Society of Medical Education & Research

2025, Volume 5, Issue 2, Page No: 24-38
ISSN: 3108-4826

Journal of Medical Sciences and Interdisciplinary Research

A Spatiotemporal Neural Network Framework for EEG-Based Emotion Recognition in Depression Assessment

Tomas Jan Novak[1]*, Petr Martin Dvorak[1]

[1]Department of Management, Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic.

***E-mail** ✉ t.novak.muni@outlook.com

Abstract

Depression is primarily characterized by emotional dysfunction, reflected in heightened negative emotions and diminished positive emotions. Consequently, accurate emotion recognition has become a crucial approach for assessing depression. Among the various signals employed for emotion recognition, electroencephalogram (EEG) signals have gained considerable attention due to their rich spatiotemporal information across multiple channels. In this study, EEG data were first preprocessed using filtering and Euclidean alignment. For feature extraction, time-frequency features were obtained through short-time Fourier transform and Hilbert–Huang transform, while spatial features were extracted using convolutional neural networks. Temporal relationships were then explored using a bi-directional long short-term memory network. Prior to convolution operations, EEG features were transformed into 3D tensors based on the unique topology of EEG channels. The proposed framework demonstrated strong performance on two emotion datasets: SEED and the Emotional BCI dataset from the 2020 WORLD ROBOT COMPETITION. When applied to depression recognition, the method achieved an accuracy exceeding 70% under five-fold cross-validation. Moreover, using a subject-independent protocol on the SEED dataset, the approach attained state-of-the-art performance, surpassing previous studies. Overall, we present a novel EEG-based emotion recognition framework for depression detection, offering a robust algorithm suitable for real-time clinical applications.

Keywords: Long-short term memory network (LSTM), Emotion recognition, Depression, Convolutional neural network (CNN), Electroencephalogram (EEG)

Introduction

Emotion recognition is a central focus of affective computing and has been extensively applied in depression detection [1, 2]. Emotions play a critical role in shaping human experiences, influencing both their quality and scope [3]. With advances in brain–computer interfaces (BCIs) and artificial intelligence, EEG-based emotion recognition has emerged as a prominent research area. EEG signals are particularly valuable because they carry rich emotion-related information, exhibit high temporal resolution, and are difficult to intentionally mask [4–6]. These properties make EEG highly suitable for real-time emotion monitoring, which has important applications in domains such as fatigue detection during driving [7], depression monitoring [8], and real-time observation of critically ill patients [9].

Previous studies have revealed the neural basis of emotional processing. Brain regions implicated in emotion include the orbitofrontal cortex, insular cortex, and anterior and posterior cingulate cortices, while the amygdala links perception with automatic emotional responses and memory formation [3]. Amygdala activity is more strongly associated with negative emotions, and right frontal lobe activation has been linked to affective states such as fear and disgust [10]. Specifically, fear is primarily related to amygdala activity [11], anger to the

Access this article online https://smerpub.com/

Received: 12 April 2025; **Accepted:** 06 August 2025

How to cite this article: Novak TJ, Dvorak PM. A Spatiotemporal Neural Network Framework for EEG-Based Emotion Recognition in Depression Assessment. J Med Sci Interdiscip Res. 2025;5(2):24-38. https://doi.org/10.51847/A2pBOYHJW1

orbitofrontal cortex and anterior cingulate cortex [12], sadness to the amygdala and right temporal pole [13], and disgust to the anterior insula and anterior cingulate cortex [14]. Furthermore, alpha-band power and hemispheric asymmetries reflect emotional states [15–17], gamma-band oscillations correlate with happiness and sadness, and lateralized reductions in alpha activity in temporal lobes correspond to joy and sorrow (left hemisphere—sadness, right hemisphere—happiness) [18, 19].

Extracting emotion-specific EEG features that maximize inter-class differences while minimizing intra-class variability is essential, particularly for cross-database applications. Common EEG feature extraction techniques span multiple domains. In the time domain, methods include Hjorth parameters (activity, mobility, and complexity) [20], higher-order crossover features [21], and magnitude squared coherence [22]. Frequency-domain approaches often use power spectral density [23, 24], while time-frequency analyses employ time-frequency spectrum (TFS) features [25]. Other approaches include autoregressive modeling [26], asymmetric spatial patterns [27], entropy measures (differential entropy [7], sample entropy [28], approximate entropy [29]), maximum relevance minimum redundancy (mRMR) [30], common spatial patterns (CSP) [31], filter bank CSP (FBCSP) [32], and fractal-based measures such as Higuchi fractal dimension [33]. Features are usually extracted from delta, theta, alpha, beta, and gamma frequency bands [34]. Combining multiple complementary features enables more accurate characterization and classification of EEG signals.

Developing reliable models often requires large amounts of labeled calibration data, a process that is both time-intensive and laborious, limiting practical deployment of emotional BCIs. Reducing or eliminating this calibration to achieve plug-and-play functionality is therefore a critical challenge. Transfer learning can help by leveraging annotated data from auxiliary users to construct models for new users [35]. However, individual differences in neural responses to identical stimuli necessitate adaptation of data distributions to minimize these discrepancies [36]. To address this, we employ an unsupervised distribution alignment method, Euclidean alignment (EA), to standardize data across users [37].

Deep learning has shown great potential in improving EEG-based emotion recognition by learning hierarchical, high-level features. Techniques such as deep belief networks [7], recurrent neural networks [38], graph convolutional networks [39], transfer learning [40], and adversarial networks [41] have all been applied successfully. Nevertheless, most studies are limited to subject-dependent or single-database experiments, which do not fully reflect real-world applicability. Cross-database recognition remains a major challenge due to differences in feature distributions between training and test datasets, often resulting in performance degradation [42]. In this work, we leverage deep learning's capacity for high-level feature learning to address this cross-database problem.

The key contributions of this study are as follows:

- (1) We propose a comprehensive recognition framework for emotional EEG data, covering preprocessing, feature extraction, classification, and cross-database evaluation.
- (2) For feature engineering, we introduce a time-frequency-spatial extraction method that integrates TFS, CNN, and BiLSTM to capture multi-dimensional, informative features.
- (3) We adopt an unsupervised alignment strategy to map EEG data from different databases into a common space. Considering the topological configuration of EEG electrodes, TFS features are transformed into three-dimensional tensors to preserve spatial relationships.

The remainder of this paper is organized as follows: Section 2 describes the emotion datasets, preprocessing steps, and experimental setup. Section 3 reports results on the emotion databases and their application to depression recognition. Section 4 discusses the methodology and findings, and Section 5 concludes the paper.

Materials and Methods

Animal study

As illustrated in **Figure 1**, this section primarily describes the emotion databases and the methodologies for preprocessing and feature engineering. The processes include filtering, downsampling, Euclidean alignment (EA), short-time Fourier transform (STFT), Hilbert–Huang transform (HHT), transformation of one-dimensional sequences into three-dimensional tensors, and a spatiotemporal feature extraction framework integrating convolutional neural networks (CNN) with BiLSTM.

