Heliyon 10 (2024) e36411

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Multimodal insights into granger causality connectivity: Integrating physiological signals and gated eye-tracking data for emotion recognition using convolutional neural network

Javid Farhadi Sedehi, Nader Jafarnia Dabanloo [*], Keivan Maghooli, Ali Sheikhani

Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

ARTICLE INFO

Keywords:
Convolutional neural network (CNN)
Effective connectivity
EEG-ECG
Eye-tracking data
Emotion recognition

ABSTRACT

This study introduces a groundbreaking method to enhance the accuracy and reliability of emotion recognition systems by combining electrocardiogram (ECG) with electroencephalogram (EEG) data, using an eye-tracking gated strategy. Initially, we propose a technique to filter out irrelevant portions of emotional data by employing pupil diameter metrics from eye-tracking data. Subsequently, we introduce an innovative approach for estimating effective connectivity to capture the dynamic interaction between the brain and the heart during emotional states of happiness and sadness. Granger causality (GC) is estimated and utilized to optimize input for a highly effective pre-trained convolutional neural network (CNN), specifically ResNet-18. To assess this methodology, we employed EEG and ECG data from the publicly available MAHNOB-HCI database, using a 5-fold cross-validation approach. Our method achieved an impressive average accuracy and area under the curve (AUC) of 91.00 % and 0.97, respectively, for GC-EEG-ECG images processed with ResNet-18. Comparative analysis with state-of-the-art studies clearly shows that augmenting ECG with EEG and refining data with an eye-tracking strategy significantly enhances emotion recognition performance across various emotions.

1. Introduction

Emotions are the vibrant colors that paint the canvas of our daily lives, influencing our thoughts, decisions, and social interactions in profound ways. They not only shape our perception of the world but are also fundamental to our mental and physical well-being. The ability to recognize and understand emotions is essential across various fields, including psychology, marketing, and the rapidly evolving realm of artificial intelligence [1–3]. Thanks to advances in emotion recognition technologies, we are now enhancing customer service, developing more intuitive AI systems, and refining mental health diagnostics.

Researchers tap into the human psyche using sophisticated techniques such as electrocardiography (ECG) [4–6] and electroencephalography (EEG) [7–11]. While EEG tracks the brain's electrical activity to identify patterns associated with different emotions, ECG monitors heart rate variability to gauge stress or relaxation levels. These tools not only deepen our understanding of emotions' physiological roots but also aid in creating responses that resonate on a human level. Further exploring the brain's complexity, connectivity analysis offers a window into how different brain regions interact, revealing insights into various mental states. This analysis is categorized into structural, effective, and functional connectivity [12]. Structural connectivity examines the anatomical

* Corresponding author.
E-mail address: jafarnia@srbiau.ac.ir (N. Jafarnia Dabanloo).

https://doi.org/10.1016/j.heliyon.2024.e36411
Received 28 May 2024; Received in revised form 14 August 2024; Accepted 14 August 2024
Available online 15 August 2024

links between brain regions, using techniques like diffusion tensor imaging (DTI) [12,13]. This provides a roadmap of the brain's physical connections, vital for understanding its architecture and identifying potential abnormalities in neurological and psychiatric disorders. However, its scope is limited as it doesn't fully capture the dynamic interactions between brain areas, and it can be affected by imaging artifacts. On the other hand, functional connectivity (FC) [12,14] focuses on the statistical relationships between different brain regions' activities, measured via methods like functional magnetic resonance imaging (fMRI) or EEG. Although useful, FC does not establish causation and can be influenced by noise and other confounding factors, which limits its ability to discern the structural underpinnings of brain function. Effective connectivity (EC) [12,15], the third category, explores the directional influence one brain region has over another, shedding light on the causality and dynamics of these interactions. This technique is pivotal in understanding how information flows within the brain, supporting the study of cognitive processes and disorders. By unraveling these complex neural relationships, EC provides crucial insights that guide both scientific inquiry and practical application in understanding the human mind.

Convolutional Neural Networks (CNNs) have become a cornerstone in EEG research, lauded for their adept handling of complex spatial and temporal data [16,17]. Their rise to prominence is driven by their capability to analyze high-dimensional EEG data efficiently, allowing researchers to extract meaningful patterns directly from the raw signals without intensive manual feature engineering. This feature is crucial in neuroscience, where precise interpretation of brain signals is paramount. CNNs are versatile, used for tasks ranging from detecting specific wave patterns and classifying emotional states [7–11,18,19] to diagnosing neurological disorders such as schizophrenia [20], depression [21–23] and Parkinson's disorder [24]. The widespread use of CNNs is also supported by advanced computational tools and libraries that simplify their implementation, broadening their accessibility to a diverse range of neuroscientific researchers and practitioners. In the process of utilizing CNNs, creating 2D images often involves converting single-channel EEG (or ECG) data into time-frequency representations (TFR). While traditional techniques like the continuous wavelet transform (CWT) [7,20,25–28], synchrosqueezing wavelet transform (SSWT) [10], and Smoothed pseudo-Wigner-Ville distribution (SPWVD) [18,20] focus on single-channel data, newer methods leveraging multiple EEG channels, such as brain connectivity analysis, are gaining traction for their broader applications and deeper analytical capabilities. These multi-channel techniques provide a richer, more detailed view of the brain's network interactions compared to single-channel methods. Time-frequency representation methods such as CWT, SSWT, and WVD excel in analyzing how frequency components of EEG and ECG signals evolve over time, offering a detailed view of brain activity at the micro level. These techniques are particularly adept at tracking temporal frequency changes linked to various cognitive and physiological states. Conversely, brain connectivity methods, which examine interactions across multiple EEG channels, offer a macroscopic view of brain function. This approach is especially valuable for its ability to uncover complex patterns and network dynamics across different brain regions during specific tasks or conditions. The comprehensive insights provided by brain connectivity are critical, especially for understanding collaborative brain region activities, which play an essential role in diagnosing and managing neurological conditions. In essence, while TFR methods deliver precise frequency analysis, brain connectivity techniques provide a holistic view of the brain's networked activity, paving the way for groundbreaking discoveries in EEG research.

The contributions of this study are multifaceted and significant in advancing the field of emotional signal processing:

Eye gate strategy: This novel approach is employed to identify and exclude irrelevant segments from emotional EEG and ECG signals. In this strategy, the most common feature of pupil diameter is used to identify and remove bad parts of emotional stimulation.

Multichannel information flow analysis: By implementing the eye gate strategy, this study explores the interaction and information flow between multiple EEG and ECG channels during tasks that involve emotional processing. Specifically, the EC method of GC is applied to data extracted from 32 EEG channels and 3 ECG channels, representing two emotional states—happiness and sadness—sourced from the public MAHNOB-HCI database. This is notably the first instance of integrating ECG with EEG to analyze connectivity during emotional tasks within this database.

