2023, doi: 10.1016/j.medengphy.2023.103956.
[51] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Ann. Biomed. Eng.*, vol. 44, no. 5, pp. 1587–1597, 2016, doi: 10.1007/s10439-015-1444-y.
[52] T. F. Zaidi and O. Farooq, "EEG sub-bands based sleep stages classification using Fourier Synchrosqueezed transform features," *Expert Syst. Appl.*, vol. 212, p. 118752, Feb. 2023, doi: 10.1016/j.eswa.2022.118752.
[53] M. Sharma, D. Goyal, P. V. Achuth, and U. R. Acharya, "An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank," *Comput. Biol. Med.*, vol. 98, pp. 58–75, Jul. 2018, doi: 10.1016/j.compbiomed.2018.04.025.
[54] J. Zhang, R. Yao, W. Ge, and J. Gao, "Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG," *Comput. Methods Programs Biomed.*, vol. 183, p. 105089, Jan. 2020, doi: 10.1016/j.cmpb.2019.105089.
[55] Z. Jia, Y. Lin, J. Wang, X. Wang, and Y. Zhang, "SalientSleepNet: Multimodal Salient Wave Detection Network for Sleep Staging," 2021.
[56] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013, doi: 10.1109/TPAMI.2012.59.
[57] Z. Qiu, T. Yao, and T. Mei, "Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 5534–5542. doi: 10.1109/ICCV.2017.590.
[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770–778.

[59] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," *Comput Methods Programs Biomed*, vol. 124, 2016.
[60] L. Song, Y. Li, and N. Lu, "ProfileSR-GAN: A GAN based Super-Resolution Method for Generating High-Resolution Load Profiles." arXiv, Jan. 26, 2022. doi: 10.48550/arXiv.2107.09523.
[61] E. Khalili and B. Mohammadzadeh Asl, "Automatic Sleep Stage Classification Using Temporal Convolutional Neural Network and New Data Augmentation Technique from Raw Single-Channel EEG," *Comput. Methods Programs Biomed.*, vol. 204, p. 106063, Jun. 2021, doi: 10.1016/j.cmpb.2021.106063.
[62] M. Sharma, P. Makwana, R. S. Chad, and U. R. Acharya, "A novel automated robust dual-channel EEG-based sleep scoring system using optimal half-band pair linear-phase biorthogonal wavelet filter bank," *Appl. Intell.*, vol. 53, no. 15, pp. 18681–18699, Aug. 2023, doi: 10.1007/s10489-022-04432-0.
[63] G.-Q. Zhang *et al.*, "The National Sleep Research Resource: towards a sleep data commons," *J. Am. Med. Inform. Assoc. JAMIA*, vol. 25, no. 10, pp. 1351–1358, Oct. 2018, doi: 10.1093/jamia/ocy064.
[64] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215-220, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.
[65] M. G. Terzano *et al.*, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, no. 6, pp. 537–553, Nov. 2001, doi: 10.1016/S1389-9457(01)00149-6.
[66] S. F. Quan *et al.*, "The Sleep Heart Health Study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, Dec. 1997.
[67] M. Dutt, S. Redhu, M. Goodwin, and C. W. Omlin, "SleepXAI: An explainable deep learning approach for multi-class sleep stage identification," *Appl. Intell.*, vol. 53, no. 13, pp. 16830–16843, Jul. 2023, doi: 10.1007/s10489-022-04357-8.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

6.1. Thesis summary

This study focuses on developing more robust and generative deep learning methods to classify sleep stages automatically. The research background and objectives are presented at first, after which, the literature review and gaps of problems are followed. Based on existing research, this thesis proposed three deep learning models for sleep stage classification performance improvement, and the comparisons between the existing methods and the proposed algorithms were conducted in the attached paper files.

6.2. Conclusion

In this study, deep learning methods, including a GCN, a 3D-CNN, and a combined model are explored to build robust systems for sleep stage classification. Compared with traditional machine learning methods, 1D-CNN and 2D-CNN, which fail to explore functional connections between brain regions during brain activities, these three proposed models not only can learn functional connections but also improve classification performance with a less execution time.

The GCN model, namely the JK-STGCN model, is an improved version of the GraphSleepNet model. A novel adaptive adjacency matrix learning method is designed for this model to learn the correlations among brain regions during brain activities. Temporal features are extracted from two EOG channels, six EEG channels, one EMG channel (Chin EMG), and one ECG channel through a feature extractor named FeatureNet and fed into this GCN model. The removal of spatial attention mechanisms and temporal attention mechanisms increases the training efficiency, which increases the possibility of clinical implementation and the deployment of edge devices. Experiments are conducted on the ISRUC-S3 and ISRUC-S1 subsets. The overall accuracy, the F1-score, and Cohen kappa reached 0.831, 0.814, and 0.782, respectively, which are much better in terms of performance compared to those Euclidean-input deep learning methods and the existing STGCN methods. Ablation experiments are also conducted on the same datasets to investigate the contribution of each module in this model. The generalization of this model is evaluated in the

mixed dataset of ISRUC-S3 and ISRUC-S1, whose results demonstrate that the JK-STGCN model has high robustness on both healthy and unhealthy subjects.

The 3DSleepNet method, namely, the 3D-CNN model in Chapter 4 (published in IEEE Transactions), is a logic idea from the CNNs and GCNs, consisting of two 3D-CNN streams and one 2D-CNN stream. 3D convolutional operations are used to extract spatial-temporal features and spatial-spectral-temporal features from temporal inputs and temporal-frequency inputs, respectively. 2D convolutional operations are also utilized in the proposed model to extract spatial-spectral features from frequency inputs. Temporal features are extracted through down-sampling from EEGs, EOGs, and EMG. Temporal-frequency features are extracted through short-term differential entropy of 9 crossed frequency bands: 0.5-4 Hz, 2-6 Hz, 4-8 Hz, 6-11 Hz, 8-14 Hz, 11-22 Hz, 14-31 Hz, 22-40 Hz and 31-49 Hz, for each selected channels. In the meantime, spatial-spectral features are extracted through calculating the power spectral density of each epoch on the same frequency bands. The partial dot-product attention mechanism is designed to help the 3DSleepNet model to pay attention to the most relevant information, while the LSTM layer is added to enable this model to learn transition rules among neighbouring sleep epochs. To increase the training efficiency, several 3D-CNN layers are replaced by pseudo-3D convolutional layers. The overall accuracy achieves 0.832, and the F1-score and Cohen's kappa reach 0.814 and 0.783, respectively, on ISRUC-S3. The performance is competitive compared with the state-of-the-art baselines, especially compared to the JK-STGCN model, while its training time decreases heavily. Ablation experiments are conducted on the ISRUC-S3 subset to evaluate the contributions of modules and channels. The experimental results demonstrate that each module helps to improve the classification results. It also illustrates that the 3DSleepNet model has a lower performance if only EEGs are used. This phenomenon also indicates the significance of EOG and EMG signals in sleep stage classification task.

