Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 12, No. 1, 2024, 83-93.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining(JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Enhancing Emotion Classification via EEG Signal Frame Selection

Masoumeh Esmaeili and Kourosh Kiani*

Electrical and Computer Engineering Department, Semnan University, Semnan, Iran.

Article Info

Article History:
Received 20 November 2023
Revised 21 January 2024
Accepted 17 March 2024

DOI:10.22044/jadm.2024.13861.2502

Keywords:
Electroencephalogram (EEG), Emotion Recognition, Common Spatial Pattern (CSP), Transformer Encoder, Segmentation, Genetic Algorithm

**Corresponding author: Kourosh.kiani@semnan.ac.ir (K. Kiani).*

Abstract

The classification of emotions using electroencephalography (EEG) signals is inherently challenging due to the intricate nature of brain activity. Overcoming inconsistencies in EEG signals and establishing a universally applicable sentiment analysis model are essential objectives. This study introduces an innovative approach to cross-subject emotion recognition, employing a genetic algorithm (GA) to eliminate non-informative frames. Then, the optimal frames identified by the GA undergo spatial feature extraction using common spatial patterns (CSP) and the logarithm of variance. Subsequently, these features are input into a Transformer network to capture spatio-temporal features, and the emotion classification is executed using a fully connected (FC) layer with a Softmax activation function. Therefore, the innovations of this paper include using a limited number of channels for emotion classification without sacrificing accuracy, selecting optimal signal segments using the GA, and employing the Transformer network for high-accuracy and high-speed classification. The proposed method undergoes evaluation on two publicly accessible datasets, SEED and SEED-V, across two distinct scenarios. Notably, it attains mean accuracy rates of 99.96% and 99.51% in the cross-subject scenario, and 99.93% and 99.43% in the multi-subject scenario for the SEED and SEED-V datasets, respectively. Noteworthy is the outperformance of the proposed method over the state-of-the-art (SOTA) in both scenarios for both datasets, thus underscoring its superior efficacy. Additionally, comparing the accuracy of individual subjects with previous works in cross subject scenario further confirms the superiority of the proposed method for both datasets.

1. Introduction

The significance of emotion recognition has increased, given its applications in various fields such as brain-computer interfaces (BCI), mental health diagnosis, and fatigue and mental workload detection. Emotion recognition primarily follows two main approaches: the analysis of behavioral signals, including text, speech, and facial expressions, and the analysis of physiological signals derived from the eyes, brain, and heart utilizing tools like electrooculogram (EOG), electroencephalogram (EEG), and electrocardiogram (ECG). Physiological signals, emerging spontaneously and evading easy control by individuals, offer valuable avenues for studying emotion recognition. Particularly, EEG stands out for emotion recognition because it is highly reliable and accurately reflects real brain activity. However, the intrinsic complexity and variability of brain activity pose challenges in identifying emotions from EEG signals. Developing a cross-subject model for sentiment detection based on EEG, which remains robust against various artifacts and is applicable to diverse individuals, poses a significant challenge. Addressing internally generated artifacts, such as eye blinking (primarily occurring below 4 Hz), and externally generated artifacts from mains electricity (above 40 or 50 Hz), numerous studies have employed a band-pass filter within the 4-45 Hz range [1]. Despite these efforts, certain uninformative parts persist in the signal, resisting removal via these preprocessing methods. These parts can adversely affect the accuracy of emotion classification. Furthermore, emotional EEG signals are characterized by extended lengths, not only reducing the precision of emotion detection algorithms but also prolonging their processing time. In a study conducted by Hsu et al. [2], active segment selection is executed through continuous wavelet transform (CWT) and Student's two-sample t-statistics to identify optimal time frames in the time-frequency domain. However, the proposed approach, emphasizing the time domain, involves the initial segmentation of EEG signals, followed by the removal of irrelevant frames using a genetic algorithm (GA). Deep learning networks, extensively employed across diverse domains [3-9], have proven effective in addressing individual variability within EEG signals. Sartipi et al. [10] integrated Transformers (TF) and adversarial discriminative domain adaptation (ADDA) techniques for cross-subject emotion recognition, utilizing ADDA to minimize disparities in EEG data collected from various subjects. Wang et al. [11] utilized a convolutional neural network (CNN) named SACNN for cross-subject emotion recognition. They trained the network using the top 10 channels with the highest accuracy, identified through separate training on EEG data from each of the 62 channels. In their study, Luan et al. [12] introduced Bi-CapsNet, a model tailored for cross-subject EEG emotion recognition, incorporating a Bi-hemispheric Capsule Network. Their methodology includes a long short-term memory (LSTM) layer to capture the asymmetry in emotional expression between the left and right hemispheres of the human brain. While CNNs effectively identify spatial patterns, they encounter challenges in capturing the inherent temporal connections in EEG signals due to their local feature learning mechanism [13]. Conversely, recurrent networks such as LSTM excel in capturing temporal relationships and managing the non-stationary characteristics of signals [14]. Nevertheless, these models encounter difficulties in managing prolonged connections. The attention mechanism, proven effective in EEG classification, especially in noisy environments, empowers the model to prioritize essential elements within the EEG signal, thereby improving its ability to understand extensive connections. Moreover, it has demonstrated enhancements in diverse EEG classification tasks [15] by enabling the model to focus exclusively on the most relevant signal segments. While the Transformer has its advantages, it faces difficulties in effectively capturing relationships between different EEG channels. On the other hand, the common spatial pattern (CSP) [16, 17] proves adept at representing interconnections among EEG channels. Hence, this study introduces a cross-subject approach for classifying emotions based on EEG signals. After signal segmentation and optimal frame selection through GA, the focus shifts to extracting spatio-temporal features using both CSP and a Transformer network (its encoder part) [18]. These extracted spatio-temporal features are then fed into the Softmax activation function for emotion classification. The principal contributions of this article can be summarized as follows:

1- Reduction of channels from 62 to 14 channels without compromising accuracy.

2- Feature selection based on GA

3- Utilizing the Transformer model as a classifier for rapid and highly accurate emotion categorization.

The rest of this paper is structured as follows: Section 2 reviews related work, Section 3 introduces the datasets and the proposed model. The results and analysis are discussed in Section 4. Finally, Section 5 presents the conclusion.

