Computer Methods and Programs in Biomedicine 244 (2024) 107992

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

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

Xiaopeng Ji [a,*], Yan Li [a], Peng Wen [b], Prabal Barua [c], U Rajendra Acharya [a]

[a] School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[b] School of Engineering, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[c] Cogninet Brain Team, Sydney, NSW 2010, Australia

ARTICLE INFO

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

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.

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

* Corresponding author.
E-mail addresses: xiaopeng.ji@usq.edu.au (X. Ji), yan.li@usq.edu.au (Y. Li), paul.wen@usq.edu.au (P. Wen), prabal.barua@usq.edu.au (P. Barua), rajendra.acharya@usq.edu.au (U.R. Acharya).

https://doi.org/10.1016/j.cmpb.2023.107992
Received 5 September 2023; Received in revised form 9 December 2023; Accepted 19 December 2023
Available online 27 December 2023

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 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 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.

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, Nchannel 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 Nchannel 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 byτ. 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].

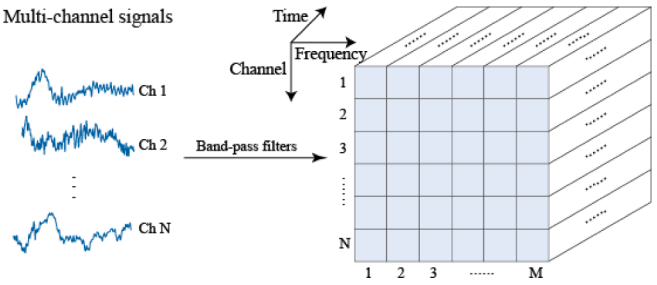

Fig. 2. For each of the N channels, M filters are applied to extract pertinent band waves. Subsequently, 3D temporal features are extracted from the filtered band waves by means of down-sampling.

2.2. 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]:

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

where w_{ijm}^{pqr} is the (p, q, r)-th value of the kernel connected to the m-th

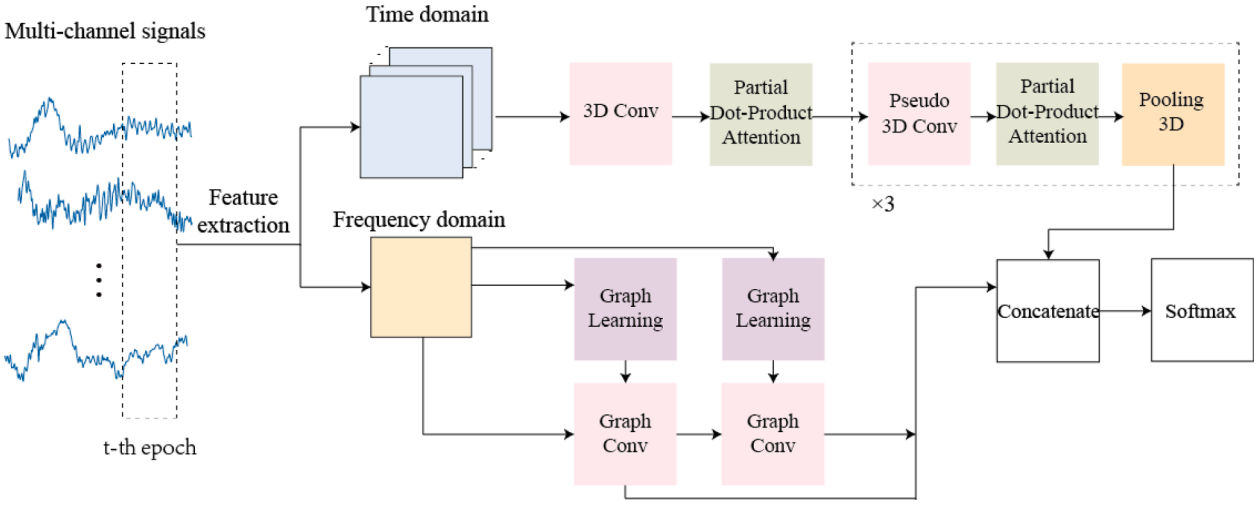

Fig. 1. The holistic structure of MixsleepNet encompasses the incorporation of time domain and frequency domain features, which are input into the model. Time domain features are acquired through the down-sampling of the initial signals, while frequency domain features are derived via differential entropy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

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:

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

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:

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

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 $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_F)^T \in \mathbb{R}^{F\times1}$ are iteratively updated by minimizing the ensuing loss function:

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

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:

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

where the parameterβ represents the strength of L2 regularization applied to the adjacency matrix *A*. $\mathscr{L}_{\text{adaptive\_graph}}$ denotes the loss function defined in equation (5) and $\mathscr{L}_{\text{cross\_entropy}}$ 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:

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

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.