The MixSleepNet method, namely, the combined model of the GCN and 3D-CNN in Chapter 5 aims to combine the advantages of both the 3D-CNN and GCN, where the 3D-CNN focuses on temporal and spatial dimensional features and the GCN focuses on frequency and spatial dimensional features. The classification experiments on ISRUC-S1 and S3 from two experts' labels demonstrate that this combined model can achieve excellent results for sleep stage identification. Further experiments on ISRUC-S3 highlight an intriguing phenomenon that the precision and

recall of the proposed model consistently converge to the outcomes of the more effective model categories. For example, the precision of the awake stage of all CNN-based models, including DeepSleepNet, TinySleepNet, and SalientSleepNet, achieves 0.920, 0.929, and 0.909, respectively, overperforming all GCN models, whose best result is 0.893. While the combined model achieves 0.908, which is closer to the CNNs results than GCNs. For the N3 stage, GCNs have a better result than CNNs but the proposed model also overperforms all CNNs. To validate the contributions of the types of signals (EEG, EOG, ECG or EMG), extra experiments are conducted on the ISRUC-S3 subset. The experimental results show that the classification performance of this combined model depends on the number of PSGs and the types of PSGs, where EOGs make a great contribution to enhancing the classification performance of almost all the stages, especially the REM stage, while EMG and ECG channels can improve the overall classification performance, especially for the N1 stage. Ablation experiments are conducted to further investigate the impact of modules in this model. The experimental results show that the combined branches with partial dot-product layers achieved the best performance.

Overall, this thesis introduces the development of deep learning methods for sleep stage classification, where the GCN and 3D-CNN based models are applied to improve the classification performance, execution time, and generalization.

6.3. Future work

In terms of the JK-STGCN model, an inevitable disadvantage is that GCN models require more computing resources than CNNs, especially multi-channel bio-signals are utilized to enhance the prediction capability. Therefore, a natural idea to reduce the computing complexity is to develop a single-channel GCN, which decreases the dimensions of inputs and computing steps. The new architecture of the GCN is also worthy to build for improving the classification performance.

Concerning the 3D-CNN models, the extracted features, especially the temporal features of band waves occupy a large storage and memory for computation, which limits their applications in clinical settings. A possible solution is to develop a new presentation of 3D temporal features, which requires fewer resources.

Even though we have attempted to give the possible explainability of the combined model, there are still challenges remained to be addressed. For example,

the correlations between the classification performance and PSGs used are explained, but the functional connections in each stage are still not visible. In the future, we intend to explore the possibility of using explainable artificial intelligence methods to visualize the features and functional connections responsible for various sleep stages.

REFERENCES

Abdulla, S., Diykh, M., Siuly, S., Ali, M., 2023. An intelligent model involving multi-channels spectrum patterns based features for automatic sleep stage classification. Int. J. Med. Inf. 171, 105001. https://doi.org/10.1016/j.ijmedinf.2023.105001

Aboalayon, K.A.I., Faezipour, M., 2014. Multi-class SVM based on sleep stage identification using EEG signal, in: 2014 IEEE Healthcare Innovation Conference (HIC). Presented at the 2014 IEEE Healthcare Innovation Conference (HIC), pp. 181–184. https://doi.org/10.1109/HIC.2014.7038904

Aboalayon, K.A.I., Ocbagabir, H.T., Faezipour, M., 2014. Efficient sleep stage classification based on EEG signals, in: IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014. Presented at the IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014, pp. 1–6. https://doi.org/10.1109/LISAT.2014.6845193

Acharya, U.R., Bhat, S., Faust, O., Adeli, H., Chua, E.C.-P., Lim, W.J.E., Koh, J.E.W., 2015. Nonlinear dynamics measures for automated EEG-based sleep stage detection. Eur. Neurol. 74, 268–287. https://doi.org/10.1159/000441975

Alickovic, E., Subasi, A., 2018. Ensemble SVM Method for Automatic Sleep Stage Classification. IEEE Trans. Instrum. Meas. 67, 1258–1265. https://doi.org/10.1109/TIM.2018.2799059

Al-Nuaimi, A.H.H., Jammeh, E., Sun, L., Ifeachor, E., 2018. Complexity Measures for Quantifying Changes in Electroencephalogram in Alzheimer's Disease. Complexity 2018, e8915079. https://doi.org/10.1155/2018/8915079

Al-Salman, W., Li, Y., Oudah, A.Y., Almaged, S., 2023. Sleep stage classification in EEG signals using the clustering approach based probability distribution features coupled with classification algorithms. Neurosci. Res. 188, 51–67. https://doi.org/10.1016/j.neures.2022.09.009

Al-Salman, W., Li, Y., Wen, P., 2019. Detecting sleep spindles in EEGs using wavelet fourier analysis and statistical features. Biomed. Signal Process. Control 48, 80–92. https://doi.org/10.1016/j.bspc.2018.10.004

Ansari-Asl, K., Chanel, G., Pun, T., 2007. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood, in:

2007 15th European Signal Processing Conference. Presented at the 2007 15th European Signal Processing Conference, pp. 1241–1245.

Aserinsky, E., Kleitman, N., 1953. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. Science 118, 273–274. https://doi.org/10.1126/science.118.3062.273

B, V.P., Chinara, S., 2021. Automatic classification methods for detecting drowsiness using wavelet packet transform extracted time-domain features from single-channel EEG signal. J. Neurosci. Methods 347, 108927. https://doi.org/10.1016/j.jneumeth.2020.108927

Ball, T., Kern, M., Mutschler, I., Aertsen, A., Schulze-Bonhage, A., 2009. Signal quality of simultaneously recorded invasive and non-invasive EEG. NeuroImage 46, 708–716. https://doi.org/10.1016/j.neuroimage.2009.02.028

Biswal, S., Kulas, J., Sun, H., Goparaju, B., Westover, M.B., Bianchi, M.T., Sun, J., 2017. SLEEPNET: Automated Sleep Staging System via Deep Learning. https://doi.org/10.48550/arXiv.1707.08262

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92. Association for Computing Machinery, New York, NY, USA, pp. 144–152. https://doi.org/10.1145/130385.130401

Bramer, M., 2013. Ensemble Classification, in: Bramer, M. (Ed.), Principles of Data Mining, Undergraduate Topics in Computer Science. Springer, London, pp. 209–220. https://doi.org/10.1007/978-1-4471-4884-5_14

Carrozzi, M., Accardo, A., Bouquet, F., 2004. Analysis of sleep-stage characteristics in full-term newborns by means of spectral and fractal parameters. Sleep 27, 1384–1393. https://doi.org/10.1093/sleep/27.7.1384

Casson, A.J., 2019. Wearable EEG and beyond. Biomed. Eng. Lett. 9, 53–71. https://doi.org/10.1007/s13534-018-00093-6