2. Related Work

In this section, we offer a brief overview of recent advancements in EEG-based emotion recognition and related domains, categorizing them from two perspectives:

1-Feature representation: This part provides a concise review of diverse feature representations employed in EEG-based emotion recognition and related fields. An increasing number of research efforts have adopted features such as differential entropy (DE) and power spectral density (PSD) to extract information from emotional EEG signals. For instance, Li et al. [19] conducted a study where subsequent to extracting temporal variation and spatial topological information using single-channel DE and cross-channel functional connectivity, this information was input into a novel convolutional graph attention network for the extraction of higher-level graph structural information. In the model developed by Iyer et al. [20], DE is computed for each of the five frequency bands in EEG signals. Subsequently, they constructed a hybrid model that integrates CNN and LSTM components to achieve precise emotion detection. Kanuboyina et al. [21] introduced the PCA-ANN model, employing fast Fourier transform (FFT) with PSD to generate feature vectors from EEG signals. Following this, principal component analysis (PCA) was applied to reduce the dimensionality of the extracted features, and artificial neural network (ANN) was employed for emotion recognition. In a different approach, Zong et al. [22] presented the FCAN–XGBoos model, which extracts both DE and PSD features from four frequency bands of EEG signals. However, DE, akin to PSD, lacks the capability to capture spatial information across different EEG channels. Furthermore, both methodologies exhibit sensitivity to noise, and considering the inherent noise present in EEG signals, this sensitivity may lead to less resilient feature extraction. Wavelet transform is a frequently employed technique for EEG-based emotion classification in the literature [23]. However, it may not comprehensively capture intricate temporal details, as its ability to represent rapid changes in the signal may be constrained by the chosen wavelet basis and decomposition levels. Additionally, some studies opt for the use of raw EEG signals. For example, Han et al. [24] implemented E2ENNet, an end-to-end network. They segmented signals using a sliding window of 1-second duration, subsequently passing these frames through Conv2D, DepthwiseConv2D, SeparableConv2D, and LSTM layers for spatial and temporal feature extraction.

2- Models: This section offers a concise overview of recent models employed in EEG-based emotion recognition, categorized into two groups: machine learning and deep learning models [25]. Conventional machine learning methodologies include support vector machine (SVM), k-nearest neighbor (K-NN), Linear Discriminant Analysis (LDA), logistic regression (LR), ensemble learning and Naïve Bayes [26, 27]. As exemplified, Chen et al. [28] proposed an approach for emotion classification based on EEG utilizing SVM and Stacked Auto-Encoder (SAE). Their approach focused on feature extraction, specifically emphasizing the calculation of energy means for detail coefficients. The effectiveness of these features was validated, establishing a foundation for practical implementation in human-computer interaction (HCI). In Li et al.'s study [29], the extraction of brain rhythm codes was undertaken, and four classifiers (K-NN, SVM, LDA, and LR) were employed to identify the single optimal channel-specific feature yielding the highest accuracy in emotion recognition. This feature selection process provides valuable insights for creating portable BCI devices that are capable of recognizing emotions. In Kamble et al.'s research [30], they incorporated a technique based on dual decomposition into an ensemble learning framework. Following preprocessing, a set of 31 statistical features was derived from the dual-decomposed EEG signal. These features were subsequently utilized as inputs for five distinct ensemble learning algorithms, namely bagging, random forest, rotation forest, adaptive boosting, and extreme gradient boosting. Notably, among these algorithms, bagging demonstrated superior performance in the context of emotion recognition. Goshvarpour et al. [31] examined the influence of the number of selected features on emotion recognition, employing Naïve Bayes and K-NN. Their investigation revealed that opting for nine top-ranked features led to 6-NN outperforming other K-NN classification methods on the SEED-IV dataset. Conversely, Naïve Bayes demonstrated superior performance when only five features were selected. However, while traditional machine learning methods exhibit commendable accuracy in emotion recognition, they face challenges in extracting deeper features. Moreover, certain complexities in EEG-based emotion recognition remain unaddressed by these conventional approaches. Notably, the variability of EEG signals across individuals presents a formidable challenge in establishing a cross-subject approach. However, recently introduced deep learning methodologies have demonstrated efficacy in overcoming these challenges. Among the prominent techniques for emotion recognition, CNN, graph convolutional network (GCN), deep belief network (DBN), and recurrent networks including recurrent neural network (RNN), gated recurrent units (GRU), LSTM, and BiLSTM stand out. In recent investigations, conventional CNNs have been utilized to capture spatial information among distinct EEG channels [32]. Immanuel et al. [33] proposed a model incorporating PCA for feature dimensionality reduction, where the selected features are input into the DCNNER classifier—a deep convolutional neural network—to discern the presence of stress in an individual. To account for correlations among various EEG channels and address long-term dependencies and contextual information within the signals, Chao et al. [34] introduced a framework known as deep belief-conditional random field (DBN-CRF). Notably, they applied the conditional random field methodology to DBNs for the first time in the context of emotion detection. Additionally, several studies have employed GCNs to extract spatial features [35]. However, networks such as CNN, DBN, and GCN may not inherently address temporal dependencies within EEG signals as effectively as recurrent networks. Jehosheba Margaret et al.'s [36] research aims to assess the efficacy of a novel hybrid BiLSTM model in classifying 16 emotions within a three-dimensional model. The study further compares the performance of this model with other deep learning approaches, including CNN, LSTM, Hybrid Bi-LSTM, and Hybrid Bi-GRU, to identify the most suitable model. FFT is employed to convert time-domain data into the frequency domain in all models. Experimental results revealed that Hybrid GRU achieved the highest accuracy, followed by Hybrid BiLSTM. However, recurrent networks face limitations in extracting spatial features from EEG signals. Additionally, Transformers demonstrate superior capability in capturing long-range dependencies due to their self-attention mechanism. Consequently, in the proposed approach, after employing a GA to determine optimal frames, each frame undergoes CSP application followed by the logarithm of variance to extract spatial features. Subsequently, these spatial features are sequentially input into a Transformer to extract spatio-temporal features.

3. Proposed Method and Datasets

This section offers insights into the utilized datasets, the applied preprocessing methods, the proposed model, details concerning the model's parameters, and information about the hardware employed in the study.

