A Multi-Domain Adaptive Graph Convolutional Network for EEG-based Emotion Recognition

Rui Li
Shanghai Jiao Tong University
Shanghai, China
realee@sjtu.edu.cn

Yiting Wang
Shanghai Jiao Tong University
Shanghai, China
yw883@cornell.edu

Bao-Liang Lu*
Shanghai Jiao Tong University
Shanghai, China
bllu@sjtu.edu.cn

ABSTRACT

Among all solutions of emotion recognition tasks, electroencephalogram (EEG) is a very effective tool and has received broad attention from researchers. In addition, information across multimedia in EEG often provides a more complete picture of emotions. However, few of the existing studies concurrently incorporate EEG information from temporal domain, frequency domain and functional brain connectivity. In this paper, we propose a Multi-Domain Adaptive Graph Convolutional Network (MD-AGCN), fusing the knowledge of both the frequency domain and the temporal domain to fully utilize the complementary information of EEG signals. MD-AGCN also considers the topology of EEG channels by combining the inter-channel correlations with the intra-channel information, from which the functional brain connectivity can be learned in an adaptive manner. Extensive experimental results demonstrate that our model exceeds state-of-the-art methods in most experimental settings. At the same time, the results show that MD-AGCN could extract complementary domain information and exploit channel relationships for EEG-based emotion recognition effectively.

CCS CONCEPTS

• **Human-centered computing** → *HCI design and evaluation methods*; • **Computing methodologies** → *Artificial intelligence; Philosophical/theoretical foundations of artificial intelligence; Cognitive science.*

KEYWORDS

affective computing; EEG-based emotion recognition; adaptive graph convolutional network; functional brain connectivity

ACM Reference Format:
Rui Li, Yiting Wang, and Bao-Liang Lu. 2021. A Multi-Domain Adaptive Graph Convolutional Network for EEG-based Emotion Recognition. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3474085.3475697

*Corresponding author.

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3475697

1 INTRODUCTION

Emotion recognition plays an important role in affective brain computer interface (aBCI) [23] and mental health assessment [2]. Although it is straightforward for humans to understand others' emotions, this task is still a challenge in affective computing. Moreover, many mental disorders are related to emotions, and thus accurately evaluating patients' emotion states will enhance the treatment of neurological diseases [2]. In consequence, this field has drawn growing attention from researchers in recent years. Many modalities and their combinations have been adopted to analyze human emotions, including facial expressions, eye movements [16, 17, 33–35], skin conductance responses, electrocardiograms (ECGs) and electroencephalographs (EEGs) [25, 36, 37], etc.. Among them, EEG signals are collected by placing electrodes on the scalp to measure voltage fluctuations from the cortex in the brain [22]. It has been proven that EEG signals have the ability to reveal delicate changes in emotions with high time resolution [5], which is more objective and precise when analyzing emotion states [37].

Traditional EEG features can be divided into three categories: the time domain, frequency domain and time-frequency domain. In the time domain, the most widely used EEG features are Hjorth Features, Fractal Dimension (FD) and Higher Order Crossings (HOC) [8]. EEG signals are discrete time series indicating that the time domain could contain essential information for emotion recognition [15]. Due to the non-stationary characteristics of EEG signals as well as the interference of noise and artifacts of raw EEG data, frequency domain features such as Power Spectral Density (PSD) [4] and Differential Entropy (DE) [3] , and time-frequency domain features such as Hilbert-Huang Spectrum (HHS) [8] have shown good performance in EEG-based emotion recognition. In addition, growing evidences [7, 20, 27, 30] have shown that functional brain connectivity is associated with multiple psychophysiological disorders of cognitive deficit disease. Mauss [19] proposed that emotional procedures should be recognized as involving distributed circuits instead of specific brain regions in isolation. Therefore, the functional brain connectivity network should not be neglected while recognizing emotions.

Existing studie have achieved many compelling outcomes based on deep neural networks. Zhang *et al.* [31] introduced a sparse dynamic graph convolutional neural network model with frequency domain features. Li *et al.* [15] utilized intrinsic spatial relationships of EEG channels by a hierarchical neural network model. Zhong *et al.* [38] proposed a regularized graph neural network to exploit the inter-channel relationship of EEG channels with DE features for EEG-based emotion recognition. Although their work accomplishes outstanding ability for recognizing emotions, few of the existing

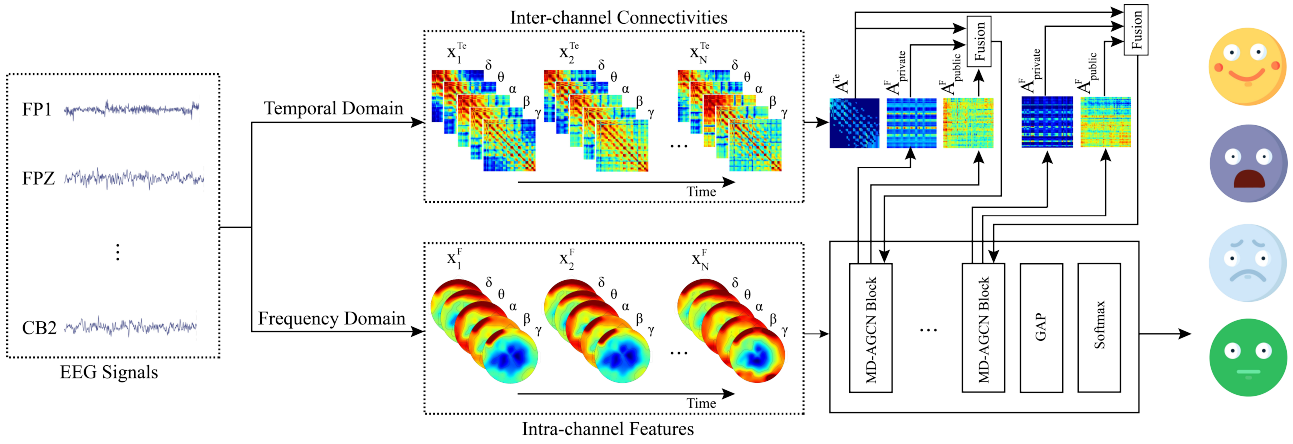

Figure 1: The overall process of our proposed MD-AGCN for EEG-based emotion recognition. The MD-AGCN utilizes complimentary information from two domains of EEG signals. The functional brain connectivity calculated in the temporal domain and the frequency functional brain connectivity learned by the model are fused to recognize the emotion classes.

studies combine the temporal domain, frequency domain of EEG signals and emotion-related functional brain connectivity together.