Chaganti, S.Y., Nanda, I., Pandi, K.R., Prudhvith, T.G.N.R.S.N., Kumar, N., 2020. Image Classification using SVM and CNN, in: 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). Presented at the 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), IEEE, Gunupur, India, pp. 1–5. https://doi.org/10.1109/ICCSEA49143.2020.9132851

Chambon, S., Galtier, M.N., Arnal, P.J., Wainrib, G., Gramfort, A., 2018. A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series. IEEE Trans. Neural Syst. Rehabil. Eng. 26, 758–769. https://doi.org/10.1109/TNSRE.2018.2813138

Chandra, M.A., Bedi, S.S., 2021. Survey on SVM and their application in imageclassification. Int. J. Inf. Technol. 13, 1–11. https://doi.org/10.1007/s41870-017-0080-1

Chen, L., Zhao, Y., Zhang, J., Zou, J., 2015. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. Expert Syst. Appl. 42, 7344–7355. https://doi.org/10.1016/j.eswa.2015.05.028

Chen, X., Zheng, Y., Niu, Y., Li, C., 2020. Epilepsy Classification for Mining Deeper Relationships between EEG Channels based on GCN, in: 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL). Presented at the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), pp. 701–706. https://doi.org/10.1109/CVIDL51233.2020.00050

Cheng, J., Zhang, F., Xiang, D., Yin, Q., Zhou, Y., 2022. PolSAR Image Classification With Multiscale Superpixel-Based Graph Convolutional Network. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3079438

Choubey, H., Pandey, A., 2021. A combination of statistical parameters for the detection of epilepsy and EEG classification using ANN and KNN classifier. Signal Image Video Process. 15, 475–483. https://doi.org/10.1007/s11760-020-01767-4

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, in: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018

Covert, I.C., Krishnan, B., Najm, I., Zhan, J., Shore, M., Hixson, J., Po, M.J., 2019. Temporal graph convolutional networks for automatic seizure detection, in:

Machine Learning for Healthcare Conference, Machine Learning for Healthcare Conference. PMLR, pp. 160–180.

Dang-Vu, T.T., Schabus, M., Desseilles, M., Sterpenich, V., Bonjean, M., Maquet, P., 2010. Functional Neuroimaging Insights into the Physiology of Human Sleep. Sleep 33, 1589–1603. https://doi.org/10.1093/sleep/33.12.1589

Darmatasia, Fanany, M.I., 2017. Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM), in: 2017 5th International Conference on Information and Communication Technology (ICoIC7). Presented at the 2017 5th International Conference on Information and Communication Technology (ICoIC7), pp. 1–6. https://doi.org/10.1109/ICoICT.2017.8074699

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/arXiv.1810.04805

Diba, A., Fayyaz, M., Sharma, V., Karami, A.H., Arzani, M.M., Yousefzadeh, R., Van Gool, L., 2017. Temporal 3D ConvNets: New Architecture and Transfer Learning for Video Classification. https://doi.org/10.48550/arXiv.1711.08200

Dimitriadis, S.I., Laskaris, N.A., Del Rio-Portilla, Y., Koudounis, G.Ch., 2009. Characterizing Dynamic Functional Connectivity Across Sleep Stages from EEG. Brain Topogr. 22, 119–133. https://doi.org/10.1007/s10548-008-0071-4

Dimitrovski, I., Kitanovski, I., Kocev, D., Simidjievski, N., 2023. Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. ISPRS J. Photogramm. Remote Sens. 197, 18–35. https://doi.org/10.1016/j.isprsjprs.2023.01.014

Diykh, M., Li, Y., 2016. Complex networks approach for EEG signal sleep stages classification. Expert Syst. Appl. 63, 241–248. https://doi.org/10.1016/j.eswa.2016.07.004

Diykh, M., Li, Y., Abdulla, S., 2020. EEG sleep stages identification based on weighted undirected complex networks. Comput. Methods Programs Biomed. 184, 105116. https://doi.org/10.1016/j.cmpb.2019.105116

Djemal, R., AlSharabi, K., Ibrahim, S., Alsuwailem, A., 2017. EEG-Based Computer Aided Diagnosis of Autism Spectrum Disorder Using Wavelet, Entropy, and ANN. BioMed Res. Int. 2017, e9816591. https://doi.org/10.1155/2017/9816591

Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P.M., Guo, Y., 2017. Mixed neural network approach for temporal sleep stage classification. IEEE Trans. Neural Syst. Rehabil. Eng. 26, 324–333. https://doi.org/10.1109/TNSRE.2017.2733220

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.

Duan, L., Li, Mengying, Wang, C., Qiao, Y., Wang, Z., Sha, S., Li, Mingai, 2021. A Novel Sleep Staging Network Based on Data Adaptation and Multimodal Fusion. Front. Hum. Neurosci. 15. https://doi.org/10.3389/fnhum.2021.727139

Duman, F., Erdamar, A., Erogul, O., Telatar, Z., Yetkin, S., 2009. Efficient sleep spindle detection algorithm with decision tree. Expert Syst. Appl. 36, 9980–9985. https://doi.org/10.1016/j.eswa.2009.01.061

Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P., 2015. Convolutional networks on graphs for learning molecular fingerprints, in: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Eichenlaub, J.-B., Bertrand, O., Morlet, D., Ruby, P., 2014. Brain Reactivity Differentiates Subjects with High and Low Dream Recall Frequencies during Both Sleep and Wakefulness. Cereb. Cortex 24, 1206–1215. https://doi.org/10.1093/cercor/bhs388

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., Guan, C., 2021. An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 809–818. https://doi.org/10.1109/TNSRE.2021.3076234

Fang, Y., Xia, Y., Chen, P., Zhang, J., Zhang, Y., 2023. A dual-stream deep neural network integrated with adaptive boosting for sleep staging. Biomed. Signal Process. Control 79, 104150. https://doi.org/10.1016/j.bspc.2022.104150

Fell, J., Röschke, J., Schäffner, C., 1996. Surrogate data analysis of sleep electroencephalograms reveals evidence for nonlinearity. Biol. Cybern. 75, 85–92. https://doi.org/10.1007/BF00238742

Fraiwan, L., Lweesy, K., Khasawneh, N., Fraiwan, M., Wenz, H., Dickhaus, H., 2011. Time Frequency Analysis for Automated Sleep Stage Identification in Fullterm and Preterm Neonates. J. Med. Syst. 35, 693–702. https://doi.org/10.1007/s10916-009-9406-2

Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H., Dickhaus, H., 2012. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Comput. Methods Programs Biomed. 108, 10–19. https://doi.org/10.1016/j.cmpb.2011.11.005

Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 55, 119–139. https://doi.org/10.1006/jcss.1997.1504

Funahashi, K., Nakamura, Y., 1993. Approximation of dynamical systems by continuous time recurrent neural networks. Neural Netw. 6, 801–806. https://doi.org/10.1016/S0893-6080(05)80125-X

