Bipartite Graph Adversarial Network for Subject-Independent Emotion Recognition

Marzieh Niaki, Shyamal Y. Dharia, Yangjun Chen, Camilo E. Valderrama

***Abstract*— Emotions play a vital role in connecting and sharing with others. However, individuals with emotional disorders face challenges in expressing their emotions, affecting their social lives. Current artificial intelligence tools support this problem by enabling the development of methods that recognize emotions from electroencephalographic (EEG) signals. However, the high variability across individuals poses challenges in developing emotion recognition methods that generalize well across different subjects. Previous studies have addressed this issue using domain adversarial neural networks (DANN), in which differences in EEG among individuals are minimized. Although DANN has shown a potential to reduce domain variance, previous studies have little explored the inclusion of layer-specific components to further advance towards that goal. This study addressed this limitation by incorporating bipartite (BP) graphs in a DANN architecture to reduce variability further. We evaluated our model on five benchmark datasets for emotion recognition (SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER) comprising a total of 62 individuals. Our model yielded an accuracy of 82.1%, 77.3%, 85.8%, 90.7%, and 87.6% for the SEED-V, SEED-IV, SEED, SEED-FRA, and SEED-GER datasets, respectively. Notably, these accuracies are either higher or comparable to the current state-of-the-art models. Furthermore, our model identified that the frontal, temporal, and parietal EEG channels are crucial for detecting emotions evoked by audiovisual stimuli.**

***Index Terms*— Deep Learning, Domain Adaptation, Electroencephalogram (EEG), Emotion Recognition, Graph Neural Networks**

I. Introduction

Emotions are complex cognitive processes and physiological states that arise in response to stimuli, such as experiences, thoughts, or interactions. Emotions encompass personal feelings, thought processes, behaviors, physical responses, and methods of expression. Given the relevance of emotion in human life, recognizing emotions is critical across various fields [1]. For instance, in healthcare applications, recognizing emotions supports management of sleep disorders [2], depression [3], attention and autism disorders [4], and panic disorder [5]. In marketing, emotion recognition can help to reveal consumer preferences, allowing businesses to refine their strategies [6]. Finally, in human-computer interaction, emotion recognition enables machines to interpret and replicate human emotional behavior in various applications [7].

Humans can recognize emotions by observing cues, such as speech, body gestures, and facial expressions [8]. However, these external indicators can be easily manipulated, thus raising concerns about their reliability for accurately identifying emotions. Since the brain regulates emotions, a more reliable method to recognize emotions is through neuroimaging techniques. Among these techniques, electroencephalography (EEG) has shown promise, as it can capture electrical activity associated with neuronal processes by using electrodes placed on the scalp. This capability supports the development of machine learning models that associate cortical electrical activity with emotional states [9].

Although EEG has the potential to support emotion recognition methods, EEG signals differ widely between individuals. This condition challenges the usability of EEG-based emotion recognition systems because, once a model is developed, it might not be effective for new users. In the machine learning field, this problem is known as the domain shift problem, occurring when the training data distribution does not match that of the test data [10]. For EEG signals, the significant variability between individuals means that the patterns learned from the training individuals may not fully apply to new individuals, thereby reducing the predictive performance of the models.

To address the domain shift problem, researchers have widely adopted Domain Adversarial Neural Networks (DANN) [11] due to their efficacy in reducing the variance between the training data (source domain) and test data (target domain). DANN aims to generate domain-invariant features by employing adversarial learning strategies, where a domain classifier works against a feature extractor module to minimize the discrepancy across the source and target domains [11].

Previous studies have shown the potential of DANN to address the domain shift problem inherent in subject-independent approaches [12]–[14]. The effectiveness of DANN is based on its ability to generate features that not only support emotion recognition but also have a similar distribution between the source and target domains. However, relying only on DANN to address the domain shift problem limited the exploration of layer-specific components or unique network architectures

This work was supported by the NSERC under Discorvery Grant RGPIN-2024-05575. Marzieh Niaki, Shyamal Y. Dharia and Camilo E. Valderrama contribute equally (Corresponding author: Camilo E. Valderrama)

All authors are affiliated with the Department of Applied Computer Science, University of Winnipeg, Winnipeg, MB, Canada. Camilo E. Valderrama is also affiliated with the Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. (e-mail: pouresmaeilniaki-s@webmail.uwinnipeg.ca; dharia-s@webmail.uwinnipeg.ca; y.chen@uwinnipeg.ca; c.valderrama@uwinnipeg.ca).

that could further mitigate this problem. Therefore, the use of layer-specific components remains unexplored in emotion recognition for subject-independent approaches, thus providing an opportunity to enhance subject-independent emotion recognition approaches.

One way to incorporate layer-specific components into deep learning architectures is to use bipartite (BP) graphs. The potential of BP graphs has been shown in video processing for classification [15], where BP graph has reduced the discrepancies between source and target domains. Specifically, BP graphs represent samples from source and target domains as nodes and their similarity as edges, thus mapping the source and target feature distribution to a common feature space, in which DANN principles can act more suitably.

In this study, we adapt the work proposed in [15] for video classification, introducing an architecture that integrates DANN with BP graphs and transformer encoders to address the domain shift problem in EEG-based emotion recognition. Our model comprises two modules, one for capturing spatial relationships between EEG channels and another for aligning patterns across the temporal dimension. To evaluate the effectiveness of our approach in overcoming the domain shift problem, we test it on five benchmark datasets for emotion recognition (SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER), which collectively comprise data from 62 subjects across three nationalities (Chinese, French, and German).

Our contributions are summarized as follows:

- Introducing a novel architecture incorporating dual-level BP graphs to extract domain-invariant features, enabling subject-independent emotion recognition.
- Achieving a performance matching or surpassing that of current state-of-the-art methods, showing the robustness of BP graphs in generating domain-invariant features for subject-independent emotion recognition.
- Identifying key EEG channels for emotion recognition by analyzing the features extracted by BP graphs across five independent datasets.

II. Related Work

A. Strategies to address domain shift problem in emotion recognition

Previous studies have addressed the domain shift problem in emotion recognition tasks by using DANN to align feature distributions across source and target domains through adversarial training [12]–[14]. To design the feature extractor of the DANN, transformer-based architectures have gained prominence due to their ability to capture complex spatial-temporal dependencies in EEG signals via self-attention mechanisms. For example, Liu et al. [16] proposed a transformer-based DANN framework for emotion recognition that jointly processes EEG and eye movement data. Similarly, Li et al. [17] introduced a knowledge distillation-based lightweight DANN model, in which a transformer-based teacher network guides a lightweight Bi-LSTM student model to enhance temporal-spatial feature learning.

Researchers have also used graph neural networks (GNNs) to enhance the feature extraction of DANN models. GNN allows learning spatial dependencies between EEG channels, preserving topological structures and frequency-specific insights critical for emotion decoding. For instance, Chen et al. [13] showed that including GNN can enhance emotion recognition on cross-subject approaches. Similarly, Shi et al. [14] proposed a Functional Connectivity Patterns Learning network (FCPL) using a GNN to model functional connectivity between brain regions to capture fine-grained emotional connectivity patterns.

In addition to DANN, previous studies have explored other strategies to address the domain shift problem in EEG-based emotion recognition. One of these strategies is to use functional connectivity (FC) analysis, which captures inter-channel relationships in the temporal and frequency domains. As reported by [18], [19], FC-derived brain connectivity features enhanced performance on subject-independent emotion recognition tasks. Another strategy consists of using autoencoders to dynamically model inter-channel dependencies. This was explored in [20], where autoencoders were used to learn latent embedding features that support emotion recognition across multiple sessions for the same user. Lastly, multimodal approaches that integrate EEG signals with eye-tracking signals have been shown to enhance generalization in subject-independent approaches [21]. However, these approaches require access to additional modalities, which may not be practical for EEG-only applications.

Although DANN, FC, and multimodal strategies have all shown promise in improving subject-independent emotion recognition using EEG signals, these strategies primarily focus on input-level feature alignment, inter-channel EEG correlations, and cross-modal fusion with eye-tracking signals. As such, previous studies have ignored including layer-specific components that could further address the domain shift problem in EEG-based emotion recognition. In other fields, BP graphs have shown to have the potential to be this layer-specific component that can help alleviate the domain shift problem [15]. However, this idea has not been explored in emotion recognition methods, thus leaving room to explore using BP graphs to align complex EEG features and enhance domain-invariant learning.

B. Important EEG channels for emotion recognition

In addition to developing emotion recognition methods with high predictive performance, it is also important to provide insight into which EEG channels are more relevant for emotion recognition across subjects. The latter is because using fewer EEG channels reduces computational requirements and supports the development of more comfortable EEG caps [9].

Previous studies have identified important EEG channels by examining energy distribution, based on differential entropy (DE) features, across the brain cortex [22]–[24]. According to these analyses, happy stimuli increase activation in the temporal lobe, fearful stimuli reduce activation in the occipital region, and neutral stimuli activate the parietal and frontal lobes [22]. Happy stimuli were also shown to produce stronger

activation than other emotions, especially in the temporal lobe [23], as well as in the lateral temporal and prefrontal lobes [24]. In addition to analyzing DE distribution, other studies have used feature selection methods to identify significant EEG channels for emotion recognition. For instance, Apicella et al. [25] found that channels Fp_1, Fp_2, F_3, and F_4 were most important for detecting valence, while P_3 and P_4 were most informative for arousal.

Despite significant efforts to identify relevant EEG channels, previous studies have focused on analyzing DE features prior to training deep learning models. This limits the capacity of deep learning to uncover, throughout the training process, the relevant patterns associated with emotional recognition. In order to address this limitation, in our study, we analyze the domain-invariant features learned by the BP graph during the training process, thus providing information into which EEG channels were more determinant of the emotion predictions.

