2023 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-8867-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IJCNN54540.2023.10191227

Cross-Subject Decision Confidence Estimation from EEG Signals Using Spectral-Spatial-Temporal Adaptive GCN with Domain Adaptation

Rong-Fei Gu[1,†], Rui Li[1,†], Wei-Long Zheng[1], and Bao-Liang Lu[1,2,*], *Fellow, IEEE*
[1]Center for Brain-Like Computing and Machine Intelligence
Department of Computer Science and Engineering
Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
Brain Science and Technology Research Center
Shanghai Jiao Tong University, Shanghai, 200240, China
[2]RuiJin-Mihoyo Laboratory, Clinical Neuroscience Center, RuiJin Hospital
Shanghai Jiao Tong University School of Medicine, Shanghai, 200020, China

***Abstract*—The study of the human decision-making process has long been a valuable field for both scientific research and practical application. Towards knowing and taking control of the decision-making process, evaluating the reliability of human decisions objectively plays an important role. Various studies have demonstrated that the confidence level of humans during the decision-making process is an important factor that reflects the correctness of decisions. In literature, several deep learning based methods have been developed to estimate decision confidence using Electroencephalography (EEG). Among these approaches, the spectral-spatial-temporal adaptive graph convolutional neural network (SST-AGCN) stands out. However, SST-AGCN focuses on specific subjects, and may lead to less efficiency in cross-subject situations, which are more common in application scenarios. In this paper, we propose a deep learning model called SST-AGCN with Domain Adaptation (SST-AGCN-DA) for cross-subject decision confidence estimation. To examine the effectiveness of our proposed model, we compare our SST-AGCN-DA with the original SST-AGCN, three typical domain adaption algorithms in the field, and the SST-AGCN with Domain Generalization (SST-AGCN-DG), which is another transfer learning model we developed in this paper. We conduct cross-subject confidence estimation experiments on an EEG dataset collected under a text-based decision-making task. The averaged results of leave-one-out cross-validation come out that the F1-scores of our proposed SST-AGCN-DA and SST-AGCN-DG are 79.45% and 77.04%, respectively, while the original SST-AGCN and the best of the existing domain adaptation algorithms are 74.15% and 74.25%, respectively.**

***Index Terms*—decision confidence estimation, electroencephalography (EEG), cross-subject, domain adaptation, domain generalization**

I. Introduction

The fast-developing science and technology enable automated machines to substitute human forces in various positions, but critical decision-making processes still require human participation. Therefore, it becomes more important to evaluate the reliability of human decision-making in time. To realize the real-time evaluation of the decisions, researchers attempt at estimating the level of human decision confidence during the decision-making process. Decision confidence is the subconscious estimation of the subject that has been shown to increase accordingly with the possibility for the decision to be correct [1].

To precisely estimate decision confidence, several kinds of physiological data have been investigated. In traditional works, researchers have worked on functional magnetic resonance imaging (fMRI) [2], [3] and event-related potential (ERP) [4], commonly acquired under experiments around techniques like psychological, and have revealed the critical regions that affect decision confidence [5]. These methods are poor in applicability because of the high professionalism of the experiments and the various experimental restrictions to data collection.

In recent years, labeled datasets have been developed, including other forms of psychological data collected under life-like tasks. These works cover the above shortages while promoting advanced models for supervised learning in decision confidence estimation. Among the various kinds of psychological data, Electroencephalography (EEG) stands out because of its objectivity. Li *et al.* developed an image-based experiment containing tasks close to realistic scenarios like detecting objects from remote sensing images and created an EEG dataset with data labeled by five levels of decision confidence [6]. The experiment environment and the time limit are delicately designed to simulate real-world situations as well as the task targets. Base on this dataset, a deep-

This work was supported in part by grants from STI 2030-Major Projects+2022ZD0208500, National Natural Science Foundation of China (Grant No. 61976135), Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX), Shanghai Pujiang Program (Grant No. 22PJ1408600), SJTU Global Strategic Partnership Fund (2021 SJTU-HKUST), Shanghai Marine Equipment Foresight Technology Research Institute 2022 Fund (No. GC3270001/012) and GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine.

[†]R. F. Gu and R. Li are co-first authors of the article.

[*]Corresponding author (bllu@sjtu.edu.cn)