Gao, Y., Fu, X., Ouyang, T., Wang, Y., 2022. EEG-GCN: Spatio-Temporal and Self-Adaptive Graph Convolutional Networks for Single and Multi-View EEG-Based Emotion Recognition. IEEE Signal Process. Lett. 29, 1574–1578. https://doi.org/10.1109/LSP.2022.3179946

Garces Correa, A., Laciar Leber, E., 2010. An automatic detector of drowsiness based on spectral analysis and wavelet decomposition of EEG records, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. Presented at the 2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010), IEEE, Buenos Aires, pp. 1405–1408. https://doi.org/10.1109/IEMBS.2010.5626721

Garcés Correa, A., Orosco, L., Laciar, E., 2014. Automatic detection of drowsiness in EEG records based on multimodal analysis. Med. Eng. Phys. 36, 244–249. https://doi.org/10.1016/j.medengphy.2013.07.011

Ghimatgar, H., Kazemi, K., Helfroush, M.S., Aarabi, A., 2019. An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model. J. Neurosci. Methods 324, 108320. https://doi.org/10.1016/j.jneumeth.2019.108320

Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E., 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex

physiologic signals. Circulation 101, E215-220. https://doi.org/10.1161/01.cir.101.23.e215

Goshtasbi, N., Boostani, R., Sanei, S., 2022. SleepFCN: A Fully Convolutional Deep Learning Framework for Sleep Stage Classification Using Single-Channel Electroencephalograms. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 2088–2096. https://doi.org/10.1109/TNSRE.2022.3192988

Gu, Y., Liang, Z., Hagihira, S., 2019. Use of Multiple EEG Features and Artificial Neural Network to Monitor the Depth of Anesthesia. Sensors 19, 2499. https://doi.org/10.3390/s19112499

Gunnarsdottir, K.M., Gamaldo, C.E., Salas, R.M.E., Ewen, J.B., Allen, R.P., Sarma, S.V., 2018. A Novel Sleep Stage Scoring System: Combining Expert-Based Rules with a Decision Tree Classifier, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3240–3243. https://doi.org/10.1109/EMBC.2018.8513039

Gupta, K., Bajaj, V., 2023. Deep learning models-based CT-scan image classification for automated screening of COVID-19. Biomed. Signal Process. Control 80, 104268. https://doi.org/10.1016/j.bspc.2022.104268

Hamdan, Y.B., Sathish, 2021. Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition. J. Inf. Technol. Digit. World 3, 92–107. https://doi.org/10.36548/jitdw.2021.2.003

Hassan, A.R., Bashar, S.K., Bhuiyan, M.I.H., 2015. On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram, in: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). Presented at the 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2238–2243. https://doi.org/10.1109/ICACCI.2015.7275950

Hassan, A.R., Bhuiyan, M.I.H., 2017. Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. Comput. Methods Programs Biomed. 140, 201–210. https://doi.org/10.1016/j.cmpb.2016.12.015

Hassan, A.R., Bhuiyan, M.I.H., 2016. Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and

bootstrap aggregating. Biomed. Signal Process. Control 24, 1–10. https://doi.org/10.1016/j.bspc.2015.09.002

Hassan, A.R., Bhuiyan, M.I.H., 2015. Automatic sleep stage classification, in: 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT). Presented at the 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT), pp. 211–216. https://doi.org/10.1109/EICT.2015.7391948

Hassan, A.R., Hassan Bhuiyan, M.I., 2016. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. Biocybern. Biomed. Eng. 36, 248–255. https://doi.org/10.1016/j.bbe.2015.11.001

Hassan, A.R., Subasi, A., 2017. A decision support system for automated identification of sleep stages from single-channel EEG signals. Knowl.-Based Syst. 128, 115–124. https://doi.org/10.1016/j.knosys.2017.05.005

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hjorth, B., 1970. EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. 29, 306–310. https://doi.org/10.1016/0013-4694(70)90143-4

Hou, R., Chen, C., Sukthankar, R., Shah, M., 2019. An Efficient 3D CNN for Action/Object Segmentation in Video. https://doi.org/10.48550/arXiv.1907.08895

Hou, Y., Jia, S., Lun, X., Zhang, S., Chen, T., Wang, F., Lv, J., 2021. Deep Feature Mining via Attention-based BiLSTM-GCN for Human Motor Imagery Recognition. https://doi.org/10.48550/arXiv.2005.00777

Hou, Y., Jia, S., Zhang, S., Lun, X., Shi, Y., Li, Y., Yang, H., Zeng, R., Lv, J., 2020. Deep Feature Mining via Attention-based BiLSTM-GCN for Human Motor Imagery Recognition. ArXiv Prepr. ArXiv200500777.

Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., Hsu, C.-Y., 2013. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing 104, 105–114. https://doi.org/10.1016/j.neucom.2012.11.003

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2019. Squeeze-and-Excitation Networks. https://doi.org/10.48550/arXiv.1709.01507

Huang, C.-S., Lin, C.-L., Ko, L.-W., Liu, S.-Y., Sua, T.-P., Lin, C.-T., 2013. A hierarchical classification system for sleep stage scoring via forehead EEG signals, in: 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB). Presented at the 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), pp. 1–5. https://doi.org/10.1109/CCMB.2013.6609157

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2018. Densely Connected Convolutional Networks. https://doi.org/10.48550/arXiv.1608.06993

Huang, L., Ma, D., Li, S., Zhang, X., WANG, H., 2019. Text Level Graph Neural Network for Text Classification. https://doi.org/10.48550/arXiv.1910.02356

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci. 454, 903–995. https://doi.org/10.1098/rspa.1998.0193

Hüsken, M., Stagge, P., 2003. Recurrent neural networks for time series classification. Neurocomputing 50, 223–235. https://doi.org/10.1016/S0925-2312(01)00706-8

Ivanenko, A., Gururaj, B.R., 2009. Classification and Epidemiology of Sleep Disorders. Child Adolesc. Psychiatr. Clin. 18, 839–848. https://doi.org/10.1016/j.chc.2009.04.005

James, C.J., Hesse, C.W., 2005. Independent component analysis for biomedical signals. Physiol. Meas. 26, R15–R39. https://doi.org/10.1088/0967-3334/26/1/R02

Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3D Convolutional Neural Networks for Human Action Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 221–231. https://doi.org/10.1109/TPAMI.2012.59

Ji, X., Li, Y., Wen, P., 2022. Jumping Knowledge Based Spatial-Temporal Graph Convolutional Networks for Automatic Sleep Stage Classification. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 1464–1472. https://doi.org/10.1109/TNSRE.2022.3176004

Jia, Z., Lin, Y., Cai, X., Chen, H., Gou, H., Wang, J., 2020a. SST-EmotionNet: Spatial-Spectral-Temporal based Attention 3D Dense Network for EEG Emotion Recognition, in: Proceedings of the 28th ACM International Conference on

Multimedia, Proceedings of the 28th ACM International Conference on Multimedia. pp. 2909–2917.

Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., Zhou, Y., Li-wei, H.L., 2021a. Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 1977–1986.

Jia, Z., Lin, Y., Wang, J., Wang, X., Xie, P., Zhang, Y., 2021b. SalientSleepNet: Multimodal Salient Wave Detection Network for Sleep Staging. https://doi.org/10.48550/arXiv.2105.13864

Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., Zhao, Y., 2020b. Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI. pp. 1324–1330.

Jiang, H., Diao, Z., Shi, T., Zhou, Y., Wang, F., Hu, W., Zhu, X., Luo, S., Tong, G., Yao, Y.-D., 2023. A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation. Comput. Biol. Med. 157, 106726. https://doi.org/10.1016/j.compbiomed.2023.106726

Jiang, X., Bian, G.-B., Tian, Z., 2019. Removal of Artifacts from EEG Signals: A Review. Sensors 19, 987. https://doi.org/10.3390/s19050987

Kales, A., Rechtschaffen, A., 1968. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. US Department of Health, Education and Welfare, Public Health Service ….

Kanwal, S., Uzair, M., Ullah, H., Khan, S.D., Ullah, M., Cheikh, F.A., 2019. An Image Based Prediction Model for Sleep Stage Identification, in: 2019 IEEE International Conference on Image Processing (ICIP). Presented at the 2019 IEEE International Conference on Image Processing (ICIP), pp. 1366–1370. https://doi.org/10.1109/ICIP.2019.8803026

Kaur, R., Kumar, R., Gupta, M., 2023. Deep neural network for food image classification and nutrient identification: A systematic review. Rev. Endocr. Metab. Disord. https://doi.org/10.1007/s11154-023-09795-4

Kausar, A., Sharif, M., Park, J., Shin, D.R., 2018. Pure-CNN: A Framework for Fruit Images Classification, in: 2018 International Conference on Computational Science and Computational Intelligence (CSCI). Presented at the 2018 International

Conference on Computational Science and Computational Intelligence (CSCI), pp. 404–408. https://doi.org/10.1109/CSCI46756.2018.00082

Kayikcioglu, T., Maleki, M., Eroglu, K., 2015. Fast and accurate PLS-based classification of EEG sleep using single channel data. Expert Syst. Appl. 42, 7825–7830. https://doi.org/10.1016/j.eswa.2015.06.010

Khalighi, S., Sousa, T., Pires, G., Nunes, U., 2013. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. Expert Syst. Appl. 40, 7046–7059. https://doi.org/10.1016/j.eswa.2013.06.023

Khalighi, S., Sousa, T., Santos, J.M., Nunes, U., 2016. ISRUC-Sleep: A comprehensive public dataset for sleep researchers. Comput. Methods Programs Biomed. 124, 180–192. https://doi.org/10.1016/j.cmpb.2015.10.013

Killgore, W.D.S., Jankowski, S., Henderson-Arredondo, K., Lucas, D.A., Patel, S.I., Hildebrand, L.L., Huskey, A., Dailey, N.S., 2023. Functional connectivity of the default mode network predicts subsequent polysomnographically measured sleep in people with symptoms of insomnia. NeuroReport 34, 734. https://doi.org/10.1097/WNR.0000000000001949

Kipf, T.N., Welling, M., 2017. Semi-Supervised Classification with Graph Convolutional Networks. https://doi.org/10.48550/arXiv.1609.02907

Koley, B., Dey, D., 2012. An ensemble system for automatic sleep stage classification using single channel EEG signal. Comput. Biol. Med. 42, 1186–1195. https://doi.org/10.1016/j.compbiomed.2012.09.012

Krakovská, A., Mezeiová, K., 2011. Automatic sleep scoring: a search for an optimal combination of measures. Artif. Intell. Med. 53, 25–33. https://doi.org/10.1016/j.artmed.2011.06.004

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105.

Kumar, Y., Dewal, M.L., Anand, R.S., 2014. Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network. Signal Image Video Process. 8, 1323–1334. https://doi.org/10.1007/s11760-012-0362-9

Kuo, C.-E., Chen, G.-T., Liao, P.-Y., 2021. An EEG spectrogram-based automatic sleep stage scoring method via data augmentation, ensemble convolution neural network, and expert knowledge. Biomed. Signal Process. Control 70, 102981. https://doi.org/10.1016/j.bspc.2021.102981

Lajnef, T., Chaibi, S., Ruby, P., Aguera, P.-E., Eichenlaub, J.-B., Samet, M., Kachouri, A., Jerbi, K., 2015. Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. J. Neurosci. Methods 250, 94–105. https://doi.org/10.1016/j.jneumeth.2015.01.022

Li, C., Qi, Y., Ding, X., Zhao, J., Sang, T., Lee, M., 2022. A Deep Learning Method Approach for Sleep Stage Classification with EEG Spectrogram. Int. J. Environ. Res. Public. Health 19, 6322. https://doi.org/10.3390/ijerph19106322

Li, M., Chen, W., Zhang, T., 2016. Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM. Biocybern. Biomed. Eng. 36, 708–718. https://doi.org/10.1016/j.bbe.2016.07.004

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M., 2014. Medical image classification with convolutional neural network, in: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV). Presented at the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 844–848. https://doi.org/10.1109/ICARCV.2014.7064414

Li, Y., Yingle, F., Gu, L., Qinye, T., 2009. Sleep stage classification based on EEG Hilbert-Huang transform, in: 2009 4th IEEE Conference on Industrial Electronics and Applications. Presented at the 2009 4th IEEE Conference on Industrial Electronics and Applications, pp. 3676–3681. https://doi.org/10.1109/ICIEA.2009.5138842

Liu, C., Tan, B., Fu, M., Li, J., Wang, J., Hou, F., Yang, A., 2021. Automatic sleep staging with a single-channel EEG based on ensemble empirical mode decomposition. Phys. Stat. Mech. Its Appl. 567, 125685. https://doi.org/10.1016/j.physa.2020.125685

Liu, T., Yang, D., 2021. A Densely Connected Multi-Branch 3D Convolutional Neural Network for Motor Imagery EEG Decoding. Brain Sci. 11, 197. https://doi.org/10.3390/brainsci11020197

Liu, X., Zhao, Z., Zhang, Y., Liu, C., Yang, F., 2022. Social Network Rumor Detection Method Combining Dual-Attention Mechanism With Graph Convolutional Network. IEEE Trans. Comput. Soc. Syst. 1–12. https://doi.org/10.1109/TCSS.2022.3184745

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows, in:

Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. https://doi.org/10.48550/arXiv.1411.4038

Loomis, A.L., Harvey, E.N., Hobart, G., 1936. Electrical potentials of the human brain. J. Exp. Psychol. 19, 249–279. https://doi.org/10.1037/h0062089

Loomis, A.L., Harvey, E.N., Hobart, G.A., 1937. Cerebral states during sleep, as studied by human brain potentials. J. Exp. Psychol. 21, 127. https://doi.org/10.1037/h0057431

Malekzadeh, M., Hajibabaee, P., Heidari, M., Zad, S., Uzuner, O., Jones, J.H., 2021. Review of Graph Neural Network in Text Classification, in: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). Presented at the 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0084–0091. https://doi.org/10.1109/UEMCON53757.2021.9666633

Malghan, P.G., Hota, M.K., 2020. A review on ECG filtering techniques for rhythm analysis. Res. Biomed. Eng. 36, 171–186. https://doi.org/10.1007/s42600-020-00057-9

Marvin, M., Seymour, A.P., 1969. Perceptrons. Camb. MA MIT Press 6, 318–362.