In this paper, we propose a Multi-Domain Adaptive Graph Convolutional Network (MD-AGCN) to thoroughly exploit the complementary knowledge between different domains of EEG signals and the topological structure of EEG channels. We construct an emotional graph of the brain where the vertexes of the graph represented by EEG channels are linked by inter-channel edges to perform functional brain connectivity as the topology of the graph. Furthermore, emotion-associated functional brain connectivity can be learned by our model in an adaptive manner.

The main contributions of this paper are as follows:

- We propose a Multi-Domain Adaptive Graph Convolutional Network that fuses the complementary information of EEG signals in the temporal domain and the frequency domain with the topological structure of EEG channels.
- The inter-channel correlation and the intra-channel information of EEG channels are integrated for emotion recognition in our model. Moreover, topological structure can be adaptively learned to reveal emotion-related functional brain connectivity.
- We conduct extensive experiments on three datasets: SEED [36], SEED-IV [35], and SEED-V [12, 33]. Experimental results show significant improvement of our model on classification problems with more than three classes.

2 RELATED WORK

In this section, we review the related work in terms of EEG-based emotion recognition, graph convolutional neural networks and emotion-related functional connectivity.

2.1 EEG-based Emotion Recognition

Classifiers for EEG-based emotion recognition have been developed substantially from traditional machine learning algorithms to deep learning methods. Wang *et al.* [26] implemented a support vector machine (SVM) in the frequency domain to identify different classes of emotions. Bahari *et al.* [1] practiced a non-linear k nearest neighbor classifier (kNN) based on a recurrence plot. Deep learning methods provide computational models comprised of various processing layers that perform automatic feature extraction [10]. A deep belief network was proposed to identify critical frequency bands and channels that achieved higher accuracy than traditional machine learning methods [36]. Furthermore, researchers began to exploit spatial information to utilize the topological structure of EEG channels. Zhang *et al.* [32] proposed a recurrent neural network (RNN) to learn spatial-temporal information from EEG signals. Jia *et al.* [9] proposed an attention 3D dense network with fusing EEG features in the time domain and the frequency domain. Although their work accounts for one or more types of features, they do not consider the topological structure of EEG electrodes with time and frequency domains simultaneously.

2.2 Graph Convolutional Neural Network

Graph convolutional neural networks (GCNNs) extend spectral theory in traditional CNN algorithms, and have been applied in various fields. For example, Shi *et al.* [24] proposed an adaptive graph convolutional network for skeleton-based action recognition that achieves significant improvement in the field by adding adaptive graph convolutional layers and second-order information. Due to the outstanding performance of graph convolutional neural networks, an increasing number of researchers extend this algorithm to emotion recognition because it is an efficient way to utilize the inner relationship of EEG channels [25]. Song *et al.* [25] proposed a dynamic graph convolutional neural network (DGCNN) to dynamically learn the intrinsic relationships between different EEG channels. In his method, the adjacency matrix is used to learn more discriminative features while training. Zhang *et al.* [31] added a sparseness constraint to the DGCNN model to improve the performance. The study of Zhong *et al.* [38] explored important brain regions and inter-channel relations for EEG-based emotion recognition by a regularized graph neural network. Yin *et al.* [29] fused

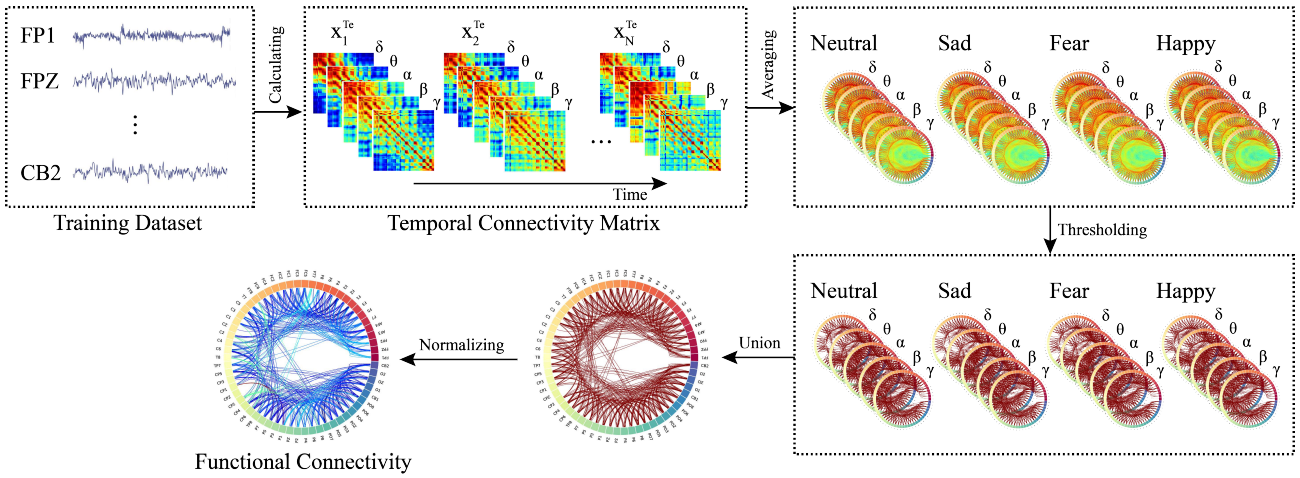

Figure 2: Construction of the functional brain connectivity on the temporal domain. Temporal connectivity matrices are calculated from the EEG signals in the training dataset, from which emotion-relevant connections are picked by averaging and applying thresholds over each emotion class. The final temporal functional brain connectivity is constructed by uniting and normalizing the emotion-relevant connections.

LSTM and graph convolutional neural networks with frequency features.

2.3 Emotion Related Functional Connectivity

Brain connections have always been used in the fields of neuroscience and neuroimaging to study and explore the nature of the brain. Murias *et al.* [20] found that there exists an obviously robust EEG connection among patients with autism spectrum disorder under a resting state. Yin *et al.* [30] concluded that functional brain connections based on EEG signals are usually slower and inefficient in patients with schizophrenia. Ho *et al.* [7] pointed out that depression in adolescence is related to the erratic increase in network connections based on the default mode shown by functional magnetic resonance imaging. Whitton *et al.* [27] considered that the increase in the high frequency band based on the functional connection of EEG signals could be a neural model of the recurrent course of major depression.