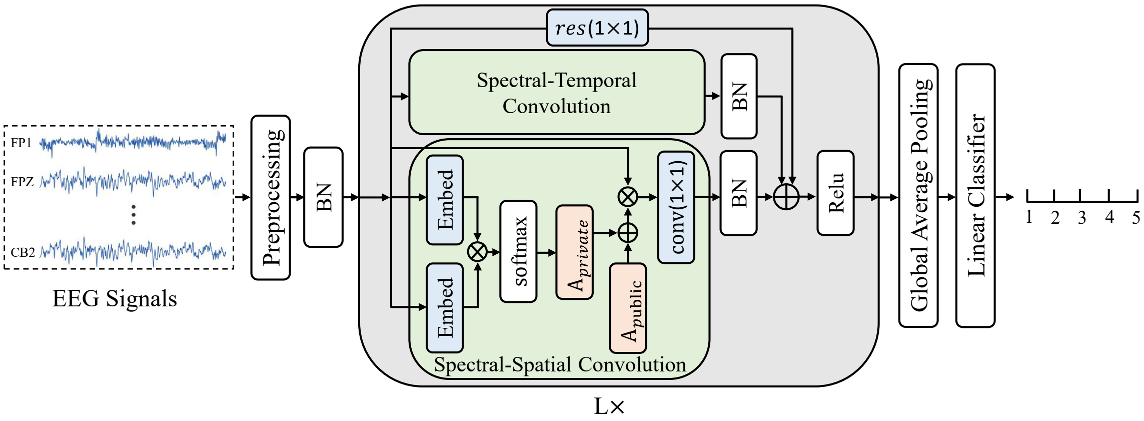

Fig. 1. Structure of the SST-AGCN model, composed of preprocessing, feature extraction, and decision confidence level classification steps. The input samples are the EEG signals collected from the subjects during the decision-making process. The preprocessing step calculates primary features from the raw EEG signals. In the feature extraction step, each of the L SST-AGCN blocks concurrently extracts features from the spectral-spatial and the spectral-temporal aspects. The classifier finally predicts the decision confidence level base on the extracted feature.

learning based approach was proposed [7] and advanced neural networks like graph convolutional network proposed by Liu *et al.* [8] efficiently predict the level of decision confidence. These works demonstrated the effectiveness of EEG in decision confidence estimation. Moreover, Li *et al.* put forward a text-based experiment to further explore the ability of EEG in recognizing decision confidence levels based on text exams [9]. A spectral-spatial-temporal adaptive graph convolutional neural network (SST-AGCN) is designed in this work [9]. The SST-AGCN model fully utilizes information from the EEG data in the dimensions of spectral, spatial, and temporal, and is shown to be the most effective model in text-based decision confidence estimation.

Although the above studies greatly promote the accuracy of decision confidence estimation, they are not suitable for cross-subject scenarios and thus infeasible in practical applications. For the existing subject-dependent methods, a specific model is trained for each subject, which promotes the accuracy of confidence estimation within individuals. However, these models rely too much on the labeled data of the target subjects, while in application it is usually impractical to first obtain adequate labeled samples of the subjects. So to improve the confidence estimation models for practical application, we adopt the idea in transfer learning, namely domain adaptation, which can eliminate the distribution differences between the feature extracted from the massive known data collected before application and the limited data from the target. We thus formulate the spectral-spatial-temporal adaptive graph convolutional neural network with domain adaptation (SST-AGCN-DA). We also include another transfer learning technique called domain generalization for comparison, and form the spectral-spatial-temporal adaptive graph convolutional neural network with domain generalization (SST-AGCN-DG). These models, together with several other domain adaptation algorithms, are tested on cross-subject decision confidence estimation based on the EEG dataset collected under the text-based decision-making task.

II. Related Work

A. Spectral-Spatial-Temporal Adaptive Graph Convolutional Neural Network

Among the various models for decision confidence estimation, the spectral-spatial-temporal adaptive graph convolutional neural network (SST-AGCN) [9] achieves leading performance in subject-dependent tasks, and our work further explores the potential of it in cross-subject tasks. The SST-AGCN model is originate from the idea of graph convolutional neural network (GCN) [10], which realizes feature extraction based on graph information. However, the original GCN model requires the input graph manually defined in advance, while it is infeasible in dealing with data of unknown functional connections. To cope with this circumstance, Shi *et al.* suggested the adaptive GCN model [11], where the topological structure can be learned by the backpropagation algorithm. Although the aim of the work is skeleton-based action recognition, the work gives inspiration to the SST-AGCN model to adapt to the human brain and the EEG signal.

The structure of the SST-AGCN model is shown in Figure 1. The main component of the feature extraction module is the L SST-AGCN blocks, where spectral-spacial convolution and spectral-temporal convolution are applied simultaneously within each block to the input. The residual structure [12] and the batch normalization operations retain the original information and ensure the stability of the model. The input of the first SST-AGCN block is the spectral feature, defined as $B_{in} \in \mathbb{R}^{S_{in} \times T \times C}$. S_{in}, T and C are the scales of spectral, temporal and spatial dimensions. The output of the block can be expressed as:

$$\tilde{B} = \sigma(BN(\tilde{B}_{ss}) + BN(\tilde{B}_{st}) + residual(B_{in})), \tag{1}$$

where $\tilde{B}_{ss} \in \mathbb{R}^{S_{out} \times T \times C}$ is the output of the spectral-spatial convolution module while $\tilde{B}_{st} \in \mathbb{R}^{S_{out} \times T \times C}$ is the output of the spectral-temporal convolution module. σ refers to the Relu activation function. After the L SST-AGCN blocks for feature extraction, the global average pooling layer and the linear output layer complete the downstream classification step.

To implement graph convolution and utilize spatial features, the spectral-spatial module adaptively learns the weighted adjacency matrix that carries the connection information among the EEG channels. The matrix is composed of the public matrix A_{pub} and the private matrix A_{pri}, both belong to $\mathbb{R}^{C \times C}$ where C is the number of channels. A_{pub} is a shared trainable parameter that holds the general brain function connectivity and reflects the neural patterns during decision-making. A_{pri} is extracted for each sample to measure the correlation between two EEG channels. It can be calculated from:

$$A_{pri} = softmax(E_\theta^T E_\tau), \tag{2}$$

where $E_\theta \in \mathbb{R}^{(S_e T) \times C}$ and $E_\tau \in \mathbb{R}^{(S_e T) \times C}$ are reconstructed embedded features separately obtained through convolution on input B_{in} using 1×1 kernels. Because of the complexity of human brain, the function connection graph for EEG data is fully connected. Thus the calculation for convolution can be inferred from the original GCN model:

$$\tilde{B}_{ss} = W B_{in}(A_{pub} + A_{pri}), \tag{3}$$

where W represents the convolution kernel of size 1×1, number of input channels as S_{in} and number of output channels as S_{out}.