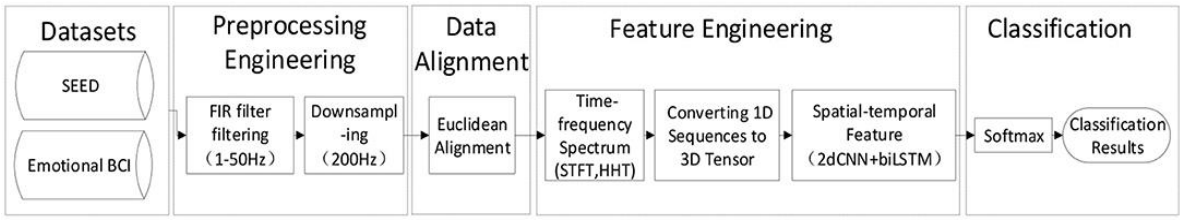


Figure 1. Framework for cross-database emotion recognition based on EEG signals. The training data is derived from the Emotional BCI Competition dataset of the 2020 WORLD ROBOT COMPETITION–BCI Control Brain Robot Contest.

Emotion database

This study utilizes the SEED dataset [43], which contains EEG recordings from 15 participants across three sessions each. Video clips were used to evoke three emotional states: positive, neutral, and negative. The recordings were downsampled to 200 Hz and filtered with a bandpass of 0–75 Hz. Segments were created at 1-second intervals, producing 152,730 samples in total.

The additional dataset is the training portion of the Emotional BCI Competition Database from the 2020 World Robot Competition—BCI-Controlled Brain Robot Contest. It involves 23 participants, with two sessions each (A and B), and video-induced emotions labeled as happy, sad, and neutral. Acquired at a 100 Hz sampling rate, the EEG data were divided into 1-second segments, generating 156,520 samples overall.

Table 1 provides a comparison of the two datasets, highlighting variations in emotional labels, participant numbers, session counts, and total samples. For consistency in later steps, the happy, sad, and neutral labels in the Emotional BCI Competition Database were aligned with positive, negative, and neutral emotions, respectively.

Table 1. Details of the two experimental databases.

Dataset	SEED	Emotional BCI Competition Database
Emotion Categories	Positive, neutral, negative	Happy, sad, neutral
Number of Channels	62	62
Number of Subjects	15	23
Number of Sessions per Subject	3	2
Samples in Positive/Happy Class	49,680	49,110
Samples in Neutral Class	52,650	50,722
Samples in Negative/Sad Class	50,400	56,694
Total Samples	152,730	156,526

Preprocessing engineering

EEG signals captured from the scalp are often affected by artifacts such as eye movements, blinking, or other muscular activity, which can obscure the underlying neural signals. To address this, preprocessing is required to enhance the quality of the recorded data. This involves organizing or converting the raw EEG, removing corrupted or incomplete segments, and segmenting the continuous recordings into meaningful units while preserving clean data. Band-pass filtering is typically applied to suppress noise from multiple sources, with finite impulse response (FIR) filters being preferred due to their ability to avoid phase distortion [44].

Euclidean alignment (EA) is then used to reduce variability across subjects. By projecting each user's EEG data into a transformed space, EA minimizes differences in the second-order statistics of the mean covariance matrices among users. This procedure is applied to both auxiliary and new users. As a result, the aligned data from different users share a similar average covariance structure, enabling models trained on one set of users to generalize more effectively to others.

Data alignment

EA provides a simple, fully unsupervised approach for harmonizing EEG data across individuals or domains [35]. The method begins by computing the mean of the spatial covariance matrices across all subjects.

$$\bar{R} = \frac{1}{N}\sum_{n=1}^{N} X_n (X_n)^T \tag{1}$$

Then the alignment is carried out by

$$\tilde{X}_n = \bar{R}^{-\frac{1}{2}} X_n \tag{2}$$

Here, $Xn \in R^{c \times t}$ represents the *n*-th EEG trial, where *c* denotes the number of channels and *t* the number of time samples. After alignment, the EEG trials are whitened, ensuring that each subject's mean spatial covariance matrix becomes the identity matrix [45]. This procedure reduces variability across subjects, producing a more consistent EEG distribution that is essential for reliable cross-database recognition.

Time-frequency spectrum

EEG signals are inherently non-linear and non-stationary, meaning their statistical characteristics, such as spectral density, vary significantly over time. Traditional spectral analysis cannot capture these time-varying components nor provide precise time-frequency localization. Time-frequency analysis, however, allows the examination of how the spectral content of a non-stationary EEG signal evolves over time, producing a joint time-frequency distribution (TFD) of signal power [46]. In this study, two techniques—short-time Fourier transform (STFT) [47–49] and Hilbert–Huang transform (HHT) [50, 51]—are employed for time-frequency spectrum (TFS) analysis, following the approach described by Song *et al*. [25].

For STFT, the TFS is computed as:

$$TFS_{STFT}(t,f) = \left| \int_{-\infty}^{+\infty} \omega(\tau - t) \times (\tau) e^{-j2\pi f\tau}\, d\tau \right|^2 \quad (3)$$

where x(t) is the EEG time series and w(τ–t) represents the short-time analysis window.

The Hilbert–Huang spectrum is derived using HHT, which involves three main steps. First, empirical mode decomposition (EMD) is applied to decompose the signal into a set of intrinsic mode functions (IMFs):

$$x(t) = \sum_{i=1}^{K} IMF_i(t) + r_K(t) \quad (4)$$

Here, rK(t) is the residual signal, which is either constant or monotonic, and each IMF satisfies specific conditions. Next, the Hilbert transform is applied to each IMF to obtain its corresponding Hilbert spectrum, representing the IMF in the time-frequency domain. An analytic signal for each IMF, reconstructed from a conjugate pair (IMF_k,IMF_k*) is expressed as:

$$Z_k = IMF_k(t) + jIMF_k^* = A_k(t)e^{j\theta_k(t)} \quad (5)$$

where $A_k(t)$ denotes the instantaneous amplitude $Z_k(t)$ and $\theta_k(t)$ the instantaneous phase of $IMF_k(t)$. Finally, the Hilbert spectra of all IMFs are combined to obtain the spectrum of the original signal. The original time series x(t) can then be reconstructed as:

$$x(t) = \sum_{i=1}^{K} A_k(t) e^{j2\pi \int f_k(t)\mathrm{d}t} \quad (6)$$

and the instantaneous frequency is calculated as:

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i}{dt} \quad (7)$$

Thus, the squared amplitude $A^2{}_k\ (t)$ together with the instantaneous frequency $f_k(t)$ forms the time-frequency spectrum of the EEG signal.