Deep feature extraction and representation: Deep learning techniques are utilized to extract significant features from combined EEG-ECG signal windows. The study introduces a unique way to visualize multichannel EC through images that represent the complex interactions in two fundamental emotional states: sadness and happiness.

These advancements provide a more accurate and nuanced understanding of emotional states through bioelectrical signals, potentially leading to improved applications in fields such as neuromarketing, psychiatric diagnosis, and personalized therapy.

1.1. Related studies

Li et al. [29] explored FC by utilizing the phase locking value (PLV) derived from 32 EEG channels across two databases, DEAP and MAHNOB-HCI. They employed graph regularized extreme learning machine and support vector machine (SVM) techniques to classify three emotional states—negative, positive, and neutral—achieving accuracies of 62 % and 68 % for each database, respectively. Moon et al. [30] developed a CNN-based framework that utilizes a connectivity matrix crafted from brain connectivity measures including Pearson correlation coefficient (PCC), PLV, and transfer entropy (TE). This matrix was structured using two novel approaches that gauge similarity and dissimilarity among EEG signals from different spatial locations. Tested on the DEAP database, the system attained an accuracy of 80.73 % using the PLV measure, highlighting that the effective representation of emotional valence in the connectivity matrix significantly boosts classification accuracy. Bagherzadeh et al., in Ref. [9] estimated the partial directed coherence (PDC) measure from EC maps and used them to fine-tune six pre-trained CNN architectures to classify emotional states. Among these, ResNet-18 stood out, achieving the highest accuracies of 94.27 % for the DEAP database, 95.25 % for the MAHNOB-HCI database, and 96.00 % for the DREAMER database, underlining the effectiveness of EC-based CNN models in emotion detection from EEG signals. These techniques have also been employed in detecting depression. Also, Bagherzadeh et al., in another study [8], leveraged two EC measures, PDC and direct directed transfer function (dDTF), to differentiate four emotional states along with neutrality, based on the valence-arousal model. Using the ResNet-50 architecture, they achieved impressive accuracies of 99.43 % and 96.26 % on the MAHNOB-HCI database with alpha-dDTF and alpha-PDC images, respectively. Another study [11] fused three EC measures—TE, PDC, and dDTF—to create novel images for use with combined pre-trained CNN and LSTM models. They reached accuracies of 98.76 % and 98.86 % in recognizing five emotional states from the DEAP and MAHNOB-HCI databases, respectively, based on a two-dimensional valence-arousal model.

Pan et al. in Ref. [31], proposed a deep learning-based multimodal emotion recognition (MER) framework called Deep-Emotion, which integrates features from facial expressions, speech, and EEG to enhance MER performance. Deep-Emotion consists of three branches: the facial branch uses an improved GhostNet network, the speech branch employs a lightweight fully convolutional neural network (LFCNN), and the EEG branch utilizes a tree-like LSTM (tLSTM) model for feature extraction. Decision-level fusion combines the results from all branches for comprehensive recognition. Experiments on the CK+, EMO-DB, and MAHNOB-HCI datasets demonstrate the superiority of the Deep-Emotion method. In Ref. [32], Pan et al., proposed a novel spatio-temporal self-constructing graph neural network (ST-SCGNN) for cross-subject emotion recognition. It combines activation and connection pattern features for spatio-temporal feature generation. The self-constructing module dynamically updates the graph structure of the neural network. Experiments using the SEED and SEED-IV datasets showed average accuracies of 85.90 % and 76.37 %, surpassing state-of-the-art metrics.

Other EEG signal processing methods are.

2. Material and methods

2.1. Multimodal MAHNOB-HCI dataset

The MAHNOB-HCI dataset [33] is an innovative collection of EEG and ECG data harvested from 27 diverse, healthy individuals, who participated in the study by engaging with 20 unique video clips. The dataset encompasses data from 16 females and 11 males with varied educational backgrounds, ranging in age from 19 to 40 years (26.06 ± 4.39). Data was recorded under supervision of Imperial College of London. It was recorded over 32 EEG channels using the advanced Biosemi Active II system, adhering to the standard 10–20 electrode placement system at a frequency of 256 Hz. The EEG channels included are Fp1, Af3, F3, F7, Fc5, Fc1, C3, T7, Cp5, Cp1, P3, P7, Po3, O1, Oz, Pz, Fp2, Af4, Fz, F4, F8, Fc6, Fc2, Cz, C4, T8, Cp6, Cp2, P4, P8, Po4, and O2. Additionally, three ECG channels were simultaneously recorded from the upper right and left corners of the chest, and the left corner of the abdomen, all at the same sampling frequency.

The research protocol consisted of three stages designed to elicit distinct emotional responses: it began with a neutral video to balance any pre-existing emotional biases, followed by the viewing of emotionally charged video clips, and concluded with a self-assessment phase. Each participant's involvement lasted about 50 min, with individual assessments taking around two and a half minutes each. The lengths of the video clips varied, ranging from approximately 34 to 117 s, to fully engage the participants and elicit a range of emotions. Participants identified their emotions using concepts of valence and arousal, ultimately categorizing them into four emotional classes: happiness, disgust, fear, and sadness. This approach allowed for a nuanced examination of emotional responses

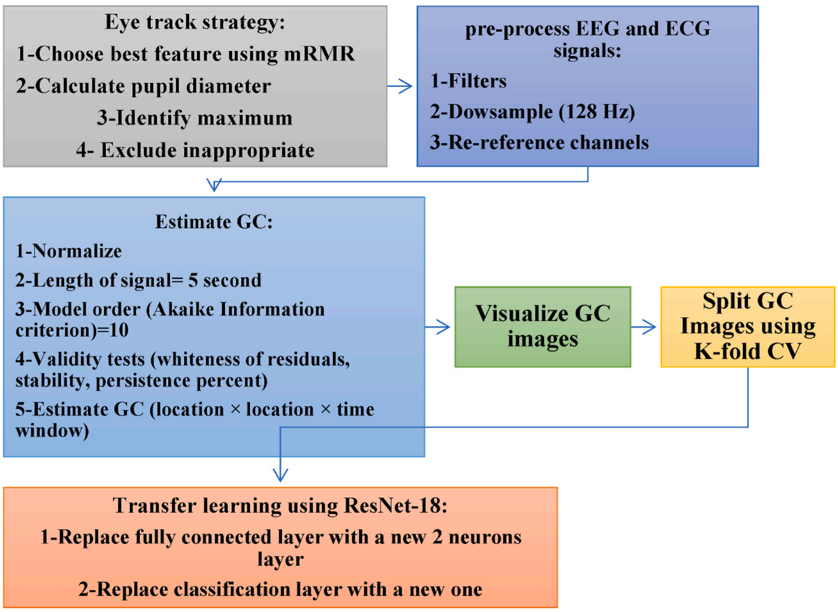


Fig. 1. Flowchart of the proposed emotion recognition based on GC images and ResNet-18 from multimodal MAHNOB-HCI database.

based on both valence and arousal scores.