3.1. EEG datasets associated with emotional states

In this study, experiments were conducted using two publicly available emotional EEG datasets, namely SEED and SEED-V, to assess the learning efficacy of the proposed model. The SEED dataset [37, 38], also known as the SJTU Emotion EEG Dataset, gathers emotional EEG data from 15 subjects, including seven males and eight females. Participants were exposed to film clips eliciting three distinct emotional tendencies: negative, positive, and neutral. The SEED-V dataset [39] is a multimodal emotion recognition dataset, incorporating EEG and eye movement (EM) signals. An enhancement of the original SEED dataset, it broadens emotional categories to include Neutrality, Happiness, Sadness, Fear, and Disgust. This dataset comprises 16 subjects, including six males and ten females, who viewed 15 film clips designed to evoke these five emotions. Detailed information about these datasets is outlined in Table 1.

Table 1. The details of datasets employed for the proposed method.

Dataset	Subjects	Channels	Sampling Rate (HZ)	Classes
SEED	15	62	1000	3
SEED-V	16	62	1000	5

3.2. Preprocessing

In the SEED dataset, the signals were recorded starting 5 seconds before the onset of the video clips and continued for 15 seconds after their conclusion. During the preprocessing stage, these signal parts are cropped. As EEG is a non-invasive method that captures brain activity from the scalp, the signals are susceptible to various artifacts. Participant actions such as scalp muscle contractions or blinking can introduce artifacts like EOG, typically characterized by low frequencies below 4 Hz. Additionally, the presence of similar devices near the EEG recording setup can introduce noise to the signals. Fortunately, the frequencies of these artifacts are often above 50 Hz and can be easily mitigated using a low-pass filter. Hence, many studies on EEG-based emotion recognition commonly filter signals between 4 and 45 Hz. Furthermore, recognizing the significance of alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-50 Hz) frequencies in emotions, this paper employs a band-pass filter to exclude frequencies higher than 50 Hz and lower than 8 Hz. For this purpose, a Butterworth filter has been employed in this paper. It allows the specification of the filter type, filter order and the cutoff frequencies. The type of this filter is configured as “band-pass” with an order of 3, where the cutoff frequencies are set to 8Hz and 50Hz. To ensure the isolation of signals from each brain lobe and prevent contamination from other lobes, a common average reference (CAR) filter is implemented. This filter not only enhances the spatial resolution of EEG signals but also improves their signal-to-noise ratio (SNR) by eliminating specific artifacts. Using all EEG channels would increase the computational load due to the expansion of the feature matrix dimensions. Additionally, not all EEG channels equally contribute to emotional brain activity, with certain brain lobes, particularly the temporal and frontal lobes, being more involved during EEG-based emotional experiences [40-42]. Inclusion of irrelevant channels from other lobes may introduce noise, decreasing the performance of emotion recognition systems. Therefore, it is essential to selectively choose active channels. In this study, only the channels of the frontal lobe, including FPZ, AF3, FP1, FP2, AF4, FZ, F1, F3, F5, F7, F2, F4, F6, F8, are chosen. Additionally, before injecting spatial features into the Transformer, normalization is conducted on them. MinMaxScaleris used for this purpose. It is a feature scaling technique commonly used in machine learning, which transforms the data by scaling each feature to a specified range, typically between 0 and 1. The purpose is to ensure that all features contribute equally to the analysis and prevent some features from dominating due to their larger scales.

3.3. The proposed method

In EEG signals, especially those related to emotions, a notable limitation is their high resolution, capturing brain activity at the same speed it occurs. Recorded through numerous electrodes on the scalp, these signals result in a substantial volume. This volume not only lessens the speed of classification process but can also, at times, compromise accuracy. Evolutionary algorithms serve as a technique to address dimensionality reduction in EEG brain signal analysis for channel selection [43] and feature selection [44], demonstrating superior performance compared to other feature selection methods. Thus, this article proposes a method that utilizes the GA to eliminate unnecessary frames. Before applying GA, the signals are initially segmented into 50 equal frames. This segmentation is performed statically by sliding a window with a duration of 0.7 second (Figure 1 part (a)). Subsequently, after applying some preprocessing steps, spatial feature vectors for each frame of a trial are extracted using CSP and log of variance (Figure 1 part (b)). The utilization of the log of variance emerges as the appropriate technique for feature extraction following the implementation of the CSP method. This choice is motivated by the fact that the CSP method transforms the data into a space where high separation between classes is achieved, akin to the PCA method but in a supervised manner. Therefore, extracting the variance feature from data that has achieved high separation is the optimal choice. In the GA, an initial population is established by generating binary chromosomes with randomly assigned values. The number of genes in a chromosome is equal to the number of frames in a trial. Subsequently, based on the genes of each chromosome, the feature vectors corresponding to the frames of a trial are either retained (when the value is 1) or removed (when the value is 0). A separate classifier is considered for each chromosome, and for each chromosome, the final feature vectors related to all trials are divided into training data (80% of the data) and test data (20% of the data). The classifier associated with this chromosome is then trained using its corresponding training data. Subsequently, the fitness value of this chromosome is calculated based on the classification of test data using the following formula:

$$Fitness = \frac{\text{The number of correctly classified data}}{\text{Total number of test data}} \quad (1)$$

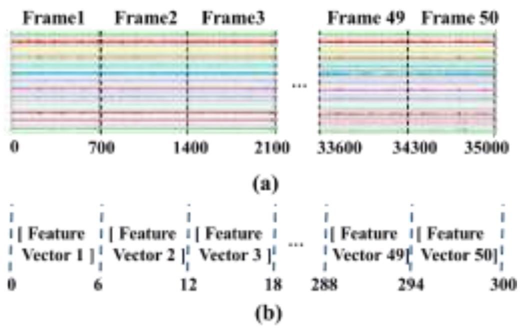

Figure 1. (a) Static segmentation of a trial into 50 equal frames each with a duration of 700 milliseconds, (b) The extraction of spatial features from individual frames.

After executing this operation for all chromosomes and calculating the fitness value for each, crossover and mutation operations are applied to generate a new population. This new population subsequently replaces the previous one. This iterative process is repeated multiple times, and the stopping criterion is either that the fitness values of the chromosomes remain unchanged from one population to another or the number of iterations reaches a predefined limit. The proposed GA model (Figure 2) comprises a population of 68 chromosomes, considering 100 epochs, with crossover and mutation rates set at 99% and 1% respectively.