T transforming 1d feature sequences into 3d tensors

EEG signals are notoriously difficult to interpret due to their high noise levels and the subtle, often non-obvious relationships between the signals and specific brain activities. Many existing studies simplify EEG data as linear sequences, neglecting the complex interdependencies between neighboring channels or the benefits of representing EEG in alternative formats, such as image-like structures [52].

Considering the natural topological arrangement of EEG electrodes, illustrated in **Figure 2**, the one-dimensional sequence $St = [s1f, ..., scf, ..., sCf]$—where scf denotes the TFS feature extracted from the c-th electrode at frequency f—can be reorganized into a three-dimensional tensor $Tn \in \mathbb{R}H \times W \times F$. In this representation, H and W define the spatial dimensions (height and width) corresponding to the electrode layout, while F represents the number of features per channel in the nth EEG trial. The mapping from the 1D feature vector to the 3D tensor (H, W, F) captures both the spatial structure of the electrodes and the extracted feature information.

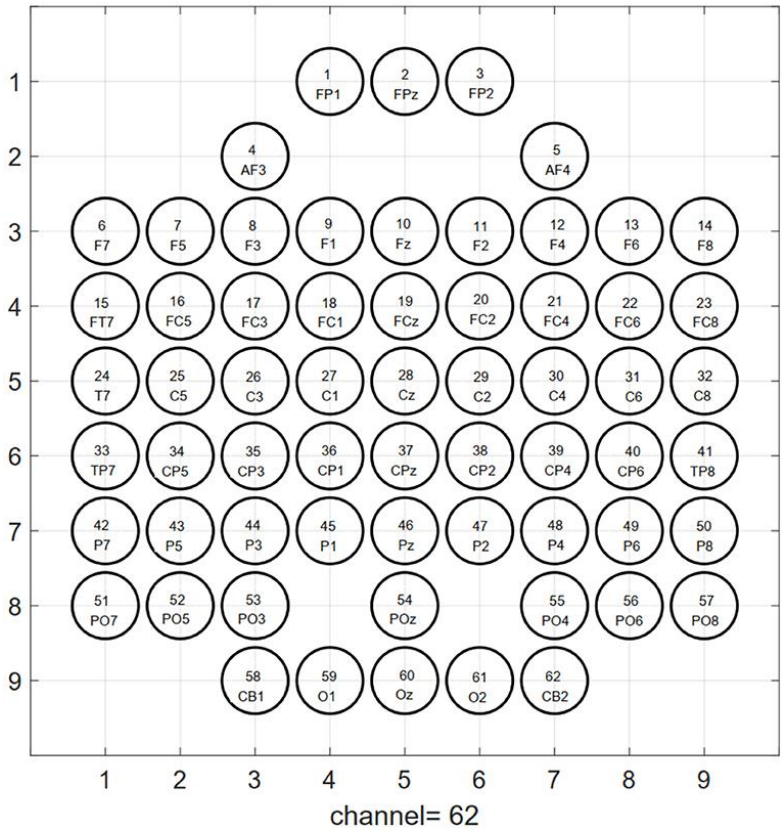


Figure 2. Topological arrangement of 64 EEG electrodes represented on a 2D plane. Each circle corresponds to an electrode, with its serial number and label shown inside. The left and right mastoid electrodes (M1 and M2) serve as reference electrodes during signal acquisition and are therefore excluded from the input used for emotion recognition.

$$T_n(H,W,f) = \begin{bmatrix} 0 & 0 & 0 & s_f^1 & s_f^2 & s_f^3 & 0 & 0 & 0 \\ 0 & 0 & s_f^4 & 0 & 0 & 0 & s_f^5 & 0 & 0 \\ s_f^6 & s_f^7 & s_f^8 & s_f^9 & s_f^{10} & s_f^{11} & s_f^{12} & s_f^{13} & s_f^{14} \\ s_f^{15} & s_f^{16} & s_f^{17} & s_f^{18} & s_f^{19} & s_f^{20} & s_f^{21} & s_f^{22} & s_f^{23} \\ s_f^{24} & s_f^{25} & s_f^{26} & s_f^{27} & s_f^{28} & s_f^{29} & s_f^{30} & s_f^{31} & s_f^{32} \\ s_f^{33} & s_f^{34} & s_f^{35} & s_f^{36} & s_f^{37} & s_f^{38} & s_f^{39} & s_f^{40} & s_f^{41} \\ s_f^{42} & s_f^{43} & s_f^{44} & s_f^{45} & s_f^{46} & s_f^{47} & s_f^{48} & s_f^{49} & s_f^{50} \\ s_f^{51} & s_f^{52} & s_f^{53} & 0 & s_f^{54} & 0 & s_f^{55} & s_f^{56} & s_f^{57} \\ 0 & 0 & s_f^{58} & s_f^{59} & s_f^{60} & s_f^{61} & s_f^{62} & 0 & 0 \end{bmatrix} \quad (8)$$

BiLSTM+2dCNN

We proposed a cascaded deep convolutional recurrent neural network (CDCRNN) framework, illustrated in **Figure 3**, to effectively capture the spatiotemporal characteristics of EEG signals. The model takes as input the 3D tensor Tn, which encodes both spatial and temporal information. Spatial features are first extracted using a 2D CNN, followed by a BiLSTM layer to model temporal dependencies. The output from the final time step of the BiLSTM is passed through a fully connected layer, and a softmax layer is applied at the end to generate the final emotion prediction.

