# An RSB-GNN-based EEG Approach for Exploring Students' Affective States in E-learning


Ting Li
*Sino-French Engineer School*
*Beihang University*
Beijing, China
caro3766@buaa.edu.cn

Chuantao Yin*
*Sino-French Engineer School*
*Beihang University*
Beijing, China
chuantao.yin@buaa.edu.cn

Yanmei Chai
*School of Information*
*Central University of Finance and Economics*
Beijing, China
ymchai@cufe.edu.cn

Hui Chen
*Department of Planning and Finance*
*Beihang University*
Beijing, China
chenhui@buaa.edu.cn

Wenge Rong
*School of Computer Science and Engineering*
*Beihang University*
Beijing, China
w.rong @buaa.edu.cn

Yuanxin Ouyang
*School of Computer Science and Engineering*
*Beihang University*
Beijing, China
oyyx @buaa.edu.cn



***Abstract*—This research paper describes an original method of affective state recognition in the e-learning field. Emotion plays a key role in both knowledge-building and mental health. In recent years, more and more emotion researchers have broken through traditional questionnaires to utilize physiological data to monitor students' cognitive and affective states. Among these methods, electroencephalography (EEG) can directly reflect the physiological activities of the brain and has unique potentials and advantages over others. For analyzing this multi-channel noised signal, current emotion recognition methods mainly use deep learning methods to learn the spatial or temporal representation of each channel, and then process the classification through a multimodal fusion strategy, while emotional expression highly relies on brain functional connectivity.**

**In this research work, for the EEG-based learning-centered affective state recognition, we adopted a novel residual shrinkage block (RSB) to construct the graph neural network (GNN). During the feature extraction, the RSB is designed to obtain the features of interest and reduce the influence of artifact noises for recognition. GNN considers the biological topology among different brain regions to capture relations among different EEG channels. Extensive experiments on the CAL dataset prove that the performance of the proposed model is superior to current deep learning methods.**

**Prior research may use the findings of this study to empower adaptive self-regulated learning environments through the automated recommendation of learning strategies, learning contents, and emotion regulation strategies according to students' learning-centered affective states, to further improve their learning performance as well as mental health. On the other hand, teachers or online course designers can use emotional feedback to adjust the learning materials and the pace of the instruction according to students' needs and preferences.**



***Keywords—emotional learning, adaptive computer learning, confusion, electroencephalogram (EEG), graph neural network (GNN)***



This work is supported by the National Natural Science Foundation of China (No. 62377002).


## I. Introduction

With the development of information and communication technologies, e-learning, which means learning using electronic technologies outside of traditional classrooms, has become a new trend, especially in engineering and computer education. Compared with the traditional learning mode, e-learning enables learners to transcend temporal and spatial constraints, granting access to a wealth of online learning resources. However, it is important to note that this mode of learning primarily involves interaction with digital learning materials, such as instructional videos, which may potentially lead to adverse effects on students' mental well-being due to the absence of face-to-face communication and the physical separation of educators and peers.

Emotion plays a key role in both mental health and knowledge-building. Studies have found that emotions affect all aspects of learning, such as attention [1], memory processes [2], motivation [3], and cognitive problem-solving [4], thus affecting learning outcomes. Traditional research on affective states in the education field has focused on the study of test anxiety and the causal attribution of success and failure as antecedents of emotions in classical learning circumstances. For e-learning environments, the emotions detected or used are basic emotions, non-basic emotions, learning-centered emotions, trait emotions, or a combination of two or three of them [5], but there is no consensus on a specific or definitive list of affective states among experts and researchers [6]. Reference [7] identified the primary affective states of students during complex learning using educational technology as flow/engagement frustration, confusion, and boredom. Among these affective states, flow/engagement, confusion, and boredom appeared relatively frequently, whereas happiness, curiosity, and frustration appeared less frequently [8]. Reference [9] investigated the impact of cognitive-affective states of students during the use of three different computer-based learning environments and observed that boredom was very persistent across the learning environments, concentration and confusion were the most common states, frustration was less persistent, and joy and surprise were less common. The literature [10] states that anxiety, enjoyment, anger/frustration, boredom, confusion, and curiosity/interest are the most independently researched emotions in Technology-Based Learning Environments (TBLEs) in studies from 1965 to 2018.

Traditional emotion research uses self-report to assess emotions during learning, for example, the Academic Emotion Questionnaire (AEQ) [11] has been widely used to measure learning-related affective states. Besides, with the development of computer vision and machine learning, researchers measure e-learning emotions based on various physiological and non-physiological data, including heart rate variability (HRV) [12], electroencephalography (EEG), eye-tracking [13], facial expression [14], forum discussions [15], behavior data [16]. Among these methods, EEG-based emotion recognition has the potential and benefits different from others. EEG signals have good temporal resolution and are an objective and direct means of detecting internal emotional states [17]. Therefore, EEG-based emotion recognition has received increasing attention in the fields of computer science, education, biomedicine, psychology, and other disciplines.

Our previous work developed the E-learning Epistemic Emotion Scale (EEES) to assess the affective states that learners experience in e-learning environments, with emphasis on their cognitive process. EEES contains flow, confusion, fatigue, frustration, curiosity, and boredom. This research mainly aims to propose an effective methodology for EEG-based emotion recognition, in consideration of the topology characteristic of EEG signals.

We construct a novel deep learning framework by adopting residual shrinkage blocks (RSB) to the graph neural network (GNN). Experiments on the CAL dataset prove the advantages and potentials of the proposed method in comparison with baseline models. This paper mainly contains the methodology and experiment results of this approach.