Therefore, it is important to explore the functional brain network that is related to EEG-based emotion recognition. In addition, the functional connectivity of EEG channels is informative for emotion recognition, and some studies have proposed various methods to address this proposition. Wu *et al.* [28] developed a critical subnetwork method that uses topological features to identify emotion states based on EEG signals. Their results reveal that topological features of the temporal domain outperform those of the frequency domain in the critical subnetwork method. Wu *et al.* [28] exploited the topological properties of the brain network, however, temporal and frequency domains were separated. Zhong *et al.* [38] proposed a regularized graph neural network to investigate inter-channel relations for emotion recognition with pre-computed frequency features from EEG signals, which utilized the frequency domain information from EEG data. Although the aforementioned existing studies achieved excellent results, none of them combined EEG information from the temporal domain and the frequency domain with topological structure of EEG channels, which might not fully take advantage of complicated EEG signals.

3 PRELIMINARIES

We define the EEG connectivity matrices, which containing N samples in the temporal domain as $\mathrm{X}^{Te} = \left(\mathrm{x}_1^{Te}, \mathrm{x}_2^{Te}, \ldots, \mathrm{x}_N^{Te}\right) \in \mathbb{R}^{N\times F\times V\times V}$, where V denotes the number of EEG channels, and F represents the set of frequency bands $(\delta, \theta, \alpha, \beta, \gamma)$, which is transformed by the filters from the original signal in the temporal domain. $V \times V$ is the dimension of the connectivity matrix for one sample in one frequency band. The EEG features in the frequency domain are defined as $\mathrm{X}^{F} = \left(\mathrm{x}_1^{F}, \mathrm{x}_2^{F}, \ldots, \mathrm{x}_N^{F}\right) \in \mathbb{R}^{N\times F\times V}$, where N is the number of samples after the feature extraction. F also denotes the five frequency bands, but it is converted by a short-term Fourier transform (STFT) in the frequency domain.

We construct an emotional brain graph that is represented as $\mathrm{G} = (V^*, E^*)$. V^* denotes the set of EEG channels, the vertexes in this graph, where $V = |V^*|$. E^* denotes the set of edges between vertexes in V^*. The EEG data X^{Te} and X^{F} represent the information on V^*. We define a weighted adjacency matrix $\mathrm{A} \in \mathbb{R}^{V\times V}$ that represents the set of edges E^*, which also means the functional brain connectivity. A consists of A^{Te} and A^{F}, where A^{Te} is the adjacency matrix calculated from X^{Te} in the temporal domain, and A^{F} denotes the adjacency matrix learned from X^{F} in the frequency domain.

4 METHODOLOGY

4.1 Model Overview

We propose a Multi-domain Adaptive Graph Convolutional Network that fuses the information of EEG signals from two domains

(the temporal domain and the frequency domain) to improve the efficiency of emotion recognition and analyze the patterns of emotion-associated functional brain connectivity. The overall process of our method is shown in Figure 1. In the temporal and frequency domains, we employ the original EEG signals in the time series and the differential entropy (DE) features for EEG-based emotion recognition as inputs, respectively.

4.2 Temporal Domain

4.2.1 Data Preprocessing. The raw EEG data were downsampled to 200 Hz to reduce the amount of calculation, and noise reduction operations were also been down. After that, the EEG data were divided into five frequency bands through five band filters (δ: 1-4 Hz, θ: 4-8 Hz, α: 8-14 Hz, β: 14-31 Hz, and γ: 31-50 Hz). The data after preprocessing is represented as $\mathrm{P}^{Te} = \left(\mathrm{p}_1^{Te}, \mathrm{p}_2^{Te}, \ldots, \mathrm{p}_S^{Te}\right) \in \mathbb{R}^{S\times F\times V}$, where S is the number of samples in the time series, F represents the five frequency bands, and V is the number of EEG channels.

4.2.2 Temporal Functional Brain Connectivity. Inspired by the effective method of extracting temporal EEG features [28], we construct the functional connectivity presented by the weighed adjacency matrix A^{Te} in the temporal domain, as shown in Figure 2. Pearson's correlation coefficient [11] is used as the connectivity index. The connectivity matrices of EEG channels are calculated from P^{Te} in each frequency band using non-overlapped windows, which are denoted as $\mathrm{X}^{Te} = \left(\mathrm{x}_1^{Te}, \mathrm{x}_2^{Te}, \ldots, \mathrm{x}_N^{Te}\right) \in \mathbb{R}^{N\times F\times V\times V}$, where $V \times V$ denotes the dimension of the connectivity matrix and N is the size of the samples after processing. We set a proportional threshold to pick out emotion-relevant connections according to the recognition accuracy. Finally, those connections are united and normalized to construct the functional brain connectivity. Assume that $\mathrm{Y} = \{y^i\}_{i=1}^N$ ($y^i \in L$) is the label set, where L is a set of emotion classes. The details of constructing the functional brain connectivity A^{Te} are presented as Algorithm 1.

4.3 Frequency Domain

4.3.1 Data Preprocessing. We apply extensively used DE features in the frequency domain, which are extracted in five frequency bands (δ: 1-4 Hz, θ: 4-8 Hz, α: 8-14 Hz, β: 14-31 Hz, and γ: 31-50 Hz). Specifically, the raw EEG data are firstly down-sampled to 200 Hz to speed up the calculation. Then a short-time Fourier transform (STFT) with a non-overlapped Hanning window is employed to calculate the DE features in the frequency domain. The extracted DE features are represented as $\mathrm{X}^F = \left(\mathrm{x}_1^F, \mathrm{x}_2^F, \ldots, \mathrm{x}_N^F\right) \in \mathbb{R}^{N\times F\times V}$. Furthermore, we transformed X^F into samples $\mathrm{X}^{F'} = \left(\mathrm{x}_1^{F'}, \mathrm{x}_2^{F'}, \ldots, \mathrm{x}_N^{F'}\right) \in \mathbb{R}^{N\times F\times T\times V}$ with an overlapping window size of T. For each sample, $\mathrm{x}_i^{F'} \in \mathbb{R}^{F\times T\times V}$.