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., Poria, S., 2023. A review of deep learning techniques for speech processing. Inf. Fusion 99, 101869. https://doi.org/10.1016/j.inffus.2023.101869

Memar, P., Faradji, F., 2018. A Novel Multi-Class EEG-Based Sleep Stage Classification System. IEEE Trans. Neural Syst. Rehabil. Eng. 26, 84–95. https://doi.org/10.1109/TNSRE.2017.2776149

Michielli, N., Acharya, U.R., Molinari, F., 2019. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. Comput. Biol. Med. 106, 71–81. https://doi.org/10.1016/j.compbiomed.2019.01.013

Minsky, M.L., Papert, S.A., 1988. Perceptrons, expanded ed. ed. MIT Press, London.

Mohammed Alqahtani, T., 2023. Big Data Analytics with Optimal Deep Learning Model for Medical Image Classification. Comput. Syst. Sci. Eng. 44, 1433–1449. https://doi.org/10.32604/csse.2023.025594

Mohan, Y., Chee, S.S., Xin, D.K.P., Foong, L.P., 2016. Artificial neural network for classification of depressive and normal in EEG, in: 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). Presented at the 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 286–290. https://doi.org/10.1109/IECBES.2016.7843459

Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C.M., White, P.R., 2014. Signal processing techniques applied to human sleep EEG signals—A review. Biomed. Signal Process. Control 10, 21–33. https://doi.org/10.1016/j.bspc.2013.12.003

Mou, L., Lu, X., Li, X., Zhu, X.X., 2020. Nonlocal Graph Convolutional Networks for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 58, 8246–8257. https://doi.org/10.1109/TGRS.2020.2973363

Narayan, V., Mall, P.K., Alkhayyat, A., Abhishek, K., Kumar, S., Pandey, P., 2023. Enhance-Net: An Approach to Boost the Performance of Deep Learning Model Based on Real-Time Medical Images. J. Sens. 2023, e8276738. https://doi.org/10.1155/2023/8276738

Nguyen, T., Grishman, R., 2018. Graph Convolutional Networks With Argument-Aware Pooling for Event Detection. Proc. AAAI Conf. Artif. Intell. 32. https://doi.org/10.1609/aaai.v32i1.12039

Nguyen-Ky, T., Wen, P., Li, Y., 2013. Consciousness and Depth of Anesthesia Assessment Based on Bayesian Analysis of EEG Signals. IEEE Trans. Biomed. Eng. 60, 1488–1498. https://doi.org/10.1109/TBME.2012.2236649

Nguyen-Ky, T., Wen, P., Li, Y., Gray, R., 2011. Measuring and Reflecting Depth of Anesthesia Using Wavelet and Power Spectral Density. IEEE Trans. Inf. Technol. Biomed. 15, 630–639. https://doi.org/10.1109/TITB.2011.2155081

Norman, R.G., Pal, I., Stewart, C., Walsleben, J.A., Rapoport, D.M., 2000. Interobserver agreement among sleep scorers from different centers in a large dataset. Sleep 23, 901–908.

Ohayon, M.M., 2002. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med. Rev. 6, 97–111. https://doi.org/10.1053/smrv.2002.0186

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas. https://doi.org/10.48550/arXiv.1804.03999

Omidvar, M., Zahedi, A., Bakhshi, H., 2021a. EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers. J. Ambient Intell. Humaniz. Comput. 12, 10395–10403. https://doi.org/10.1007/s12652-020-02837-8

Omidvar, M., Zahedi, A., Bakhshi, H., 2021b. EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers. J. Ambient Intell. Humaniz. Comput. 12, 10395–10403. https://doi.org/10.1007/s12652-020-02837-8

Pei, W., Li, Y., Siuly, S., Wen, P., 2022. A hybrid deep learning scheme for multi-channel sleep stage classification. Comput. Mater. Contin. 71, 889–905. https://doi.org/10.32604/cmc.2022.021830

Peker, M., 2016. A new approach for automatic sleep scoring: Combining Taguchi based complex-valued neural network and complex wavelet transform. Comput. Methods Programs Biomed. 129, 203–216. https://doi.org/10.1016/j.cmpb.2016.01.001

Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P.J., Igel, C., 2021. U-Sleep: resilient high-frequency sleep staging. Npj Digit. Med. 4, 1–12. https://doi.org/10.1038/s41746-021-00440-5

Perslev, M., Jensen, M.H., Darkner, S., Jennum, P.J., Igel, C., 2019. U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging. https://doi.org/10.48550/arXiv.1910.11162

Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., De Vos, M., 2019. Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification. IEEE Trans. Biomed. Eng. 66, 1285–1296. https://doi.org/10.1109/TBME.2018.2872652

Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., Vos, M.D., 2018. Automatic Sleep Stage Classification Using Single-Channel EEG: Learning Sequential Features with Attention-Based Recurrent Neural Networks, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2018 40th Annual International Conference of the IEEE

Engineering in Medicine and Biology Society (EMBC), pp. 1452–1455. https://doi.org/10.1109/EMBC.2018.8512480

Phan, H., Mikkelsen, K., Chén, O.Y., Koch, P., Mertins, A., De Vos, M., 2022. SleepTransformer: Automatic Sleep Staging with Interpretability and Uncertainty Quantification. IEEE Trans. Biomed. Eng. 69, 2456–2467. https://doi.org/10.1109/TBME.2022.3147187

Polat, K., Güneş, S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl. Math. Comput. 187, 1017–1026. https://doi.org/10.1016/j.amc.2006.09.022

Puthankattil, S.D., Joseph, P.K., 2012. Classification of eeg signals in normal and depression conditions by ann using rwe and signal entropy. J. Mech. Med. Biol. 12, 1240019. https://doi.org/10.1142/S0219519412400192

Qassim, H., Verma, A., Feinzimer, D., 2018. Compressed residual-VGG16 CNN model for big data places image recognition, in: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). Presented at the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 169–175. https://doi.org/10.1109/CCWC.2018.8301729

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M., 2020. U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection. Pattern Recognit. 106, 107404. https://doi.org/10.1016/j.patcog.2020.107404