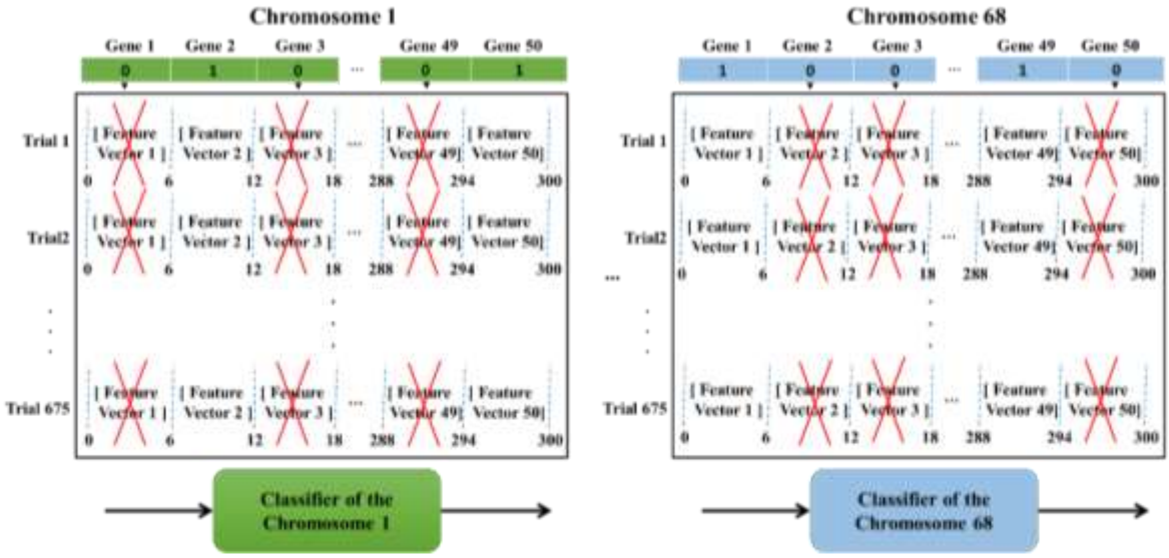

Figure 2. Block diagram of the proposed frame selection method using the GA.

Once optimal frames of emotional EEG signals are determined through GA, their spatial feature vectors are sequentially input into a Transformer network to extract features enriched both temporally and spatially. The number of Transformer layers is set to 2, and the number of heads is set to 16. These spatiotemporal features are then fed into a fully connected (FC) layer, followed by a Softmax activation function for classification. Therefore, this paper integrates traditional methods (CSP and GA) with deep learning models (Transformer). CSP not only enhances signal separability but is also employed for spatial feature extraction through the utilization of the logarithm of variance. The overall architecture of the proposed method is illustrated in Figure 3.

3.4. Model parameters and hardware specifications

A part of the proposed methodology, encompassing preprocessing steps like the CAR filter, frequency filtering, and CSP method, was implemented using MATLAB R2016a 64-bit. The MATLAB software operated on a personal computer running a 64-bit Windows 10 operating system with 4GB of RAM and an Intel(R) Core (TM) i5-2430M CPU. Additionally, GA and Transformer algorithms were executed using Python. The execution took place on Google Colab, which is equipped with 13GB of RAM. The learning rate for this network was configured to 0.00001, and the number of epochs was set to 1000.

4. Results and analysis

The proposed method undergoes evaluation from two perspectives: cross-subject and multi-subject. The subsequent section provides the experimental results, accompanied by a comparison between these results and the current state-of-the-art (SOTA).

4.1. Multi-subject scenario

In a multi-subject scenario, following a common practice in various studies, the data is split into two parts – one assigned for training and the other for testing, regardless of individual subjects. In this study, 80% of the data is allocated for training purposes, while the remaining 20% is reserved for testing. To ensure a robust performance assessment, the proposed method employs k-fold cross-validation, with k configured to 10. This approach allows for a comprehensive assessment of the model's effectiveness. A comparative analysis between the proposed model and SOTA models in EEG-based emotion recognition, performed on the SEED and SEED-V datasets, highlights the significant performance of the proposed method, as illustrated in Table 2.
Remarkably, the proposed model demonstrates exceptional performance on the SEED-V dataset, overcoming challenges that earlier models faced in achieving high accuracy. The key element contributing to this improved accuracy stems from the identification and elimination of irrelevant frames by the GA. In addition, it is evident that the accuracy achieved on the SEED dataset surpasses that on the SEED-V dataset. This difference can be attributed to the smaller number of classes and individuals in the SEED dataset. As a result, the proposed model demonstrates a notably quicker convergence time on the SEED dataset in comparison to the SEED-V dataset.

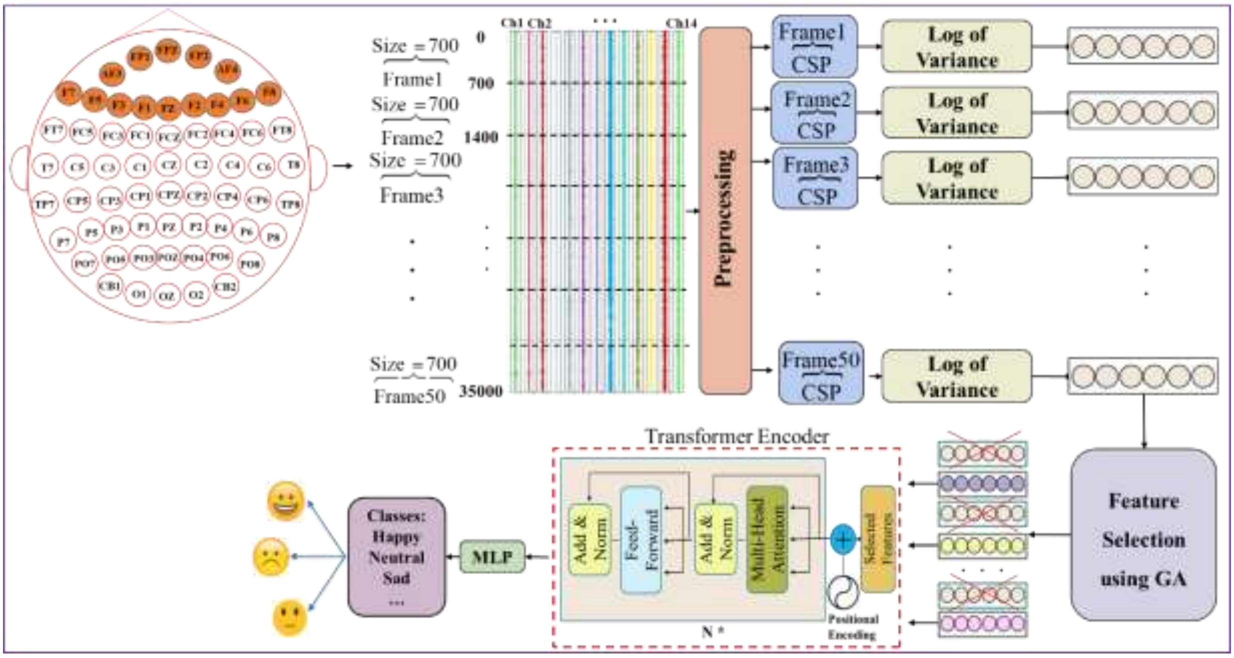

Figure 3. The overall architecture of the proposed model: After identifying the optimal frames using GA, the spatial features extracted from these frames are sequentially input into a Transformer and then into an MLP for emotion classification.