III. Methodology

A. Datasets

In this work, we used five different datasets. All of these datasets contained EEG signals collected from subjects while watching video clips. Each video was aimed to evoke a specific emotion. The EEG signals were collected using a 62-channel EEG system at a sampling rate of 1,000 Hz. To enhance computational efficiency while preserving relevant information, the data were downsampled to 200 Hz.

1) SEED: The SEED dataset [26] comprised data from 15 subjects (7 male and 8 female). Each participant underwent experiments in three sessions, each spaced about a week apart. In each session, the EEG data was recorded while subjects reacted to 15 movie clips, intended to elicit negative, neutral, and positive emotional responses. A total of 15 EEG signals were collected for each stimulus, totaling 45 EEG signals for each subject.

2) SEED-IV: The SEED-IV dataset [27] contained EEG signals evoked with video clips from four emotions: happy, sad, fear, and neutral. The dataset included data from 15 subjects (6 male and 9 female), each participating in three separate sessions on different days. Each session comprised 24 trials, at which participants watched six film clips for each emotion. As a result, 72 EEG signals were collected for each subject.

3) SEED-V: The SEED-V dataset [28] contained EEG signals associated with five emotions: happy, sad, fear, disgust, and neutral. A total of 16 subjects (6 male and 10 female) participated in this study. Each subject participated in three sessions, watching a total of three movie clips per emotion at each session (i.e., 15 movie clips per session). As a result, 45 EEG signals were collected for each subject.

4) SEED-FRA: The SEED-FRA dataset [29] includes EEG signals recorded from eight native French subjects. Each subject participated in three experimental sessions. During each session, the participant watched 21 video clips, each representing one of three emotional categories: positive, neutral, or negative. Consequently, a total of 63 EEG signals were collected per subject.

5) SEED-GER: The SEED-GER dataset [29] consists of EEG signals from eight native German subjects. Each subject participated in three sessions. Each of these sessions included a total of 18 video clips, corresponding to three emotional stimuli: positive, neutral, and negative. Consequently, a total of 54 EEG signals were recorded for each subject.

B. EEG preprocessing

To remove noise and artifacts from raw EEG signals, a bandpass filter from 0.5 to 50 Hz was applied. This range satisfied the Nyquist frequency (100 Hz) and preserves the key brain wave frequencies: delta (δ : 0.5−4 Hz), theta (θ : 4−8 Hz), alpha (α : 8−12 Hz), beta (β : 12−30 Hz), and gamma (γ : 30−50 Hz).

To further remove noise and artifacts from the EEG signals, the signals were modeled as a linear dynamic system as:

$$x_t = z_t + w_t, \quad z_t = Az_{t-1} + v_t, \tag{1}$$

where x_t was the observed EEG signal at time t, z_t was the latent component corresponding to the actual neural activity, w_t was the white noise of the observation ($w_t \sim \mathcal{N}(\bar{w}, Q)$), A was the state transition matrix, and v_t was the white noise associated with neural sources ($v_t \sim \mathcal{N}(\bar{v}, R)$). To solve these equations, they can be expressed in terms of Gaussian conditional distributions, as follows:

$$\begin{aligned} p(x_t|z_t) &= \mathcal{N}(x_t|z_t + \bar{w}, Q), \\ p(z_t|z_{t-1}) &= \mathcal{N}(z_t|Az_{t-1} + \bar{v}, R). \end{aligned} \tag{2}$$

To solve Eq. 2, the initial state was defined as $p(z_1) = \mathcal{N}(z_1|\pi_0, S_0)$, leading to the parametrization of the model as $\omega = \{\bar{w}, Q, A, \bar{v}, R, \pi_0, S_0\}$. This model was solved using the Expectation-Maximization (EM) algorithm, as described in [30].

C. EEG segmentation

The EEG signals were segmented using a four-second window, providing a frequency resolution of 0.25 Hz ($\frac{1}{4}$). This resolution allowed each segment to capture two full cycles of the lowest frequency, ensuring an accurate representation of the slowest brain wave activity.

D. Differential Entropy

At each 4-second, spectral characteristics were calculated from the EEG signal using differential entropy (DE) in the frequency bands δ, θ, α, β, and γ. The DE is an extension of Shannon entropy for continuous random variables, which measures the uncertainty or randomness as:

$$DE = -\int_{-\infty}^{\infty} P(x)\ln(P(x))\,dx, \tag{3}$$

where $P(x)$ is the probability density function.

Assuming that EEG signals obey a Gaussian distribution $x \sim N(\mu, \sigma)$, the DE can be approximated as follows:

$$\begin{aligned} DE &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx \\ &\approx \frac{1}{2}\ln(2\pi e\sigma^2), \end{aligned} \tag{4}$$

where σ^2 denotes the variance of the segment.

DE was calculated for each channel, resulting in 5 values per channel (one for each frequency band), totaling 310 features per 4-second segment (62 channels $\times$ 5 frequency bands $=$ 310).

E. Preprocessed dataset dimensions

After calculating the DE for each four-second segment, the features corresponding to the same video clip were concatenated, forming a three-dimensional structure for each clip with dimensions (W, C, F). Here, W represented the number of 4-second segments, C was the number of EEG channels, and F was the dimensionality of the feature vector F, where F was a vector containing the five DE features corresponding to the frequency bands ($F = [f_\delta, f_\theta, f_\alpha, f_\beta, f_\gamma]$). Since the durations of the video clips varied, the number of 4-second segments, W, differed across videos. To standardize the input dimensions, W was set to match the longest video. The maximum values for W were 66, 64, 74, 75, and 105 for SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER, respectively.

The three-dimensional structures were then stacked to create four-dimensional input of shape (N, W, C, F), where N denotes the number of samples (subjects × video clips per subject). For SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER these shapes were: $(675, 66, 62, 5)$, $(1080, 64, 62, 5)$, $(720, 74, 62, 5), (945, 75, 62, 5), (810, 74, 105, 5)$, respectively.

F. Network Architecture

Figure 1 shows the architecture used for emotion prediction. The model was structured with two modules to process the EEG features. The first module, known as the spatial module, captures the relationships between EEG channels based on their DE features. The second module, known as the temporal module, captured temporal dependencies by modeling variations in feature representations across the 4-second EEG segments.

Both spatial and temporal modules have a transformer encoder with a BP graph. The transformer captured complex dependencies between the elements: the EEG channels for the spatial module, and the 4-second segments for the temporal module. The BP graph post-processed the embedding features given by the transformer to align features across source and target domains.

1) Spatial Module: To process the relationships between EEG channels, we first reshaped the dimensions of the input data *(B, W, C, F)* into *(B $\times$ W, C, F)*, where B is the batch size, W the number of 4-second segments per video, C the number of channels, and F the DE features.

a) Spatial Transformer Encoder: The reshaped data was fed into a transformer encoder, which used an attention mechanism to capture complex patterns between EEG channels. This attention mechanism was implemented using multi-head attention layer, where each attention head computed its attention weights using the scaled dot-product attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{5}$$

The query (Q), key (K), and value (V) matrices corresponded to the transformer inputs with dimensions *(B $\times$ W, C, F)*. The matrix QK^T measured the similarities between EEG channels based on the DE feature values, while the softmax function normalized these values, ensuring they sum to 1 across each EEG channel. This correlation weight shows how strongly the activity in one channel relates to another, allowing the model to focus on interactions or dependencies across EEG channels. The final multiplication with the matrix V generated new DE features for each EEG channel by linearly combining the DE features of other EEG channels, weighted by the correlations given by QK^T.

The outputs from all attention heads were then concatenated and projected back into the original feature space using the multi-head attention mechanism:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O, \tag{6}$$

where W^O was the output projection matrix.

The output of the multi-head layer was then added to the input using a residual connection, followed by layer normalization for stability and improved convergence. The final step of the transformer encoder was a feed-forward layer with an expansion factor of 2 and a dropout rate of 0.2.

b) Spatial bipartite graph: The output of the transformer encoder was inputted into a BP graph. This BP graph was used to align features from the source and target domains in order to bridge the domain gap and improve the model's ability to make cross-domain predictions. Specifically, we defined a bipartite graph $G_s = (V_s, V_t, E_{st})$, where V_s and V_t were the output of the spatial transformer for the source and target samples. The edge set E_{st} represented the connections between the source and target nodes, measuring their similarity between the corresponding EEG features.

To calculate the similarity between the DE features of the source and target nodes, we first permuted the dimensions of the node features V_s and V_t, from *(B $\times$ W, C, F)* to *(C , B $\times$ W, F)*. This permutation allowed us to measure similarities between nodes for each EEG channel independently. To compute the similarity between the DE features of the source and target nodes, we calculated the pairwise absolute differences between the feature vectors. Given a source node x_i and a target node x_j, the difference was computed as:

$$x_{ij} = |x_i - x_j|.$$

This procedure resulted in a structure of dimensions *(C , B $\times$ W, B $\times$ W, F)*.

To learn similarity scores between the node pairs, the computed differences were processed through a convolutional neural network (CNN), yielding the normalized edge weights, A, as:

$$A = \sigma\left(F(|x_i - x_j|; \omega)\right),$$

where $F(\cdot; \omega)$ was a two-layer CNN parameterized by ω. The first layer of the CNN consisted of κ filters, while the second layer had a single filter. Both layers employed a kernel size and stride of 1. The activation function, σ, was the sigmoid function to constrain the values of A between 0 and 1.