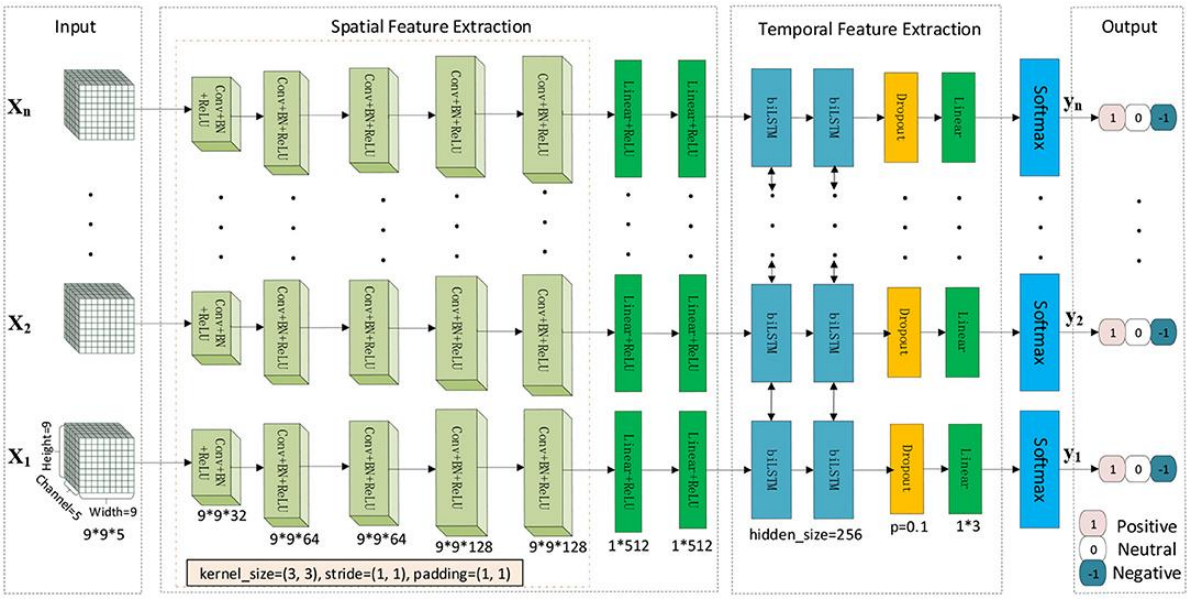


Figure 3. Cascaded 2D CNN + BiLSTM architecture.

In this study, a 2D CNN combined with BiLSTM was designed to capture rich spatiotemporal representations from multi-channel EEG data. Each EEG segment, lasting 1 second, is treated as an individual sample. Time-frequency features were extracted for each segment using STFT and HHT, and these features were then fed into the deep network for further representation learning. Each 1-second sample, denoted as X_i (i = 1, 2, …, n), is treated as a spatial image with five channels. Five convolutional layers with 3×3 kernels and ReLU activations were used to learn local non-linear spatial patterns, followed by fully connected layers to capture global spatial features. Since spatial information alone is often insufficient for representing temporal signals discriminatively, a BiLSTM layer was employed to model temporal dependencies and enhance the representation.

Experimental settings

The raw EEG signals were first processed using a 50th-order FIR band-pass filter in the 1–50 Hz range. For consistency with the SEED dataset, the Emotional BCI data were downsampled to 200 Hz, followed by EA (artifact elimination). The relative energy of five frequency bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma (30–50 Hz)—was extracted for each electrode channel using both STFT and HHS methods. This yielded 310 features per sample (5 frequency bands × 62 electrodes), which were then reshaped into a 3D tensor of size 9×9×5.

The resulting 3D tensors were input to the spatiotemporal network for training. The network was trained with a batch size of 32, a frame length of 12 seconds, and 100 epochs. Cross-entropy was used as the loss function, while the SGD optimizer was applied with an initial learning rate of 0.005. The learning rate was updated according to: $lr = init\_lr*(0.95^{epoch//10})$ where `init_lr` is the initial learning rate.

Results and Discussion

Emotion recognition performance

To evaluate the performance of the proposed emotion recognition framework, three experimental protocols were designed. In all three protocols, the training and test sets are strictly non-overlapping, and the test labels are never used during training. In the first two protocols, training and test sets are drawn from different databases, while the third protocol uses a leave-one-subject-out approach. Considering class imbalance, both accuracy and weighted F1-score were reported. The model was also applied to depression recognition using five-fold cross-validation.

In the first protocol, the entire Emotional BCI competition database was used for training, and the SEED dataset was used for testing. Recognition results for two types of manually extracted features are summarized in **Table 2**, showing that features extracted using STFT achieved better performance. To further

illustrate class-wise predictions, confusion matrices were analyzed. **Figure 4** presents confusion matrices for STFT and HHS features, showing that neutral emotion consistently achieved the highest recognition rate among the three emotion classes. Overall, recognition rates with STFT features were higher than those with HHS features.

Table 2. Recognition results on the emotional database.

Experiment	Source Dataset → Target Dataset	Training Performance		Validation Performance	
		Accuracy (%)	F1 Score	Accuracy (%)	F1 Score
1	Emotional BCI Competition Database → SEED	83.56	0.84	83.60	0.84
2	SEED → Emotional BCI Competition Database	74.33	0.72	70.26	0.70
3	Leave-One-Subject-Out	81.58	0.80	79.29	0.77

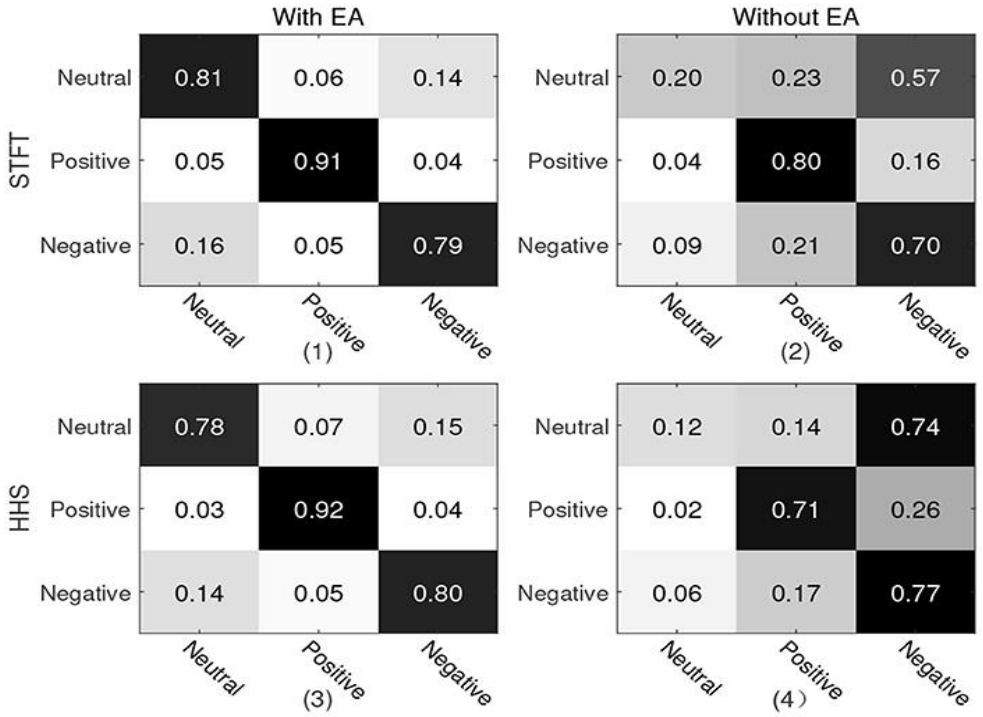


Figure 4. Confusion matrix for Protocol 1. The y-axis indicates the actual labels, and the x-axis shows the predicted labels. Panels (1) and (2) correspond to results using STFT features, while panels (3) and (4) correspond to HHS features. Panels (1) and (3) include the Euclidean Alignment (EA) module, whereas panels (2) and (4) present results without EA.