## II. Related Work

### A. EEG data collection and public dataset

EEG signal acquisition can be performed by invasive and non-invasive methods. Invasive methods have a higher signal-to-noise ratio and accuracy compared to non-invasive methods, but the disadvantage is that the electrodes need to be surgically implanted in the cranial cavity, and the electrodes penetrate the cerebral cortex to acquire the signals, which is difficult and risky. Non-invasive methods are most commonly used in research, and signal acquisition can be accomplished efficiently with the help of wearable EEG caps or headsets that place electrodes along the scalp. Fig. 1 [18] represents the electrode positions of the International 10/20 system. The numbers 10 and 20 indicate the distance between adjacent electrodes (10 or 20 percent of the total front-back or right-left distance of the skull). Each site has a letter to identify the lobe and a number to identify the hemisphere location. F stands for Frontal, T for Temporal, C for Central (although there is no central lobe, the C letter is used for identification purposes), P for Parietal, and O for Occipital. z (zero) refers to an electrode placed on the midline.

For emotion recognition based on EEG signals, in recent years researchers have generally used video clips as trial stimuli, and relevant public datasets have also been established. SEED dataset [19] was established in 2015, which selected 15 Chinese movie clips of about 4 minutes in duration as the stimuli, recorded the EEG signals, facial expression videos, and eye-movement data of 15 participants, and categorized affective states as positive, neutral, and negative emotions. Researchers further developed and released a four-category (happy, sad, fearful, and neutral) version of SEED-IV [20] in 2019 and a five-category (happy, sad, fearful, disgusted, and neutral) version of SEED-V [21] in 2022. Another widely used dataset is the DEAP dataset [22], which recorded the EEG as well as other physiological signals of 32 participants while watching 40 music video clips of one minute. These participants then rated each video clip in detail according to arousal, valence, preference, dominance, and familiarity. Other datasets that use video clips as stimuli, such as AMIGOS [23], DREAMER [24], and ASCERTAIN [25], have also been widely used by researchers. With the growing interest in the study of epistemic emotions (emotions associated with knowledge acquisition [26]), researchers have established the CAL dataset [27] using Raven's Standard Progressive Matrices as the stimuli, focusing on the recognition of confusion.

### B. EEG-based emotion recognition

The general process of EEG-based emotion recognition contains data collection, data preprocessing, feature extraction, feature selection, training of classification model, and model application.

EEG signals have high sensitivity and are easily interfered with by various noises in the acquisition process. EEG signal preprocessing mainly removes artifacts and other noises from the EEG signal, such as electrooculogram (EOG), electromyogram (EMG), cardiac artifacts, skin electrical responses, electromagnetic interferences, and power frequency disturbances [28]. Since the frequency of EEG signals is generally in the range of 0-100 Hz, artifacts beyond the frequency range of EEG signals can be efficiently filtered out by

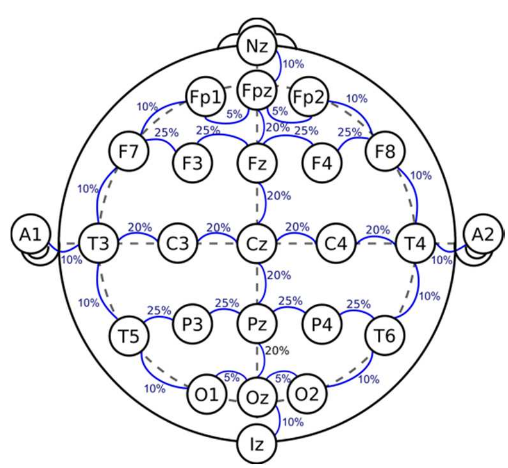


Fig. 1 The International 10/20 System

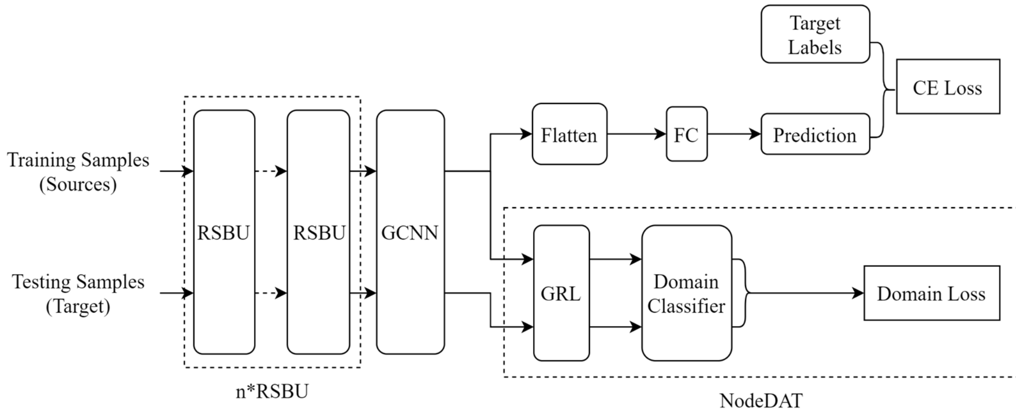


Fig. 2 Overall Structure of the Proposed Deep Learning Method

using appropriate band-pass filters. 4-45 Hz is the most commonly used frequency range for band-pass filters [29]. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) can also be used to remove noises in EEG data.

The key to EEG-based emotion recognition lies in the extraction of EEG features from raw EEG signals that can characterize different emotional states, which is essential to subsequent data processing. Time-domain features (TDF), frequency domain features (FDF), and time-frequency domain features (TFDF) are the three basic types of features [29]. For EEG-based emotion recognition, traditional machine learning classifiers such as K-Nearest Neighbor (KNN) [30], Support Vector Machine (SVM) [31], and Random Forest (RF) [32], take these features as input.