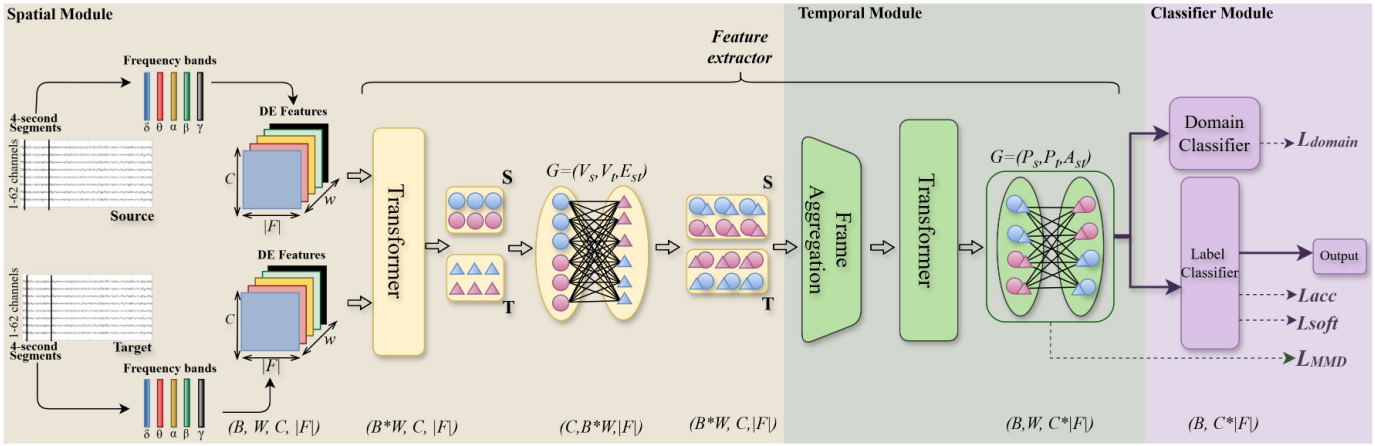


Fig. 1: Proposed DANN architecture incorporating BP graphs. The feature extractor comprises spatial and temporal modules, each utilizing transformers and BP graphs. Extracted features were fed into the domain and label classifiers, with four distinct loss functions applied to align the source and target domains. The shapes at the bottom of each block illustrate the reshaping operations performed at each layer. Here, B is the batch size, W is the number of 4-second window segments, C is the number of EEG channels, and $|F|$ is the set of DE features from five frequency bands.

The resulting edge weights A were further normalized using L1 normalization across both rows and columns to ensure that each source node's edge weights sum up to 1:

$$\tilde{A}_i = \frac{A_i}{\|A_i\|_1}, \quad E_{st} = \frac{\tilde{A}_j}{\|\tilde{A}_j\|_1}$$

Once the edge weights were computed, the source and target features were aggregated using edge matrix E_{st} as:

$$\tilde{V}_s = E_{st}V_t, \quad \tilde{V}_t = (E_{st})^T V_s$$

The aggregated features were then concatenated with the original node features, and processed through another CNN to update the node embeddings:

$$V_s \leftarrow H([V_s; \tilde{V}_s]; \phi), \quad V_t \leftarrow H([V_t; \tilde{V}_t]; \phi),$$

where $H(\cdot; \phi)$ was a CNN of two layers parameterized by ϕ, responsible for performing the feature fusion. The first layer of G consists of κ filters, and the second layer has F filters. Both layers use a kernel size and stride of 1. After computing the spatial alignment of the source and target features, the output of the spatial BP was permuted to *(B* $\times$ *W, C, F)*.

2) Temporal Module: The spatial features were input into the temporal module to capture variations across the 4-second segments. To achieve this, we first reshaped the input from *(B* $\times$ *W, C, F)* to *(B, W, C* $\times$ *F)*. This reshaping organized the temporal information, W windows, along the second dimension, a requirement for input to the transformer encoder.

a) Temporal transformer encoder: Before inputting the three-dimensional input into the Transformer encoder, we applied positional encoding across the temporal dimension to help the model understand the sequential order of the 4-second segments. The output of the positional encoder was then fed into a transformer encoder composed of a multi-head attention layer and a feed-forward layer.

For the temporal transformers, the attention multiplication QK^T measured the similarities between the 4-second segments based on the spatial features extracted from each segment. The final multiplication with the matrix V then updated the spatial features of each 4-second segment by incorporating the spatial features of other segments, weighted by the correlation scores.

b) Temporal bipartite graph : Analogous to the spatial module, the output of the temporal transformer was passed through a BP graph, $G_t = (P_s, P_t, A_{st})$, where P_s and P_t were the output of the temporal transformers for the source and target samples, respectively, and A_{st} were the edges measuring similarities between the nodes. To perform the alignment of the features across each temporal segment independently, P_s and P_t were permuted to dimensions *(W , B, C* $\times$ *F)*. The similarity computation and fusion of temporal features followed the same procedure as described for the spatial BP graph. The output of the temporal BP was reshaped back to its original dimension, namely *(B , W, C* $\times$ *F)*.

c) Temporal average pooling: In the final step of the temporal module, we applied average pooling across the window dimension. This process reduced the output to a two-dimensional structure with dimensions *(B, C* $\times$ *F)*.

3) Domain-Adversarial Neural Networks (DANN) : To further enhance the model's performance across different domains and ensure the generation of domain-invariant features, we trained our model using a DANN strategy. Thus, the spatial and temporal modules were established as the feature extractor. The output of the feature extractor was fed into the label classifier and the domain classifier.

The label classifier consisted of a dropout layer and a fully connected (FCN) layer with a number of outputs equal to the number of emotions: 3 for SEED, SEED-FRA, and SEED-GER; 4 for SEED-IV, and 5 units for SEED-V. Similarly, the domain classifier also included a dropout layer, and a FCN with 2 outputs ('0': source, '1':target).

4) Loss functions: To train our model, we employed three different loss functions. The first loss function corresponded to the losses of the label and domain classifiers. The second loss

function focused on avoiding an even distribution of the target outputs. Finally, the third loss function aimed to minimize the discrepancy between the feature distributions of the source and target domains.

a) DANN loss functions: For the DANN, we used categorical cross-entropy loss for the label prediction as:

$$L_{\text{cce}} = -\frac{1}{B_s}\sum_{i=1}^{B_s} y_i \log(C_y(x_{s_i})), \tag{7}$$

where B_s was the batch size for the source data, y_i was the true labels, x_{s_i} represented the i-th data sample in the source batch, and C_y was the classifier's output function.

Similarly, for the domain classifier, the binary cross-entropy was calculated as:

$$L_{\text{domain}} = -\frac{1}{B}\sum_{k=1}^{B} d_k \log(C_d(x_k)), \tag{8}$$

where B was the combined batch size of source and target data, d_k were the domain labels, x_k was the k-th data sample, and C_d was the domain classifier's output function.

b) Soft entropy loss function: To reduce ambiguity or evenly situation in the target output, we employed a soft entropy loss function by applying Shannon's entropy on the target output as:

$$L_{\text{soft}} = -\frac{1}{B_t}\sum_{j=1}^{B_t} C_y(n_{t_j}) \log(C_y(n_{t_j})), \tag{9}$$

where B_t was the batch size for the target data, n_{t_j} was the j-th data sample in the target batch, and C_y represents the classifier's output for the target samples.

c) Domain Discrepancy Reduction: Additionally, we applied the maximum mean discrepancy (MMD) loss function to minimize the domain discrepancy between the source and target domains. The MMD loss was calculated using Radial Basis Function (RBF) kernels with different σ values (specifically, 1, 2, 3, 4, and 16), which was essential for capturing multi-scale structure in the data distribution:

$$L_{\text{MMD}} = \left\| \frac{1}{N_s}\sum_{i=1}^{N_s} \psi(x_i^s) - \frac{1}{N_t}\sum_{j=1}^{N_t} \psi(x_j^t) \right\|_{\mathcal{H}_k}^2 \tag{10}$$

where $\psi(\cdot)$ represented the feature map into a space induced by the RBF kernel, x_i^s and x_j^t were the samples from the source and target domains, respectively.

d) Overall Loss Function: The total loss function for training the model was a weighted sum of the categorical cross-entropy for task and domain classification, soft entropy, and MMD losses, with the weights for the soft entropy loss and the MMD loss each set to 0.1:

$$L_{\text{total}} = L_{\text{cce}} + L_{\text{domain}} + 0.1 \cdot L_{\text{soft}} + 0.1 \cdot L_{\text{MMD}} \tag{11}$$

IV. Experiment Details

A. Execution Environment

Our method was implemented using PyTorch with Python version 3.10.10. The model was trained on an NVIDIA RTX A6000 GPU. We set the epochs to 60, batch size to 32 and used weighted Adam optimizer (AdamW) as the network optimizer.

B. Model Evaluation

To evaluate the performance of our proposed model, we used the Leave-One-Subject-Out Cross-Validation (LOSOCV), thus evaluating our model under the subject-independent approach. For each subject, we calculated the accuracy for each emotion class. We also calculated the overall accuracy by taking the average across the emotions.

To optimize each model's hyperparameters, at each iteration of the LOSOCV, a grid search using 20 epochs was performed on a validation set composed by samples of two subjects different from the test subject. The grid search space was defined as follows:

- Learning rate: $\{10^{-4}, 5\times10^{-4}, 10^{-3}\}$
- Weight decay: $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
- Number of heads for the spatial multi-head attention layer: $\{1, 5\}$ (embedding dimension: 5)
- Number of heads for the temporal multi-head attention layer: $\{2, 5\}$ (embedding dimension: 310)
- Number of filters, κ, used in the CNN to compute the edge matrix and perform feature fusion within the BP graphs: $\{8, 16, 32, 64\}$
- Dropout rate for the label and domain classifiers: $\{0.3, 0.5, 0.7\}$

C. Feature Visualization with t-SNE

To qualitatively assess the discriminative capability of the learned representations, we used t-distributed Stochastic Neighbor Embedding (t-SNE) to visualize the high-dimensional feature space in two dimensions. Thus, we applied t-SNE with a perplexity of 30 to the features outputting the temporal module, as these were the features used for emotion and domain prediction. This visualization was performed for each dataset (SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER) to examine the degree of class separability and the impact of different model components.