For Protocol 2, the SEED dataset served as the training set, and the Emotional BCI Competition Database was used for testing. Recognition performance for the two manually extracted feature types is reported in **Table 2**. STFT features outperformed HHS features by 4.07%, although the overall accuracy was about 9% lower than in Protocol 1. Confusion matrices for STFT and HHS features are shown in **Figure 5**. STFT features produced relatively balanced recognition across all three emotion categories, while with HHS features, positive emotion was recognized considerably more accurately than neutral or negative emotions.

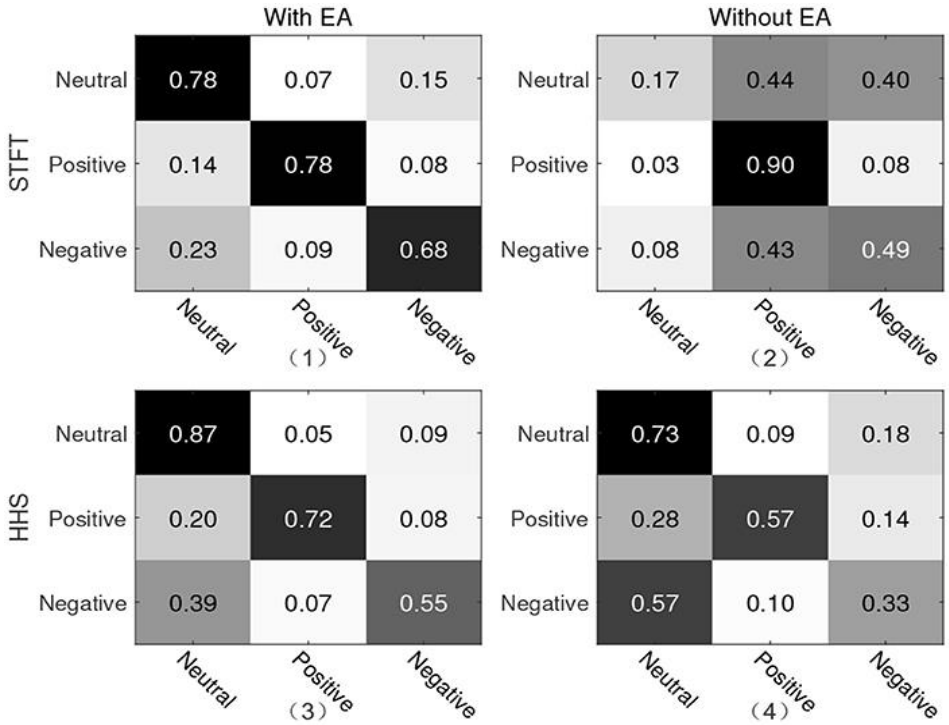


Figure 5. Confusion matrix for Protocol 2. The y-axis represents the actual labels, and the x-axis shows the predicted labels. Panels (1) and (2) depict recognition results using STFT features, while panels (3) and (4) correspond to HHS features. Panels (1) and (3) include the Euclidean Alignment (EA) module, whereas panels (2) and (4) show results without EA.

In the third protocol, recognition outcomes for both databases are summarized in **Table 2**, reporting accuracy and weighted F1-score. The results are organized by database, averaging the performance of all subjects within each dataset, and the overall mean accuracy across all subjects is also calculated. STFT features consistently yielded slightly higher recognition performance compared with HHS features.

To assess the contribution of the EA module, an ablation study was conducted. As illustrated in **Figures 4 and 5, and Table 3**, including EA substantially improved recognition accuracy in both Protocols 1 and 2. Additionally, to investigate the temporal dependencies in EEG emotion frames, experiments were performed using frame lengths ranging from 8 to 32 with a step size of 4. These tests were conducted under Protocols 1 and 2, and the results are presented in **Figure 6**.

Table 3. Results of the ablation study on the EA module.

Domain Adaptation Direction	Feature Extraction Method	Emotion Alignment (EA)	F1 Score	Accuracy (%)
Emotional BCI Competition Database → SEED	STFT	With EA	0.84	83.56
		Without EA	0.54	57.29
	HHS	With EA	0.84	83.60
		Without EA	0.49	53.84
SEED → Emotional BCI Competition Database	STFT	With EA	0.72	74.33
		Without EA	0.48	52.40
	HHS	With EA	0.70	70.26
		Without EA	0.53	53.46

Bold values indicate identical experimental conditions and highlight the maximum performance when comparing results with or without the EA module.

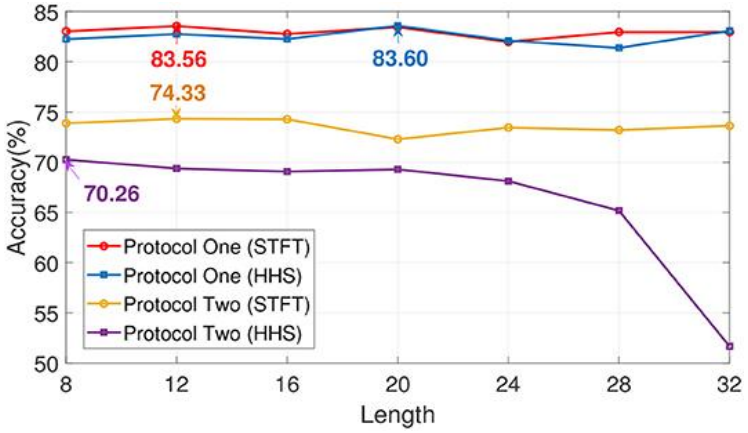


Figure 6. Influence of time window length on recognition performance, with time measured in seconds.

Depression recognition results

For depression detection, we employed the MODMA multi-modal dataset, which provides 128-channel ERP recordings from 53 participants—24 diagnosed with major depressive disorder and 29 healthy controls—ranging in age from 16 to 52 years [53–55]. Considering the electrode layout shown in **Figure 7**, the extracted 3D tensors have dimensions of 21×19×5.

Participants were divided into five folds for cross-validation. To account for unequal group sizes, the first three folds each included 5 depressed and 6 healthy subjects, while the final fold comprised 4 depressed and 5 healthy subjects. Recognition outcomes for each fold are summarized in **Figure 8.**

The ERP task was a dot-probe paradigm, in which cue stimuli consisted of three types of emotional-neutral face pairs: Happy-Neutral ("hcue"), Fear-Neutral ("fcue"), and Sad-Neutral ("scue"). In addition to evaluating overall accuracy, we analyzed recognition performance for each specific cue type. Among them, the highest recognition rate was achieved for the "hcue" condition, reaching 71.14%.