With the development of machine learning, especially deep learning, more and more researchers adopt deep learning methods to process EEG data for emotion recognition. Since deep learning methods can use raw data as input, data preprocessing steps may not be necessary for them; meanwhile, in most cases, deep learning methods can automatically compute features of EEG [33].

Convolution Neural Network (CNN) [34] and Recurrent Neural Network (RNN) [35] are performant for EEG-based emotion recognition. Long Short Term Memory Neural Network (LSTM) [36] is an optimization of recurrent neural networks, which provides excellent performance in EEG-based emotion recognition due to its unique neural network structure [37]. A combination of CNN and LSTM can also improve the accuracy of emotion recognition [38]. Emotional expression highly relies on brain functional connectivity. Graph Convolutional Neural Networks (GCNNs) are extended from traditional CNNs, which combine convolution with spectral theory. Compared with CNNs, GCNNs are more have more advantages in extracting discriminative features of signals in the discrete spatial domain. The ability of GCNN to characterize the intrinsic relationships between different nodes of a graph provides a potential way to explore the relationships between multiple EEG channels during EEG-based emotion recognition [39].

## III. Research Method/Methodology

In this study, we proposed a Residual Shrinkage Building Graph Neural Network (RSB-GNN) to synthesize the advantages of graph convolutional neural networks and residual shrinkage networks. The node-domain adversarial training (NodeDAT) was also added to improve the robustness of the model. Fig. 2 represents the overall structure of the proposed method.

To investigate the problem of emotion recognition of multi-channel EEG signals, this study adopts a graph representation inspired by the success of the graph convolutional neural network model. In this approach, each channel of EEG is regarded as a node, while the connection between two different nodes corresponds to an edge of the graph. This approach contributes to a deeper understanding of the process of EEG emotion recognition and provides new ideas for related applications. Meanwhile, artifacts and noise of the EEG signals are ignored during the training of the GCNN model because the GCNN is in the form of an end-to-end classifier. Therefore, there is a bottleneck in the classification performance of GCNNs. The denoising process plays an important role in the classification of EEG signals due to their nonlinear and non-smooth characteristics. With the development of deep neural networks, Deep Residual Shrinkage Networks (DRSNs) have been proposed and applied to fault diagnosis [40], which extract discriminative features from highly noised signals through a soft thresholding process. DRSN has also been applied to improve the recognition of motor imagery EEG signals [41].

### A. *Residual Shrinkage Building Unit*

Deep Residual Shrinkage Networks (DRSNs) [40] incorporate soft thresholding in deep residual networks (ResNets). Soft thresholding is an essential step in the signal-denoising method. The key task in signal denoising is to design a filter that converts the noisy information into near-zero features, converts the useful information into features with larger absolute values, and then applies soft thresholding to convert the near-zero features to zero. The soft threshold function is represented by the following:

$$y=\begin{cases} x-\tau, & x>\tau \\ 0, & -\tau \le x \le \tau \\ x+\tau, & x<-\tau \end{cases} \tag{1}$$

where $x$ and $y$ are the inputs and outputs, respectively, and $\tau$ is a positive parameter representing the threshold. The partial derivative of the soft threshold function is 0 or 1, so it can prevent the problems of gradient vanishing and gradient explosion.

Residual Shrinkage Building Unit (RSBU) is a basic element of DRSN. It adds a soft threshold function and an attention unit, where the attention unit includes an absolute value function, a GAP, two fully connected (FC) layers, two activation functions (ReLU, Sigmoid), and one batch normalization (BN). In our proposal, the RSBU is designed to obtain the features of interest and reduce the influence of artifact noises for EEG-based emotion recognition.

Fig. 3 represents the structure of the RSBU [40], where the main part retains the identity shortcut in the residual module. *K* is the number of convolution kernels in the convolutional layer, *M* is the number of neurons in the fully connected network, *C*, *W,* and *1* are the number of channels, width, and height of the feature map, respectively, and /2 indicates that the step size of the convolution is 2.

## B. Graph Convolutional Neural Networks

For EEG-based emotion recognition using GCNN, each channel of EEG signals is considered as a node in the graph. The topology structure of EEG channels is denoted by a weighted adjacency matrix $A \in \mathbb{R}^{n \times n}$, where $n$ denotes the number of channels. Each element of the adjacency matrix represents the connectivity between the channels. Inspired by [42], to facilitate more discriminative EEG feature extraction, the adjacency matrix of the graph is also used as a learnable parameter of the model. In order to reduce overfitting, the matrix is modeled as a symmetric matrix using only $n(n+1)/2$ rather than $n^2$ parameters.

In the proposed method, the adjacency matrix is initialized as in [43] :

$$A_{i,j}=\begin{cases} \min(1,\frac{\delta}{d_{i,j}^2}), & \text{if } d_{i,j} \le \tau \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

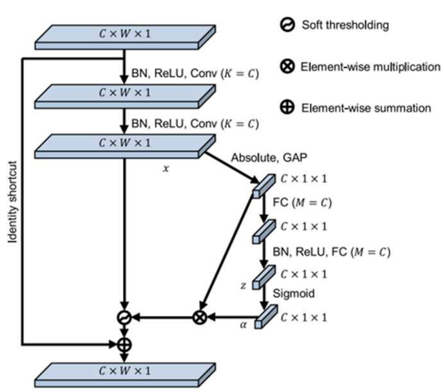


Fig. 3 Structure of the RSBU

Here, $\tau$ and $\delta$ are two constants, and $d_{i,j}$ denotes the Euclidean distance between the *i*-th and *j*-th channels. We chose $\tau$ =2 and $\delta$ =5 such that about 20% of the elements in the adjacency matrix parameters are not zero.