2.2. Proposed ER system using GC maps and transfer learning from multimodal signals

The eye gate strategy is introduced to enhance the recognition of emotional states from multimodal signals. Pupil diameter has been chosen using the minimum-redundancy-maximum-relevance (mRMR) feature selection method [34] among pool of features. Also, this feature, serves as a common feature for detecting patterns of sadness and happiness within MAHNOB-HCI data. Once discriminative patterns are identified, relevant segments of EEG and ECG signals are isolated, as illustrated in Fig. 1. Then, these signals undergo noise and artifact removal through filtering, followed by a reduction in sampling frequency to decrease computational load and costs. Subsequently, the EEG signals are re-referenced to the average of all channels. The interaction between multichannel EEG signals and ECG signals is then examined within two primary emotional classes: happiness and sadness. These interactions are evaluated using the GC effective connectivity method. To ensure accurate GC estimation, signal stationarity is verified by segmenting the signals into smaller parts. GC estimation is carried out based on a specific model order and validated against predefined criteria. The resulting GC measures are visualized as color-scale images, which are divided into five-folds for fine-tuning using the pre-trained CNN model ResNet-18. This technique employs transfer learning, leveraging the knowledge of CNN models to address our new problem. The process involves replacing the fully connected layer with one containing two neurons, corresponding to the number of emotional classes. Additionally, the classification layer linked to the previous fully connected layer is updated accordingly.

2.3. Remove inappropriate parts of EEG and ECG using eye-tracking strategy

First, multiple features such as mean and standard deviation of fixation duration and pupil diameter (in X and Y directions), fixation frequency and maximum fixation duration are extracted [35]. Then, the pupil diameter has been chosen among these features using the mRMR method. This feature selection approach systematically identifies a set of features characterized by minimal mutual redundancy while maximizing the mutual information with the outcome variable (class label) [34].

During investigation pupil diameter values from sad and happy data, we found that during sadness this feature decrease and during happiness it increases. So, we used this as a good metric to remove irrelevant parts from data. For example, when the subject watched a sad clip and pupil diameter increases so the subject goes out from sadness. Therefore, corresponding times in EEG and ECG signals are determined and irrelevant parts are removed. Fig. 2 shows this issue. Pupil diameter (D) is calculated by Eq (1) as follows:

$$D = 2 \times \sqrt{\frac{A}{\pi}} \tag{1}$$

where, A is area of the pupil.

2.4. Estimate GC effective connectivity

Effective connectivity is a useful tool to analyze multichannel EEG signal with the aim of revealing interaction between brain channels and therefore discriminate brain states. Each effective connectivity method estimates this interaction based on its point of view. GC is an appropriate effective connectivity method in emotion recognition [35–39]. This method is used to reveal interaction between 32 EEG channels and 3 ECG channels during two emotional states of happiness and sadness.

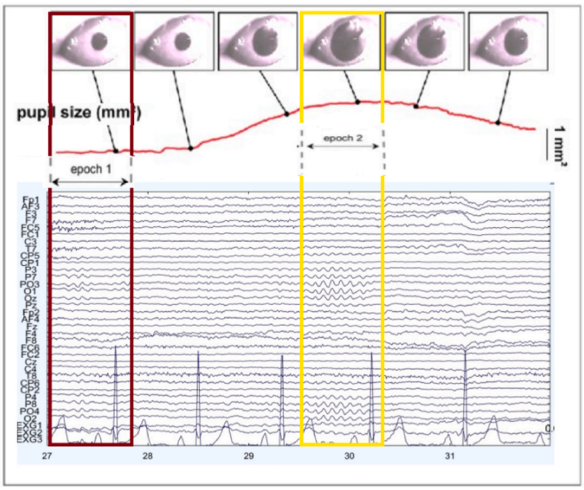


Fig. 2. An example from a participant watching clip with sad label. As it can be observed, when pupil diameter increases the subject passed sadness and emotional state is changed. So, this time cane be removed from EEG and ECG signals.

When examining two signals concurrently, represented as $x(t)$ and $y(t)$, if the predictive capability of the first signal significantly improves by incorporating historical data from the second signal as opposed to relying solely on its own information, then the second signal is considered causal to the first [40]. In the context of the univariate autoregressive model (AR), we define Eq (1), where $a_{i,j}$ represents the model parameters (typically estimated via the least squares method), P denotes the order of the AR model, and u_i signifies the residuals associated with the model. It is noteworthy that the prediction of each signal (x and y) is based solely on its respective past data [40].

$$x(n) = \sum_{k=1}^{P} a_{x,k}x(n-k) + \sum_{k=1}^{P} u_x u(n)\ y(n) = \sum_{k=1}^{P} a_{y,k}y(n-k) + \sum_{k=1}^{P} u_y u(n) \tag{1}$$

Eq (2) calculate the GC from y to x:

$$GC_{y\rightarrow x} = \ln\left(\frac{V_{x|\overline{x}}}{V_{x|\overline{x},\overline{y}}}\right) \tag{2}$$

The variances of these residuals are computed as $V_{x|\overline{x},\overline{y}} = var\left(\mu_{yx}\right)$ and $V_{x|\overline{x}} = var\left(\mu_y\right)$. The GC value ranges from zero to infinity. As previously mentioned, GC offers the advantage of asymmetry, allowing it to discern effective connectivity. However, it is a linear parametric technique, dependent on the autoregressive model of order p, where p denotes the model order. The estimation of GC is conducted using MATLAB 2023b with the HERMES connectivity toolbox, version 2020, and the model order is preset to 10.

2.5. Transfer learning and ResNet-18

Transfer learning is widely used in deep learning applications, enabling the utilization of a pre-trained network as a base to master a new task. Fine-tuning a network through transfer learning is often faster and less demanding than training a network from the ground up with randomly initialized weights. This method quickly incorporates learned features into a new task, even with a limited number of training images [8–11]. Given the wide range of available transfer learning models, it is essential to consider the trade-offs and ensure alignment with the overall objectives of our study. After testing various models to find the best match for our application, ResNet-18 [41] was selected due to its balance of storage and computational resource constraints while maintaining operational feasibility. ResNet-18 features one convolutional layer (filter size = 7 × 7), 8 consecutive residual units (each with 2 consecutive convolutional layers with a filter size of 3 × 3 and a shortcut), and one fully connected layer with 1000 neurons. Its structure is illustrated in Fig. 3. For more details about ResNet-18, see Ref. [41]. The Cross-Entropy loss function and the ADAM (Adaptive Moment Estimation) algorithm [42] were used to optimize the training process. This model had been used in deep learning studies to recognize emotional states from imaging derived EEG signals [7,9,10] and therefore, it has demonstrated it's generalization and effectiveness to solve complicated emotion recognition problems.