4.3.2 Adaptive Graph Convolutional Layer. Extending the method in adaptive graph convolutional networks for skeleton-based action recognition [24], we establish the adaptive graph convolutional model on EEG-based emotion recognition, which considers the properties of the brain. The operation of spatial graph convolution on vertex v_i in the spatial dimension can be formulated as follows [24] :

$$f_{out}\left(v_i\right) = \sum_{v_j\in\mathcal{B}_i} \frac{1}{Z_{ij}} f_{in}\left(v_j\right) \cdot w\left(l_i\left(v_j\right)\right), \tag{1}$$

where v represents the vertex of the graph, and f_{in} is the input feature map. Here, w denotes the weighting function of the convolution operation, while $\mathcal{B}_i$ represents the convolution sampling area for v_i, and l_i is the mapping function to map each vertex with a weight vector, as the sampling area $\mathcal{B}_i$ may be varied. Z_{ij} is the cardinality of the sampling area $\mathcal{B}_i$, which intends to balance the contribution of each sampling area.

Considering that the human skeleton and topological structure of the brain are different, where the EEG electrodes on the scalp do not have fixed physical connections similar to bones in the human body, the graph convolution operation that we implemented in this manuscript is as follows:

$$\mathrm{f}_{out} = \mathrm{W}\mathrm{f}_{in}\mathrm{A}^F, \tag{2}$$

where W is the $C_{out}\times C_{in}\times 1\times 1$ weight vector of 1×1 convolution operation and A^F is the weighted adjacency matrix that represents the connections between the vertexes in the frequency domain. As mentioned above, the vertex in the brain graph does not have physical connections, and we assume that each vertex has the potential to be associated with all other vertexes. Therefore, the sampling area of vertex v_i in the graph that we constructed contains all vertexes regardless of their physical distance. The input feature map f_{in} is a tensor with dimensions of $C_{in}\times T\times V$, where C_{in} denotes the number of the input channels. In the first layer, $\mathrm{f}_{in} = \mathrm{x}_i^{F'} \in \mathbb{R}^{F\times T\times V}$, where C_{in} is equal to F. Moreover, the weighted adjacency matrix A^F in the frequency domain is divided into A^F_{public} and $\mathrm{A}^F_{private}$,

Algorithm 1: The calculation of the temporal adjacency matrix A^{Te}

Input: The preprocessed data P^{Te}, the label set Y, the threshold t

Output: The temporal weighted adjacency matrix A^{Te}

Calculate the connectivity matrices X^{Te} from P^{Te} using Pearson's correlation coefficient.
for *each* $f \in F$ **do**
  for *each* $l \in L$ **do**
    Average connectivity matrices over the same emotion label l: $m_f^l = mean_{y^i==l}(\mathrm{x}_{i\_f}^{Te})$
    Filter out the strongest t percentage of connections: $\mathrm{A}_l^f = thresholding(m_f^l, t)$
  end
  Take union set of the preserved connections in all classes: $\mathrm{A}^f = union_{l\in L}(\mathrm{A}_l^f)$
end
Unite five frequency bands to obtain critical connections: $\mathrm{A}_{critical} = union_{f\in F}(\mathrm{A}^f)$
Normalize critical connections to obtain the temporal weighted adjacency matrix: $\mathrm{A}^{Te} = normalizing(\mathrm{A}_{critical})$
return A^{Te}

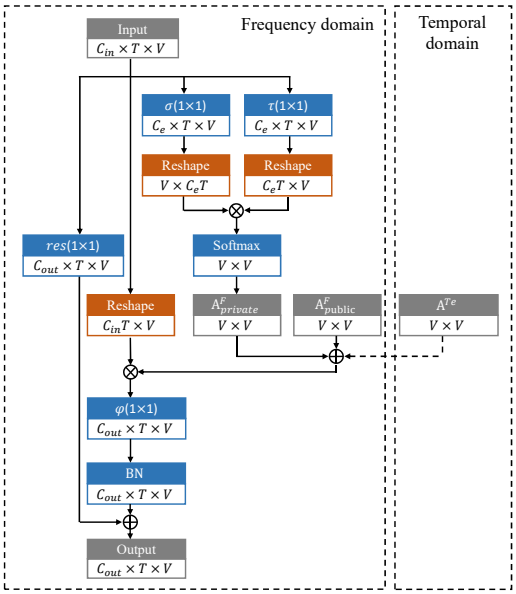

Figure 3: Structure of the adaptive graph convolutional layer. The fusion of the frequency domain (the left side) and the temporal domain (the right side) is conducted in this layer.

thus Equation 2 can be transformed into

$$\mathrm{f}_{out} = \mathrm{W}\mathrm{f}_{in}(\mathrm{A}^F_{public} + \mathrm{A}^F_{private}), \tag{3}$$

where A^F_{public} and $\mathrm{A}^F_{private}$ are weighted adjacency matrices that represent the importance of the connections between vertexes.

In particular, A^F_{public} is a $V \times V$ public adjacency matrix that is shared by all of the samples and is set to be trainable parameters. A^F_{public} is a data-driven parameter, from which the important connections between the EEG channels in each graph convolution layer of the model can be clearly illustrated. It is desired to find the general functional brain connectivity patterns for emotion recognition from the element of A^F_{public}.

$\mathrm{A}^F_{private}$ is a $V \times V$ private adjacency matrix that is obtained by measuring the similarity between two vertexes, and is unique to each sample, which means that one attention map representing the strengths of the connections between EEG channels corresponds to one sample. Concretely, the dot product is employed to measure the similarity between two vertexes. We first embed the input feature f_{in} into an embedding space using the 1×1 convolutional embedding function, i.e., σ and τ, and the dimension of f_{in} changes from $C_{in} \times T \times V$ to $C_e \times T \times V$. Two embedding features are then reshaped into $V \times C_eT$ and $C_eT \times V$ and multiplied into $V \times V$ to obtain the private adjacency matrix. Finally, the matrix is normalized into $0-1$ by the $softmax$ operation. The calculation of $\mathrm{A}^F_{private}$ can be formulated as follows:

$$\mathrm{A}^F_{private} = softmax\left(\mathrm{f}^T_{in}\mathrm{W}^T_{\sigma}\mathrm{W}_{\tau}\mathrm{f}_{in}\right), \tag{4}$$

where W_{σ} and W_{τ} are the weight vectors of the embedding functions. The left side of Figure 3 illustrates the detailed calculation process of the adaptive graph convolutional layer in the frequency domain. $\varphi(1 \times 1)$ in Figure 3 corresponds to the weight vector W in Equation 3, and the residual branch [6] achieved by the 1×1 convolution operation is introduced to ensure the stability of the network.