Between adjacent layers in GNNs, the feature transformation can be written as

$$H^{l+1}=f(H^l,A), \tag{3}$$

where $l$ denotes the number of layers, and $f$ denotes the function to learn during training. GCNN is an extension of the traditional CNN method by combining CNN with spectral theory [44], and the feature transformation in GCNN [45] can be written as:

$$H^{l+1}=\sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^lW^l), \tag{4}$$

where $D$ denotes the diagonal degree matrix of $A$ and σ denotes a non-linear function.

## C. Node-wise Domain Adversarial Training

Inspired by [43], the node-wise domain adversarial training (NodeDAT) method was adopted to reduce the discrepancies between source and target domains, i.e., training and testing sets, respectively. Specifically, a domain classifier is proposed to classify each node representation into either source domain or target domain. During optimization, the deep learning model aims to confuse the domain classifier by learning domain-invariant representations. NodeDAT method can improve the generalizability of trained classifiers, especially in cross-subject classification settings.

Given labeled source/training data $X^S \in \mathbb{R}^{N \times n \times d}$, and unlabeled target/testing data $X^T \in \mathbb{R}^{N \times n \times d}$, which in practice $X^T$ can be either oversampled or down-sampled to have the same number of samples as $X^S$, the domain classifier aims to minimize the sum of the following two binary cross-entropy losses:

$$\Phi_D = -\sum_{i=1}^{N}\sum_{j=1}^{n}\left(\log(p_j(0|X_j^S,\theta_D)) + \log(p_j(1|X_i^T,\theta_D))\right), \tag{5}$$

where $\theta_D$ denotes the parameters of the domain classifier, 0 and 1 denote source and target domains, respectively. Intuitively, the domain classifier is trained to classify source data as 0 and target data as 1. A gradient reversal layer (GRL) was implemented to act like an identity layer in the forward propagation and to reverse the gradients of the domain classifier during backpropagation. The reversed gradients are further scaled by a GRL scaling factor $\beta$ which gradually increases from 0 to 1 as the training progresses. The gradually increasing $\beta$ allows the domain classifier to be less sensitive to noisy inputs at the early stages of the training process. The calculation of $\beta$ can be represented as:

$$\beta=\frac{2}{1+e^{-10p}}-1, \tag{6}$$

where $p$ represents the progress of training and varies from 0 to 1.

The overall loss of our model can be calculated as:

$$\Phi=\Phi_{CE}+\Phi_D+\alpha\|\Theta\|, \tag{7}$$

where $\alpha$ is the coefficient of regularization, $\Theta$ denotes all training parameters, and $\Phi_{CE}$ can be calculated as:

$$\Phi_{CE} = -\sum_{i=1}^{N} \log(p(Y_i | X_i, \Theta)), \tag{8}$$

where $X \in \mathbb{R}^{N \times n \times k}$ represents the input signal, $n$ is the number of channels, $k$ is the length of the signal and $Y \in \mathbb{R}^{N}$ denotes emotion labels.

## IV. Experimental Settings

### A. Dataset

The EEES includes six emotions: flow, curiosity, confusion, tiredness, frustration, and boredom, of which the role of most of the emotions for e-learning has been relatively clear, with the two positive affective states flow and curiosity being beneficial to online learning, and fatigue, frustration, and boredom having a negative impact on learning performance. However, whether confusion is beneficial for learning remains controversial [46]. According to the confusion dynamic model, confusion is considered a transitional emotion. When confusion is resolved, the learner's emotions follow a cycle of engagement-confusion-engagement; frustration is triggered if the learner is unable to regain the state of flow; and prolonged unresolved confusion eventually leads to boredom [47]. It can be seen that the development of confusion encompasses almost all other emotions in the model and is more critical in online learning environments. Whether confusion produces positive or negative outcomes depends on its resolution, which provides an opportunity for e-learning smart systems to improve the learning experience and outcomes by intervening in the learning process. Therefore, the identification of confusion is important for research in e-learning.

However, as mentioned earlier, the widely used datasets in the field of EEG signal-based emotion recognition usually use movie clips or music videos as stimulus sources to study emotions in general contexts rather than in learning contexts, and most of the emotions studied are two-dimensional categorizations or basic emotion models. The recognition of these emotions is also relatively well-studied. The study of emotions in cognitive activities, i.e. epistemic emotions, is still in its early stages and there are almost no publicly available datasets. The CAL dataset chosen for this study, officially released in 2023, fills this gap by designing experiments to actively and accurately elicit confusing emotions in reasoning activities [27].

CAL dataset [27] employed OpenBCI as the EEG collector. Eight channels (Fp1, Fp2, C3, C4, T5, T6, O1, O2) EEG data with a sampling rate of 250 Hz were collected for each of the 23 subjects while trying to solve Raven's Progressive Matrices. The Raven's test is a nonverbal psychological test used to measure human intelligence and abstract reasoning. 48 questions of high difficulty were chosen to ensure that confusion was induced. The labeling technique combines subjective (self-reported feeling) and objective labeling (whether or not the answer was correct). There were a total of four categories of labels, of which, if a participant answered incorrectly and expressed confusion in his/her self-report, the EEG signal was labeled as "confused"; if the participant answered correctly and was not confused, the corresponding EEG was labeled as "non-confused"; if the participant answered the test item correctly and reported confusion, this was categorized as "guess"; if the participant answered the test item incorrectly but did not show confusion, it was categorized as "think-right".

### B. EEG data preprocessing

First, the raw EEG signal is processed using a 4-45 Hz bandpass filter as proposed in [29]. To eliminate the effect of amplitude size on the subsequent model training, the filtered EEG signals of each channel were normalized to fall in the interval of [-1,1].