2.6. Evaluation statistics

To evaluate the recognition performance of our GC-ResNet-18 approach, we employ 5-fold cross-validation (CV). In this setup, images from each emotion category are split into four folds for fine-tuning ResNet-18, with the remaining fold used for testing. This process is repeated five times to ensure each image is tested. We calculate the average accuracy and the area under the curve (AUC) as metrics to assess performance in our two-class emotion recognition problem [43]. Eq (1) and Eq (2) calculated these measures, respectively.

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \tag{2}$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{3}$$

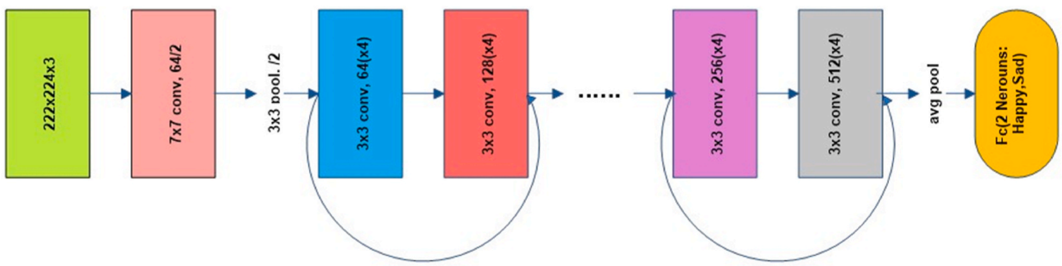


Fig. 3. Simplified structure of ResNet-18. ResNet-18 is designed based on residual units which includes 2 consecutive convolutional layers and a shortcut.

where.

- TP (True Positive) refers to the correct identification of Class 1 over Class 2,
- TN (True Negative) refers to the correct rejection of Class 1 in favor of Class 2,
- FN (False Negative) indicates an incorrect rejection of Class 1 in favor of Class 2,
- FP (False Positive) indicates an incorrect identification of Class 1 over Class 2.

This evaluation method is applied across two emotional classes. Specifically, Class 1 represents happiness and Class 2 represents sadness.

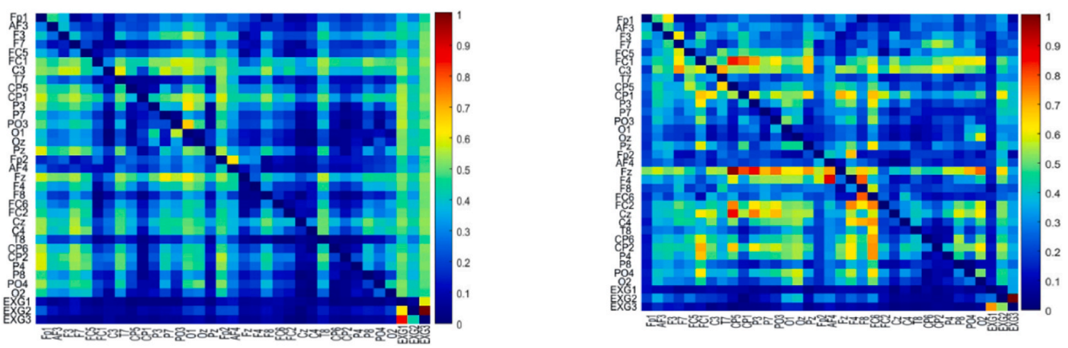
3. Results

EEG data from 32 channels and 3 ECG channels from the MAHNOB-HCI database, which included 27 participants, was initially processed using the EEGLAB toolbox in MATLAB software (version 2023b), as detailed in the EEG signal preprocessing section. Data from three subjects were excluded due to technical errors. To calculate the Granger causality (GC) connectivity measure across the processed EEG channels from 24 participants, the HERMES toolbox within the MATLAB environment was utilized. The process started with determining the window length based on the variance ratio test to select stationary segments, resulting in a consistent window of 5000 ms for analysis across the database. A sliding window of 2500 ms was used to generate numerous images for analysis with pre-trained CNN models. An MVAR model with an order of 10, chosen according to the Akaike Information Criterion (AIC) [12], was consistently applied. The model's appropriateness was confirmed through checks for residual whiteness, consistency, and stability [12]. After these validations, the GC measure was computed, resulting in a 35 × 35 × 25 array (location × location × time window). This GC data was visualized using color scales and saved in.jpg format (default size in MATLAB is 875 × 656). These images were then used as input for mentioned pre-trained CNN, producing 2853 images for sadness and 4391 for happiness. Fig. 4 displays GC images from one participant showing happiness (a) and sadness (b), where information flow is depicted with cooler blue tones for lower values and warmer red tones for higher values. The first 32 arrays on each axis represent EEG channels (ranging from Fp1 to O2), and the last three are ECG channels (EXG1: upper right corner of the chest, EXG2: upper left corner of the chest, and EXG1: left side of the abdomen). According to Fig. 4 (b), this GC image reveal the most interactions between EEG and EXG2 channels than the other two ECGs. Indeed, this person during this sadness window has interactions between second ECG channel and some of frontal, parietal and occipital channels such as Af3, Fc5, O1, Po3. Also, ECG channels have most interaction among them. This is similar for happiness (Fig. 4 (a)), but interactions among ECG and EEG channels are weak.

Subsequently, as we have used ResNet-18 through transfer learning approach, the fully connected layer is replaced by a new layer with two neurons to classify sadness and happiness. Also, a new classification layer related to new fully connected layer is applied. Then, GC images were used to fine-tune ResNet-18 through 5-fold CV method. The re-train process was repeated five-times based on ADAM optimizing method and each time average accuracy and AUC values are calculated. hyperparameters were set as follows: an initial learning rate of 0.0008, a squared gradient decay factor of 0.99, a mini-batch size of 32, and a maximum of 160 epochs. The stopping criterion was loss value lower than 0.00001 in five consecutive epochs. All analytical procedures and coding were executed in MATLAB software version 2023b on a laptop powered by an Intel(R) Core(TM) i7-10610U CPU @1.80 GHz, GPU Nvidia Quadro P520.