Qiu, X., Wang, S., Wang, R., Zhang, Y., Huang, L., 2023. A multi-head residual connection GCN for EEG emotion recognition. Comput. Biol. Med. 163, 107126. https://doi.org/10.1016/j.compbiomed.2023.107126

Qiu, Z., Yao, T., Mei, T., 2017. Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks, in: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, pp. 5534–5542. https://doi.org/10.1109/ICCV.2017.590

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106. https://doi.org/10.1007/BF00116251

Radha, M., Garcia-Molina, G., Poel, M., Tononi, G., 2014. Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2014 36th Annual International

Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1876–1880. https://doi.org/10.1109/EMBC.2014.6943976

Rahman, M.M., Bhuiyan, M.I.H., Hassan, A.R., 2018. Sleep stage classification using single-channel EOG. Comput. Biol. Med. 102, 211–220. https://doi.org/10.1016/j.compbiomed.2018.08.022

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. https://doi.org/10.48550/arXiv.1506.02640

Rényi, A., 1961. On measures of entropy and information, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. University of California Press, pp. 547–562.

Rhee, S., Seo, S., Kim, S., 2018. Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification. https://doi.org/10.48550/arXiv.1711.05859

Rodríguez-Sotelo, J.L., Osorio-Forero, A., Jiménez-Rodríguez, A., Cuesta-Frau, D., Cirugeda-Roldán, E., Peluffo, D., 2014. Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques. Entropy 16, 6573–6589. https://doi.org/10.3390/e16126573

Roebuck, A., Monasterio, V., Gederi, E., Osipov, M., Behar, J., Malhotra, A., Penzel, T., Clifford, G.D., 2013. A review of signals used in sleep analysis. Physiol. Meas. 35, R1. https://doi.org/10.1088/0967-3334/35/1/R1

Rohan, R., Kumari, L.V.R., 2021. Classification of Sleep Apneas using Decision Tree Classifier, in: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). Presented at the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1310–1316. https://doi.org/10.1109/ICICCS51141.2021.9432197

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev. 65, 386–408. https://doi.org/10.1037/h0042519

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. nature 323, 533–536. https://doi.org/10.1038/323533a0

Salama, E.S., El-Khoribi, R.A., Shoman, M.E., Shalaby, M.A.W., 2018. EEG-based emotion recognition using 3D convolutional neural networks. Int J Adv Comput Sci Appl 9, 329–337.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2009a. The Graph Neural Network Model. IEEE Trans. Neural Netw. 20, 61–80. https://doi.org/10.1109/TNN.2008.2005605

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2009b. Computational Capabilities of Graph Neural Networks. IEEE Trans. Neural Netw. 20, 81–102. https://doi.org/10.1109/TNN.2008.2005141

See, A.R., Liang, C.-K., 2011. A study on sleep EEG Using sample entropy and power spectrum analysis, in: 2011 Defense Science Research Conference and Expo (DSR), 2011 Defense Science Research Conference and Expo (DSR). IEEE, pp. 1–4.

Sekkal, R.N., Bereksi-Reguig, F., Ruiz-Fernandez, D., Dib, N., Sekkal, S., 2022. Automatic sleep stage classification: From classical machine learning methods to deep learning. Biomed. Signal Process. Control 77, 103751. https://doi.org/10.1016/j.bspc.2022.103751

Şen, B., Peker, M., Çavuşoğlu, A., Çelebi, F.V., 2014. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. J. Med. Syst. 38, 18. https://doi.org/10.1007/s10916-014-0018-0

Sen, D., Mishra, B.B., Pattnaik, P.K., 2023. A Review of the Filtering Techniques used in EEG Signal Processing, in: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI). Presented at the 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 270–277. https://doi.org/10.1109/ICOEI56765.2023.10125857

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shi, W., Rajkumar, R., 2020. Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1711–1719.

Shiao, H.-T., Cherkassky, V., Lee, J., Veber, B., Patterson, E.E., Brinkmann, B.H., Worrell, G.A., 2017. SVM-Based System for Prediction of Epileptic Seizures From iEEG Signal. IEEE Trans. Biomed. Eng. 64, 1011–1022. https://doi.org/10.1109/TBME.2016.2586475

Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. IEEE Access 9, 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition.

Siuly, N.A., Li, Y., Wen, P., 2013. Identification of motor imagery tasks through CC-LR algorithm in brain computer interface. Int. J. Bioinforma. Res. Appl. 9, 156. https://doi.org/10.1504/IJBRA.2013.052447

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., Payen, J.-F., 2018. A convolutional neural network for sleep stage scoring from raw single-channel EEG. Biomed. Signal Process. Control 42, 107–114. https://doi.org/10.1016/j.bspc.2017.12.001

Supratak, A., Dong, H., Wu, C., Guo, Y., 2017. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 25, 1998–2008. https://doi.org/10.1109/TNSRE.2017.2721116

Supratak, A., Guo, Y., 2020. TinySleepNet: An Efficient Deep Learning Model for Sleep Stage Scoring based on Raw Single-Channel EEG, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Presented at the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 641–644. https://doi.org/10.1109/EMBC44109.2020.9176741

Šušmáková, K., Krakovská, A., 2008. Discrimination ability of individual measures used in sleep stages classification. Artif. Intell. Med. 44, 261–277. https://doi.org/10.1016/j.artmed.2008.07.005

Sweeney, K.T., Ward, T.E., McLoone, S.F., 2012. Artifact Removal in Physiological Signals—Practices and Possibilities. IEEE Trans. Inf. Technol. Biomed. 16, 488–500. https://doi.org/10.1109/TITB.2012.2188536

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in:

2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tagluk, M.E., Sezgin, N., Akin, M., 2010. Estimation of Sleep Stages by an Artificial Neural Network Employing EEG, EMG and EOG. J. Med. Syst. 34, 717–725. https://doi.org/10.1007/s10916-009-9286-5

Taran, S., Bajaj, V., 2019. Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method. Comput. Methods Programs Biomed. 173, 157–165. https://doi.org/10.1016/j.cmpb.2019.03.015

Tsinalis, O., Matthews, P.M., Guo, Y., Zafeiriou, S., 2016. Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. ArXiv Prepr. ArXiv161001683.