In other datasets where movie clips or music videos were used as the stimuli, the EEG signals captured for each participant were of the same length because each participant watched the same length of the video, whereas, during the CAL dataset acquisition, each participant did not have the same length of time for answering the questions, which resulted in different lengths of EEG signals being captured. Meanwhile, thus CAL aimed to trigger confusion, almost half (48%) of EEG data are labeled as "confused", while "guess" only accounts for 4%. To solve the above problems, the segmentation method was applied during data preprocessing. The filtered EEG data was divided into 1 s sliding windows in the light of the literature. To tackle the issue of data imbalance, we set different step sizes: 1 s for confused, 0.5 s for non-confused, 0.5 s for think-right, and 0.15 s for guess.

Finally, a total of 22,842 samples of the 8×250 dimension was obtained, of which "confused" had 7656 samples, "non-confused" had 4927 samples, "think-right" had 6317 samples, and "guess" had 3942 samples. The overall distribution is shown in Fig. 4, and the distribution of labels for each subject is shown in Fig. 5. It can be learned by observing the label distributions that a relatively balanced and large number of sample sizes can be obtained by applying this segmentation method.

### C. Classification settings

This work focuses on two tasks: binary classification (confused and non-confused) and four-class classification (confused, non-confused, guess, and think-right). K-fold cross-validation (K=5) is used to evaluate the proposed model, and two common scenarios in emotion recognition algorithms are considered to validate the proposed method:

1) In cross-subject (or subject-independent) experiments, the differences between different subjects are emphasized to test the generalization ability of the model. There were 23 subjects in total, with the first 4 folds each containing data from 5 subjects and the last fold containing data from 3 participants.

2) In intra-subject (or subject-dependent) experiments, the model is trained across multiple subjects. The EEG data of all participants were equally divided into 5 subsets, 4 of which were taken as the training set and 1 as the validation set in each validation.

### D. Model Settings for RSB-GNN

The hardware used for the experiments in our work is a computer with a 16-core Intel Xeon Silver 4110 series CPU processor and an NVIDIA GeForce RTX 3090 graphics card, and the operating system is Ubuntu 18.04. The deep neural network used for the processing of EEG signals is implemented

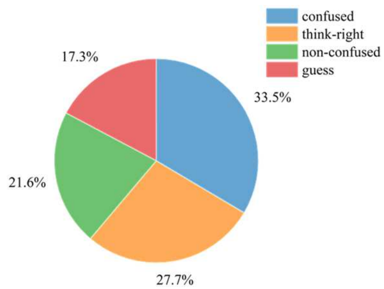


Fig. 4 Overall Label Distribution after Segmentation

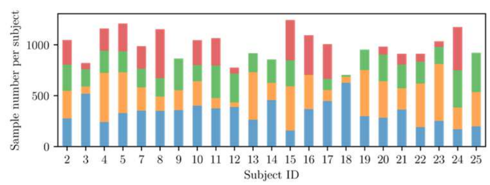


Fig. 5 Label Distribution for Each Subject

using PyTorch. The EEG signals were filtered using MNE in the data preprocessing stage. Throughout the entire process of model training, the CPU is responsible for handling a series of key tasks, including loading and preprocessing EEG signals, initializing the model, calculating evaluation metrics such as accuracy, and storing data for the optimal model. GPU is responsible for performing forward propagation and backward gradient backpropagation related to various operators in the model, such as convolution, fully connected layers, pooling, and activation functions, as well as updating parameters based on optimization strategies.

The hyperparameters of our study include two categories: model-building parameters and training parameters. Among them, the model-building parameters include the number of RSBUs, the size of the hidden layer of GCNN, and number of layers of the GCNN. The training parameters include Learning Rate (LR), L1 regularization coefficient, L2 regularization coefficient, and Dropout.

As proposed in [48], the batch size was set to 256, the maximum Epoch was set to 100, the Dropout was set to 0.5, and LR was set to 0.005; the L1 and L2 regularization coefficients were set to 0.001 for intra-subject binary classification, 0.003 for intra-subject four classification and cross-subject binary classification, and 0.005 for cross-subject four classification.

The size of the hidden layer of the GCNN was set to 80 and the number of layers to 2 [42]. The number of RSBUs was set to 3 [49]. The early stopping technique was used to prevent model overfitting, i.e., training was stopped when there was no improvement in validation set accuracy on 20 consecutive epochs.

## V. RESULTS

The performance of the model is evaluated using the average of the accuracy and F1 score in the K-fold cross-validation. The deep learning algorithms including dynamic dynamical graph convolutional neural networks (DGCNN) [42] and 1-D Deep Residual Shrinkage Network (1-D-DRSN) [49] are implemented using the PyTorch library. DGCNN can also be regarded as our model without RSBUs and NodeDAT.

### A. *Binary classification*

The results of binary classification tasks in intra-subject and cross-subject scenarios are listed in Table 1. Our model has a comparable result with the 1-D-DRSN in the intra-subject scenario and our model performs better in the cross-subject scenario.

The experimental results show that each model in the binary classification task outperforms the cross-subject experiment on the intra-subject experiment. This is due to the variability of EEG signals between subjects that makes the classification task on the cross-subject experiment more challenging. It's worth noting that the performance gap of DGCNN between the two scenarios is small, while the recognition performance of other methods has a large gap between the two scenarios. The model proposed in this study shows a large improvement in both experimental scenarios compared to the DGCNN, whereas it is comparable to the 1-D-DRSN in terms of recognition performance, but with a 2% improvement in accuracy in the cross-subject experiments. This observation illustrates the superiority of RSBUs when dealing with highly noisy data, while GCNNs have the potential to close the gap between the model's performance in both intra-subject and cross-subject experiments.

Moreover, the experimental results show that NodeDAT contributes to the robustness of the deep learning model in cross-subject scenarios by narrowing the gap between the two scenarios. It can be seen that the addition of NodeDAT provided a smaller improvement in model performance of about 1% in the intra-subject experiment scenario, while it improved by about 5% in the cross-subject experiment condition.