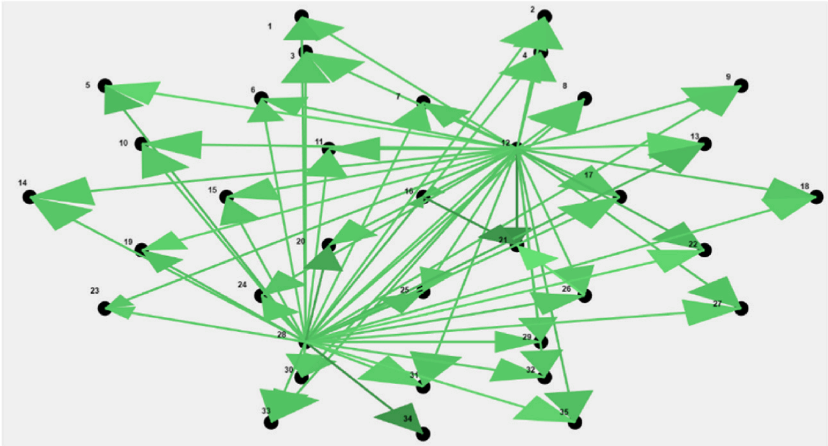
Fig. 5 shows topographic images of GC for happiness (a) and sadness (b) for subject number one at one 5-s window of analysis. This pattern of interaction between EEG channels and ECG channels was similar for most of subjects. As it can be observed from Fig. 5(a) and (b), there are coupling between Cp2 EEG channel and all ECG channel during sadness and happiness. Also, P7 is coupled with the first and third ECG channels during happiness.

Table 1 reports accuracy values achieved using the ResNET-18 on GC images. According to Table 1, the highest accuracy achieved

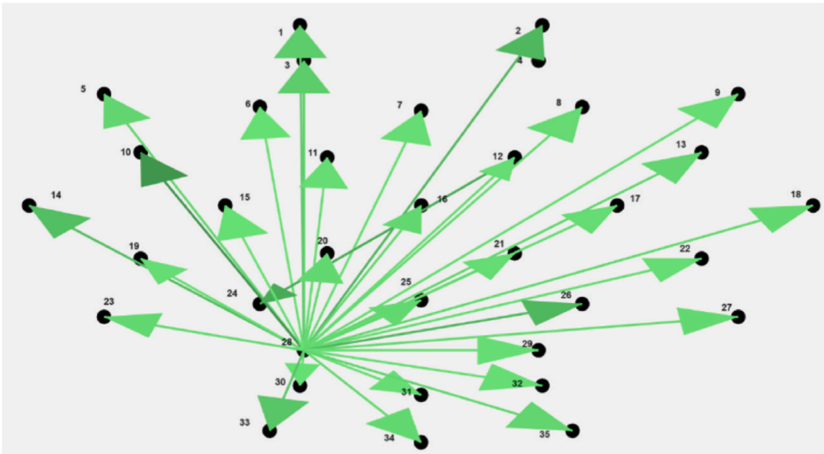


(a) (b)

Fig. 4. An example of GC images for one subject during (a) happiness and (b) sadness. The color scales illustrate the flow of information, with lower values depicted in cooler blue tones and higher values in warmer reds.



(a)



(b)

Fig. 5. Topographic representation of GC for happiness (a) and sadness (b) for first subject at one-5 second window of analysis. Order of channels from 1 to 35 are: Fp1, Af3, F3, F7, Fc5, Fc1, C3, T7, Cp5, Cp1, P3, P7, Po3, O1, Oz, Pz, Fp2, Af4, Fz, F4, F8, Fc6, Fc2, Cz, C4, T8, Cp6, Cp2, P4, P8, Po4, O2, EXG1, EXG2 and EXG3.

using GC-EEG-ECG images and ResNet-18 equal to 91.34 (std = ± 1.01). Also, the average AUC for this configuration achieved 0.97. GC images from EEG achieved the average accuracy and AUC of 89.25 % and 0.92, respectively.

Fig. 6 shows ROC for GC-ResNet-18. AUC values for happiness and sadness, for this configuration are 0.97, and 0.97, respectively.

Fig. 7 shows training progress based on percent of accuracy and loss function over epochs. This figure is result of training the ResNet-18 on GC images of 4-folds.

Table 1
Results of ResNet-18 on GC images from multichannel EEG and EEG-ECG.

Modality	Accuracy (%)(mean ± std)	AUC
EEG	89.25 ± 1.45	0.92
EEG-ECG	91.00 ± 1.01	0.97

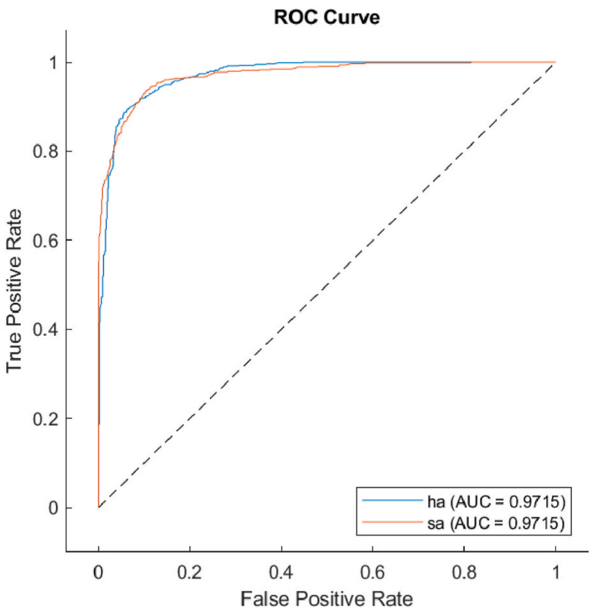


Fig. 6. ROC curve for GC images of ECG-EEG using ResNet-18.