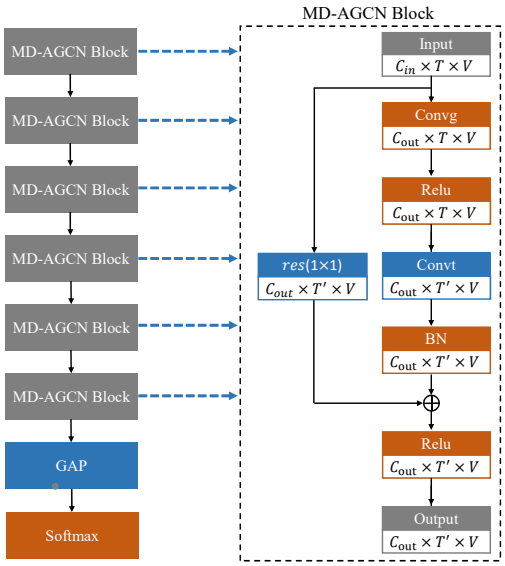

Figure 4: Components of the MD-AGCN. The MD-AGCN consists of six basic MD-AGCN blocks on the top, followed by a global average pooling layer and a $softmax$ layer to predict emotion categories as shown on the left side. The details of one MD-AGCN block are illustrated on the right.

4.4 Fusion and Classification

4.4.1 Fusion of Two Domains. To take advantage of the complimentary information of EEG associated with emotions and to identify emotions more effectively, we combine functional brain connections between EEG channels A^{Te} in the temporal domain with A^F in the frequency domain. Therefore, the graph convolution operation in Equation 3 can be transformed into:

$$\mathrm{f}_{out} = \mathrm{W}\mathrm{f}_{in}(\mathrm{A}^F_{public} + \mathrm{A}^F_{private} + \mathrm{A}^{Te}), \tag{5}$$

where A^{Te} is the temporal adjacency matrix calculated in the temporal domain. The details of the fusion of two domains are illustrated in Figure 3.

4.4.2 Multi-Domain Adaptive Graph Convolutional Network. Figure 4 demonstrates the overall architecture of the proposed Multi-Domain Adaptive Graph Convolutional Network. As illustrated on the left side of Figure 4, the MD-AGCN implemented in this paper consists of six basic MD-AGCN blocks, a global average pooling layer and a $softmax$ layer. The right part of Figure 4 shows the components of one MD-AGCN block, which contains the adaptive graph convolutional layer (Convg) as in Figure 3 and the temporal convolutional layer (Convt). For the temporal convolutional layer,

the convolution operation with the kernel size of $K_t \times 1$ is performed on the temporal dimension T of the input feature.

To retain the information of the input feature and enhance the model stability, the residual connection is introduced in the block. The batch normalization (BN) layer and the ReLU layer are also employed in the block. The algorithm of the overall process of the MD-AGCN is shown in Algorithm 2.

Algorithm 2: The process of the Multi-Domain Adaptive Graph Convolutional Network

Input: The extracted features X^F in the frequency domain, the connectivity matrices X^{Te} in the temporal domain, the label set Y, and the number of MD-AGCN blocks B in the overall MD-AGCN network

Output: The recognition accuracy acc, the weighted adjacency matrix A^{Te} in the temporal domain, the set of the weighted adjacency matrices A^{F^*} in the frequency domain

Calculate A^{Te} from X^{Te}.
Transform X^F into $\mathrm{X}^{F'}$.
Set $\mathrm{x}_i^{F'}$ in $\mathrm{X}^{F'}$ as the input of the first MD-AGCN block: $\mathrm{f}_{in}^1 = \mathrm{x}_i^{F'}$
Represent the b-th MD-AGCN block as $Block^b$:
for *each* $b \in B$ **do**
  $\mathrm{f}_{out}^b, \mathrm{A}_b^F = Block^b(\mathrm{f}_{in}^b, \mathrm{A}^{Te})$
  $\mathrm{f}_{in}^{b+1} = \mathrm{f}_{out}^b$
end
Conduct the global average pooling operation on the output of the last MD-AGCN block: $\mathrm{f}_g = GAP(\mathrm{f}_{out}^B)$
Employ the $softmax$ layer on f_g to predict the emotion categories: $\mathrm{Y}_{pre} = softmax(\mathrm{f}_g)$
Construct A^{F^*} using the learned adjacency matrix in each MD-AGCN block: $\mathrm{A}^{F^*} = (\mathrm{A}_1^F, \mathrm{A}_2^F, ..., \mathrm{A}_B^F)$
Calculate the accuracy: $acc = accuracy(\mathrm{Y}, \mathrm{Y}_{pre})$
return acc, A^{Te}, A^{F^*}

5 EXPERIMENTS

5.1 Datasets

Our proposed method is evaluated on three emotional EEG datasets (SEED [36], SEED-IV[35], and SEED-V[12, 16]). The stimuli of these datasets are film clips, which are reliable for eliciting emotions.

SEED Dataset is constituted by EEG signals of 15 subjects (8 females and 7 males) with three emotion classes: positive, neutral and negative. Each subject participated in three sessions, and each session consisted of 15 trials. Feedback was reported after each clip to guarantee that the participant's emotional reaction was the same as the emotion state of the film.

SEED-IV Dataset also has 15 subjects participating three sessions on different days, and each session contains 24 trials. Therefore, there are 72 film clips in total, including four emotion states: happy, sad, fear and neutral emotions. Similar to the SEED dataset, a self-assessment was performed after each film clip in such a way that we could confirm that the participants shared the same emotion states as those in the film clips.

SEED-V Dataset was proposed by Li *et al.* [12] and contains five emotion classes: happy, disgust, neutral, fear and sad. Sixteen subjects (10 females and 6 males) took part in the experiments for three sessions. There was a one-week or longer interval for each session. A total of 45 video clips were edited into three sessions. EEG signals and eye movement data were collected during the experiments.