### B. *Four-class classification*

Table 2 summarizes the results of the four-class classification task in intra-subject and cross-subject scenarios. The 1-D-DRSN outperforms in intra-subject scenarios with an accuracy of 55.35% and an F1 score of 0.55, while our model stands out in cross-subject scenarios with an accuracy of 44.90% and an F1 score of 0.40.

It can be seen that in the four-class classification task, there was a greater gap in the performance of the model between intra-subject and cross-subject experiments, which suggests that it is more challenging to perform finer-grained emotion recognition. The experimental results show that the model proposed in this study showed greater improvement compared to other models, especially in the cross-subjects experiment. The experimental results further demonstrate the effectiveness of NodeDAT in the cross-subjects scenario. It can be seen that the addition of NodeDAT in the intra-subject experiment scenario provides a small improvement in the model performance of about 2%, whereas it improves by about 11% in the cross-subject experiment. It can be inferred that NodeDAT effectively improves the robustness of the model in cross-subject experiments for the four-classification task.

TABLE 1 BINARY CLASSIFICATION (ACCURACY/F1 SCORE) (w/o=without)

| Method | Intra-subject | Cross-subject |
|---|---|---|
| DGCNN | 63.39/0.56 | 62.15/0.41 |
| 1-D-DRSN | **73.40/0.72** | 62.61/0.54 |
| Our model w/o NodeDAT | 72.31/0.70 | 57.62/0.57 |
| Our model | **73.46/0.71** | **64.62/0.54** |

TABLE 2 FOUR-CLASS CLASSIFICATION (ACCURACY/F1 SCORE) (w/o=without)

| Method | Intra-subject | Cross-subject |
|---|---|---|
| DGCNN | 37.64/0.23 | 36.20/0.19 |
| 1-D-DRSN | **55.35/0.55** | 34.38/0.25 |
| Our model w/o NodeDAT | 50.23/0.48 | 33.16/0.25 |
| Our model | 52.39/0.50 | **44.90/0.40** |

## VI. CONCLUSION AND DISCUSSIONS

In this work, focusing on affective states during e-learning, we proposed a novel RSB-GNN approach for EEG-based emotion recognition. RSBUs are designed to improve the processing of noisy EEG signals and GNN is used to explore the connectivity of different brain regions. NodeDAT is implemented to improve the robustness of the proposed model. The CAL dataset was chosen by analyzing epistemic affective states in e-learning, among which confusion is essential to regulation strategy. Binary and four-class classification experiments in intra-subject and cross-subject scenarios were conducted. The results on the CAL dataset demonstrated the superiority of the proposed model in the task of EEG-based emotion recognition, especially in cross-subject scenarios. The better recognition performance of RSB-GNN is most likely due to the following major points:

- Adding RSBUs makes the model more powerful for noisy data processing, such as EEG signals, by incorporating the soft threshold in residual building units;
- By learning domain-invariant representations, the NodeDAT improves the generalizability of trained classifiers in cross-subject classification settings.

Additionally, the adjacency matrix in the graph representation of EEG signals not only represents the connectivity of different channels but also indicates the contributions of the EEG channels to EEG emotion recognition [42]. Hence, the adjacency matrix would provide a potential way to find out what are the most contributive EEG channels in EEG emotion recognition. This approach can be interesting to reduce the complexity of calculation while processing EEG data with a larger number of channels, such as the DEAP dataset which contains EEG signals of 32 channels.

Although EEG-based emotion recognition has gained attention in education research, there is a notable absence of specific datasets focused on learning-centered emotion. Furthermore, the data scales of current EEG datasets, whether specific to education or not, remain relatively limited. This constraint may hinder the development of more robust deep neural network models, consequently impeding further advancements in this area of research. To address this, future research endeavors could explore the establishment of a dedicated EEG database on a much larger scale. Such an EEG database, encompassing a wider range of emotion categories, would not only enhance emotion research, but also benefit the evolution of intelligent e-learning systems. Leveraging the high temporal resolution of EEG, the recognition of additional types of emotion could potentially provide physiological evidence supporting the dynamic model of confusion[47]. Furthermore, the transformation of confusion into emotional feedback could serve as a valuable indicator of students' learning experiences

## ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 62377002).

## REFERENCES

[1] J. Ranellucci, K. A. Robinson, J. M. Rosenberg, Y. K. Lee, C. J. Roseth, and L. Linnenbrink-Garcia, "Comparing the roles and correlates of emotions in class and during online video lectures in a flipped anatomy classroom," (in English), *Contemporary Educational Psychology,* vol. 65, p. 101966, Apr 2021, doi: 10.1016/j.cedpsych.2021.101966.

[2] C. R. Madan, S. M. E. Scott, and E. A. Kensinger, "Positive emotion enhances association-memory," *Emotion,* vol. 19, no. 4, pp. 733-740, Jun 2019, doi: 10.1037/emo0000465.

[3] N. von der Embse, D. Jester, D. Roy, and J. Post, "Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review," *J Affect Disord,* vol. 227, pp. 483-493, Feb 2018, doi: 10.1016/j.jad.2017.11.048.

[4] C. M. Bohn-Gettler, "Getting a Grip: The PET Framework for Studying How Reader Emotions Influence Comprehension," (in English), *Discl. Process.,* Article vol. 56, no. 5-6, pp. 386-401, Jul 4 2019, doi: 10.1080/0163853x.2019.1611174.

[5] N. Mejbri, F. Essalmi, M. Jemni, and B. A. Alyoubi, "Trends in the use of affective computing in e-learning environments," (in English), *Education and Information Technologies,* vol. 27, no. 3, pp. 3867-3889, Apr 2022, doi: 10.1007/s10639-021-10769-9.