Based on the spectral feature, the spectral-spatial convolution module extract the spatial feature of the EEG data, and in the spectral-temporal convolution module, the temporal feature is further explored. The graph structure enables the model to learn with the information of channel-wise relationship. So to inform the model with temporal influence of the adjacent frames in the time series, convolution is conducted to the sequential data of each channel, providing $\tilde{B}_{st} = Conv_t(B_{in})$. The kernel size is $K_t \times 1$, and K_t denotes the number of neighboring frames that are related with the current frame.

Combining the above equations, the output of each SST-AGCN block can be calculated, and the final extracted feature can be applied in decision confidence estimation.

B. Transfer Learning

Models trained by the above method perform well with specific subjects in decision confidence estimation, but when applied to new subjects with few labeled EEG data, the performance of estimation will decline. This decline is caused by individual differences between subjects. So we introduce transfer learning techniques into the SST-AGCN model to remove the personalized information and enhance the generality of the learned models among different subjects.

Domain adaptation is a progressed idea in the field of transfer learning. Domain $D = \{\mathscr{X}, P(X)\}(X \in \mathscr{X})$ refers to the feature space $\mathscr{X}$ and the distribution $P(X)$ of a certain group of samples [13]. Adaptation is to make the model trained on the source domain $D_S = \{\mathscr{X}_\mathscr{S}, P(X_S)\}$ apply to the target domain $D_T = \{\mathscr{X}_\mathscr{T}, P(X_T)\}$ with few samples. Various domain adaptation methods have been suggested, and some are used in EEG-based tasks. Pan *et al.* proposed transfer component analysis to transform the feature space [13]. Zheng *et al.* conducted transductive parameter transfer in emotion recognition based on EEG [14]. The idea of adversarial network is also adopted in various domain adaptation methods to blur the distinction between features from the source and the target domains. Li *et al.* introduced the domain-adversarial neural network (DANN) into cross-subject emotion recognition based on EEG [15]. The DANN model create the adversarial relationship between the feature extractor and the domain discriminator to eliminate domain information. The popular adversarial discriminative domain adaptation (ADDA) [16] is also widely applied in EEG based tasks. In the ADDA model, a target domain feature extractor is fine-tuned from the source domain feature extractor to remove specific information of the target domain. Zhao *et al.* proposed the plug-and-play domain adaptation and make representation partition to enable fast model transfer [17]. Moreover, remarkable results in EEG-based cross-subject emotion recognition have been achieved by the Wasserstein generative adversarial network domain adaptation (WGANDA) model [18].

Although domain adaptation methods are effective in most cross-subject tasks, there may still be circumstances when no information from the target domain is reachable before training. Therefore, researchers put forward the idea of domain generalization. The domain generalization method aims at removing the differences among all the subjects to retain only the target-related information. Compared with domain adaptation methods, domain generalization methods may be less effective in tasks with a certain amount of samples from the target, but the models trained under domain generalization methods may have better potential in generality. Ma *et al.* proposed the domain generalization DANN model (DG-DANN) [19] based on the DANN model, and Jia *et al.* proposed a spatial-temporal graph convolutional network with domain generalization to classify sleep stages [20].

III. Methods

A. Spectral-Spatial-Temporal Adaptive Graph Convolutional Neural Network with Domain Adaptation

To improve the cross-subject decision confidence estimation and eliminate the domain deviation caused by data distribution differences among different subjects, we take advantage of the adversarial module similar to the DANN model to construct the SST-AGCN with domain adaptation (SST-AGCN-DA) model, as shown in Figure 2. The model is composed of the feature extractor G_f base on the SST-AGCN model, the decision confidence level predictor G_y, and the domain predictor G_d. A gradient reversal layer (GRL) [21] is added between G_f and G_d to form the adversarial relationship. G_y and G_d aim at correctly predict the level of decision confidence and domain of the feature generated by G_f, while G_f is optimized to remove domain-related information from