D. Feature distribution across cortical areas

To identify the relevant EEG channels, we analyzed the features produced by the feature extractor after passing the input through the spatial and temporal modules. Given that the data input had dimensions $(N, W, 62, 5)$, the feature extractor outputted an element with dimensions $(N, 62\times5)$, where N denoted the number of samples (subjects × video clips per subject), 62 corresponded to the EEG channels and 5 to the frequency bands.

To visualize common activation patterns associated with each emotion and frequency band, the extracted feature outputs were reshaped to a tensor of dimensions $(S, M, 62, 5)$, where S was the number of subjects in each dataset (SEED: 15, SEED-IV: 15, SEED-V: 16, SEED-FRA: 8, SEED-GER: 8), and M was the number of video samples per subject (SEED: 45, SEED-IV: 72, SEED-V: 45, SEED-FRA: 63, SEED-GER: 54). To ensure consistency across subjects, features were normalized individually for each subject. Subsequently, for each emotion category, the mean feature values were computed across all subjects for each EEG channel and frequency band

pair. These average values were then visualized using topographic maps (topomaps) to identify cortical regions exhibiting stronger feature values across different emotional states.

E. Finding relevant channels for emotion recognition

To identify the relevant EEG channels for distinguishing among emotions, we applied the Friedman test [31] at a significance level of 0.05 to evaluate whether the features extracted from each EEG channel and frequency band significantly differed across emotions. We performed the Friedman test individually for each of the five datasets.

To determine the relevant EEG electrodes across the five data sets, we identified the EEG channels for each frequency band that resulted in statistically significant emotional differences for at least three datasets. Finally, to derive an overall group of relevant EEG channels across frequency bands, we identified those that were significant in at least three of the five frequency bands.

F. Ablation Study

To evaluate the contribution of each component to emotion prediction, we conducted an ablation study by systematically removing key modules from the proposed architecture. A total of six ablation experiments were performed. In the first experiment, both BP graphs were removed, and the model was trained using only the features extracted by the transformer modules. The second and third experiments assessed the role of spatial components: the second excluded only the spatial BP graph, while the third removed the entire spatial module. Similarly, the fourth and fifth experiments examined the temporal components by first removing only the temporal BP graph and then omitting the entire temporal module. Finally, in the sixth experiment, the model was trained without the DANN component to assess the impact of excluding the domain loss.

V. Results

A. Model performance

Table I presents the model performance on the SEED-V dataset. The model achieved an overall accuracy of 82.1% and a standard deviation of 5.5%, indicating stable and consistent performance across subjects. The class with the lowest performance was sad emotion, with an accuracy of 73.6% and a standard deviation of 10.5. In particular, as shown in Figure 2a, the model misclassified sad emotion with 'neutral' 11.8% of the time. In contrast, the highest accuracy was achieved for disgust, with an accuracy of 93.8% and a standard deviation of 7.0%.

For SEED-IV, the accuracy was lower than that of SEED-V, resulting in an overall 77.3% (Table II). However, the standard deviation was also low (3.2%), suggesting that although accuracy declined, the model produced a stable performance across all subjects. Similarly to SEED-V results, the lowest performance was for a negative emotion: fear, with an accuracy of 72.7% and a standard deviation of 7.9%. On the other hand, the highest performing class was happy, achieving 85.9% accuracy with a standard deviation of 3.6%. The confusion matrix for SEED-IV (Figure 2b) showed a similar trend to SEED-V, where the least performing class, fear, was misclassified as neutral 17.4% of the time.

TABLE I: Emotion performance for the 16 subjects in the SEED-V dataset. The last column displays the average performance across all emotions. The last two rows present the mean and standard deviation for each emotion across subjects, along with the 95% confidence interval.

Subject	Disgust (%)	Fear (%)	Sad (%)	Neutral (%)	Happy (%)	Overall (%)
1	100.0	88.9	88.9	88.9	88.9	91.1
2	88.9	77.8	100.0	77.8	77.8	84.4
3	100.0	88.9	66.7	88.9	77.8	84.4
4	100.0	77.8	66.7	88.9	77.8	82.2
5	100.0	88.9	77.8	88.9	77.8	86.7
6	88.9	77.8	66.7	77.8	55.6	73.3
7	88.9	88.9	88.9	88.9	77.8	86.7
8	100.0	77.8	88.9	66.7	77.8	82.2
9	88.9	100.0	66.7	66.7	77.8	80.0
10	88.9	66.7	55.6	66.7	88.9	73.3
11	88.9	77.8	66.7	77.8	88.9	80.0
12	100.0	88.9	55.6	88.9	88.9	84.4
13	100.0	88.9	77.8	100.0	55.6	84.4
14	77.8	77.8	66.7	66.7	77.8	73.3
15	100.0	88.9	77.8	88.9	88.9	88.9
16	88.9	77.8	66.7	77.8	77.8	77.8
Mean ± SD	93.8 (7.0)	83.3 (12.7)	73.6 (10.5)	81.3 (10.3)	78.5 (9.9)	82.1 (5.5)
95% CI	90.0-97.5	79.0-87.7	66.8-80.4	75.6-86.9	73.0-84.0	79.2-85.0

TABLE II: Emotion performance for the 15 subjects in the SEED-IV dataset. The last column displays the average performance across all emotions. The last two rows present the mean and standard deviation for each emotion across subjects, along with the 95% confidence interval.

Subject	Neutral (%)	Sad (%)	Fear (%)	Happy (%)	Overall (%)
1	61.1	77.8	72.2	88.9	75.0
2	77.8	72.2	72.2	88.9	77.8
3	72.2	83.3	66.7	83.3	76.4
4	77.8	83.3	66.7	83.3	77.8
5	72.2	72.2	66.7	88.9	75.0
6	72.2	72.2	72.2	83.3	75.0
7	66.7	88.9	66.7	94.4	79.2
8	77.8	83.3	66.7	83.3	77.8
9	72.2	88.9	77.8	83.3	80.6
10	72.2	72.2	94.4	88.9	81.9
11	55.6	66.7	66.7	88.9	69.4
12	88.9	83.3	72.2	83.3	81.9
13	61.1	72.2	83.3	83.3	75.0
14	88.9	72.2	72.2	83.3	79.2
15	83.3	77.8	66.7	83.3	77.8
Mean ± SD	73.3 (9.7)	77.8 (7.0)	72.2 (7.9)	85.9 (3.6)	77.3 (3.2)
95% CI	68.0-78.7	73.9-81.6	67.9-76.6	84.0-87.9	75.5-79.1

Finally, Table III, IV, and V show the performance of the model for the SEED, SEED-FRA, and SEED-GER datasets, respectively. These datasets resulted in the highest accuracy rates across the five, yielding an overall accuracy of 85.4% for SEED, 90.7% for SEED-FRA, and 87.6% for SEED-GER. As shown in Figures 2c, 2d, and 2e, the distributions among the classes varied between these three datasets. SEED achieved the highest accuracy for the neutral class and the lowest for the negative class, which was consistent with the pattern observed in SEED-V and SEED-V. In contrast, SEED-FRA achieved the highest accuracy rate for the negative class (100%) and the lowest for the positive class (85%). For the SEED-GER, the positive class also resulted in the lowest accuracy (80.8%), whereas the neutral class obtained the highest accuracy (98.3%).

TABLE III: Emotion performance for the 15 subjects in the SEED dataset. The last column displays the average performance across all emotions. The last two rows present the mean and standard deviation for each emotion across subjects, along with the 95% confidence interval.

Subject	Neutral (%)	Positive (%)	Negative (%)	Overall (%)
1	80.0	100.0	93.3	91.1
2	73.3	93.3	93.3	86.7
3	80.0	93.3	66.7	80.0
4	80.0	86.7	93.3	86.7
5	73.3	100.0	80.0	84.4
6	86.7	93.3	93.3	91.1
7	66.7	86.7	86.7	80.0
8	66.7	93.3	86.7	82.2
9	73.3	100.0	100.0	91.1
10	60.0	80.0	86.7	75.6
11	73.3	93.3	86.7	84.4
12	80.0	100.0	80.0	86.7
13	80.0	93.3	93.3	88.9
14	73.3	100.0	93.3	88.9
15	80.0	100.0	86.7	88.9
Mean ± SD	75.1 (6.9)	94.2 (6.1)	88.0 (8.0)	85.8 (4.7)
95% CI	71.3-78.9	90.8-97.6	83.5-92.5	83.2-8.4

TABLE IV: Emotion performance for the 8 subjects in the SEED-FRA dataset. The last column displays the average performance across all emotions. The last two rows present the mean and standard deviation for each emotion across subjects, along with the 95% confidence interval.

Subject	Neutral (%)	Positive (%)	Negative (%)	Overall (%)
1	100.0	76.2	85.7	87.3
2	100.0	71.4	81.0	84.1
3	100.0	95.2	71.4	88.9
4	100.0	90.5	85.7	92.1
5	100.0	85.7	85.7	90.5
6	100.0	100.0	90.5	96.8
7	100.0	85.7	90.5	92.1
8	100.0	90.5	90.5	93.7
Mean ± SD	100.0 (0.0)	86.9 (9.4)	85.1 (6.5)	90.7 (3.9)
95% CI	100.0-100.0	79.0-94.8	79.7-90.5	87.4-94.0

TABLE V: Emotion performance for the 8 subjects in the SEED-GER dataset. The last column displays the average performance across all emotions. The last two rows present the mean and standard deviation for each emotion across subjects, along with the 95% confidence interval.