5.2 Implementation Details

To make our results comparable, we follow the same common experimental settings as the prior studies on three datasets. For the SEED dataset, we employ the same experimental settings as in [13, 15, 25, 32, 36, 38], which used the first 9 trials as the training set and the remaining 6 trials as the test set in each session. The performance of the model is evaluated by the average accuracy and standard deviation over the sessions. For the SEED-IV dataset, we conducted the experiments as in [13, 35, 38]. The first 16 trials were the training data, and the remaining 8 trials were the test data of each session. Our model was also evaluated by the average accuracy and standard deviation over the sessions. For the SEED-V dataset, similar to [12, 33], three-fold cross validation is applied for each subject.

In the temporal domain, the parameter of the threshold t to calculate A^{Te} is set to 20%. The DE features $\mathrm{X}^F \in \mathbb{R}^{N \times F \times V}$ are transformed into $\mathrm{X}^{F'} \in \mathbb{R}^{N \times F \times T \times V}$ by an overlapping window with the size of T to keep the same sample size N as the compared experiments, where T is set to 5 in the experiments and V is the number of EEG channels, which is equal to 62. The experiments are conducted in MATLAB [18] and PyTorch [21] deep learning framework. The learning rate of our model is set to 0.01, and the batch size is 32. We select the cross-entropy as the loss function in our experiments. The channel size of each MD-AGCN block ranges from 32 to 128 for each experiment.

5.3 Baseline Models

- SVM[12, 38]: A support vector machine with the linear kernel. The SEED-V dataset is also tested on this model.
- DBN[36]: The deep belief network is a probabilistic generative model with a deep architecture.
- STRNN[32]: The spatial-temporal recurrent neural network is a unified spatial-temporal dependency model with the features learned from both the spatial and temporal information in the signal sources.
- DGCNN[25]: The dynamical graph convolutional neural network applies a graph representation method to explore multichannel EEG emotion recognition in a dynamic way.
- BiDANN[14]: The bi-hemispheres domain adversarial neural network investigates both the right and left sides of the hemisphere EEG features and maps them to discriminative feature spaces separately.
- BiHDM[13]: The bi-hemispheres discrepancy model focuses on the asymmetric differences of the right and left hemispheres for EEG emotion recognition.

- R2G-STNN[15]: A region to global spatial-temporal neural network model learns discriminative spatial-temporal EEG features.
- RGNN[38]: The regularized graph neural network explores the topology of the EEG channels with two regularizers.

5.4 Results Analysis and Comparison

The comparison of our model and baseline models on the SEED, SEED-IV and SEED-V datasets is presented in Table 1. It is worthwhile to note that we do not compare our results with those models that do not follow the common experimental settings that are similar to others, such as the SST-EmotionNet [9] and the SparseDGCNN [31]. They divide the training set and test set in different ways. Concretely, the samples were shuffled in SST-EmotionNet [9], which is not appropriate in the EEG-based emotion recognition task. The EEG features of adjacent time periods are more similar, and the samples from one clip can be divided into both the test set and the training set if the samples are shuffled, which should be avoided. For SparseDGCNN [31], they perform leave-one-clip-out validation which take only one clip as test set. In this way, there is only one emotion class in the test set, which is also not very appropriate. Therefore, it is not reasonable to compare their classification

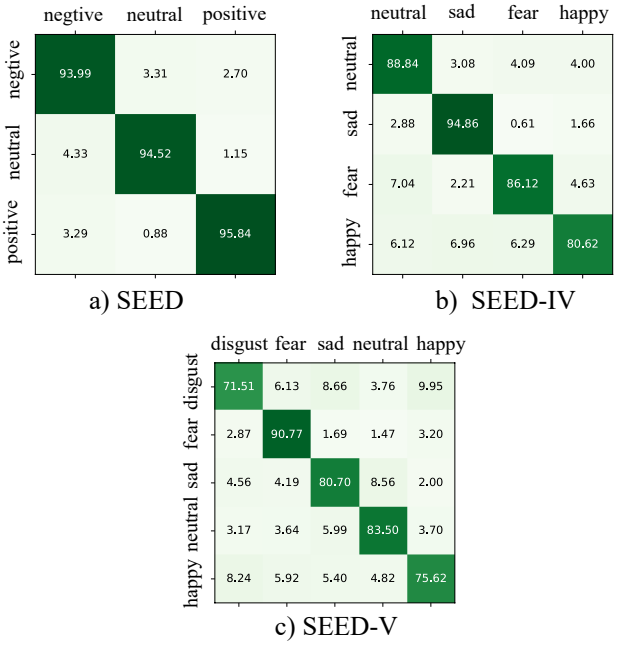

Figure 5: Confusion matrices of MD-AGCN. (a) Confusion matrix on the SEED dataset. (b) Confusion matrix on the SEED-IV dataset. (c) Confusion matrix on the SEED-V dataset. Each column represents the predicted class that our model outputs and each row serves as the target class.

results with ours. Our model achieves state-of-the-art results on all of the SEED, SEED-IV and SEED-V datasets compared with the models that follow the similar common experimental settings. Specifically, the recognition accuracy of our model reaches 94.81% with a standard deviation of 4.52% on the SEED dataset. On the SEED-IV dataset, our model achieves the highest accuracy of 87.63% with the lowest standard deviation of 5.77%, whose performance is much better than other models. For the SEED-V dataset, our model also performs better than the existing approaches using EEG signals only, with the accuracies and standard deviations of 80.77% and 6.61%, respectively. In general, our MD-AGCN model achieves the highest accuracies with the lowest standard deviations over all baseline models.