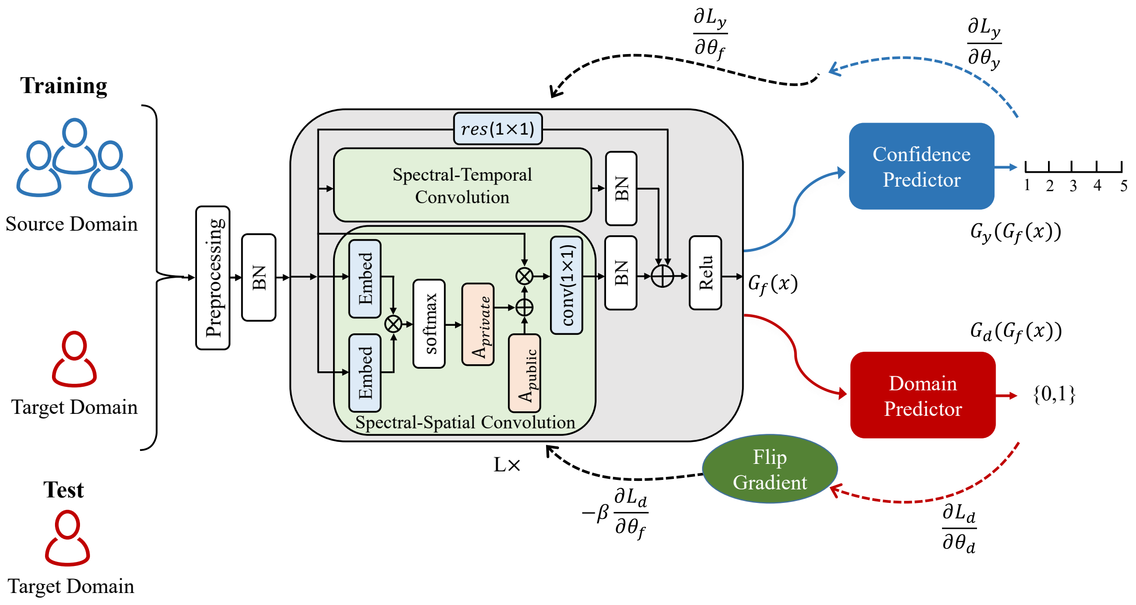

Fig. 2. The overall process of the spectral-spatial-temporal adaptive graph convolutional neural network with domain adaptation. The training set includes all the EEG signals from the source domain and the target domain. They are labeled by the domains to which they belong to train the domain predictor, while only the confidence levels from the source domain are applied to train the confidence predictor. The test set is composed of both the EEG signals from the target domain and their confidence levels to evaluate the precision of the confidence predictor. Solid arrows represent the forward propagation path and dashed arrows represent the backward propagation path. The flip gradient module reverses the optimization target of the SST-AGCN module to remove domain-related information from the extracted feature and brings in a weight coefficient to balance the training process.

the feature to support G_y and confuse G_d. In the domain adaptation method, the domains are the source domain and the target domain.

To be precise, in the forward propagation step, we denote $x_{r,i}$ as the ith input EEG feature from domain $r \in \{s, t\}$ where s and t represent the source and target domain. G_f extracts the feature $\hat{x}_{r,i} = G_f(x_{r,i})$ from the input and sends it to downstream tasks including decision confidence estimation and domain prediction. The outputs of G_y and G_d are separately $\hat{y}_{r,i} = G_y(\hat{x}_{r,i})$ and $\hat{d}_{r,i} = G_d(\hat{x}_{r,i})$ while the ground truth decision confidence level and domain are $y_{r,i}$ and $d_{r,i}$. Since the ground truth labels of confidence level for samples from the target domain may be unavailable, G_y only take into consideration the correctness of predicting the level of samples from the source domain. Then in the backpropagation step, the predicted labels are compared with the known ground truth labels to optimize the parameters of the SST-AGCN-DA model. We apply cross entropy loss in both confidence level and domain prediction. The loss functions are:

$$\mathscr{L}_y = -\frac{1}{|s|}\Sigma_{i=1}^{|s|} y_{s,i} \log \hat{y}_{s,i}, \tag{4}$$

$$\mathscr{L}_d = -\frac{1}{|s| + |t|}(\Sigma_{i=1}^{|s|} d_{s,i} \log \hat{d}_{s,i} + \Sigma_{i=1}^{|t|} d_{t,i} \log \hat{d}_{t,i}). \tag{5}$$

Since the domain prediction loss is reversed between G_f and G_d, the optimization of G_f will be maximizing $\mathscr{L}_d$, and the overall loss function of the SST-AGCN-DA model can be expressed as:

$$\mathscr{L} = \mathscr{L}_y - \beta\mathscr{L}_d, \tag{6}$$

where β is the hyperparameter that controls the balance between decision confidence level prediction and domain prediction.

The optimized parameters θ_f, θ_y and θ_d of the three modules should then be:

$$\hat{\theta}_f, \hat{\theta}_y = argmin_{\theta_f,\theta_y}\mathscr{L}(\theta_f, \theta_y, \hat{\theta}_d) \tag{7}$$

$$\hat{\theta}_d = argmax_{\theta_d}\mathscr{L}(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \tag{8}$$

During training the SST-AGCN-DA model, especially the domain predictor G_d, samples from at least two domains are required to form the source and the target domains. Considering this fact, our SST-AGCN-DA model only applies to cross-subject scenarios instead of intra-subject ones. Meanwhile, our model faces another problem caused by separating the two domains. The domain predictor may suffer from the imbalanced dataset since the target domain usually refers to a specific subject, while the source domain is possibly composed of all the rest subjects. If not handled appropriately, the huge difference between the sample size of the two domains is prone to bring the domain predictor into local optimum, and the effectiveness of the feature extractor G_f in removing domain-related information becomes doubtful. To relieve the problem of the imbalanced dataset, we apply random oversampling to the target domain during the training process in this paper. The