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 1.

In this study, we assessed the classification performance by calculating several evaluation metrics, including *Accuracy* (ACC), *F*1-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 2. 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 be made available on GitHub (https://github.com/XiaopengJi-USQ) upon publication of the paper.

Table 1
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	10-70 Hz
ECG	X2	-

Table 2
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

3.2. 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 3 and Table 4, 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.

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,

Table 3
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 4
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.

Table 5
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.

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 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 5 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 5 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. 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.

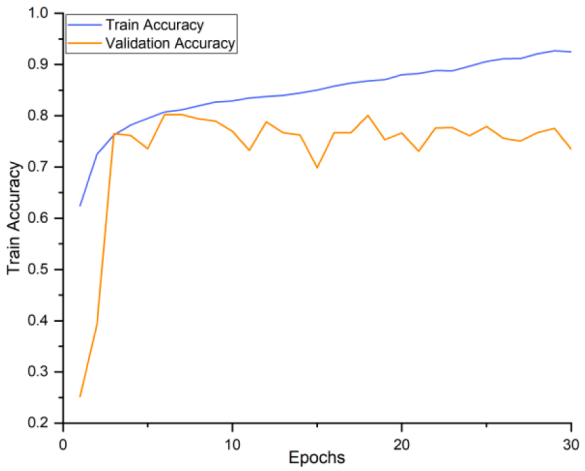

Fig. 4. An example of training curve and validation curve on the ISRUC-S3.

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 6. 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 7. The classification performance of each experiment is presented in Table 7 and Fig. 5. The single channel selections are 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

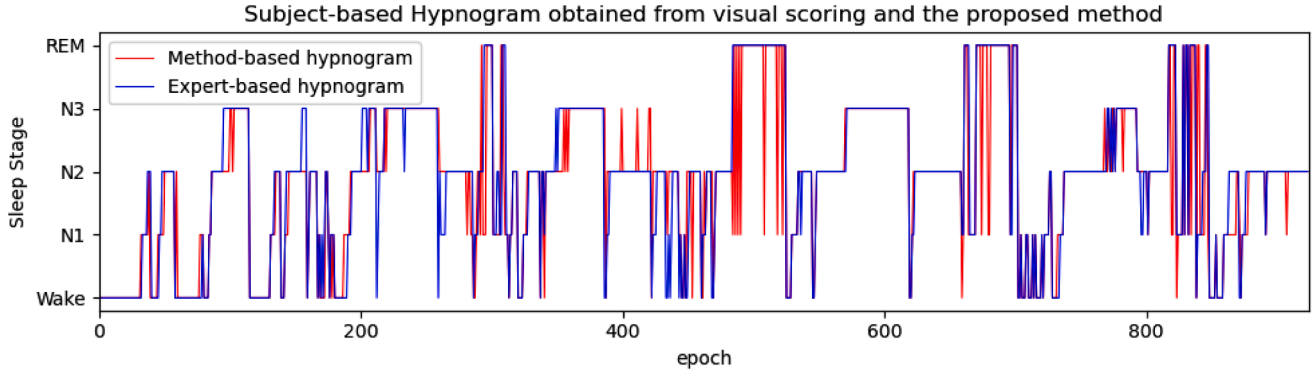

Fig. 3. Comparison between manual score annotated by expert 1 (blue) and automated scoring generated by the proposed model (red) for subject one from the ISRUC-S3 dataset.

Table 6
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

Table 7
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**

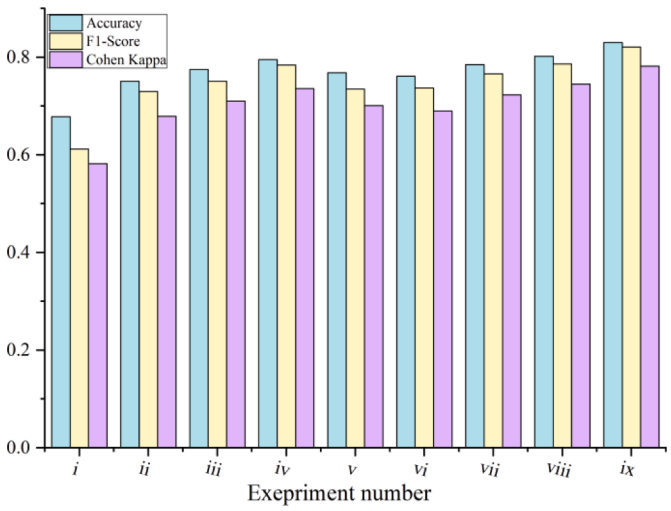

Fig. 5. Comparison of performance for different experiments.