Table 1: The classification accuracies (mean/std) compared with the performance of state-of-the-art models on the SEED, SEED-IV and SEED-V datasets

Model	SEED[36]	SEED-IV[35]	SEED-V[12]
SVM[12, 36, 38]	83.99/09.72	56.61/20.05	69.5/10.28
DBN[36, 38]	86.08/08.34	66.77/07.38	-
STRNN[32]	89.50/07.63	-	-
DGCNN[25, 38]	90.40/08.49	69.88/16.29	-
BiDANN[14, 38]	92.38/07.04	70.29/12.63	-
BiHDM[13]	93.12/06.06	74.35/14.09	-
R2G-STNN[15]	93.38/05.96	-	-
RGNN[38]	94.24/05.95	79.37/10.54	-
MD-AGCN	**94.81/04.52**	**87.63/05.77**	**80.77/06.61**

Figure 5 presents the confusion matrices of our model, which indicates the ability to recognize each type of emotion on different datasets. For the SEED dataset, our model can recognize positive and neutral emotions (95.84% and 94.52%, respectively) better than negative states (93.99%). For the SEED-IV dataset, our model achieves much better performance on the sad emotion with an accuracy of 94.86%, and is less likely to confuse the sad state with the fear state, as it has only 0.61% of the samples misclassified. For the SEED-V dataset, the disgust state is the most difficult state to recognize for our model, with only 71.51% correctness, while the fear state is significantly easier to classify than the other four emotions with an accuracy of 90.77% .

Two factors cloud contribute to the main improvement in the performance of our models: 1) We take fully advantage of EEG signals by fusing the temporal domain, the frequency domain and the topological structure together. 2) We adaptively update the weighted adjacency matrix in such a way that the functional brain connectivity is learned by the model itself from the EEG data for the emotion recognition.

5.5 Ablation Study

The ablation study is implemented to verify the contribution of the temporal domain on all three datasets. Table 2 shows the performance of our model with and without the temporal domain. After removing the temporal domain input, the adaptive graph convolutional network with the frequency domain, namely FD-AGCN, achieves accuracies of 93.4%, 85.76% and 78.25% with standard deviations of 4.93%, 6.35% and 6.91% on the three datasets. It is obvious that the accuracy of the model with two domains (MD-AGCN) is higher than the accuracy of the model that uses only the frequency domain. MD-AGCN improves the accuracies by 1.41%, 1.87% and

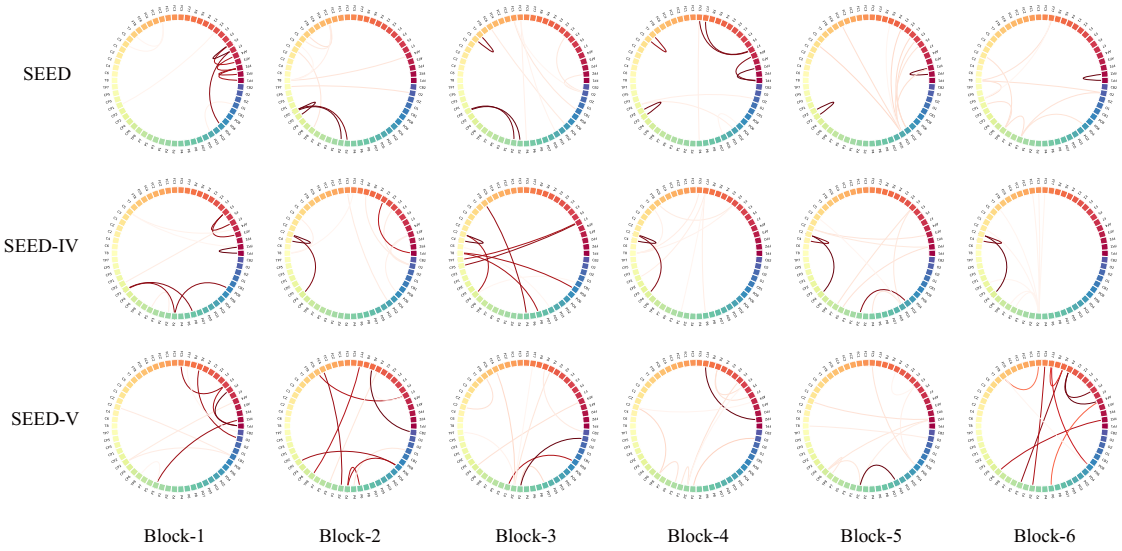

Figure 6: The functional brain connectivities learned in the MD-AGCN are visualized by the top 10 connections between EEG channels. The rows present MD-ACGN blocks, while the columns show the different datasets. A darker color of the line represents a larger edge weight of the adjacency matrix, which also indicates a stronger connection between EEG channels.

Table 2: Ablation study for the classification accuracies (mean/std) on the SEED, SEED-IV and SEED-V datasets

Model	SEED	SEED-IV	SEED-V
FD-AGCN	93.40/04.93	85.76/06.35	78.25/06.91
MD-AGCN	**94.81/04.52**	**87.63/05.77**	**80.77/06.61**

2.52% on the SEED, SEED-IV and SEED-V datasets, respectively, compared with FD-AGCN. The results show that the information in the temporal domain can improve the performance of the model. Accounting for the temporal domain and the frequency domain of the EEG signals is more effective and can improve the classification accuracy.

5.6 Visualization of the Learned Graphs

Figure 6 separately visualizes the functional brain connectivities learned in each MD-AGCN block from the three datasets, which are merged from both the temporal weighted adjacency matrix and the frequency weighted adjacency matrix learned by our model. It is obvious that in the first block, the connections aggregate mainly on the frontal regions for all three datasets. Because the low-level and evident features are usually learned at the lower layer and this phenomenon is consistent with the existing studies [38], our model is proven to have the ability to learn the common information as learned in other models. The experimental results also demonstrate that the frontal connections might be important for EEG-based emotion recognition. Furthermore, more complicated connections, especially those across hemispheres, appear as the training becomes deeper. The results indicate that global inter-channel connections are also important for emotion recognition and that our model can process global connectivities with the deep layers.

6 CONCLUSIONS

In this paper, we proposed a Multi-Domain Adaptive Graph Convolutional Network for EEG-based emotion recognition. The model combines the temporal domain and the frequency domain of the EEG signals with the topological structure of the EEG channels to effectively utilize the complementary information in the EEG data. Emotion-related functional brain connectivity can be learned in an adaptive manner by our model, and the data-driven MD-AGCN makes the model more flexible and more applicable. Extensive experiments on the SEED, SEED-IV and SEED-V datasets demonstrate the extraordinary performance of our model in comparison with various competitive baseline models, especially on problems with more than three emotion classes.