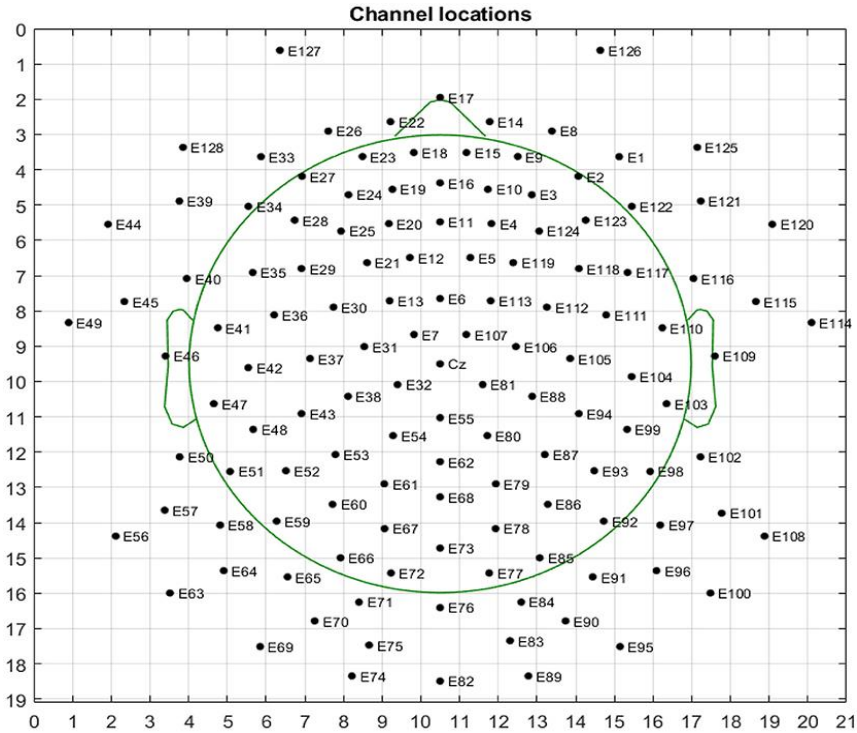
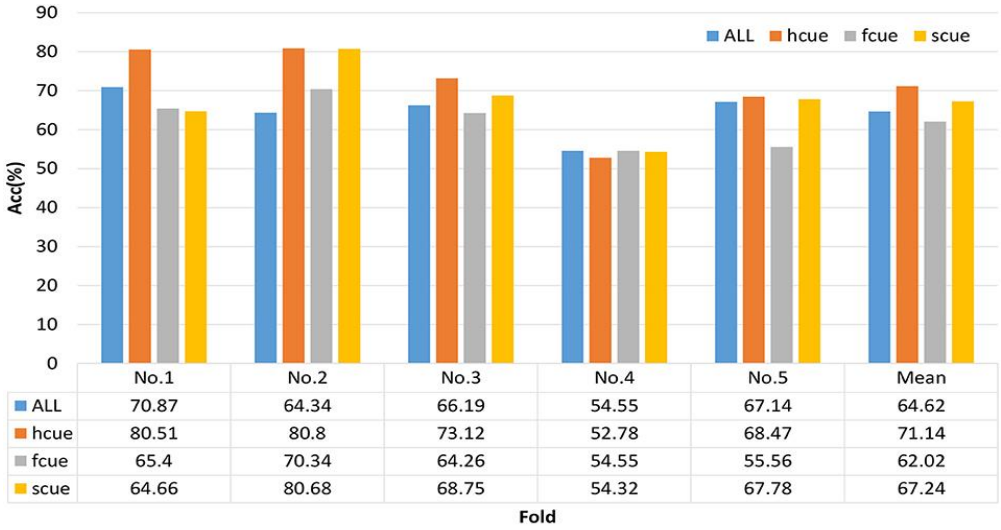


Figure 7. Two-dimensional representation of the 128-electrode EEG layout. Each circle corresponds to an electrode, with its serial number and label shown inside.



	No.1	No.2	No.3	No.4	No.5	Mean
ALL	70.87	64.34	66.19	54.55	67.14	64.62
hcue	80.51	80.8	73.12	52.78	68.47	71.14
fcue	65.4	70.34	64.26	54.55	55.56	62.02
scue	64.66	80.68	68.75	54.32	67.78	67.24

Figure 8. Illustrates the classification performance of the spatiotemporal neural network applied to the depression dataset.

This study presents a comprehensive pipeline for EEG-based emotion recognition, achieving an overall accuracy exceeding 80%. In the preprocessing stage, an unsupervised Euclidean Alignment (EA) method is applied to map data from different databases into a common feature space. Time-frequency features are then extracted using STFT, followed by spatial feature learning via CNN and temporal modeling using BiLSTM. Prior to the CNN operations, the one-dimensional EEG time series is transformed into a three-dimensional tensor based on the spatial arrangement of the electrodes, ensuring that inter-channel correlations are fully preserved. The 2D CNN captures spatial patterns, while the BiLSTM models the temporal relationships across the extracted features.

Analysis of Protocol 1's confusion matrix indicates that the recognition rates for the three emotion categories are relatively balanced under both STFT and HHS feature extraction methods, with positive emotion achieving the highest accuracy. In Protocol 2, neutral emotion shows the highest recognition rate using HHS features, whereas negative emotion is recognized least accurately. Overall, there is approximately a 9% drop in accuracy when comparing Protocols 1 and 2, likely due to the greater diversity of samples in the Emotional BCI database (as seen in **Table 1**, where the first 15 participants and the last eight participants differ across the three emotion categories).

Across all protocols, the TFS features extracted via STFT consistently outperform those obtained with HHS in terms of both accuracy and weighted F1-score. **Figures 9 and 10** visualize the features extracted by STFT and HHS, respectively. The three emotion categories are displayed across five frequency bands, with each spectrum arranged according to the electrode layout shown in **Figure 2**. Notably, the two methods yield different high-frequency patterns: for STFT, the relative energy in the beta (14–30 Hz) and gamma (30–50 Hz) bands is high, showing clear distinctions among the three emotion categories. In contrast, HHS features show relatively high relative energy for positive and neutral emotions in these bands, while negative emotions consistently exhibit low energy across all frequencies. This explains why recognition performance with HHS is generally lower than STFT, although HHS still performs well for positive and negative emotion classification.