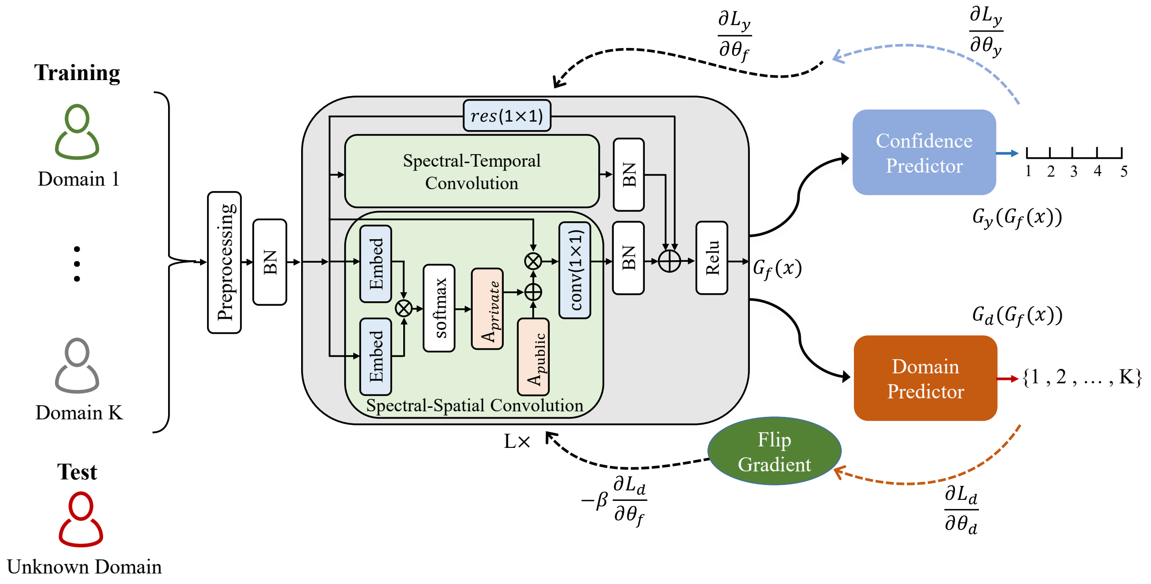

Fig. 3. The overall process of spectral-spatial-temporal adaptive graph convolutional neural network with domain generalization. The training set includes only the EEG signals from the known subjects. The subject information and the confidence levels respectively form the corresponding labels that support the training process of the multi-class domain predictor and the confidence predictor. The test set is composed of the EEG signals and confidence levels obtained from new subjects to evaluate the model with completely no information related to the application target. The solid arrows, the dashed arrows, and the flip gradient act similarly as in the SST-AGCN-DA model.

results of the experiment section demonstrate the superiority of the SST-AGCN-DA model, and other strategies to cope with the imbalanced dataset still need to be explored.

B. Spectral-Spatial Sdaptive Graph Convolutional Neural Network with Domain Generalization

The SST-AGCN-DA model is capable of removing the differences between the source domain and the target domain, and to further cope with unknown target domain, we introduce the SST-AGCN with domain generalization (SST-AGCN-DG) model shown in Figure 3. Instead of confusing the feature extracted from the source and target domain, the SST-AGCN-DG model aims at eliminating all the personalized information and retaining all the task-related general information. This is achieved by viewing samples from each subject as a domain, which has been proven to be effective in other tasks [22]. Once the model is able to extract only the non-redundant information from each domain, the generality of this model will support any unknown new domain.

Similar to SST-AGCN-DA, the SST-AGCN-DG model occupies the feature extractor G_f based on SST-AGCN, the decision confidence level predictor G_y, and the domain predictor G_d. The gradient reversal layer is also applied between G_f and G_d to form the adversarial relationship. However, G_d in this model is the multi-class classifier instead of the binary classifier previously applied which is the significant difference between the domain adaptation method and the domain generalization method. G_d tries to tell features from each domain apart while G_f brings disturbance. During forward propagation, we denote $x_{r,i}$ as the ith input EEG feature from the rth subject D_r, or the rth domain as well. The domains are marked by $D=\{D_1, D_2 ... D_r ...\}$. The signs for the extracted features, the predicted confidence levels, and the predicted domains are similarly transferred from the definition in SST-AGCN-DA. Then in back propagation, the loss function of confidence estimation and domain prediction based on cross entropy loss are defined as:

$$\mathscr{L}_y = -\frac{1}{\Sigma_{r=1}^{D}|D_r|}\Sigma_{r=1}^{D}\Sigma_{i=1}^{|D_r|} y_{r,i}\log\hat{y}_{r,i}, \tag{9}$$

$$\mathscr{L}_d = -\frac{1}{\Sigma_{r=1}^{D}|D_r|}\Sigma_{r=1}^{D}\Sigma_{i=1}^{|D_r|} d_{r,i}\log\hat{d}_{r,i}. \tag{10}$$

The overall loss function and the parameter optimization of the SST-AGCN-DG model follow that of the SST-AGCN-DA model, which are listed in Equation (6)−(8).

Although the SST-AGCN-DG model is also not applicable in intra-subject scenarios, the dataset is balanced for G_d since each subject makes up a unique domain. However, the difficulty of training an effective domain predictor increases with the number of subjects, which is also the number of categories of the classification task for G_d. When facing practical applications with varying numbers of subjects, the adversarial relationship between G_f and G_d becomes too sensitive to hyper-parameters like β in Equation (6) to provide stable results. Thus the SST-AGCN-DG model is theoretically less general than the SST-AGCN-DA model. Another approach to relieving this problem is to utilize implicit labels to redefine the domain of the subjects. The implicit labels could be gender, age bracket, degree of education, etc. The appropriate domain

TABLE I

Experimental results of cross-subject decision confidence levels classification on the EEG dataset collected under the text exam task (%). Models related to the SST-AGCN structure are included. The highest average value and the lowest standard deviation among the models in each evaluation criteria are shown in bold font. Our SST-AGCN-DA takes the lead, and our SST-AGCN-DG performs better than the pure SST-AGCN.