Subject	Neutral (%)	Positive (%)	Negative (%)	Overall (%)
1	77.8	100.0	83.3	87.0
2	83.3	100.0	88.9	90.7
3	94.4	100.0	83.3	92.6
4	94.4	94.4	77.8	88.9
5	75.0	100.0	75.0	83.3
6	83.3	100.0	83.3	88.9
7	83.3	91.7	66.7	80.6
8	83.3	100.0	83.3	88.9
Mean ± SD	84.4 (7.0)	98.3 (3.3)	80.2 (6.9)	87.6 (3.9)
95% CI	78.6-90.2	95.5-100.0	74.5-92.5	83.2-8.4

B. Comparison with previous studies

Table VI presents a comparison of our model's performance with previous studies. On the SEED-V dataset, our model yielded a higher accuracy than the current state-of-the-art (SOTA) model. For the SEED-IV and SEED datasets, our model produced results that were comparable to those of the SOTA models, falling short by less than 2%. While our model did not outperform the SOTA models, it achieved a lower standard deviation than the SOTA models, indicating its ability to maintain better consistent performance across different subjects.

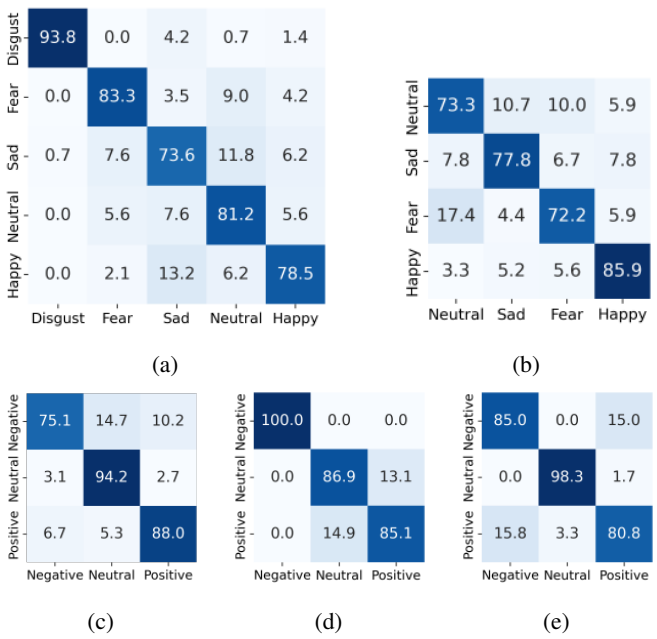


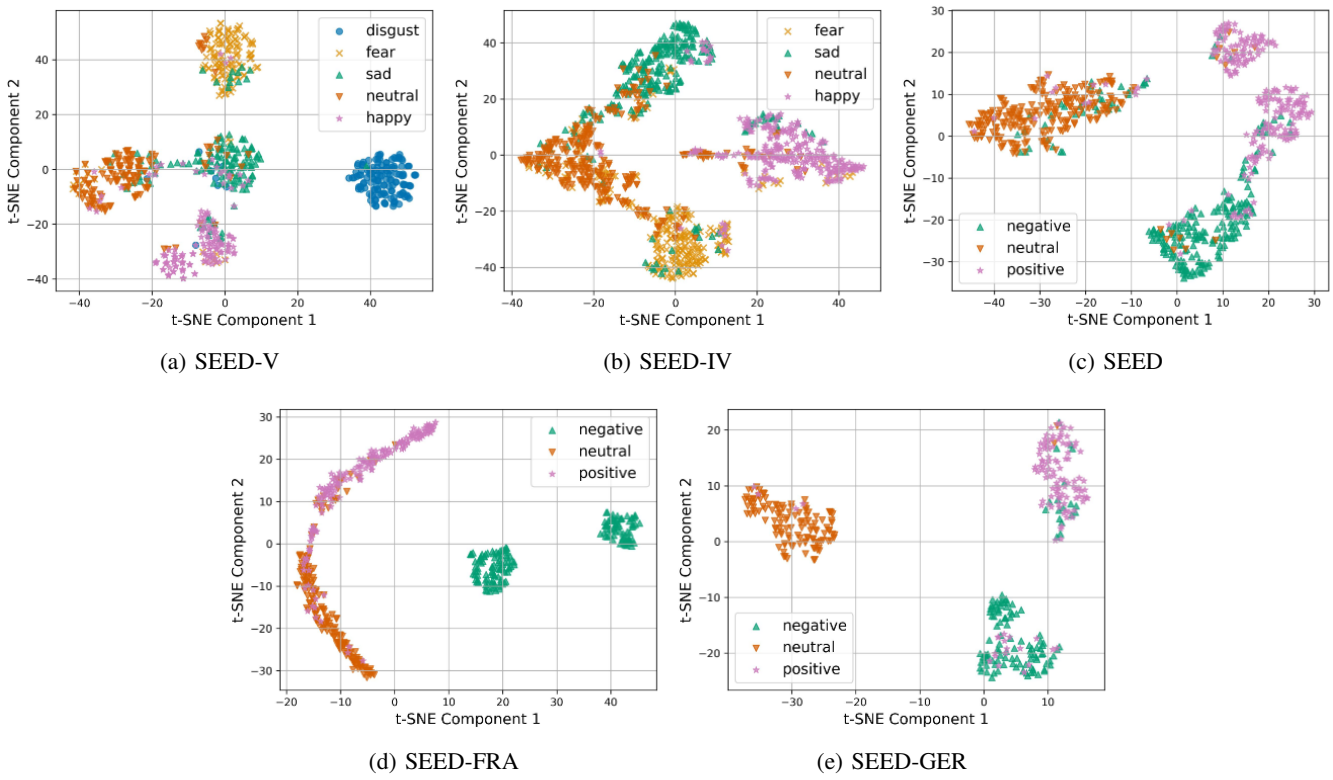
Fig. 2: Average confusion matrices across subjects for: (a) **SEED-V**, (b) **SEED-IV**, (c) **SEED**, (d) **SEED-FRA**, (e) **SEED-GER** datasets.

TABLE VI: Models comparison between previous emotion recognition methods and our approach (last row) on SEED, SEED-IV, SEED-V, SEED-FRA, and SEED-GER datasets. Performance is reported as accuracy mean/standard deviation.

Models	SEED	SEED-IV	SEED-V	SEED-FRA	SEED-GER
SVM [32]	56.7/16.2	37.9/12.5	23.7/8.2	50.1/10.3	55.6/12.1
BiDANN [33]	83.2/9.6	65.6/10.4	–	–	–
BDGLS [34]	–	–	59.6/4.8	–	–
DGCNN [35]	79.9/9.0	52.8/9.2	41.9/6.7	–	–
A-LSTM [36]	72.1/10.8	55.0/9.3	40.3/8.7	–	–
P-GCNN [37]	–	–	64.8/9.8	–	–
IAG [38]	86.3/6.9	–	59.7/9.4	–	–
RGNN [39]	85.3/6.7	73.8/8.0	66.3/16.7	–	–
BiHDM [40]	85.4/7.5	69.0/8.7	–	–	–
GECNN [41]	82.4/–	–	66.8/8.2	–	–
BiHDM w/o DA [42]	81.5/9.7	67.4/8.2	–	–	–
PGCN [43]	–	76.9/7.1	71.4/9.4	–	–
GMSS [42]	86.5/6.2	73.5/7.4	–	–	–
ResNet-18 [44]	–	76.7	78.1	75.0	81.3
DNN [29]	79.4/5.3	70.8/9.2	58.0/8.5	64.2/7.5	65.9/10.0
Graph-LSTM-DANN [45]	79.3/5.8	69.5/9.6	60.7/15.3	–	–
Stacked Graph-LSTM [46]	81.5/7.8	74.6/8.3	78.1/13.7	73.0/5.0	65.6/6.0
Ours	**85.8/4.7**	**77.3/3.2**	**82.1/5.5**	**90.7/3.9**	**87.6/3.9**

C. Feature Visualization with t-SNE

Figure 3 shows the two-dimensional t-SNE representations of the features obtained after processing the input through both the spatial and temporal modules. The extracted features formed distinct clusters for each emotion class across the datasets. However, for some emotion classes, there was an overlap in the clusters. In SEED-V, the sad class exhibited significant overlap, with some samples misclassified as neutral or fear. A similar pattern was observed in SEED-IV, where sad, fear, and neutral showed some degree of misclassification among these emotions. In SEED, the highest cluster overlap occurred between negative and neutral. In contrast, SEED-FRA and SEED-GER exhibited more distinct and well-separated clusters, with minimal overlap among the three emotion classes.



(a) SEED-V (b) SEED-IV (c) SEED

(d) SEED-FRA (e) SEED-GER

Fig. 3: t-SNE visualization of feature embeddings for the (a) **SEED-V**, (b) **SEED-IV**, (c) **SEED**, (d) **SEED-FRA**, and (e) **SEED-GER** datasets.

TABLE VII: Ablation study evaluating the effect of removing different components from the deep learning model shown in Figure 1. The last two columns show the average impact of component removal across the five datasets, along with the rank position among all ablation experiments.