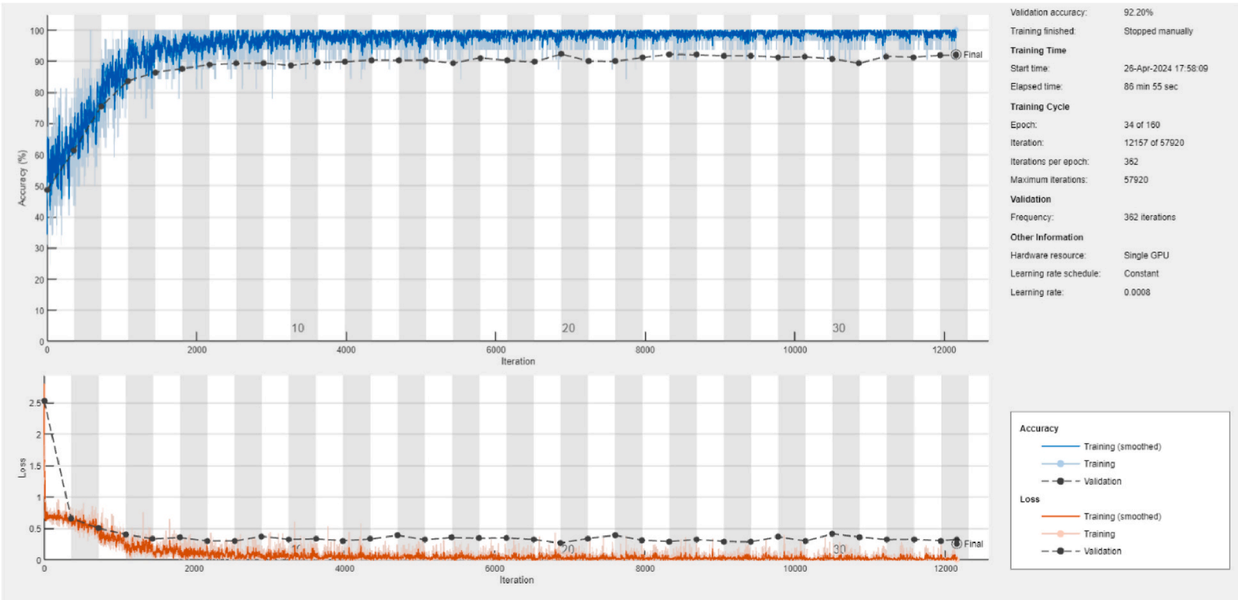


Fig. 7. up: accuracy curves and down: loss function curves over epochs for ResNet-18 on GC images of 4-folds.

4. Discussion

4.1. Identifying irrelevant data sections

The primary objective of this study was to identify irrelevant sections of data related to elicited emotions, specifically sadness and happiness. Pupil diameter was selected using the mRMR method among common features (fixation duration in X and Y directions and etc.) as a key metric for this purpose. The findings revealed a clear pattern: pupil diameter increased when subjects experienced happiness and decreased during sadness. This differentiation allowed for the exclusion of irrelevant data segments, ensuring that only pertinent EEG and ECG data were analyzed.

4.2. Exploring EEG and ECG interactions

The second objective focused on exploring the interactions between EEG and ECG channels during emotional states using GC measures. This study stands out for its innovative approach of jointly estimating GC measures from both EEG and ECG data for emotion recognition tasks. This dual-modality analysis provides a more nuanced understanding of emotional processing by capturing both the brain's electrical activity and the autonomic nervous system's response.

While previous studies have leveraged EEG and photoplethysmogram (PPG) to estimate connectivity metrics such as Generalized Partial Directed Coherence (gPDC) and dDTF in contexts like deceptive interviews [44,45], the current study highlights the added value of integrating EEG with ECG. This combination enriches our understanding of the physiological underpinnings of emotions. Recent research underscores the advantages of such multimodal approaches, advocating for their potential to offer a comprehensive view of emotional states.

The integration of EEG and ECG signals is particularly advantageous for real-time emotion recognition systems, which are increasingly used in fields such as neuromarketing and mental health monitoring. EEG provides rapid detection capabilities by measuring direct brain responses to emotional stimuli, while ECG offers a stable and simple means to monitor the autonomic nervous system's reactions. Together, these modalities enhance the robustness and usability of emotion recognition systems in practical applications.

According to Figs. 4 and 5, there are notable interactions between the Af3, Fc5, O1, Po3, Cp2, and P7 EEG channels and ECG channels during states of sadness and happiness. The frontal region, in particular, has been identified in physiological studies as the primary area for emotional processing. These findings align with previously introduced channels in studies [7,9,10,46], which also highlight the significant role of these regions in emotional responses.

The observed interactions suggest that the frontal lobe, especially the regions corresponding to the Af3 and Fc5 channels, is heavily involved in processing emotions. Additionally, the occipital (O1), parietal (Po3, P7), and central (Cp2) regions also show significant engagement, indicating a widespread network of brain areas contributing to emotional experiences. This comprehensive network's involvement underscores the complexity of emotional processing, integrating both cortical and autonomic responses as reflected in the ECG channels.

Moreover, the consistency of these channels with those identified in previous studies reinforces the reliability of these regions in emotional processing research. This alignment with earlier findings enhances the robustness of our results and suggests potential biomarkers for specific emotional states. As such, these channels could serve as critical focal points for further research into the neural underpinnings of emotions and their physiological correlates.

4.3. Comparison with recent emotion recognition studies on same database

Table 2 presents a comparison of our results with those from recent emotion recognition studies using the same database (MAHNOB-HCI). The differences between our study and others include the number of emotion classes, evaluation criteria, modalities, and the data segments processed. We have enhanced our study by augmenting it with ECG data to provide deeper insights into the interactions between the brain (EEG) and heart (ECG) during emotional responses. According to this table, our results surpass those of all other studies compared. Overall, our proposed emotion recognition system outperforms the current state-of-the-art.

Table 2 presents a meticulous comparison of our results with those from recent emotion recognition studies that also used the MAHNOB-HCI database. Our research is distinct in several ways: we incorporated fewer emotion classes (two compared to up to four in other studies), used the combination of EEG and ECG data (while others used either EEG, ECG, or facial expressions combined with EEG), and implemented a 5-fold cross-validation evaluation method, which is similar to the methods used in some of the other studies. Our technique employs GC with a ResNet-18 architecture, providing an innovative approach to emotion recognition by capturing the nuanced interactions between brain and heart activities during emotional responses.

The table illustrates that our method achieves a higher accuracy of 91.34 % (±1.19), which is notably superior to other methods referenced. For instance, study [47] which uses a multivariate synchrosqueezing transform with a deep CNN on EEG data, achieves an accuracy of 88.17 % under a hold-out evaluation criterion. Similarly, the recent study [47] combines face and EEG data and achieves accuracies of 75.21 % for valence and 74.17 % for arousal using a 10-fold cross-validation. In contrast, methods focusing solely on ECG data, like in studies [4,49], achieved lower accuracies under similar or less stringent evaluation criteria.

This comparative data underscores the effectiveness of integrating EEG and ECG modalities and highlights the robustness of our system against both single-modality and dual-modality approaches, affirming our system's position as a new benchmark in the field of emotion recognition technology.

The advantages of this study include.

- Exclude irrelevant parts from emotional data using eye-tracking gate strategy.
- Estimating the coupling of EEG and ECG during emotional processing to enhance the performance of emotion recognition systems.
- Jointly analyzing the effective connectivity of multichannel EEG and ECG data to explore their roles in processing emotions such as sadness and happiness.
- Utilizing a transfer learning approach to classify emotional states from physiological data, compensating for the limited number of initial training samples.

However, this study has certain limitations.