Table 2. Comparing the mean accuracy (%) of the proposed method in 10-fold cross-validation with SOTA emotion recognition models on SEED and SEED-V datasets (multi-subject scenario)

Ref.	Models	SEED	SEED-V
Cizmeci et al. [45]	Capsule Network	98.21	-
Alotaibi et al. [46]	DNN	96.95	-
Gong et al. [47]	Transformer & CNN	98.47	-
Zhang et al. [48]	DWT and CNN	-	82.06
Jadhav et al. [49]	CNN	97.09	88.23
Ours	**Transformer**	**99.93**	**99.43**

4.2. Cross subject scenario

Extending emotion recognition technology to real-life scenarios faces challenges due to individual variability. An ideal model should be applicable to various subjects without sacrificing accuracy. To tackle this challenge, Transfer Learning has gained considerable attention and practical application in research [50]. Therefore, in the cross-subject scenario of the proposed method, the Transformer network is initialized with weights from the multi-subject model. Subsequently, a cross-validation task utilizing the Leave-One-Out approach is executed, wherein the proposed method is trained using data from all individuals except one. The data from the excluded subject is then employed as the test data. This process is iteratively repeated for all individuals, and to ensure result reliability, 10 experiments were conducted per individual. Mean classification accuracies were computed from these experiments, offering insights into the performance of the proposed model for each subject. The cumulative accuracy across all subjects and experiments contributes to the overall accuracy in the cross-subject scenario. Applying Transfer Learning in the cross-subject scenario facilitated faster convergence of the network and raised its accuracy. This procedure was conducted on two datasets, SEED and SEED-V, with Table 3 presenting the average accuracy for each subject. Average accuracies range from 99.56% to 100% in the SEED dataset and from 99.11% to 100% in the SEED-V dataset. Notably, experimental results highlight comparatively lower performance for the second participant in the SEED dataset and the twelfth and sixteenth participants in the SEED-V dataset compared to other participants. To highlight the excellent performance of the proposed method, the accuracy of each individual is compared with that of the SOTA models on both datasets. As evident from Figure 4 parts (a) and (b), the proposed method outperformed all other models for every participant across both datasets. This emphasizes the effectiveness and superior performance of the proposed model.

Table 3. The performance of the proposed method across various individuals in the SEED and SEED-V datasets using 10 experiments (cross-subject scenario).

Subject	SEED	SEED-V	Subject	SEED	SEED-V
SBJ 1	100	99.78	**SBJ 9**	100	99.33
SBJ 2	99.56	99.33	**SBJ 10**	100	99.33
SBJ 3	99.98	99.78	**SBJ 11**	100	99.55
SBJ 4	100	99.78	**SBJ 12**	100	99.11
SBJ 5	99.98	99.55	**SBJ 13**	100	100
SBJ 6	100	99.33	**SBJ 14**	100	99.33
SBJ 7	100	99.78	**SBJ 15**	99.99	99.55
SBJ 8	100	99.55	**SBJ 16**	-	99.11
Mean				***99.96***	***99.51***

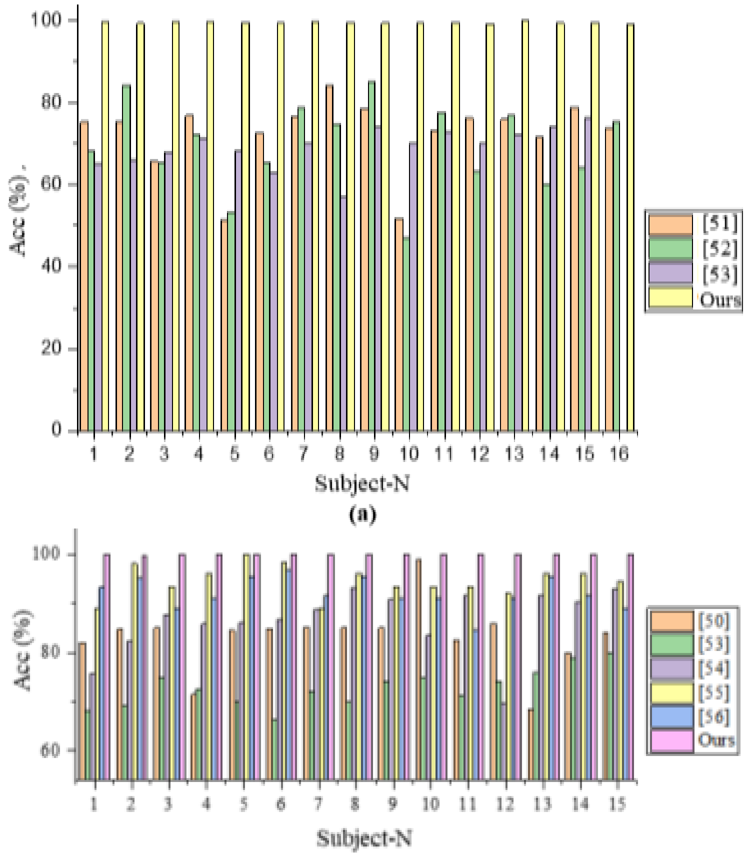

Figure 4. Comparative analysis of individual accuracy in the proposed method versus SOTA models under cross-subject scenario (a) SEED-V Dataset, (b) SEED Dataset.

Across 10 experiments, the proposed method attained an average accuracy of 99.96% on the SEED dataset and 99.51% on the SEED-V dataset. Table 4 compares the overall mean accuracy across all subjects in the cross-subject scenario with SOTA models. As evident, the proposed method has significantly outperformed SOTA models. The effectiveness of this method can be attributed to various preprocessing techniques and the integration of both traditional and modern methodologies. These include utilizing CSP for improved signal separability, employing a Transformer for efficient extraction of temporal features in extended emotional signals, and notably, implementing a GA to identify the most informative segments of EEG signals.