[6] Y. P. Xu, Y. Y. Li, Y. S. Chen, H. G. Bao, and Y. Q. Zheng, "Spontaneous visual database for detecting learning-centered emotions during online learning," (in English), *Image and Vision Computing,* vol. 136, p. 104739, Aug 2023, doi: 10.1016/j.imavis.2023.104739.

[7] S. K. D'Mello, A. C. Strain, A. Olney, and A. Graesser, "Affect, Meta-affect, and Affect Regulation During Complex Learning," in *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven Eds., (Springer International Handbooks of Education. New York, NY: Springer New York, 2013, ch. Chapter 44, pp. 669-681.

[8] S. D'Mello, "A Selective Meta-Analysis on the Relative Incidence of Discrete Affective States During Learning With Technology," (in English), *Journal of Educational Psychology,* Article vol. 105, no. 4, pp. 1082-1099, Nov 2013, doi: 10.1037/a0032674.

[9] R. S. J. D. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments," (in English), *International Journal of Human-Computer Studies,* vol. 68, no. 4, pp. 223-241, Apr 2010, doi: 10.1016/j.ijhcs.2009.12.003.

[10] K. Loderer, R. Pekrun, and J. C. Lester, "Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments," (in English), *Learning and Instruction,* vol. 70, p. 101162, Dec 2020, doi: 10.1016/j.learninstruc.2018.08.002.

[11] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld, and R. P. Perry, "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ)," (in English),

*Contemporary Educational Psychology,* vol. 36, no. 1, pp. 36-48, Jan 2011, doi: 10.1016/j.cedpsych.2010.10.002.

[12] S. Bounyong, R. Yoshida, and M. Yoshioka, "Epistemic Emotion Detection by Video-based and Heart Rate Variability Features for Online Learning," presented at the 2023 IEEE International Conference on Consumer Electronics (ICCE), 2023.

[13] R. Murali, C. Conati, and R. Azevedo, "Predicting Co-occurring Emotions in MetaTutor when Combining Eye-Tracking and Interaction Data from Separate User Studies," presented at the LAK23: 13th International Learning Analytics and Knowledge Conference, 2023.

[14] J. C. Hung, K. C. Lin, and N. X. Lai, "Recognizing learning emotion based on convolutional neural networks and transfer learning," (in English), *Applied Soft Computing,* vol. 84, p. 105724, Nov 2019, doi: 10.1016/j.asoc.2019.105724.

[15] S. Liu, S. Liu, Z. Liu, X. Peng, and Z. Yang, "Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement," *Computers & Education,* vol. 181, 2022, doi: 10.1016/j.compedu.2022.104461.

[16] Y. Sun and S. Z. Wang, "Study on the Distance Learners' Academic Emotions Using Online Learning Behavior Data," (in English), *Mobile Information Systems,* vol. 2022, pp. 1-12, Aug 28 2022, doi: 10.1155/2022/9785591.

[17] M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso, "Multimodal data as a means to understand the learning experience," (in English), *International Journal of Information Management,* vol. 48, pp. 108-119, Oct 2019, doi: 10.1016/j.ijinfomgt.2019.02.003.

[18] S. M. Alarcao and M. J. Fonseca, "Emotions Recognition Using EEG Signals: A Survey," (in English), *Ieee Transactions on Affective Computing,* vol. 10, no. 3, pp. 374-393, Jul-Sep 2019, doi: 10.1109/Taffc.2017.2714671.

[19] Z. Wei-Long and L. Bao-Liang, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Transactions on Autonomous Mental Development,* vol. 7, no. 3, pp. 162-175, 2015, doi: 10.1109/tamd.2015.2431497.

[20] W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions," *IEEE Trans Cybern,* vol. 49, no. 3, pp. 1110-1122, Mar 2019, doi: 10.1109/TCYB.2018.2797176.

[21] W. Liu, J. L. Qiu, W. L. Zheng, and B. L. Lu, "Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition," (in English), *Ieee Transactions on Cognitive and Developmental Systems,* vol. 14, no. 2, pp. 715-729, Jun 2022, doi: 10.1109/Tcds.2021.3071170.

[22] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis Using Physiological Signals," (in English), *Ieee Transactions on Affective Computing,* vol. 3, no. 1, pp. 18-31, Jan-Mar 2012, doi: 10.1109/T-Affc.2011.15.

[23] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," (in English), *Ieee Transactions on Affective Computing,* vol. 12, no. 2, pp. 479-493, Apr-Jun 2021, doi: 10.1109/Taffc.2018.2884461.

[24] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics,* vol. 22, no. 1, pp. 98-107, Jan 2018, doi: 10.1109/JBHI.2017.2688239.

[25] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors," (in English), *IEEE Transactions on Affective Computing,* vol. 9, no. 2, pp. 147-160, Apr-Jun 2018, doi: 10.1109/Taffc.2016.2625250.

[26] K. Nerantzaki, A. Efklides, and P. Metallidou, "Epistemic emotions: Cognitive underpinnings and relations with metacognitive feelings," *New Ideas in Psychology,* vol. 63, 2021, doi: 10.1016/j.newideapsych.2021.100904.

[27] T. Xu, J. Wang, G. Zhang, L. Zhang, and Y. Zhou, "Confused or not: decoding brain activity and recognizing confusion in reasoning learning using EEG," *Journal of Neural Engineering,* vol. 20, no. 2, p. 026018, Mar 17 2023, doi: 10.1088/1741-2552/acbfe0.

[28] C. Yu and M. Wang, "Survey of emotion recognition methods using EEG information," *Cognitive Robotics,* vol. 2, pp. 132-146, 2022, doi: 10.1016/j.cogr.2022.06.001.