Tsiouris, K.M., Pezoulas, V.C., Zervakis, M., Konitsiotis, S., Koutsouris, D.D., Fotiadis, D.I., 2018. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. Comput. Biol. Med. 99, 24–37. https://doi.org/10.1016/j.compbiomed.2018.05.019

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is All you Need, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Vatankhah, M., Akbarzadeh-T., M.-R., Moghimi, A., 2010. An intelligent system for diagnosing sleep stages using wavelet coefficients, in: The 2010 International Joint Conference on Neural Networks (IJCNN). Presented at the 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, Barcelona, Spain, pp. 1–5. https://doi.org/10.1109/IJCNN.2010.5596732

Vural, C., Yildiz, M., 2010. Determination of Sleep Stage Separation Ability of Features Extracted from EEG Signals Using Principle Component Analysis. J. Med. Syst. 34, 83–89. https://doi.org/10.1007/s10916-008-9218-9

Wan, S., Gong, C., Zhong, P., Du, B., Zhang, L., Yang, J., 2020. Multiscale Dynamic Graph Convolutional Network for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 58, 3162–3177. https://doi.org/10.1109/TGRS.2019.2949180

Wang, W., Xiao, Q., Xue, J., Hu, Z., 2023. DynamicSleepNet: a multi-exit neural network with adaptive inference time for sleep stage classification. Front. Physiol. 14. https://doi.org/10.3389/fphys.2023.1171467

Wang, Y., Huang, Z., McCane, B., Neo, P., 2018. EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition, in: 2018 International Joint Conference on Neural Networks (IJCNN). Presented at the 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. https://doi.org/10.1109/IJCNN.2018.8489715

Wang, Z., Yang, J., Sawan, M., 2021. A Novel Multi-scale Dilated 3D CNN for Epileptic Seizure Prediction, in: 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS). Presented at the 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), pp. 1–4. https://doi.org/10.1109/AICAS51828.2021.9458571

Weber, F., Dan, Y., 2016. Circuit-based interrogation of sleep control. Nature 538, 51–59. https://doi.org/10.1038/nature19773

Welch, P., 1967. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoustics 15, 70–73. https://doi.org/10.1109/TAU.1967.1161901

Weng, Z., Qin, Z., Tao, X., Pan, C., Liu, G., Li, G.Y., 2023. Deep Learning Enabled Semantic Communications with Speech Recognition and Synthesis. IEEE Trans. Wirel. Commun. 1–1. https://doi.org/10.1109/TWC.2023.3240969

Wu, H., Wang, S., Fang, H., 2021. LP-UIT: A Multimodal Framework for Link Prediction in Social Networks, in: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Presented at the 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 742–749. https://doi.org/10.1109/TrustCom53373.2021.00108

Wu, T., Kong, X., Zhong, Y., Chen, L., 2022. Automatic detection of abnormal EEG signals using multiscale features with ensemble learning. Front. Hum. Neurosci. 16, 943258. https://doi.org/10.3389/fnhum.2022.943258

Xie, Songyun, Li, Yabing, Xie, X., Wang, W., Duan, X., 2017. The Analysis and Classify of Sleep Stage Using Deep Learning Network from Single-Channel EEG Signal, in: Liu, D., Xie, Shengli, Li, Yuanqing, Zhao, D., El-Alfy, E.-S.M. (Eds.),

Neural Information Processing, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 752–758. https://doi.org/10.1007/978-3-319-70093-9_80

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., Jegelka, S., 2018. Representation Learning on Graphs with Jumping Knowledge Networks, in: Proceedings of the 35th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 5453–5462.

Yildirim, O., Baloglu, U.B., Acharya, U.R., 2019. A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals. Int. J. Environ. Res. Public. Health 16, 599. https://doi.org/10.3390/ijerph16040599

Yu, S., Chen, X., Wang, B., Wang, X., 2012. Automatic sleep stage classification based on ECG and EEG features for day time short nap evaluation, in: Proceedings of the 10th World Congress on Intelligent Control and Automation. Presented at the Proceedings of the 10th World Congress on Intelligent Control and Automation, pp. 4974–4977. https://doi.org/10.1109/WCICA.2012.6359421

Yuan, W., Xiang, W., Si, K., Yang, C., Zhao, L., Li, J., Liu, C., 2023. Multi-channel EEG-based sleep staging using brain functional connectivity and domain adaptation. Physiol. Meas. 44, 105007. https://doi.org/10.1088/1361-6579/ad02db

Yubo, Z., Yingying, L., Bing, Z., Lin, Z., Lei, L., 2022. MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging. Front. Neurosci. 16. https://doi.org/10.3389/fnins.2022.973761

Yulita, I.N., Fanany, M.I., Arymuthy, A.M., 2017. Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification. Procedia Comput. Sci. 116, 530–538. https://doi.org/10.1016/j.procs.2017.10.042

Zhang, J., Wu, Y., 2017. A New Method for Automatic Sleep Stage Classification. IEEE Trans. Biomed. Circuits Syst. 11, 1097–1110. https://doi.org/10.1109/TBCAS.2017.2719631

Zhao, B., Lu, H., Chen, S., Liu, J., Wu, D., 2017. Convolutional neural networks for time series classification. J. Syst. Eng. Electron. 28, 162–169. https://doi.org/10.21629/JSEE.2017.01.18

Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., Sun, L., 2019. A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification. IEEE Trans. Neural Syst. Rehabil. Eng. 27, 2164–2177. https://doi.org/10.1109/TNSRE.2019.2938295

Zhao, Y., Lin, X., Zhang, Z., Wang, X., He, X., Yang, L., 2023. STDP-based adaptive graph convolutional networks for automatic sleep staging. Front. Neurosci. 17, 1158246. https://doi.org/10.3389/fnins.2023.1158246

Zhao, Y., Yang, J., Lin, J., Yu, D., Cao, X., 2020. A 3D Convolutional Neural Network for Emotion Recognition based on EEG Signals, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–6.

Zhou, H., Zhang, Shanghang, Peng, J., Zhang, Shuai, Li, J., Xiong, H., Zhang, W., 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. https://doi.org/10.48550/arXiv.2012.07436

Zhu, G., Li, Y., Wen, P., 2014a. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE J. Biomed. Health Inform. 18, 1813–1821. https://doi.org/10.1109/JBHI.2014.2303991

Zhu, G., Li, Y., Wen, P.P., Wang, S., 2014b. Analysis of alcoholic EEG signals based on horizontal visibility graph entropy. Brain Inform. 1, 19–25. https://doi.org/10.1007/s40708-014-0003-x

Zhu, T., Luo, W., Yu, F., 2020. Convolution- and Attention-Based Neural Network for Automated Sleep Stage Classification. Int. J. Environ. Res. Public. Health 17, 4152. https://doi.org/10.3390/ijerph17114152

Zoubek, L., Charbonnier, S., Lesecq, S., Buguet, A., Chapotot, F., 2007. Feature selection for sleep/wake stages classification using data driven methods. Biomed. Signal Process. Control, IFAC Symposia on Biomedical Systems Modelling & Control 2, 171–179. https://doi.org/10.1016/j.bspc.2007.05.005

APPENDIX A

The completed source code of Paper 1 was uploaded to Github: https://github.com/XiaopengJi-USQ/JK-STGCN.

APPENDIX B

The source code of the 3DSleepNet model file for Paper 2 was uploaded to Github: https://github.com/XiaopengJi-USQ/3DSleepNet.

APPENDIX C

The source code of the MixSleepNet model for Paper 3 was uploaded to Github: https://github.com/XiaopengJi-USQ/MixSleepNet.