Experiments	SEED-V		SEED-IV		SEED		SEED-FRA		SEED-GER		Average	Ranking
	Acc./Std. (%)	Var. (%)	Acc./Std. (%)	Var. (%)	Acc./Std. (%)	Var. (%)	Acc./Std. (%)	Var. (%)	Acc./Std. (%)	Var. (%)		
BP+DANN	82.1/5.5		77.3/3.2		85.8/4.7		90.7/3.9		87.6/3.9			
w/o BP	44.9/6.7	-45.3	57.5/3.8	-25.6	77.8/6.1	-9.3	79.4/5.6	-12.5	61.0/7.2	-30.4	-24.6	2
w/o spatial BP	44.9/8.3	-45.3	55.6/6.8	-28.1	81.8/6.9	-4.7	81.5/6.3	-10.1	83.0/6.0	-5.3	-18.7	3
w/o spatial module	64.6/5.4	-21.3	56.0/5.4	-27.5	68.1/12.4	-20.6	86.9/5.7	-4.2	88.5/4.0	1.1	-14.5	4
w/o temporal BP	79.2/4.7	-3.6	71.4/3.6	-7.7	82.4/5.9	-4.0	87.1/3.1	-3.9	87.1/3.1	-0.6	-3.9	5
w/o temporal module	60.6/14.9	-26.2	45.6/6.0	-41.1	61.0/7.4	-28.8	64.7/6.0	-28.7	49.8/9.1	-43.2	-33.6	1
w/o DANN	79.6/6.0	-3.0	75.7/5.6	-2.0	83.7/3.0	-2.4	89.5/4.5	-1.3	90.0/3.1	2.8	-1.2	6

D. Feature distribution across cortical areas

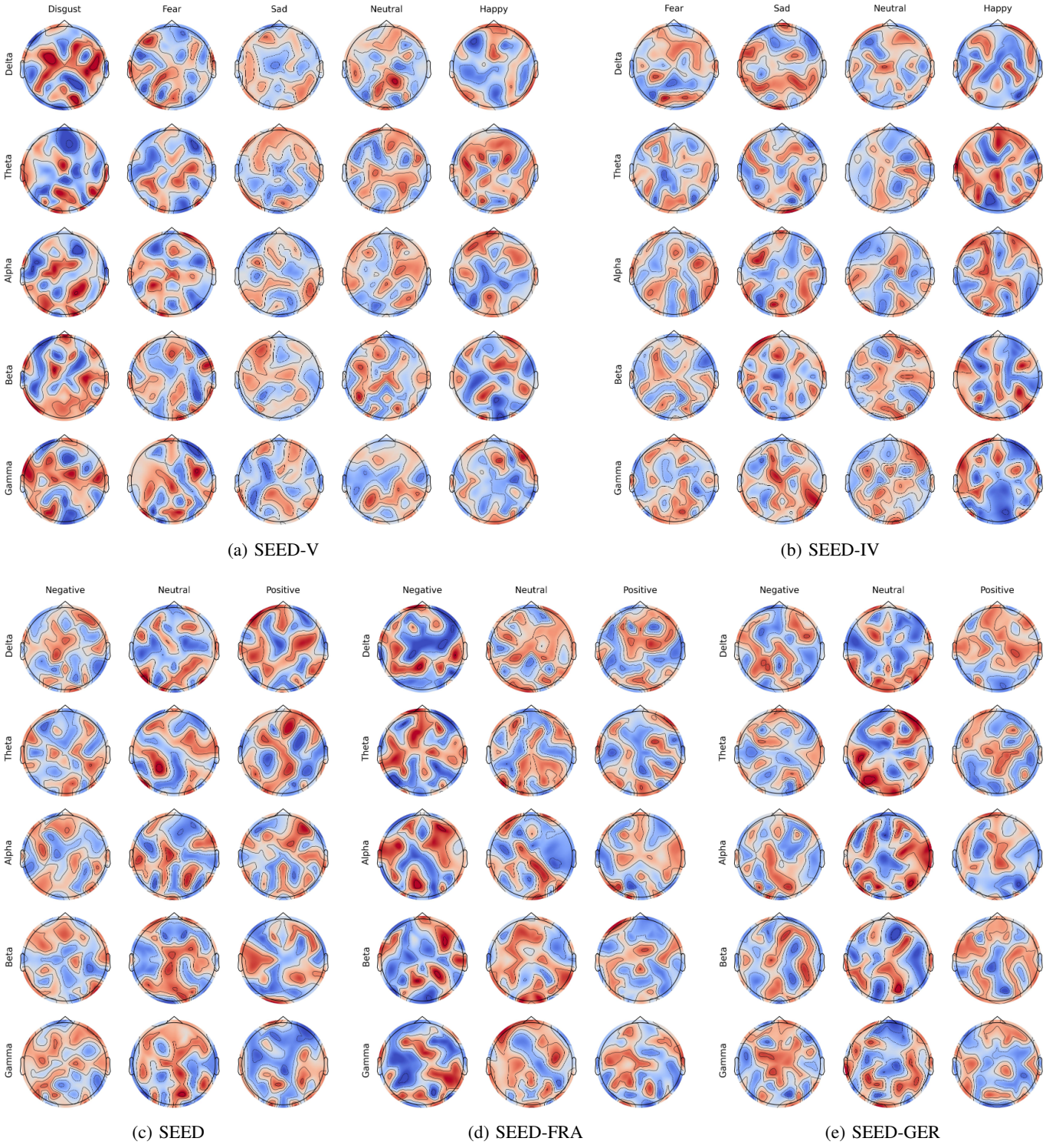
Figure 4 shows the distribution across cortical areas generated by the feature extractor composed of spatial and temporal modules. In general, there was no cortex area that dominated over the other ones. However, lower values (indicated by blue tones) were predominantly observed in the central regions, while higher values (represented by red tones) were more prominent in areas closer to the head circumference.

Regarding emotions, in SEED-V, disgust showed the highest feature values (more intense red), while sad and neutral had the lowest. In SEED-IV, the happy class exhibited the highest values, whereas fear had the lowest. For SEED, both neutral and positive emotions yielded high values, while negative ones resulted in the lowest. Finally, in SEED-FRA and SEED-GER, the highest feature values were observed for negative and neutral, respectively.

E. Most relevant EEG channels

Figure 5 highlights the EEG channels whose features significantly differed among emotions in at least three datasets (p value < 0.05, Friedman test). Channels in cyan indicate less significant p-values, while those in magenta represent the most significant values.

Among the features extracted across the five frequency bands, the most pronounced differences in emotion-related neural activity were observed primarily in the frontal, temporal, and parietal lobes. In the frontal region, significant differences were predominantly detected in electrodes covering the prefrontal (FP_1, FP_2, FP_z) and frontal (F_1, F_2, F_z) areas. For the temporal lobe, lateral EEG channels along the head circumference, including T_7, F_8, TP_7, TP_8, P_7, and P_8, were found to be particularly discriminative for emotion classification. Regarding the parietal lobe, significant differences were consistently observed in both central (C_3, C_2, CP_3, CP_4) and parietal (P_7, P_5, P_3, P_1, P_z, P_2, P_4, P_6, and



(a) SEED-V (b) SEED-IV

(c) SEED (d) SEED-FRA (e) SEED-GER

Fig. 4: Feature distribution across frequency bands for the model on the (a) **SEED-V**, (b) **SEED-IV**, (c) **SEED**, (d) **SEED-FRA**, and (e) **SEED-GER** datasets. Higher feature values, extracted from the corresponding EEG channels, are indicated by more intense red tones, while lower values are indicated in blue.

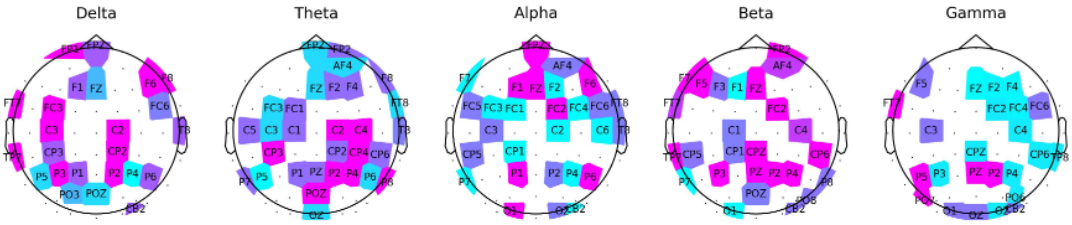


Fig. 5: EEG channels whose features values significantly differed across emotions (p-value < 0.05; Friedman test) in at least three datasets. The colors represent the average p-value for the EEG across the datasets, with cyan indicating lower significance and magenta indicating higher significance. Greater significance suggests a higher potential of the EEG channel to discriminate between emotions.

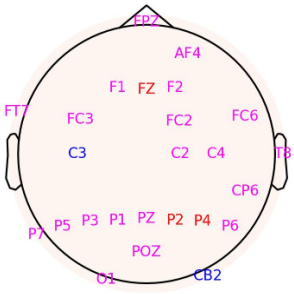


Fig. 6: Common significant EEG channels across three (magenta), four (blue), and five (red) frequency bands. The features extracted from these channels exhibited significant differences, highlighting their potential to discriminate between emotions.

P_8) electrode sites.

To further identify crucial EEG channels, Figure 6 presents the channels that were significant across three, four, or five frequency bands. Channels demonstrating significant differences in at least four frequency bands included F_Z, P_2, P_4, C_3, and CB_2. The parietal region exhibited the most significant differences across emotions and frequency bands, with ten significant channels (P_7, P_5, P_3, P_1, P_z, P_2, P_4, P_6, PO_z, and CP_6). The frontal area was also among the most influential, with eight significant channels (FT_7, FP_Z, AF_4, F_1, F_Z, F_2, FC_6, FC_3, and FC_2).

F. Ablation Study

Table VII shows the results of the ablation experiments. Removing the temporal module (temporal transformer and BP graph) caused the largest performance drop, with an average reduction of 33.6% across the five datasets. Another experiment that led to a significant accuracy decrease was the removal of both BP graphs, which resulted in a 24.5% reduction in emotion prediction accuracy. The spatial components also played a crucial role, with accuracy dropping by 18.7% when the spatial BP graph was removed, and by 14.5% when both the spatial transformer and BP graph were removed. The removal of the temporal BP graph had a minor impact, reducing the average accuracy by 3.9%. The component with the least impact on performance was the removal of the DANN strategy, which only reduced accuracy by 1.2%.

VI. Discussion

A. Effect of BP graphs on emotion prediction

Our results indicate that including BP graphs in DANN models can enhance the extraction of invariant-domain features that support the prediction of emotion under the subject-independence approach. Specifically, our proposed model predicted emotions with a higher accuracy rate than those provided by the SOTA model for the SEED-V, SEED-IV, SEED-FRA, and SEED-GER datasets and achieved comparable results for the SEED dataset. Moreover, our model achieved a lower standard deviation than the previous model, highlighting its capacity to generalize predictions across different subjects and datasets.