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 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

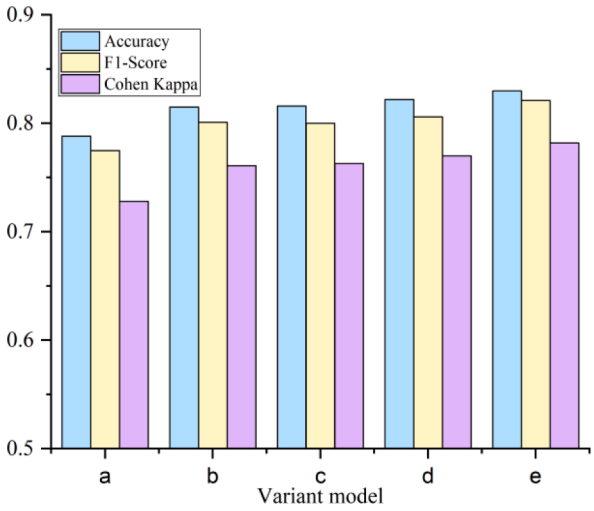

Fig. 6. Comparison of performances of different models.

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].

CRediT authorship contribution statement

Xiaopeng Ji: Writing – original draft, Methodology. **Yan Li:** Writing – review & editing. **Peng Wen:** Writing – review & editing. **Prabal Barua:** Writing – review & editing. **U Rajendra Acharya:** Writing – review & editing.

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, Y. Dan, Circuit-based interrogation of sleep control, Nature 538 (7623) (Oct. 2016) 51–59, https://doi.org/10.1038/nature19773.
[2] A. Ivanenko, B.R. Gururaj, Classification and epidemiology of sleep disorders, Child Adolesc. Psychiatr. Clin. 18 (4) (Oct. 2009) 839–848, https://doi.org/10.1016/j.chc.2009.04.005.
[3] J. Zhang, Y. Wu, A new method for automatic sleep stage classification, IEEE Trans. Biomed. Circuits Syst. 11 (5) (2017) 1097–1110, https://doi.org/10.1109/TBCAS.2017.2719631.
[4] S. Siuly, Y. Li, Discriminating the brain activities for brain–computer interface applications through the optimal allocation-based approach, Neural Comput. Appl. 26 (4) (May 2015) 799–811, https://doi.org/10.1007/s00521-014-1753-3.
[5] R.B. Berry, R. Brooks, C.E. Gamaldo, S.M. Harding, C. Marcus, B.V. Vaughn, The AASM manual for the scoring of sleep and associated events, Rules Terminol. Tech. Specif. Darien Ill. Am. Acad. Sleep Med. 176 (2012) 2012.
[6] K.A.I. Aboalayon, H.T. Ocbagabir, 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, https://doi.org/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, U.R. Acharya, Computerized detection of cyclic alternating patterns of sleep: a new paradigm, future scope and challenges, Comput. Methods Programs Biomed. (2023) 107471, https://doi.org/10.1016/j.cmpb.2023.107471.
[9] S. Khalighi, T. Sousa, G. Pires, U. Nunes, Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels, Expert Syst. Appl. 40 (17) (Dec. 2013) 7046–7059, https://doi.org/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. 9 (1) (Dec. 2018), https://doi.org/10.1038/s41467-018-07229-3. Art. no. 1.
[11] S. Roy, K. Kate, M. Hirzel, A semi-supervised deep learning algorithm for abnormal EEG identification, ArXiv Prepr (2019). *ArXiv190307822*.
[12] O. Faust, H. Razaghi, R. Barika, E.J. Ciaccio, U.R. Acharya, A review of automated sleep stage scoring based on physiological signals for the new millennia, Comput. Methods Programs Biomed. 176 (2019) 81–91, https://doi.org/10.1016/j.cmpb.2019.04.032.
[13] S.K. Khare, U.R. Acharya, Adazd-Net: automated adaptive and explainable Alzheimer's disease detection system using EEG signals, Knowl.-Based Syst 278 (Oct. 2023) 110858, https://doi.org/10.1016/j.knosys.2023.110858.
[14] S.K. Khare, S. March, P.D. Barua, V.M. Gadre, 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 99 (Nov. 2023) 101898, https://doi.org/10.1016/j.inffus.2023.101898.
[15] S.K. Khare, V. Blanes-Vidal, E.S. Nadimi, U.R. Acharya, Emotion recognition and artificial intelligence: a systematic review (2014–2023) and research recommendations, Inf. Fusion 102 (Feb. 2024) 102019, https://doi.org/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. 66 (9) (Sep. 2019) 3504–3516, https://doi.org/10.1109/TCSI.2019.2927839.
[17] I.A. Zapata, Y. Li, P. Wen, Rules-based and SVM-Q methods with multitapers and convolution for sleep EEG stages classification, IEEE Access 10 (Jan. 2022) 71299–71310, https://doi.org/10.1109/ACCESS.2022.3188286.
[18] E. Alickovic, A. Subasi, Ensemble SVM method for automatic sleep stage classification, IEEE Trans. Instrum. Meas. 67 (6) (2018) 1258–1265.
[19] M.M. Rahman, M.I.H. Bhuiyan, A.R. Hassan, Sleep stage classification using single-channel EOG, Comput. Biol. Med. 102 (Nov. 2018) 211–220, https://doi.org/10.1016/j.compbiomed.2018.08.022.
[20] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, 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. 108 (1) (2012) 10–19, https://doi.org/10.1016/j.cmpb.2011.11.005.