Table 4. Comparative analysis of the proposed model's average accuracy (%) in 10-fold cross-validation with SOTA models in emotion recognition on SEED and SEED-V datasets (cross-subject scenario).

Ref.	Models	Dataset1: SEED	Dataset2: SEED-V
Haq et al [57]	KNN, SVM and Tree Ensemble	96.7	-
Iyer et al [20]	CNN and LSTM	97.16	-
Dharia et al [51]	MDLM	-	72.3
Zhu et al [50]	MLP	83.21	60.17
Zhang et al [58]	CNN	96.36	-
Li et al [59]	Transformer	90.37	-
Esmaeili et al [60]	BiLSTM	99.93	-
Jin et al [61]	GCN	-	61.78
Ours	**Transformer**	**99.96**	**99.51**

5. Conclusion

This study presents an effective model for EEG-based emotion recognition, demonstrating significant improvements over current methodologies. The proposed method optimizes signal separation using CSP and subsequently employs the logarithm of variance for extracting spatial features. This sequential application of CSP and logarithm of variance has demonstrated efficacy in capturing spatial information. Because calculating the variance feature for data that has achieved high separation is a worthy choice. Recognizing that not all signal segments carry relevant emotional information, a GA was employed to eliminate less informative segments in post-segmentation. Following this, realizing the significance of temporal details in EEG signals, a Transformer was incorporated to extract temporal features from the remaining frames. Utilizing its attention mechanism, this network excels in extracting features from the extended sequences inherent in emotional EEG signals, surpassing recurrent networks in performance. The proposed method underwent evaluation in both cross-subject and multi-subject scenarios on the SEED and SEED-V datasets. Impressive accuracies of 99.96% and 99.51% were achieved for the cross-subject scenario, and 99.93% and 99.43% for the multi-subject scenario, on SEED and SEED-V datasets, respectively, surpassing current SOTA models for both scenarios and datasets. Furthermore, a comparative analysis of the accuracy of individual subjects with previous studies in a cross-subject scenario reaffirms the superiority of the proposed method across both datasets. This enhanced performance arises from the deliberate fusion of CSP and Transformer for effective spatiotemporal feature extraction, along with the integration of GA for identifying informative signal segments. Additionally, the utilization of Transfer Learning in the cross-subject scenario significantly improved the convergence speed of the Transformer network and enhanced its accuracy.

References

[1] S. Liu, Z. Wang., Y. An, J. Zhao, Y. Zhao, & Y. D. Zhang. “EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network”. *Knowledge-Based Systems*, vol. 265, pp. 110372, 2023.

[2] W. Y. Hsu, C. C. Lin, M. S. Ju, & Y. N. Sun, “Wavelet-based fractal features with active segment selection: Application to single-trial EEG data”, *Journal of neuroscience methods*, vol. 163, no. 1, pp. 145-160, 2007.

[3] R. Rastgoo, & K. Kiani, “Face recognition using fine-tuning of Deep Convolutional Neural Network and transfer learning”, *Journal of Modeling in Engineering*, vol. 17, no. 58, pp. 103-111, 2019.

[4] K. Kiani, R. Hematpour, & R. Rastgoo, “Automatic grayscale image colorization using a deep hybrid model”, *Journal of AI and Data Mining*, vol. 9, no. 3, pp. 321-328, 2021.

[5] N. Majidi, K. Kiani, & R. Rastgoo, “A deep model for super-resolution enhancement from a single image”, *Journal of AI and Data Mining*, vol. 8, no. 4, pp. 451-460, 2020.

[6] F. Alinezhad, K. Kiani, & R. Rastgoo, “A Deep Learning-based Model for Gender Recognition in Mobile Devices”, *Journal of AI and Data Mining*, pp. 229-236, 2023.

[7] S. Zarbafi, K. Kiani, & R. Rastgoo, “Spoken Persian digits recognition using deep learning”, *Journal of Modeling in Engineering*, vol. 21, no. 74, pp. 163-172, 2023.

[8] A. Fakhari & K. Kiani, “A new restricted boltzmann machine training algorithm for image restoration”, *Multimedia Tools and Applications*, vol. 80(2), pp. 2047-2062, 2021.

[9] A. Fakhari& K. Kiani, “An image restoration architecture using abstract features and generative models”, *Journal of AI and Data Mining*, vol. 9(1), pp. 129-139, 2021.

[10] S. Sartipi, M. Cetin, “Adversarial Discriminative Domain Adaptation and Transformers for EEG-based Cross-Subject Emotion Recognition”, *In 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE*, April 2023, pp. 1-4.

[11] Z. Wang, M. Chen, & G. Feng, “Study on Driver Cross-Subject Emotion Recognition Based on Raw Multi-Channels EEG Data”, *Electronics*, vol. 12(11), pp. 2359, 2023.

[12] X. Luan, G. Zhang, & K. Yang, “A Bi-hemisphere Capsule Network Model for Cross-Subject EEG Emotion Recognition”, *In International Conference on Neural Information Processing, Singapore: Springer Nature Singapore*, November 2022, pp. 325-336.

[13] Y. Wei, Y. Liu, C. Li, J. Cheng, R. Song, & X. Chen, “TC-Net: A Transformer Capsule Network for EEG-based emotion recognition”, *Computers in Biology and Medicine*, vol. 152, pp. 106463, 2023.

[14] M. Azarbad, H. Azami, S. Sanei, & A. Ebrahimzadeh, “A time-frequency approach for EEG signal segmentation”, *Journal of AI and Data Mining*, vol. 2, no. 1, pp. 63-71, 2014.

[15] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, Y. Zhan, "A transformer-based approach combining deep learning network and spatial–temporal information for raw EEG classification", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2126-2136, 2022.

[16] T. Yan, T. Jingtian, & G. Andong, “Multi-class EEG classification for brain computer interface based on CSP”, *In 2008 International Conference on Bio Medical Engineering and Informatics, IEEE*, vol. 2, May 2008, pp. 469-472.

[17] M. Esmaeili, M. Zahedi, “Static Partitioning of EEG Signals by GA Using Multi_CSP”, *Elixir Comp Engg 82*, pp. 32134-32138, 2015.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... & I. Polosukhin, “Attention is all you need”. *Advances in neural information processing systems*, 30, 2017.