The capacity of the model to predict emotions is reflected in tSNE visualization derived from the features extracted via BP graphs. For most emotion classes, the t-SNE plots revealed well-defined clusters, suggesting that the FCN label classifier could effectively establish non-linear boundaries to distinguish between different emotional states. Moreover, the t-SNE representations provide insight into why certain emotions resulted in higher misclassification rates than others (see Figure 2). Specifically, the overlap among negative emotions (e.g., sad and fear) and neutral in the SEED-V, SEED-IV, and SEED datasets suggests a high degree of feature similarity, potentially reducing the model's ability to discriminate between these classes. In contrast, higher classification accuracy was achieved for emotions with more distinct clusters and less overlap, such as disgust in SEED-V, happy in SEED-IV, and positive emotions in SEED. For SEED-FRA and SEED-GER, the clusters exhibited greater inter-class separability, which contributed to improved classification performance and higher accuracy in the corresponding confusion matrices.

B. Impact of cultural background on emotion recognition

Besides the t-SNE feature overlap, cultural background may explain why the prediction of negative emotions resulted in lower accuracy for SEED-V, SEED-IV, and SEED compared to SEED-FRA and SEED-GER. The SEED-V, SEED-IV, and SEED datasets comprise data from Chinese subjects, whereas SEED-FRA and SEED-GER include data from Western European subjects. Prior research suggests that cultural values influence emotion processing, shaping how individuals respond to stimuli [47], [48]. According to previous studies [49], [50],

individuals from interdependent cultures, such as Chinese, tend to suppress negative emotions to maintain a balanced (neutral) and harmonious social environment, whereas individuals from independent cultures, such as Europeans, are more likely to express their emotions openly. This cultural distinction may explain why negative emotions were more misclassified as neutral in SEED-V, SEED-IV, and SEED. Additionally, Liu et al. [29], after analyzing the cross-culture effect on emotion recognition using the SEED, SEED-FRA, and SEED-GER, found that the patterns extracted by deep learning models are more similar between German and French subjects than those from Chinese subjects. These findings suggest that distinct neural patterns, shaped by cultural backgrounds, influence emotion recognition.

In relation to previous studies, to the best of our knowledge, this study is the first to adapt strategies used in the video processing field to deal with domain-shift problems for emotion recognition. In detail, the proposed use of BP graphs at both spatial and temporal levels facilitates the extraction of domain-invariant features by effectively aligning feature distributions between source and target domains. Our findings indicate that BP graphs are viable and effective for addressing domain-shift among individuals. The effectiveness of BP graphs lies in their use of similarity measures to relate source and target samples. These similarities are then used to construct a learnable adjacency matrix that linearly transforms domain-specific features into a shared representation space, thereby enhancing generalization across different subjects.

C. Most relevant EEG channels

The features extracted from the spatial and temporal BP graphs also highlight the EEG channels most relevant for distinguishing between emotional states (see Figures 4, 5, and 6). Specifically, across frequency bands, the EEG channels exhibiting the most significant feature differences were predominantly located in the parietal, frontal, and temporal areas. The consistent differences observed across these channels suggest that they play a critical role in facilitating emotion prediction. Therefore, their inclusion in emotion recognition systems is crucial to enhance emotion recognition under subject-independent settings.

The relevance of EEG channels located in the frontal, temporal, and parietal regions for emotion recognition may be attributed to the underlying physiology of emotional processing. Emotion processing is a complex neural function that involves interactions between the prefrontal cortex (PFC) and limbic system structures, particularly the amygdala and hippocampus [51]. When an individual is exposed to an emotional stimulus, these regions collaborate to associate the stimulus with past experiences, evaluate its significance, and generate an appropriate response. However, EEG cannot directly measure the activity of deep structures such as the PFC, amygdala, or hippocampus. Instead, it captures cortical electrical activity that reflects their interactions with other brain regions. Previous studies have indicated that the amygdala-PFC interactions are observable in frontal and temporal cortex areas, such as the Fp_1, Fp_2, T_7, T_8, FT_7, and FT_8 EEG channels [52]. Likewise, amygdala-hippocampus synchrony is observed more on the central and parietal lobes [53]. Therefore, the involvement of frontal, frontotemporal, temporal, temporoparietal, and parietal regions suggests that the proposed model can capture the subcortical dynamics involved in emotion processing.

Additionally, the involvement of the parietal and central regions, which are associated with visual perception, can be linked to the use of movement-related stimuli. As noted in [54], [55], the central region is particularly significant for emotional responses to movement-based stimuli (e.g., a needle piercing a thumb). Thus, the activation observed in central and parietal areas, such as P_z and PO_z in Figure 6, indicates that these regions are critical during the early stages of visual processing, highlighting their importance in emotional processing.

The relevant identified EEG channels align with previous research that highlights the pivotal role of the frontal lobe in emotional processing [22]–[25]. However, unlike these previous studies, which determined the importance of these areas by analyzing differential entropy features without considering the patterns learned by the deep learning models, our study identifies these regions by directly analyzing the features learned by the model for emotion prediction. Therefore, our approach provides stronger evidence for the relevance of these brain areas in emotion regulation and processing.

Our findings also support our previous research [45], [56], which highlighted EEG channels in the frontal, temporal, and parietal regions as crucial for emotion processing. Since the EEG signals in this study were elicited by audiovisual stimuli, the temporal and parietal regions are likely involved in sensory processing, while the frontal region is more closely linked to the interpretation of the captured sensory information [57]–[59].

D. Ablation study

The ablation study revealed that the average accuracy across the evaluated datasets significantly decreased when BP graphs were excluded. This indicates that both the spatial and temporal modules are crucial for extracting meaningful features. By incorporating BP graphs, the predictive model can effectively align features between the source (training subjects) and the target (testing subjects) domains. Further evidence of the effectiveness of BP graphs in addressing the domain shift problem is seen in the fact that removing the DANN strategy resulted in only a minor decrease in performance. These results suggest that the proposed method relies more on the feature integration performed by BP graphs than on the adversarial training between the label and domain classifiers.

E. Technical and Clinical Implications

By comparing the identified EEG channels with commercial EEG systems to monitor emotions, such as the EMOTIV EPOC X 14-channel wireless headset [60], it is notable that there is an intersection between the channels. From the 14 EEG channels included in the EPOC X system, our approach

matches five channels: P_7, O_1, AF_4, FC_6, and T_8. Additionally, other seven channels were closed located nearby: FT_7 close to T_7, F_1 close to F_3, FC_3 close to FC_5, F_2 close to F_4, FC_2 close to F_4, P_6 close to P_8, and CB_2 close to O_2. Therefore, our study provides evidence of the reliability of these lower-density EEG channel systems for emotion applications.

Using a lower-density EEG channel system enhances their suitability for individuals with neurological diseases or older adults, as caps with fewer channels are more comfortable and easier to use. Thus, identifying a subset of EEG channels capable of reliably predicting emotions across individuals facilitates the development of emotion recognition systems. These systems hold potential for early diagnosis, intervention, and treatment of disorders such as depression, anxiety, and neurodegenerative diseases.

F. Limitations and Future Work

We note that the individuals included in this study were mentally healthy and that we did not test our method with subjects with emotional disorders. Nevertheless, our approach achieved strong prediction performance across three datasets, surpassing or matching SOTA models, demonstrating its ability to address the domain-shift problem inherent in EEG-based emotion recognition. Future research should validate these results in more diverse populations, including individuals with different backgrounds and emotional disorders, to assess the broader applicability of the method.

Future research also should focus on validating the effectiveness of the identified EEG channels for emotion recognition tasks. Demonstrating the effectiveness of using fewer EEG channels will improve the computational efficiency of EEG-based methods and improve user comfort, particularly for older adults or individuals with cognitive impairments.

Finally, given the demonstrated potential of BP graphs in mitigating domain shift, future research should explore their applicability to other EEG-based classification tasks. In particular, BP graphs may serve as a viable transfer learning framework for detecting neurological conditions such as Alzheimer's disease and epilepsy.

VII. Conclusion

This study proposed a deep learning approach that utilizes BP graphs to tackle the challenges of the domain-shift problem in subject-independent emotion recognition. The method demonstrated predictive performance that either matched or exceeded SOTA models, highlighting its effectiveness in managing cross-subject variability. Additionally, the results emphasized the significance of EEG channels in the frontal, temporal, and parietal regions for emotion recognition. These findings emphasize the need to incorporate these regions when designing EEG-based systems for predicting emotions elicited by audiovisual stimuli, thus enhancing the effectiveness of subject-independent approaches.

References

[1] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," *Information Fusion*, vol. 102, p. 102019, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523003354

[2] R. Dahl and A. Harvey, "Sleep in children and adolescents with behavioral and emotional disorders," *Sleep Medicine Clinics*, vol. 2, no. 3, pp. 501–511, 2007, sleep in Children and Adolescents. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1556407X07000513

[3] M. Sidorov and W. Minker, "Emotion recognition and depression diagnosis by acoustic and visual features: A multimodal approach," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 81–86. [Online]. Available: https://doi.org/10.1145/2661806.2661816

[4] M. Uluyagmur-Ozturk, A. R. Arman, S. S. Yilmaz, O. T. P. Findik, H. A. Genc, G. Carkaxhiu-Bulut, M. Y. Yazgan, U. Teker, and Z. Cataltepe, "ADHD and ASD classification based on emotion recognition data," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2016, pp. 810–813.

[5] A. Oussi, K. Hamid, and C. Bouvet, "Managing emotions in panic disorder: A systematic review of studies related to emotional intelligence, alexithymia, emotion regulation, and coping," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 79, p. 101835, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005791623000022

[6] M. Pluta-Olearnik and P. Szulga, "The importance of emotions in consumer purchase decisions — a neuromarketing approach," *Marketing of Scientific and Research Organizations*, vol. 44, pp. 87–104, 06 2022.