In addition, the analysis of the learned functional connectivities indicates that the connections between EEG channels in the frontal area might play an important role in EEG-based emotion recognition. The critical connections become more complicated and could cross hemispheres as the model becomes deeper. The proposed model can perform as a general framework in the future for other tasks, such as assisting in the treatment of depression.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Natural Science Foundation of China (Grant No. 61976135), SJTU Trans-Med Awards Research (WF540162605), the Fundamental Research Funds for the Central Universities, the 111 Project, and the China Southern Power Grid (Grant No. GDKJXM20185761).

REFERENCES

[1] Fatemeh Bahari and Amin Janghorbani. 2013. EEG-based emotion recognition using recurrence plot analysis and k nearest neighbor classifier. In *2013 20th Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 228–233.

[2] Andrey V Bocharov, Gennady G Knyazev, and Alexander N Savostyanov. 2017. Depression and implicit emotion processing: An EEG study. *Neurophysiologie Clinique/Clinical Neurophysiology* 47, 3 (2017), 225–230.

[3] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 81–84.

[4] Lester I Goldfischer. 1965. Autocorrelation function and power spectral density of laser-produced speckle patterns. *Josa* 55, 3 (1965), 247–253.

[5] Matti Hämäläinen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65, 2 (1993), 413.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and Pattern Recognition*. 770–778.

[7] Tiffany C Ho, Colm G Connolly, Eva Henje Blom, Kaja Z LeWinn, Irina A Strigo, Martin P Paulus, Guido Frank, Jeffrey E Max, Jing Wu, Melanie Chan, et al. 2015. Emotion-dependent functional connectivity of the default mode network in adolescent depression. *Biological Psychiatry* 78, 9 (2015), 635–646.

[8] Robert Jenke, Angelika Peer, and Martin Buss. 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing* 5, 3 (2014), 327–339.

[9] Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, and Jing Wang. 2020. SST-EmotionNet: Spatial-spectral-temporal based attention 3d dense network for EEG emotion recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*. 2909–2917.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[11] You-Yun Lee and Shulan Hsieh. 2014. Classifying different emotional states by means of EEG-based functional connectivity patterns. *PloS one* 9, 4 (2014), e95415.

[12] Tian-Hao Li, Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. 2019. Classification of five emotions from EEG and eye movement signals: Discrimination ability and stability over time. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 607–610.

[13] Yang Li, Lei Wang, Wenming Zheng, Yuan Zong, Lei Qi, Zhen Cui, Tong Zhang, and Tengfei Song. 2020. A novel bi-hemispheric discrepancy model for eeg emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems* 13, 2 (2020), 354–367.

[14] Yang Li, Wenming Zheng, Zhen Cui, Tong Zhang, and Yuan Zong. 2018. A Novel Neural Network Model based on Cerebral Hemispheric Asymmetry for EEG Emotion Recognition.. In *IJCAI*. 1561–1567.

[15] Yang Li, Wenming Zheng, Lei Wang, Yuan Zong, and Zhen Cui. 2019. From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition. *IEEE Transactions on Affective Computing* (2019).

[16] Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. 2021. Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition. *IEEE Transactions on Cognitive and Developmental Systems* (2021).

[17] Yifei Lu, Wei-Long Zheng, Binbin Li, and Bao-Liang Lu. 2015. Combining eye movements and EEG to enhance emotion recognition. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[18] P MatLab. 2018. 9.7. 0.1190202 (R2019b). *MathWorks Inc Natick MA USA* (2018).

[19] Iris B Mauss and Michael D Robinson. 2009. Measures of emotion: A review. *Cognition and Emotion* 23, 2 (2009), 209–237.

[20] Michael Murias, Sara J Webb, Jessica Greenson, and Geraldine Dawson. 2007. Resting state cortical connectivity reflected in EEG coherence in individuals with autism. *Biological Psychiatry* 62, 3 (2007), 270–273.

[21] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).

[22] Louis A Schmidt and Laurel J Trainor. 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion* 15, 4 (2001), 487–500.

[23] Maryam M Shanechi. 2019. Brain–machine interfaces from motor to mood. *Nature Neuroscience* 22, 10 (2019), 1554–1564.

[24] Lei Shi, Yifan Zhang, Jian Cheng, and Hanqing Lu. 2019. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12026–12035.

[25] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. 2018. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing* 11, 3 (2018), 532–541.

[26] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129 (2014), 94–106.

[27] Alexis E Whitton, Stephanie Deccy, Manon L Ironside, Poornima Kumar, Miranda Beltzer, and Diego A Pizzagalli. 2018. Electroencephalography source functional connectivity reveals abnormal high-frequency communication among large-scale functional networks in depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3, 1 (2018), 50–58.

[28] Xun Wu, Wei-Long Zheng, and Bao-Liang Lu. 2019. Identifying functional brain connectivity patterns for EEG-based emotion recognition. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 235–238.

[29] Yongqiang Yin, Xiangwei Zheng, Bin Hu, Yuang Zhang, and Xinchun Cui. 2021. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Applied Soft Computing* 100 (2021), 106954.

[30] Zhongliang Yin, Jun Li, Yun Zhang, Aifeng Ren, Karen M Von Meneen, and Liyu Huang. 2017. Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of EEG time series. *Biomedical Signal Processing and Control* 31 (2017), 331–338.

[31] Guanhua Zhang, Minjing Yu, Yong-Jin Liu, Guozhen Zhao, Dan Zhang, and Wenming Zheng. 2021. SparseDGCNN: Recognizing Emotion from Multichannel EEG Signals. *IEEE Transactions on Affective Computing* (2021).

[32] Tong Zhang, Wenming Zheng, Zhen Cui, Yuan Zong, and Yang Li. 2018. Spatial–temporal recurrent neural network for emotion recognition. *IEEE Transactions on Cybernetics* 49, 3 (2018), 839–847.

[33] Li-Ming Zhao, Rui Li, Wei-Long Zheng, and Bao-Liang Lu. 2019. Classification of five emotions from EEG and eye movement signals: complementary representation properties. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 611–614.

[34] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 5040–5043.

[35] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. 2018. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics* 49, 3 (2018), 1110–1122.

[36] Wei-Long Zheng and Bao-Liang Lu. 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development* 7, 3 (2015), 162–175.

[37] Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu. 2017. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Transactions on Affective Computing* 10, 3 (2017), 417–429.

[38] Peixiang Zhong, Di Wang, and Chunyan Miao. 2020. EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing* (2020).