[19] Z. Li, G. Zhang, L. Wang, J. Wei, &J. Dang, “Emotion recognition using spatial-temporal EEG features through convolutional graph attention network”, *Journal of Neural Engineering*, vol. 20, no. 1, pp. 016046, 2023.

[20] A. Iyer, S. S. Das, R. Teotia, S. Maheshwari, &R. R. Sharma, “CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings”, *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 4883-4896, 2023.

[21] S. N. V. Kanuboyina, T. Shankar, & R. R. Venkata Penmetsa, “Electroencephalograph based Human Emotion Recognition Using Artificial Neural Network and Principal Component Analysis”, *IETE Journal of Research*, vol. 69, no. 3, pp. 1200-1209, 2023.

[22] J. Zong, X. Xiong, J. Zhou, Y. Ji, D. Zhou, & Q. Zhang, “FCAN–XGBoost: A Novel Hybrid Model for EEG Emotion Recognition”, *Sensors*, vol. 23, no. 12, pp. 5680, 2023.

[23] O. Almanza-Conejo, D. L. Almanza-Ojeda, J. L. Contreras-Hernandez, & M. A. Ibarra-Manzano, “Emotion recognition in EEG signals using the continuous wavelet transform and CNNs”, *Neural Computing and Applications*, vol. 35, no. 2, pp. 1409-1422, 2023.

[24] Z. Han, H. Chang, X. Zhou, J. Wang, I. Wang, & Y. Shao, “E2ENNet: An end-to-end neural network for emotional brain-computer interface”, *Frontiers in Computational Neuroscience*, vol. 16, pp. 942979, 2022.

[25] S. Mohammadzadeh Koumleh, H. Hassanpour, M. Esmaeili, A. Gholami, “Various Deep Learning Techniques for the Applications in Polymer, Polymer Composite Chemistry, Structures and Processing”. *J. Chem. Lett*., vol. 2, no. 4, pp. 157-177, 2021.

[26] K. Kamble, & J. Sengupta, “A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals”, *Multimedia Tools and Applications*, pp. 1-36, 2023.

[27] M. Esmaeili, A. Arjomandzadeh, R. Shams & M. Zahedi, “An anti-spam system using naive Bayes method and feature selection methods”, *International Journal of Computer Applications*, vol. 165, no. 4, pp. 1-5, 2017.

[28] R. Chen, Z. Sun, X. Diao, H. Wang, J. Wang, T. Li, Y. Wang, “Happy or sad? Recognizing emotions with wavelet coefficient energy mean of EEG signals”, *Technol. Health Care*, vol. 30, pp. 937–949, 2022.

[29] J. W. Li, D. Lin, Y. Che, J. J. Lv, R. J. Chen, L. J. Wang, ... & X. Lu, “An innovative EEG-based emotion recognition using a single channel-specific feature from the brain rhythm code method”, *Frontiers in neuroscience*, vol. 17, pp. 1221512, 2023.

[30] K. S. Kamble, J. Sengupta, “Ensemble machine learning-based affective computing for emotion recognition using dual-decomposed EEG signals”, *IEEE Sensors J* 22, pp. 2496–2507, 2021.

[31] A. Goshvarpour, A. Goshvarpour, “Novel high-dimensional phase space features for EEG emotion recognition”, *Signal, Image and Video Processing*, vol. 17, no. 2, pp. 417-425, 2023.

[32] J. Li, Z. Zhang, H. He, “Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition”, *Cogn. Comput*. vol. 10, pp.368–380, 2018.

[33] R. R. Immanuel, S. K. B. Sangeetha, “ANALYSIS OF DIFFERENT EMOTIONS WITH BIO-SIGNALS (EEG) USING DEEP CNN”, *Journal of Data Acquisition and Processing*, vol. 38, no. 3, pp. 743, 2023.

[34] H. Chao, Y. Liu, “Emotion recognition from multi-channel EEG signals by exploiting the deep belief-conditional random field framework”, *IEEE Access*, vol. 8, pp. 33002-33012, 2020

[35] X. Qiu, S. Wang, R. Wang, Y. Zhang, &L. Huang, “A multi-head residual connection GCN for EEG emotion recognition”, *Computers in Biology and Medicine*, pp. 107126, 2023.

[36] M. Jehosheba Margaret, N. M. Masoodhu Banu, “Performance analysis of EEG based emotion recognition using deep learning models", *Brain-Computer Interfaces*, pp. 1-20, 2023.

[37] W. L. Zheng, B. L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks”, *IEEE Trans. Auton. Ment. Dev*. vol. 7, no. 3, pp. 162-175, 2015.

[38] R. N. Duan, J. Y. Zhu, B. L. Lu, “Differential entropy feature for EEG-based emotion classification”, *In 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE*, November 2013, pp 81-84.

[39] W. Liu, J. L. Qiu, W. L. Zheng, B. L. Lu,“ Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition”, *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp.715-729, 2021.

[40] B. Fu, F. Li, Y. Niu, H. Wu, Y. Li, G. Shi, “Conditional generative adversarial network for EEG-

based emotion fine-grained estimation and visualization", *Journal of Visual Communication and Image Representation*, vol. 74, pp. 102982, 2021.

[41] W. L. Zheng, H. T. Guo, & B. L. Lu, "Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network", *In 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE*, April 2015, pp. 154-157.

[42] F. Baradaran, A. Farzan, S. Danishvar & S. Sheykhivand, "Customized 2D CNN Model for the Automatic Emotion Recognition Based on EEG Signals", *Electronics*, vol. 12, no. 10, pp. 2232, 2023.

[43] L. A. Moctezuma, M. Molinas, "EEG Channel-Selection Method for Epileptic-Seizure Classification Based on Multi-Objective Optimization", *Front. Neurosci.* vol. 14, pp. 593, 2020.

[44] M. Esmaeili, M. Zahedi, & N. Hafezi-Motlagh, "Performance Analysis of PSO and GA Algorithms in Order to Classifying EEG Data", *Elixir*, vol. 82, 2015.

[45] H. Cizmeci, C. Ozcan, "Enhanced deep capsule network for EEG-based emotion recognition", *Signal, Image and Video Processing*, vol. 17, no. 2, pp. 463-469, 2023.

[46] F. M. Alotaibi, "An AI-Inspired Spatio-Temporal Neural Network for EEG-Based Emotional Status", *Sensors*, vol. 23, no. 1, pp.498, 2023.