[29] R. Vempati and L. D. Sharma, "A systematic review on automated human emotion recognition using electroencephalogram signals and artificial intelligence," (in English), *Results in Engineering,* vol. 18, p. 101027, Jun 2023, doi: 10.1016/j.rineng.2023.101027.

[30] W. L. Zheng, J. Y. Zhu, and B. L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," (in English), *Ieee Transactions on Affective Computing,* vol. 10, no. 3, pp. 417-429, Jul-Sep 2019, doi: 10.1109/Taffc.2017.2712143.

[31] Y. Zhang, S. H. Zhang, and X. M. Ji, "EEG-based classification of emotions using empirical mode decomposition and autoregressive model," (in English), *Multimedia Tools and Applications,* vol. 77, no. 20, pp. 26697-26710, Oct 2018, doi: 10.1007/s11042-018-5885-9.

[32] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, "A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition," (in English), *Ieee Sensors Journal,* vol. 21, no. 3, pp. 3579-3591, Feb 1 2021, doi: 10.1109/Jsen.2020.3027181.

[33] Y. Y. Zhou, S. Huang, Z. M. Xu, P. P. Wang, X. Wu, and D. Q. Zhang, "Cognitive Workload Recognition Using EEG Signals and Machine Learning: A Review," (in English), *Ieee Transactions on Cognitive and Developmental Systems,* vol. 14, no. 3, pp. 799-818, Sep 2022, doi: 10.1109/Tcds.2021.3090217.

[34] N. Kumari, S. Anwar, and V. Bhattacharjee, "Time series-dependent feature of EEG signals for improved visually evoked emotion classification using EmotionCapsNet," (in English), *Neural Computing & Applications,* vol. 34, no. 16, pp. 13291-13303, Aug 2022, doi: 10.1007/s00521-022-06942-x.

[35] E. C. Djamal and R. D. Putra, "Brain-computer interface of focus and motor imagery using wavelet and recurrent neural networks," *Telecommunication Computing Electronics and Control (TELKOMNIKA),* vol. 18, no. 4, pp. 2748-2756, 2020, doi: 10.12928/telkomnika.v18i5.14899.

[36] Y. Li, W. M. Zheng, L. Wang, Y. Zong, and Z. Cui, "From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition," (in English), *Ieee Transactions on Affective Computing,* vol. 13, no. 2, pp. 568-578, Apr-Jun 2022, doi: 10.1109/Taffc.2019.2922912.

[37] R. Ganiga, Y. Kim, R. Tulluri, and W. Choi, "Modeling EEG Signals for Mental Confusion Using DNN and LSTM With Custom Attention Layer," (in English), *Ieee Access,* vol. 11, pp. 134663-134676, 2023, doi: 10.1109/Access.2023.3337035.

[38] Y. Zhang *et al.*, "An Investigation of Deep Learning Models for EEG-Based Emotion Recognition," *Frontiers in Neuroscience,* vol. 14, p. 622759, 2020, doi: 10.3389/fnins.2020.622759.

[39] C. Li *et al.*, "EEG Emotion Recognition Based on Self-attention Dynamic Graph Neural Networks," presented at the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/36086084.

[40] M. H. Zhao, S. S. Zhong, X. Y. Fu, B. P. Tang, and M. Pecht, "Deep Residual Shrinkage Networks for Fault Diagnosis," (in English), *Ieee Transactions on Industrial Informatics,* vol. 16, no. 7, pp. 4681-4690, Jul 2020, doi: 10.1109/Tii.2019.2943898.

[41] J. C. Huang, "A novel residual shrinkage block-based convolutional neural network for improving the recognition of motor imagery EEG signals," (in English), *International Journal of Intelligent Computing and Cybernetics,* vol. 16, no. 3, pp. 420-442, Jul 12 2023, doi: 10.1108/Ijicc-05-2022-0130.

[42] T. F. Song, W. M. Zheng, P. Song, and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," (in English), *Ieee Transactions on Affective Computing,* vol. 11, no. 3, pp. 532-541, Jul-Sep 2020, doi: 10.1109/Taffc.2018.2817622.

[43] P. X. Zhong, D. Wang, and C. Y. Miao, "EEG-Based Emotion Recognition Using Regularized Graph Neural Networks," (in English), *Ieee Transactions on Affective Computing,* vol. 13, no. 3, pp. 1290-1301, Jul-Sep 2022, doi: 10.1109/Taffc.2020.2994159.

[44] M. Graña and I. Morais-Quilez, "A review of Graph Neural Networks for Electroencephalography data analysis," *Neurocomputing,* vol. 562, p. 126901, 2023, doi: 10.1016/j.neucom.2023.126901.

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

[46] J. M. Lodge, G. Kennedy, L. Lockyer, A. Arguel, and M. Pachman, "Understanding Difficulties and Resulting Confusion in Learning: An Integrative Review," *Frontiers in Education,* vol. 3, p. 49, 2018, doi: 10.3389/feduc.2018.00049.

[47] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learning and Instruction,* vol. 22, no. 2, pp. 145-157, 2012, doi: 10.1016/j.learninstruc.2011.10.001.

[48] K. Zhang, Z. Ye, Q. Ai, X. Xie, and Y. Liu, "GNN4EEG: A Benchmark and Toolkit for Electroencephalography Classification with Graph Neural Network," *arXiv preprint arXiv:2309.15515,* 2023.

[49] J. H. Wang, Y. Song, Z. M. Mao, J. J. Liu, and Q. Gao, "EEG-Based Emotion Identification Using 1-D Deep Residual Shrinkage Network With Microstate Features," (in English), *Ieee Sensors Journal,* vol. 23, no. 5, pp. 5165-5174, Mar 1 2023, doi: 10.1109/Jsen.2023.3239507.