[21] S.K. Khare, V. Bajaj, Constrained based tunable Q wavelet transform for efficient decomposition of EEG signals, Appl. Acoust. 163 (Jun. 2020) 107234, https://doi.org/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. 10 (24) (2020) 8963, https://doi.org/10.3390/app10248963.

[23] J. Zhang, Y. Wu, Complex-valued unsupervised convolutional neural networks for sleep stage classification, Comput. Methods Programs Biomed. 164 (Oct. 2018) 181–191, https://doi.org/10.1016/j.cmpb.2018.07.015.

[24] H.W. Loh, C.P. Ooi, S.G. Dhok, M. Sharma, A.A. Bhurane, U.R. Acharya, Automated detection of cyclic alternating pattern and classification of sleep stages using deep neural network, Appl. Intell. 52 (3) (Feb. 2022) 2903–2917, https://doi.org/10.1007/s10489-021-02597-8.

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

[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, U.R. Acharya, SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach, PLoS ONE 14 (5) (May 2019) e0216456, https://doi.org/10.1371/journal.pone.0216456.

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

[29] Q. Cai, Z. Gao, J. An, S. Gao, 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 68 (2) (2021) 777–781, https://doi.org/10.1109/TCSII.2020.3014514.

[30] S. Biswal, et al., SLEEPNET: automated Sleep Staging System via Deep Learning,, arXiv (Jul. 2017), https://doi.org/10.48550/arXiv.1707.08262. arXiv:1707.08262.

[31] S. Chambon, M.N. Galtier, P.J. Arnal, G. Wainrib, A. Gramfort, A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (4) (Apr. 2018) 758–769, https://doi.org/10.1109/TNSRE.2018.2813138.

[32] C. Zhao, J. Li, Y. Guo, SleepContextNet: a temporal context network for automatic sleep staging based single-channel EEG, Comput. Methods Programs Biomed. 220 (Jun. 2022) 106806, https://doi.org/10.1016/j.cmpb.2022.106806.

[33] A. Supratak, H. Dong, C. Wu, Y. Guo, DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (11) (2017) 1998–2008, https://doi.org/10.1109/TNSRE.2017.2721116.

[34] A. Supratak, 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, https://doi.org/10.1109/EMBC44109.2020.9176741.

[35] N. Michielli, U.R. Acharya, F. Molinari, Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals, Comput. Biol. Med. 106 (Mar. 2019) 71–81, https://doi.org/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, 2020, pp. 1324–1330. Proceedings of the Twenty-Ninth International Joint Conference on Artificial IntelligenceIJCAI.

[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. 29 (2021) 1977–1986.