[47] L. Gong, M. Li, T. Zhang, & W. Chen, "EEG emotion recognition using attention-based convolutional transformer neural network", *Biomedical Signal Processing and Control*, vol. 84, pp. 104835, 2023.

[48] Y. Zhang, Y. Peng, J. Li, & W. Kong, "SIFIAE: An adaptive emotion recognition model with EEG feature-label inconsistency consideration", *Journal of Neuroscience Methods*, vol. 395, pp.109909, 2023.

[49] V. Jadhav, N. Tiwari, & M. Chawla, "EEG-based Emotion Recognition using Transfer Learning Based Feature Extraction and Convolutional Neural Network", *In ITM Web of Conferences. EDP Sciences*, vol. 53,2023.

[50] L. Zhu, F. Yu, A. Huang, N. Ying, & J. Zhang, "Instance-representation transfer method based on joint distribution and deep adaptation for EEG emotion recognition", *Medical & Biological Engineering & Computing*, pp.1-15, 2023.

[51] S. Y. Dharia, C. E. Valderrama, & S. G. Camorlinga, "Multimodal Deep Learning Model for Subject-Independent EEG-based Emotion Recognition", *In 2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE,* September 2023, pp. 105-110.

[52] T. H. Li, W. Liu, W. L. Zheng, & 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),* March 2019, pp. 607-610.

[53] H. Liu, H. Guo, & W. Hu, "EEG-based emotion classification using joint adaptation networks", *In 2021 IEEE international symposium on circuits and systems (ISCAS)*, May 2021, pp. 1-5.

[54] X. Shen, X. Liu, X. Hu, D. Zhang, & S. Song, "Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition", *IEEE Transactions on Affective Computing*, 2022.

[55] W. Guo, & Y. Wang, "Convolutional gated recurrent unit-driven multidimensional dynamic graph neural network for subject-independent emotion recognition", *Expert Systems with Applications*, vol. 238, pp. 121889, 2024.

[56] W. Guo, Y. Li, M. Liu, R. Ma, & Y. Wang, "Functional connectivity-enhanced feature-grouped attention network for cross-subject EEG emotion recognition", *Knowledge-Based Systems*, vol. 283, pp. 111199, 2024.

[57] Q. M. U. Haq, L. Yao, W. Rahmaniar, F. Islam, "A hybrid hand-crafted and deep neural spatio-temporal EEG features clustering framework for precise emotional status recognition", *Sensors*, vol. 22, no. 14, pp. 5158, 2022.

[58] X. Zhang, D. Huang, H. Li, Y. Zhang, Y. Xia, & J. Liu, "Self-training maximum classifier discrepancy for EEG emotion recognition", *CAAI Transactions on Intelligence Technology*, 2023.

[59] J. Li, W. Pan, H. Huang, J. Pan, & F. Wang, "STGATE: Spatial-temporal graph attention network with a transformer encoder for EEG-based emotion recognition", *Frontiers in Human Neuroscience*, vol. 17, pp. 1169949, 2023.

[60] M. Esmaeili, K. Kiani, "Generating personalized facial emotions using emotional EEG signals and conditional generative adversarial networks", *Multimedia Tools and Applications*, pp. 1-26, 2023.

[61] M. Jin, C. Du, H. He, T. Cai, J. Li. "PGCN: Pyramidal graph convolutional network for EEG emotion recognition." *IEEE Transactions on Multimedia*, 2024.

بهبود دسته‌بندی احساسات از طریق انتخاب فریم سیگنال EEG

معصومه اسماعیلی و کوروش کیانی*

دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران.

ارسال ۲۰۲۳/۱۱/۲۰ بازنگری ۲۰۲۴/۰۱/۲۱؛ پذیرش ۲۰۲۴/۰۳/۱۷

چکیده:

دسته‌بندی احساسات با استفاده از سیگنال‌های الکتروانسفالوگرافی (EEG) به دلیل ماهیت پیچیده فعالیت مغزی، چالش‌های ذاتی دارد. غلبه بر ناهمسانی‌ها در سیگنال‌های EEG و ایجاد یک مدل تحلیل احساسات قابل استفاده عمومی، اهداف ضروری هستند. این مطالعه یک رویکرد نوآورانه برای تشخیص احساسات بین افراد را با استفاده از الگوریتم ژنتیک (GA) برای حذف فریم‌های غیراطلاعاتی معرفی می‌کند. سپس، فریم‌های بهینه توسط GA شناسایی شده، با استفاده از الگوهای فضایی مشترک (CSP) و لگاریتم انحراف معیار، استخراج ویژگی فضایی را تجربه می‌کنند. در ادامه، این ویژگی‌ها به شبکه ترنسفورمر داده می‌شوند تا ویژگی‌های فضایی-زمانی را به دست آورده و دسته‌بندی احساسات با استفاده از یک لایه کاملاً متصل (FC) با تابع فعال‌سازی Softmax انجام شود. بنابراین، نوآوری‌های این مقاله شامل استفاده از تعداد محدودی از کانال‌ها برای دسته‌بندی احساسات بدون کاهش دقت، انتخاب بخش‌های سیگنال بهینه با استفاده از GA و استفاده از شبکه ترنسفورمر برای دسته‌بندی با دقت و سرعت بالا می‌باشد. روش پیشنهادی بر دو مجموعه داده عمومی، SEED و SEED-V، در دو حالت مختلف ارزیابی می‌شود. به طور قابل توجه، میزان دقت میانگین ۹۹,۹۶٪ و ۹۹,۵۱٪ در حالت بین افراد و ۹۹,۹۳٪ و ۹۹,۴۳٪ در حالت چند افراد برای مجموعه داده‌های SEED و SEED-V به دست می‌آید. ملاحظه شود که روش پیشنهادی در هر دو حالت و برای هر دو مجموعه داده، عملکرد روش‌های موجود را بهبود می¬بخشد.علاوه بر این، مقایسه دقت افراد مختلف با کارهای قبلی در حالت بین افراد، بیشتر از همین موضوع برتری روش پیشنهادی برای هر دو مجموعه داده را تأیید می‌کند.

کلمات کلیدی: الکتروانسفالوگرام (EEG)، تشخیص احساسات، الگوی فضایی مشترک (CSP)، ترنسفورمر انکودر، تقسیم‌بندی، الگوریتم ژنتیک.