Table 2
Comparison of emotion recognition studies on MAHNOB-HCI database.

Ref	Modality	Method	Number of classes	Evaluation criterion	Accuracy (%) (m ± std)
[47]	EEG	multivariate synchrosqueezing transform, deep CNN	4	Hold out	88.17
[48]	Face, EEG	Power spectral density, SVM-recursive feature elimination, fusion	2[a]	10-fold CV	75.21 (valence), 74.17 (arousal)
[4]	ECG	Poincare map, nonlinear features, SVM, KNN	2[a]	5-fold CV	78.07 %
[49]	ECG	1DCNN-LSTM	2[a]	Hold out	73.33 (valence), 60 (arousal)
Ours	EEG-ECG	GC, ResNet-18	2	5-fold CV	91.34 ± 1.19

[a] two classes of valence and arousal, i.e., high valence vs. low valence and high arousal vs. low arousal.

- A limitation of this study concerns the analysis of heart rate variability (HRV), a commonly used metric in many ER studies that reflects changes in emotional states such as fear, anger, and disgust. HRV operates predominantly below 4 Hz, aligning with the delta frequency band in EEG, which is not specifically targeted in emotion recognition procedures. Therefore, HRV was not considered in this study.
- Additionally, the sample size poses a constraint; the database includes 27 subjects, each viewing 18 emotional video clips, which may not provide a robust sample size for extensive analysis.

5. Conclusion and future works

The study focuses on improving the accuracy and reliability of emotion recognition systems by utilizing multimodal signals. We have introduced the use of eye-tracking as a gating strategy to filter out irrelevant data segments. Additionally, we propose the integration of ECG with EEG to estimate EC, thereby representing brain-heart information flow during emotions such as happiness and sadness. For this purpose, GC is estimated and used as input for the efficient and high-performance pre-trained CNN model, ResNet-18. As a result, we achieved the highest average accuracy of 91.00 %. Our results were benchmarked against state-of-the-art studies and demonstrate that augmenting ECG with EEG enhances emotion recognition capabilities for sadness and happiness.

Looking ahead, we plan to use eye-tracking data more extensively as a filter to select the most impactful segments of EEG and ECG data for a broader range of emotions, including sadness, happiness, disgust, and fear. These emotions are fundamental and are common across different cultures, ages, and sexes. We also intend to incorporate other deep learning techniques, such as Long Short-Term Memory (LSTM) networks, to integrate temporal dynamics into our proposed emotion recognition system.

Data availability

In this study, the public available MAHNOB-HCI were analyzed. The databases can be found here: https://mahnob-db.eu/hci-tagging/.

CRediT authorship contribution statement

Javid Farhadi Sedehi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Nader Jafarnia Dabanloo:** Validation, Supervision, Methodology, Investigation, Conceptualization. **Keivan Maghooli:** Validation, Supervision, Methodology, Investigation, Conceptualization. **Ali Sheikhani:** Methodology, Investigation, Conceptualization.

Declaration of competing interest

There are no conflict interest.

Acknowledgment

We would like to thank the Science and Research Branch, Islamic Azad University, Tehran, Iran due to their support.

References

[1] M. Jafari, A. Shoeibi, M. Khodatars, S. Bagherzadeh, A. Shalbaf, D.L. García, J.M. Gorriz, U.R. Acharya, Emotion recognition in EEG signals using deep learning methods: a review, Comput. Biol. Med. (2023 Sep 9) 107450.
[2] M.M. Rahman, A.K. Sarkar, M.A. Hossain, M.S. Hossain, M.R. Islam, M.B. Hossain, J.M. Quinn, M.A. Moni, Recognition of human emotions using EEG signals: a review, Comput. Biol. Med. 136 (2021 Sep 1) 104696.
[3] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, W. Zhang, A systematic review on affective computing: emotion models, databases, and recent advances, Inf. Fusion 83 (2022 Jul 1) 19–52.
[4] M. Baghizadeh, K. Maghooli, F. Farokhi, N.J. Dabanloo, A new emotion detection algorithm using extracted features of the different time-series generated from ST intervals Poincaré map, Biomed. Signal Process Control 59 (2020 May 1) 101902.