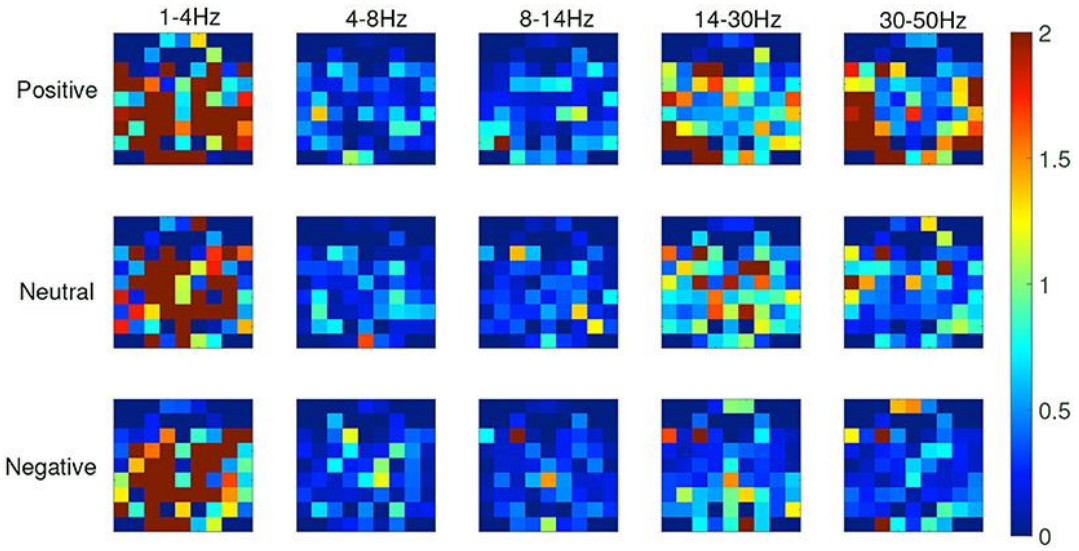


Figure 9. Relative energy maps of time-frequency spectrum (TFS) features derived using the short-time Fourier transform (STFT), arranged according to the electrode layout shown in **Figure 2**.

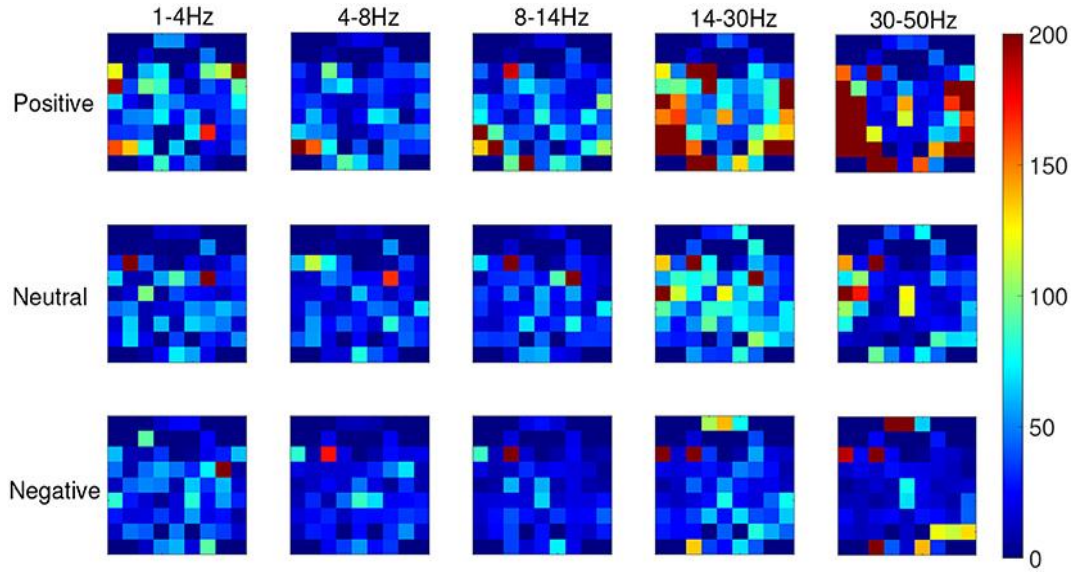


Figure 10. Displays the relative energy distribution map in the time-frequency spectrum (TFS), computed using the HHS algorithm, and arranged according to the electrode layout presented in **Figure 2.**

To additionally verify the superiority of the proposed approach, we benchmarked it against leading contemporary techniques. **Table 4** summarizes performance metrics of various subject-independent EEG-based emotion recognition models tested on the SEED dataset. These include Support vector machine with linear kernel (SVM) [56], principal component analysis using kernel method (KPCA) [57], component analysis for transfer learning (TCA) [58], parameter transfer in transductive setting (TPT) [59], adversarial neural network for domain adaptation (DANN) [60], graph convolutional neural network with dynamical structure (DGCNN) [39], domain adversarial neural network across bi-hemispheres (BiDANN) [61], BiDANN with subject adaptation (BiDANN-S) [41], spatial-temporal neural network with hierarchical design (R2G-STNN) [62], and graph that adapts per instance (IAG) [63]. The results in the table indicate that our approach delivers the top mean accuracy while exhibiting the lowest standard deviation. Distinct from the

compared baselines, our model was trained with additional data from the Emotional BCI database. This augmented training corpus improves the model's ability to generalize, highlighting the strength of our method in capturing diverse spatiotemporal features and achieving reliable emotion recognition across different datasets or participants.

Table 4. Average accuracies (Acc) and standard deviations (Std) achieved on the SEED dataset in subject-independent EEG emotion recognition tasks.

Approach	Mean Accuracy / Standard Deviation (%)
SVM [56]	56.73 / 16.29
KPCA [57]	61.28 / 14.62
TCA [58]	63.64 / 14.88
TPT [59]	76.31 / 15.89
DANN [60]	75.08 / 11.18
DGCNN [39]	79.95 / 09.02
BiDANN [61]	83.28 / 09.60
BiDANN-S [41]	84.14 / 06.87
R2G-STNN [62]	84.16 / 07.63
IAG [63]	86.30 / 06.91
Proposed method	86.42 / 05.26

Conclusion

This research developed a comprehensive pipeline encompassing preprocessing through to emotion classification using EEG signals, attaining an accuracy exceeding 80%. Notably, the application of this model to EEG-based depression detection holds substantial importance. The preprocessing stage, integrated with an unsupervised Euclidean Alignment (EA) technique, effectively aligns data from diverse databases into a unified space. Multi-domain features spanning time, frequency, and spatial dimensions are extracted by integrating three approaches: Short-Time Fourier Transform (STFT), Convolutional Neural Network (CNN), and Bidirectional Long Short-Term Memory (BiLSTM). Prior to the CNN processing, the one-dimensional time-series signals are reshaped into a three-dimensional tensor based on the topographic layout of the EEG electrodes. Looking ahead, future work will focus on developing end-to-end, real-time emotional brain-computer interfaces tailored for depression identification.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Daros AR, Haefner SA, Asadi S, Kazi S, Rodak T, Quilty LC. A meta-analysis of emotional regulation outcomes in psychological interventions for youth with depression and anxiety. Nat Hum Behav. (2021) 5:1443–57. 10.1038/s41562-021-01191-9
2. Vanderlind WM, Millgram Y, Baskin-Sommers AR, Clark MS, Joormann J. Understanding positive emotion deficits in depression: from emotion preferences to emotion regulation. Clin Psychol Rev. (2020) 76:101826. 10.1016/j.cpr.2020.101826
3. Dolan RJ. Emotion, cognition, and behavior. Science. (2002) 298:1191–4. 10.1126/science.1076358
4. Balconi M, Lucchiari C. EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. Neurosci Lett. (2006) 392:118–23. 10.1016/j.neulet.2005.09.004
5. Bekkedal MY, Rossi III J, Panksepp J. Human brain EEG indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. Neurosci Biobehav Rev. (2011) 35:1959–70. 10.1016/j.neubiorev.2011.05.001
6. Davidson PR, Jones RD, Peiris MT. EEG-based lapse detection with high temporal resolution. IEEE Trans Biomed Eng. (2007) 54:832–9. 10.1109/TBME.2007.893452
7. Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Auton Ment Dev. (2015) 7:162–75. 10.1109/TAMD.2015.2431497
8. Gross JJ, Mu noz RF. Emotion regulation and mental health. Clin Psychol (New York). (1995) 2:151–64. 10.1111/j.1468-2850.1995.tb00036.x
9. Jordan KG. Continuous EEG and evoked potential monitoring in the neuroscience intensive care unit. J Clin Neurophysiol. (1993) 10:445–75. 10.1097/00004691-199310000-00006
10. Chanel G, Kierkels JJ, Soleymani M, Pun T. Short-term emotion assessment in a recall paradigm. Int J