Model	Accuracy	Precision	Recall	F1-score
SST-AGCN	79.48/**7.94**	77.22/8.41	75.77/**8.17**	74.15/**7.59**
Our SST-AGCN-DG	81.50/9.88	79.08/9.77	77.97/9.87	77.04/10.11
Our SST-AGCN-DA	**83.64**/8.75	**81.03**/**8.20**	**80.77**/8.30	**79.45**/8.48

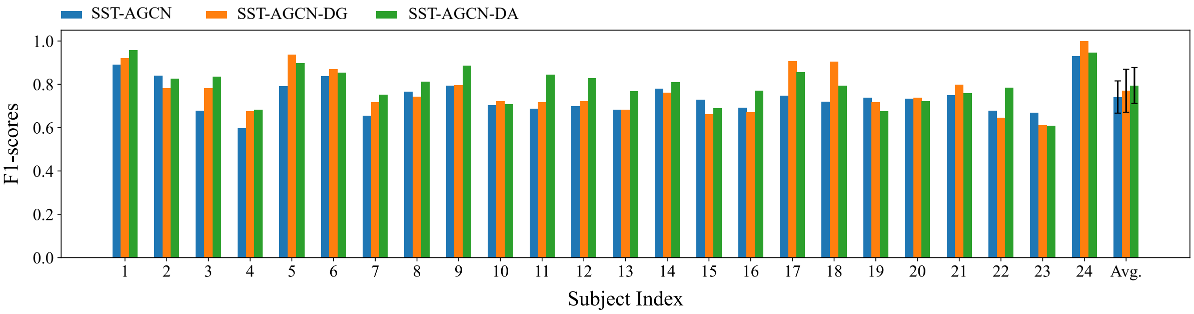

Fig. 4. F1-score for each subject as the application target. SST-AGCN-DA is more effective for most subjects and SST-AGCN-DG work on the rest. The two models work well even when the original SST-AGCN-DA has poor performance.

information used in the SST-AGCN-DG model remains to be explored in future works.

IV. Experiment

A. Dataset

To evaluate the performance of our SST-AGCN-DA and SST-AGCN-DG models in cross-subject decision confidence estimation, we conduct experiments against the original SST-AGCN model on an EEG dataset collected by Li *et al.* under a text-based decision making task [9]. During the experiment, the subjects are asked to answer text-based questions and score their level of confidence immediately after making each decision. The EEG data is recorded during the experiment. There are altogether 80 single-choice blank-filling questions in Chinese which are selected from the question bank of the Chinese high school exams, and just as in real exams, a time limit is given to each question. These measures not only make the experiment more realistic but also ensure the effectiveness of the collected EEG data. In this experiment, there are 5 levels of decision confidence ranging from certainly wrong to certainly correct, and the subjects label their decision with the 5 levels. The EEG data is collected through a 62-channel electrode cap worn by the subjects and is recorded by the ESI Neuroscan system at a frequency of 1000 Hz. The EEG data during the decision-making processes is segmented for further study. Base on this experiment, Li *et al.* created the decision confidence EEG dataset involving 24 healthy subjects, 11 males and 13 females aged between 19 to 24.

B. Experimental Setup

Base on the above EEG dataset, we adopt the leave-one-out-cross-validation method in the following experiments to evaluate the performance of the models in cross-subject scenarios. With each subject as the target domain and other subjects as the source domain, 24 independent decision confidence estimation models are trained and the results are averaged for each method.

In the preprocessing step, we first remove the eye movement artifacts from the EEG signals according to the signals from the electro-oculogram (EOG) and frontal poles zero (FPZ) channels, and filter out the noise by a band-pass filter of 0.3 HZ to 50. We smooth the feature using the linear dynamic system [23] to wipe out abnormal jitter. We retain the EEG signals collected during the decision-making process to ensure the effectiveness of the data in estimating decision confidence. To accelerate the training process and utilize the spectral feature of the raw EEG data more efficiently, we further extract the differential entropy (DE) features [24]. The superiority of DE feature has been demonstrated in other decision confidence estimation tasks [6], [7]. In this experiment, we divide the processed EEG signals into segments of 1-second length, and apply short-time Fourier transform (STFT) with 1-second Hanning window to each segment. The extracted DE features contain five frequency bands (Delta: 1-3 Hz, Theta: 4-7 Hz, Alpha: 8-13 Hz, Beta: 14-30 Hz, Gamma: 31-50 Hz).

The EEG features are of size $N \times F \times C$ where N is the number of samples, F is the number of frequency bands and C is the number of EEG channels. To include the temporal

TABLE II

EXPERIMENTAL RESULTS OF CROSS-SUBJECT DECISION CONFIDENCE LEVELS CLASSIFICATION ON THE EEG DATASET COLLECTED UNDER THE TEXT EXAM TASK (%). MODELS USING DOMAIN ADAPTATION METHODS ARE INCLUDED. THE HIGHEST AVERAGE VALUE AMONG AND THE LOWEST STANDARD DEVIATION AMONG THE MODELS IN EACH EVALUATION CRITERIA ARE SHOWN IN BOLD FONT. OUR SST-AGCN-DA PERFORMS FAR BETTER THAN THE OTHER METHODS.