[7] J. E. Solanes, L. Gracia, and J. Valls Miro, "Advances in human–machine interaction, artificial intelligence, and robotics," p. 3856, 2024.

[8] A. A. Varghese, J. P. Cherian, and J. J. Kizhakkethottam, "Overview on emotion recognition system," in *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, 2015, pp. 1–5.

[9] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, and P. Marttinen, "EEG based emotion recognition: A tutorial and review," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–57, 2022.

[10] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset shift in machine learning*. Mit Press, 2008.

[11] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[12] T. Xu, W. Dang, J. Wang, and Y. Zhou, "DAGAM: A domain adversarial graph attention model for subject independent EEG-based emotion recognition," 2022.

[13] S. Chen, Y. Wang, X. Lin, X. Sun, W. Li, and W. Ma, "Cross-subject emotion recognition in brain-computer interface based on frequency band attention graph convolutional adversarial neural networks," *Journal of Neuroscience Methods*, vol. 411, p. 110276, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027024002218

[14] C. Shi, C. L. P. Chen, S. Li, and T. Zhang, "Functional connectivity patterns learning for EEG-based emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–15, 2024.

[15] Y. Luo, Z. Huang, Z. Wang, Z. Zhang, and M. Baktashmotlagh, "Adversarial bipartite graph learning for video domain adaptation," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 19–27.

[16] J.-Y. Liu, J.-W. Liu, W.-L. Zheng, and B.-L. Lu, "Transformer-based domain adaptation for multi-modal emotion recognition in response to game animation videos," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 879–884.

[17] Z. Wang, Y. Wang, Y. Tang, Z. Pan, and J. Zhang, "Knowledge distillation based lightweight domain adversarial neural network for electroencephalogram-based emotion recognition," *Biomedical Signal Processing and Control*, vol. 95, p. 106465, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424005238

[18] X. Liu, T. Li, C. Tang, T. Xu, P. Chen, A. Bezerianos, and H. Wang, "Emotion recognition and dynamic functional connectivity analysis based on EEG," *IEEE Access*, vol. 7, pp. 143 293–143 302, 2019.

[19] C. Chen, Z. Li, F. Wan, L. Xu, A. Bezerianos, and H. Wang, "Fusing frequency-domain features and brain connectivity features for cross-subject emotion recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022.

[20] M. Pang, H. Wang, J. Huang, C.-M. Vong, Z. Zeng, and C. Chen, "Multi-scale masked autoencoders for cross-session emotion recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[21] C. Chen, Z. Li, K. I. Kou, J. Du, C. Li, H. Wang, and C.-M. Vong, "Comprehensive multisource learning network for cross-subject multimodal emotion recognition," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[22] X. Yan, W.-L. Zheng, W. Liu, and B.-L. Lu, "Identifying gender differences in multimodal emotion recognition using bimodal deep autoencoder," in *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part IV 24*. Springer, 2017, pp. 533–542.

[23] T.-H. Li, W. Liu, W.-L. Zheng, and B.-L. Lu, "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, 2019, pp. 607–610.

[24] J.-Y. Guo, Q. Cai, J.-P. An, P.-Y. Chen, C. Ma, J.-H. Wan, and Z.-K. Gao, "A transformer based neural network for emotion recognition and visualizations of crucial eeg channels," *Physica A: Statistical Mechanics and its Applications*, vol. 603, p. 127700, 2022.

[25] A. Apicella, P. Arpaia, F. Isgro, G. Mastrati, and N. Moccaldi, "A survey on EEG-based solutions for emotion recognition with a low number of channels," *IEEE Access*, vol. 10, pp. 117 411–117 428, 2022.

[26] W.-L. Zheng and B.-L. Lu, "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.

[27] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, pp. 1–13, 2018.

[28] 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," *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

[29] W. Liu, W.-L. Zheng, Z. Li, S.-Y. Wu, L. Gan, and B.-L. Lu, "Identifying similarities and differences in emotion recognition with EEG and eye movements among Chinese, German, and French people," *Journal of Neural Engineering*, vol. 19, no. 2, p. 026012, 2022.

[30] 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.

[31] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.

[32] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293–300, 1999.

[33] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, "A novel neural network model based on cerebral hemispheric asymmetry for eeg emotion recognition." in *IJCAI*, 2018, pp. 1561–1567.

[34] X.-h. Wang, T. Zhang, X.-m. Xu, L. Chen, X.-f. Xing, and C. P. Chen, "EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 1240–1244.

[35] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.

[36] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12 177–12 191, 2019.

[37] Z. Wang, Y. Tong, and X. Heng, "Phase-locking value based graph convolutional neural networks for emotion recognition," *IEEE Access*, vol. 7, pp. 93 711–93 722, 2019.

[38] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for EEG emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, pp. 2701–2708, Apr. 2020.

[39] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.

[40] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 354–367, 2020.

[41] T. Song, W. Zheng, S. Liu, Y. Zong, Z. Cui, and Y. Li, "Graph-embedded convolutional neural network for image-based EEG emotion recognition," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, pp. 1399–1413, 2021.

[42] Y. Li, J. Chen, F. Li, B. Fu, H. Wu, Y. Ji, Y. Zhou, Y. Niu, G. Shi, and W. Zheng, "GMSS: Graph-based multi-task self-supervised learning for EEG emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2512–2525, 2022.

[43] Y. Zhou, F. Li, Y. Li, Y. Ji, G. Shi, W. Zheng, L. Zhang, Y. Chen, and R. Cheng, "Progressive graph convolution network for EEG emotion recognition," *Neurocomputing*, vol. 544, p. 126262, 2023.

[44] S. Bagherzadeh, M. R. Norouzi, S. B. Hampa, A. Ghasri, P. T. Kouroshi, S. Hosseininasab, M. A. G. Zadeh, and A. M. Nasrabadi, "A subject-independent portable emotion recognition system using synchrosqueezing wavelet transform maps of EEG signals and ResNet-18," *Biomedical Signal Processing and Control*, vol. 90, p. 105875, 2024.

[45] C. E. Valderrama and A. Sheoran, "Identifying relevant eeg channels for subject-independent emotion recognition using attention network layers," *Frontiers in Psychiatry*, vol. 16, p. 1494369, 2025.

[46] A. Sheoran and C. E. Valderrama, "Impact of sex differences on subject-independent EEG-based emotion recognition models," *Computers in Biology and Medicine*, vol. 190, p. 110036, 2025.

[47] P. B. Smith and M. H. Bond, "Cultures and persons: Characterizing national and other types of cultural difference can also aid our understanding and prediction of individual variability," *Frontiers in Psychology*, vol. 10, p. 2689, 2019.

[48] F. Schunk, G. Trommsdorff, and D. König-Teshnizi, "Regulation of positive and negative emotions across cultures: does culture moderate associations between emotion regulation and mental health?" *Cognition and Emotion*, vol. 36, no. 2, pp. 352–363, 2022.

[49] M. de Greck, Z. Shi, G. Wang, X. Zuo, X. Yang, X. Wang, G. Northoff, and S. Han, "Culture modulates brain activity during empathy with anger," *NeuroImage*, vol. 59, no. 3, pp. 2871–2882, 2012.

[50] S. Huwaë and J. Schaafsma, "Cross-cultural differences in emotion suppression in everyday interactions," *International Journal of Psychology*, vol. 53, no. 3, pp. 176–183, 2018.

[51] T. Dalgleish and M. J. Power, "Cognition and emotion: Future directions," *Handbook of cognition and emotion*, pp. 799–805, 1999.

[52] S. Sonkusare, D. Qiong, Y. Zhao, W. Liu, R. Yang, A. Mandali, L. Manssuer, C. Zhang, C. Cao, B. Sun *et al.*, "Frequency dependent emotion differentiation and directional coupling in amygdala, orbitofrontal and medial prefrontal cortex network with intracranial recordings," *Molecular Psychiatry*, vol. 28, no. 4, pp. 1636–1646, 2023.

[53] M. Haaf, N. Polomac, A. Starcevic, M. Lack, S. Kellner, A.-L. Dohrmann, U. Fuger, S. Steinmann, J. Rauh, G. Nolte, and C. Mulert, "Frontal theta oscillations during emotion regulation in people with borderline personality disorder," *BJPsych Open*, vol. 10, no. 2, p. e58, 2024.

[54] C. E. MacKay, A. S. Desroches, and S. D. Smith, "An event-related potential (ERP) examination of the neural responses to emotional and movement-related images," *Cognitive Neuroscience*, vol. 15, no. 1, pp. 1–11, 2024.

[55] S. Y. Dharia, M. Hojjati, S. G. Camorlinga, S. D. Smith, and A. S. Desroches, "Leveraging machine learning and threshold-free cluster enhancement to unravel perception of emotion and implied movement," in *IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2024.

[56] F. I. Mouri, C. E. Valderrama, and S. G. Camorlinga, "Identifying relevant asymmetry features of EEG for emotion processing," *Frontiers in Psychology*, vol. 14, 2023.

[57] G. A. Calvert, P. C. Hansen, S. D. Iversen, and M. J. Brammer, "Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect," *Neuroimage*, vol. 14, no. 2, pp. 427–438, 2001.

[58] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of Neural Science*, 5th ed. New York, NY: McGraw-Hill Education, 2013.

[59] H. Saarimäki, A. Gotsopoulos, I. P. Jääskeläinen, J. Lampinen, P. Vuilleumier, R. Hari, M. Sams, and L. Nummenmaa, "Discrete neural signatures of basic emotions," *Cerebral cortex*, vol. 26, no. 6, pp. 2563–2573, 2016.

[60] Emotiv Inc., "Epoc x - 14 channel wireless EEG headset," 2024, available at https://www.emotiv.com/epoc/.