[38] X. Ji, Y. Li, P. Wen, Jumping knowledge based spatial-temporal graph convolutional networks for automatic sleep stage classification, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1464–1472, https://doi.org/10.1109/TNSRE.2022.3176004.

[39] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-S. Cheng, A rule-based automatic sleep staging method, J. Neurosci. Methods 205 (1) (Mar. 2012) 169–176, https://doi.org/10.1016/j.jneumeth.2011.12.022.

[40] G. Zhu, Y. Li, P. Wen, Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal, IEEE J. Biomed. Health Inform. 18 (6) (2014) 1813–1821, https://doi.org/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 7 (2019) 15945–15954, https://doi.org/10.1109/ACCESS.2019.2895133.

[42] P. Memar, F. Faradji, A novel multi-class EEG-based sleep stage classification system, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (1) (2018) 84–95, https://doi.org/10.1109/TNSRE.2017.2776149.

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

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

[45] M. Diykh, Y. Li, P. Wen, EEG sleep stages classification based on time domain features and structural graph similarity, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (11) (2016) 1159–1168.

[46] R. Sharma, R.B. Pachori, A. Upadhyay, Automatic sleep stages classification based on iterative filtering of electroencephalogram signals, Neural Comput. Appl. 28 (10) (2017) 2959–2978, https://doi.org/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, 10/d62rq9.

[48] A. Stochholm, K. Mikkelsen, P. Kidmose, Automatic sleep stage classification using ear-EEG. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 4751–4754, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

[49] B. Fatimah, A. Singhal, P. Singh, A multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning, Comput. Biol. Med. 148 (Sep. 2022) 105877, https://doi.org/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. 112 (Feb. 2023) 103956, https://doi.org/10.1016/j.medengphy.2023.103956.

[51] O. Tsinalis, P.M. Matthews, Y. Guo, Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders, Ann. Biomed. Eng. 44 (5) (2016) 1587–1597, https://doi.org/10.1007/s10439-015-1444-y.

[52] T.F. Zaidi, O. Farooq, EEG sub-bands based sleep stages classification using Fourier Synchrosqueezed transform features, Expert Syst. Appl. 212 (Feb. 2023) 118752, https://doi.org/10.1016/j.eswa.2022.118752.

[53] M. Sharma, D. Goyal, P.V. Achuth, 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. 98 (Jul. 2018) 58–75, https://doi.org/10.1016/j.compbiomed.2018.04.025.

[54] J. Zhang, R. Yao, W. Ge, J. Gao, Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG, Comput. Methods Programs Biomed. 183 (Jan. 2020) 105089, https://doi.org/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, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231, https://doi.org/10.1109/TPAMI.2012.59.

[57] Z. Qiu, T. Yao, T. Mei, Learning Spatio-temporal representation with pseudo-3D residual networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, Oct. 2017, pp. 5534–5542, https://doi.org/10.1109/ICCV.2017.590.

[58] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778, in Proceedings of the IEEE conference on computer vision and pattern recognition.

[59] S. Khalighi, T. Sousa, J.M. Santos, U. Nunes, ISRUC-Sleep: a comprehensive public dataset for sleep researchers, Comput. Methods Programs Biomed. 124 (2016).

[60] L. Song, Y. Li, N. Lu, ProfileSR-GAN: a GAN based super-resolution method for generating high-resolution load profiles, arXiv (2022), https://doi.org/10.48550/arXiv.2107.09523. Jan. 26.

[61] E. Khalili, 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. 204 (Jun. 2021) 106063, https://doi.org/10.1016/j.cmpb.2021.106063.

[62] M. Sharma, P. Makwana, R.S. Chad, 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. 53 (15) (Aug. 2023) 18681–18699, https://doi.org/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 25 (10) (Oct. 2018) 1351–1358, https://doi.org/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 101 (23) (Jun. 2000) E215–E220, https://doi.org/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 2 (6) (Nov. 2001) 537–553, https://doi.org/10.1016/S1389-9457(01)00149-6.

[66] S.F. Quan, et al., The sleep heart health study: design, rationale, and methods, Sleep 20 (12) (Dec. 1997) 1077–1085.

[67] M. Dutt, S. Redhu, M. Goodwin, C.W. Omlin, SleepXAI: an explainable deep learning approach for multi-class sleep stage identification, Appl. Intell. 53 (13) (Jul. 2023) 16830–16843, https://doi.org/10.1007/s10489-022-04357-8.