Model	Accuracy	Precision	Recall	F1-score
DANN	77.42/8.23	73.68/9.85	74.13/9.06	71.69/9.22
ADDA	77.89/**7.48**	74.86/9.24	73.78/**7.20**	71.93/**8.32**
WGANDA	79.11/8.18	75.01/9.12	76.88/8.84	74.25/8.90
Our SST-AGCN-DA	**83.64**/8.75	**81.03/8.20**	**80.77**/8.30	**79.45**/8.48

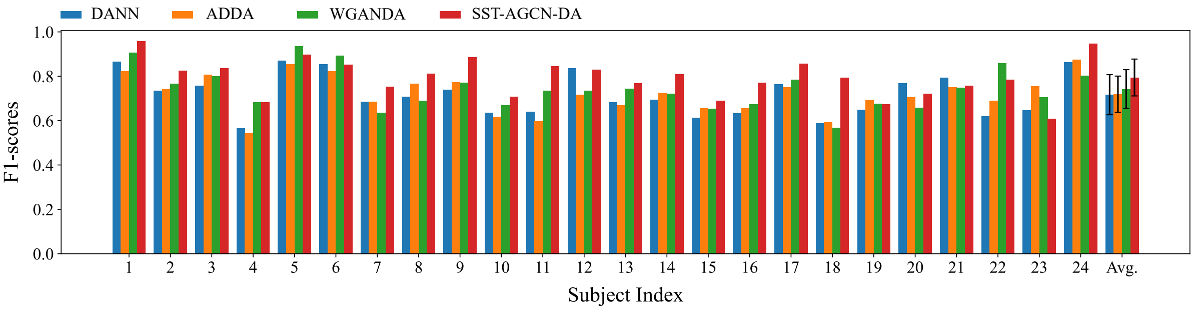

Fig. 5. F1-score for each target subject using the DANN, the ADDA, the WGANDA and the SST-AGCN-DA models. The SST-AGCN-DA model takes the lead, and the WGANDA is slightly better than the other two methods.

information in the input samples, we extend the samples into shape $N \times F \times T \times C$ by an overlapping window with the size of T. In this experiment, $F = 5$, $T = 5$ seconds and $C = 62$. We set the number of SST-AGCN blocks to be $L = 6$. The output channel size of each graph convolutional layer is selected between 30 to 120.

C. Cross-subject Confidence Estimation with SST-AGCN-DA

To exhibit the promotion brought by different transfer learning techniques, especially domain adaptation, we first compare the SST-AGCN, the SST-AGCN-DA and the SST-AGCN-DG models. For each model, we train 24 cross-subject classifiers to predict the level of decision confidence for the 24 subjects. We only adopt the extreme samples labeled by 1 and 5 to ensure the correctness of the labels. Thus, the confidence predictors g_y are binary classifiers. All the deep neural network modules in the SST-AGCN and SST-AGCN-DG models are trained on the EEG feature of the rest 23 subjects and transferred directly to the target subjects, while modules in the SST-AGCN-DA model utilize not only the other 23 subjects but also the EEG data from the target subject. These samples act as the information from the target domain and help in training the feature extractor and the domain predictor.

Table I shows the effects in estimating decision confidence in cross-subject scenarios. To evaluated the results fairly for both low and high confidence levels, we mainly judge by the F1-score. We can see that the mean F1-score and standard deviation of the classifiers based on the SST-AGCN model without using any transfer learning techniques are 74.15% and 7.59%. When the domain adaptation modules are included, the performance of SST-AGCN-DA reaches 79.45%±8.48%. Since part of the samples from the target domain are included in the training process, the feature extractor can remove the information that differentiate samples from the source and target domains against the domain classifier. Then the confidence level classifier can improve the cross-subject performance. Based on domain generalization, SST-AGCN-DG does not utilize any information of the target domain, and the effect is slightly lower than SST-AGCN-DA. Even so, the differences between domains are still eliminated in the training stage, and the SST-AGCN-DG model achieves better results than the SS-AGCN model. The averaged mean value and standard deviation are 77.04% and 10.11%. To show the generality of the performance, we also plot the F1-score for each subject, as shown in Figure 4, and reveal the same phenomenon in most subjects.

D. Comparison among Domain Adaptation Methods

The above results show the superiority of our SST-AGCN-DA model than the original SST-AGCN model in this cross-subject decision confidence estimation tasks, and the potential of the SST-AGCN-DG model is proved. We further prove the SST-AGCN-DA model to be more effective than other domain adaptation methods. The other methods we adopt are the DANN [15], the ADDA [16] and the WGANDA models [18]. We create consistent training environments for all the methods, and the results shown in Table II reveal that the SST-AGCN-DA model surpasses the other domain adaption methods. Similar to the above experiment, the F1-score of each subject acting as target is shown in Figure 5. We can conclude that the SST-AGCN blocks are necessary for efficient feature extractor, and the SST-AGCN-DA model successfully combine it with the domain adversarial structure which promotes domain adaptation. In all, the SST-AGCN-DA model is more suitable for cross-subject decision confidence estimation from EEG signals.