[5] F.E. Oğuz, A. Alkan, T. Schöler, Emotion detection from ECG signals with different learning algorithms and automated feature engineering, Signal, Image and Video Processing 17 (7) (2023 Oct) 3783–3791.
[6] T.P. Jung, T.J. Sejnowski, Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing, IEEE Transactions on Affective Computing 13 (1) (2019 May 14) 96–107.
[7] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Emotion recognition using continuous wavelet transform and ensemble of convolutional neural networks through transfer learning from electroencephalogram signal, Frontiers in Biomedical Technologies 10 (1) (2023) 47–56.
[8] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals, Cognitive Neurodynamics 16 (5) (2022 Oct) 1087–1106.
[9] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Recognition of emotional states using frequency effective connectivity maps through transfer learning approach from electroencephalogram signals, Biomed. Signal Process Control 75 (2022 May 1) 103544.
[10] S. Bagherzadeh, M.R. Norouzi, S.B. Hampa, A. Ghasri, P.T. Kouroshi, S. Hosseininasab, M.A. Zadeh, A.M. Nasrabadi, A subject-independent portable emotion recognition system using synchrosqueezing wavelet transform maps of EEG signals and ResNet-18, Biomed. Signal Process Control 90 (2024 Apr 1) 105875.
[11] S. Bagherzadeh, A. Shalbaf, A. Shoeibi, M. Jafari, R. San Tan, U.R. Acharya, Developing an EEG-based emotion recognition using ensemble deep learning methods and fusion of brain effective connectivity maps, IEEE Access 12 (2024) 50949–50965.
[12] T. Mullen, Source information flow toolbox (SIFT), Swartz Center Comput Neurosci 15 (2010 Dec 15) 1–69.
[13] D.K. Jones, Diffusion MRI, Oxford University Press, 2010 Nov 11.
[14] R.B. Buxton, Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques, Cambridge university press, 2009 Aug 27.
[15] K.J. Friston, L. Harrison, W. Penny, Dynamic causal modelling, Neuroimage 19 (4) (2003 Aug 1) 1273–1302.
[16] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, S. Zuo, EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation, IEEE Transact. Neural Networks Learn. Syst. 30 (9) (2019 Jan 10) 2755–2763.
[17] P. Thuwajit, P. Rangpong, P. Sawangjai, P. Autthasan, R. Chaisaen, N. Banluesombatkul, P. Boonchit, N. Tatsaringkansakul, T. Sudhawiyangkul, T. Wilaiprasitporn, EEGWaveNet: multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection, IEEE Trans. Ind. Inf. 18 (8) (2021 Dec 10) 5547–5557.
[18] S.K. Khare, V. Bajaj, Time–frequency representation and convolutional neural network-based emotion recognition, IEEE Transact. Neural Networks Learn. Syst. 32 (7) (2020 Jul 31) 2901–2909.
[19] S.K. Khare, V. Bajaj, An evolutionary optimized variational mode decomposition for emotion recognition, IEEE Sensor. J. 21 (2) (2020 Sep 1) 2035–2042.
[20] S.K. Khare, V. Bajaj, U.R. Acharya, SPWVD-CNN for automated detection of schizophrenia patients using EEG signals, IEEE Trans. Instrum. Meas. 70 (2021 Apr 2) 1–9.
[21] D.M. Khan, N. Yahya, N. Kamel, I. Faye, Automated diagnosis of major depressive disorder using brain effective connectivity and 3D convolutional neural network, IEEE Access 9 (2021 Jan 5) 8835–8846.
[22] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, D.P. Subha, Automated EEG-based screening of depression using deep convolutional neural network, Comput. Methods Progr. Biomed. 161 (2018 Jul 1) 103–113.
[23] H. Ke, D. Chen, T. Shah, X. Liu, X. Zhang, L. Zhang, X. Li, Cloud-aided online EEG classification system for brain healthcare: a case study of depression evaluation with a lightweight CNN, Software Pract. Ex. 50 (5) (2020 May) 596–610.
[24] M. Murugappan, W. Alshuaib, A.K. Bourisly, S.K. Khare, S. Sruthi, V. Bajaj, Tunable Q wavelet transform based emotion classification in Parkinson's disease using Electroencephalography, PLoS One 15 (11) (2020 Nov 19) e0242014.
[25] F. Tokmak, A. Subasi, S.M. Qaisar, Artificial intelligence-based emotion recognition using ECG signals, in: Applications of Artificial Intelligence Healthcare and Biomedicine, vol. 1, Academic Press, 2024 Jan, pp. 37–67.
[26] O. Ozaltin, O. Yeniay, A novel proposed CNN–SVM architecture for ECG scalograms classification, Soft Comput. 27 (8) (2023 Apr) 4639–4658.
[27] M.M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, B. BenJdira, Convolutional neural networks for electrocardiogram classification, J. Med. Biol. Eng. 38 (2018 Dec) 1014–1025.
[28] M. RashedAlMahfuz, M.A. Moni, P. Lio, S.M. Islam, S. Berkovsky, M. Khushi, J.M. Quinn, Deep convolutional neural networks based ECG beats classification to diagnose cardiovascular conditions, Biomedical engineering letters 11 (2021 May) 147–162.
[29] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, P. Xu, EEG based emotion recognition by combining functional connectivity network and local activations, IEEE Trans. Biomed. Eng. 66 (10) (2019) 2869–2881.
[30] S.E. Moon, C.J. Chen, C.J. Hsieh, J.L. Wang, J.S. Lee, Emotional EEG classification using connectivity features and convolutional neural networks, Neural Network. 132 (2020 Dec 1) 96–107.
[31] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang, S. Wang, Multimodal emotion recognition based on facial expressions, speech, and EEG, IEEE Open Journal of Engineering in Medicine and Biology 5 (2023) 396–403.
[32] J. Pan, R. Liang, Z. He, J. Li, Y. Liang, X. Zhou, Y. He, Y. Li, St-Scgnn, A spatio-temporal self-constructing graph neural network for cross-subject EEG-based emotion recognition and consciousness detection, IEEE Journal of Biomedical and Health Informatics 28 (2023) 777–788.
[33] [database] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, A multimodal database for affect recognition and implicit tagging, IEEE transactions on affective computing 3 (1) (2011 Aug 4) 42–55.
[34] A. Malekzadeh, A. Zare, M. Yaghoobi, R. Alizadehsani, Automatic diagnosis of epileptic seizures in EEG signals using fractal dimension features and convolutional autoencoder method, Big Data and Cognitive Computing 5 (4) (2021 Dec 13) 78.
[35] W.L. Zheng, W. Liu, Y. Lu, B.L. Lu, A. Cichocki, Emotionmeter: a multimodal framework for recognizing human emotions, IEEE Trans. Cybern. 49 (3) (2018 Feb 8) 1110–1122.
[36] Y. Gao, X. Wang, T. Potter, J. Zhang, Y. Zhang, Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis, J. Neurosci. Methods 346 (2020 Dec 1) 108904.
[37] T. Li, G. Li, T. Xue, J. Zhang, Analyzing brain connectivity in the mutual regulation of emotion–movement using bidirectional granger causality, Front. Neurosci. 14 (2020 May 6) 369.
[38] T. Li, G. Li, T. Xue, J. Zhang, Analyzing brain connectivity in the mutual regulation of emotion–movement using bidirectional granger causality, Front. Neurosci. 14 (2020 May 6) 369.
[39] T. Li, G. Li, T. Xue, J. Zhang, Analyzing brain connectivity in the mutual regulation of emotion–movement using bidirectional granger causality, Front. Neurosci. 14 (2020 May 6) 369.
[40] G. Niso, R. Bruña, E. Pereda, R. Gutiérrez, R. Bajo, F. Maestú, F. Del-Pozo, HERMES: towards an integrated toolbox to characterize functional and effective brain connectivity, Neuroinformatics 11 (2013 Oct) 405–434.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[42] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.
[43] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009 Jul 1) 427–437.
[44] M. Daneshi Kohan, A. Motie Nasrabadi, M.B. Shamsollahi, A. Sharifi, EEG/PPG effective connectivity fusion for analyzing deception in interview, Signal, Image and Video Processing 14 (5) (2020 Jul) 907–914.
[45] M.D. Kohan, A.M. Nasrabadi, M.B. Shamsollahi, Interview based connectivity analysis of EEG in order to detect deception, Med. Hypotheses 136 (2020 Mar 1) 109517.
[46] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, A hybrid EEG-based emotion recognition approach using wavelet convolutional neural networks and support vector machine, Basic Clin. Neurosci. 14 (1) (2023 Jan) 87.

[47] Y. Zhang, C. Cheng, Y. Zhang, Multimodal emotion recognition based on manifold learning and convolution neural network, Multimed. Tool. Appl. 81 (23) (2022 Sep) 33253–33268.
[48] Y. Huang, J. Yang, S. Liu, J. Pan, Combining facial expressions and electroencephalography to enhance emotion recognition, Future Internet 11 (5) (2019 May 2) 105.
[49] W. Mellouk, W. Handouzi, CNN-LSTM for automatic emotion recognition using contactless photoplythesmographic signals, Biomed. Signal Process Control 85 (2023 Aug 1) 104907.