E. Ablation Study

The ablation study is implicit in the above experiments. Among the three main components of the SST-AGCN-DA model, the confidence level predictor is indispensable. If the domain predictor is removed together with the flip gradient module, the model degrades to the original SST-AGCN model. If the SST-AGCN blocks are substituted by basic neural networks like the multilayer perceptron, the model becomes the DANN model. Furthermore, Thus the above experiments prove the effectiveness of combining the idea of domain adversarial in that the feature extractor can produce more target-specific features, and the efficiency of the SST-AGCN-based feature extractor is also validated.

V. Conclusions

The objective evaluation of human confidence in the decision-making process is of great value in both research and application. The SST-AGCN model has greatly improved the prediction performance of decision confidence based on individual subjects. But in practical application, effectiveness in cross-subject scenarios is more critical. In this paper, we have introduced transfer learning techniques into SST-AGCN and have proposed two cross-subject decision confidence estimation models, namely SST-AGCN-DA and SST-AGCN-DG. We utilize the decision confidence EEG dataset collected during a text-based exam to evaluate the effectiveness of our proposed models. The experimental results demonstrate that our methods can reduce the difference in data distribution among subjects and our SST-AGCN-DA achieves state-of-the-art performance on EEG-based cross-subject decision confidence estimation.

References

[1] A. Pouget, J. Drugowitsch, and A. Kepecs, “Confidence and Certainty: Distinct Probabilistic Quantities for Different Goals,” *Nature Neuroscience*, vol. 19, no. 3, pp. 366–374, 2016.

[2] D. Bang and S. M. Fleming, “Distinct Encoding of Decision Confidence in Human Medial Prefrontal Cortex,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 23, pp. 6082–6087, 2018.

[3] P. Molenberghs, F.-M. Trautwein, A. Böckler, T. Singer, and P. Kanske, “Neural Correlates of Metacognitive Ability and of Feeling Confident: a Large-scale fmri Study,” *Social Cognitive and Affective Neuroscience*, vol. 11, no. 12, pp. 1942–1951, 2016.

[4] A. Boldt, A.-M. Schiffer, F. Waszak, and N. Yeung, “Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making,” *Scientific Reports*, vol. 9, no. 1, pp. 1–17, 2019.

[5] A. R. Damasio, “The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex,” *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 351, no. 1346, pp. 1413–1420, 1996.

[6] R. Li, L.-D. Liu, and B.-L. Lu, “Measuring Human Decision Confidence from EEG Signals in an Object Detection Task,” in *International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2021, pp. 942–945.

[7] R. Li, L.-D. Liu, and B.-L. Lu, “Discrimination of Decision Confidence Levels from EEG Signals,” in *International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2021, pp. 946–949.

[8] L.-D. Liu, R. Li, Y.-Z. Liu, H.-L. Li, and B.-L. Lu, “EEG-based Human Decision Confidence Measurement Using Graph Neural Networks,” in *International Conference on Neural Information Processing*. Springer, 2021, pp. 291–298.

[9] R. Li, Y.-T. Wang, and B.-L. Lu, “Measuring Decision Confidence Levels from EEG Using a Spectral-Spatial-Temporal Adaptive Graph Convolutional Neural Network,” in *International Conference on Neural Information Processing*, 2022.

[10] M. Welling and T. N. Kipf, “Semi-supervised Classification with Graph Convolutional Networks,” in *J. International Conference on Learning Representations (ICLR 2017)*, 2016.

[11] L. Shi, Y. Zhang, J. Cheng, and H. Lu, “Two-stream Adaptive Graph Convolutional Networks for Skeleton-based Action Recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 026–12 035.

[12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain Adaptation via Transfer Component Analysis,” *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[14] W.-L. Zheng and B.-L. Lu, “Personalizing EEG-based Affective Models with Transfer Learning,” in *Proceedings of the Twenty-fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 2732–2738.

[15] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, “Cross-subject Emotion Recognition Using Deep Adaptation Networks,” in *International Conference on Neural Information Processing*. Springer, 2018, pp. 403–413.

[16] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial Discriminative Domain Adaptation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.

[17] L.-M. Zhao, X. Yan, and B.-L. Lu, “Plug-and-play Domain Adaptation for Cross-subject EEG-based Emotion Recognition,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, 2021, pp. 863–870.

[18] Y. Luo, S.-Y. Zhang, W.-L. Zheng, and B.-L. Lu, “WGAN Domain Adaptation for EEG-Based Emotion Recognition,” *International Conference on Neural Information Processing*, vol. 11305, pp. 275–286, 2018.

[19] B.-Q. Ma, H. Li, Y. Luo, and B.-L. Lu, “Depersonalized Cross-subject Vigilance Estimation with Adversarial Domain Generalization,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[20] Z. Jia, Y. Lin, J. Wang, X. Ning, Y. He, R. Zhou, Y. Zhou, W. Li, and H. Lehman, “Multi-view Spatial-Temporal Graph Convolutional Networks with Domain Generalization for Sleep Stage Classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1977–1986, 2021.

[21] Y. Ganin and V. Lempitsky, “Unsupervised Domain Adaptation by Backpropagation,” in *International Conference on Machine Learning*. PMLR, 2015, pp. 1180–1189.

[22] Y. Li, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao, “Deep Domain Generalization via Conditional Invariant Adversarial Networks,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 624–639.

[23] L.-C. Shi and B.-L. Lu, “Off-line and On-line Vigilance Estimation Based on Linear Dynamical System and Manifold Learning,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 6587–6590.

[24] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, “Differential Entropy Feature for EEG-based Emotion Classification,” in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.