Hum Comput Stud. (2009) 67:607–27. 10.1016/j.ijhcs.2009.03.005
11. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. Nature. (2005) 433:68–72. 10.1038/nature03086
12. Blair R, Morris JS, Frith CD, Perrett DI, Dolan RJ. Dissociable neural responses to facial expressions of sadness and anger. Brain. (1999) 122:883–93. 10.1093/brain/122.5.883
13. Adolphs R, Tranel D. Impaired judgments of sadness but not happiness following bilateral amygdala damage. J Cogn Neurosci. (2004) 16:453–62. 10.1162/089892904322926782
14. Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron. (2003) 40:655–64. 10.1016/S0896-6273(03)00679-2
15. Balconi M, Mazza G. Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues.: ERS/ERD and coherence measures of alpha band. Int J Psychophysiol. (2009) 74:158–65. 10.1016/j.ijpsycho.2009.08.006
16. Liu Y, Sourina O, Nguyen MK. Real-time EEG-based emotion recognition and its applications. In: Transactions on Computational Science XII. Berlin; Heidelberg: Springer; (2011). p. 256–77.
17. Jatupaiboon N, Pan-ngum S, Israsena P. Emotion classification using minimal EEG channels and frequency bands. In: The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE). Khon Kaen: IEEE (2013). p. 21–4.
18. Li M, Lu BL. Emotion classification based on gamma-band EEG. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Minneapolis, MN: IEEE; (2009). p. 1223–6.
19. Park KS, Choi H, Lee KJ, Lee JY, An KO, Kim EJ. Emotion recognition based on the asymmetric left and right activation. Int J Medical Sci. (2011) 3:201–209. 10.1016/j.neuroimage.2016.05.059
20. Hjorth B. EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol. (1970) 29:306–10. 10.1016/0013-4694(70)90143-4
21. Petrantonakis PC, Hadjileontiadis LJ. Emotion recognition from EEG using higher order crossings. IEEE Trans Learn Technol. (2009) 14:186–97. 10.1109/TITB.2009.2034649
22. Khosrowabadi R, Quek HC, Wahab A, Ang KK. EEG-based emotion recognition using self-organizing map for boundary detection. In: 2010 20th International Conference on Pattern Recognition. Istanbul: IEEE; (2010) p. 4242–5.
23. Bos DO. EEG-based emotion recognition. Influence Visual Auditory Stimuli. (2006) 56:1–17.
24. Musha T, Terasaki Y, Haque HA, Ivamitsky GA. Feature extraction from EEGs associated with emotions. Artif Life Rob. (1997) 1:15–9. 10.1007/BF02471106
25. Song T, Zheng W, Lu C, Zong Y, Zhang X, Cui Z. MPED: a multi-modal physiological emotion database for discrete emotion recognition. IEEE Access. (2019) 7:12177–91. 10.1109/ACCESS.2019.2891579
26. Hatamikia S, Maghooli K, Nasrabadi AM. The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. J Med Signals Sens. (2014) 4:194. 10.4103/2228-7477.137777
27. Huang D, Guan C, Ang KK, Zhang H, Pan Y. Asymmetric spatial pattern for EEG-based emotion detection. In: The 2012 International Joint Conference on Neural Networks (IJCNN). Brisbane, QLD: IEEE; (2012). p. 1–7.
28. Jie X, Cao R, Li L. Emotion recognition based on the sample entropy of EEG. Biomed Mater Eng. (2014) 24:1185–92. 10.3233/BME-130919
29. Hosseini SA, Naghibi-Sistani MB. Emotion recognition method using entropy analysis of EEG signals. Int J Image Graph Signal Process. (2011) 3:30. 10.5815/ijigsp.2011.05.05
30. Wang XW, Nie D, Lu BL. EEG-based emotion recognition using frequency domain features and support vector machines. In: International Conference on Neural Information Processing. Berlin; Heidelberg: Springer; (2011). p. 734–43.
31. Pan J, Li Y, Wang J. An EEG-based brain-computer interface for emotion recognition. In: 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver, BC: IEEE. (2016). p. 2063–7.
32. Koelstra S, Patras I. Fusion of facial expressions and EEG for implicit affective tagging. Image Vis Comput. (2013) 31:164–74. 10.1016/j.imavis.2012.10.002

33. Thammasan N, Fukui Ki, Numao M. Application of deep belief networks in eeg-based dynamic music-emotion recognition. In: 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver, BC: IEEE; (2016). p. 881–8.
34. Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. IEEE Trans Affect Comput. (2017) 10:374–93. 10.1109/TAFFC.2017.2714671
35. Wu D, Peng R, Huang J, Zeng Z. Transfer learning for brain-computer interfaces: a complete pipeline. arXiv preprint arXiv:200703746. (2020).31443868
36. Wu D, Xu Y, Lu BL. Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016. IEEE Trans Cogn Dev Syst. (2020) 1. 10.1109/TCDS.2020.3007453
37. He H, Wu D. Transfer learning for Brain-Computer interfaces: A Euclidean space data alignment approach. IEEE Trans Biomed Eng. (2019) 67:399–410. 10.1109/TBME.2019.2913914
38. Zhang T, Zheng W, Cui Z, Zong Y, Li Y. Spatial-temporal recurrent neural network for emotion recognition. IEEE Trans Cybern. (2018) 49:839–47. 10.1109/TCYB.2017.2788081
39. Song T, Zheng W, Song P, Cui Z. EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Trans Affect Comput. (2018) 11: 532–41. 10.1109/TAFFC.2018.2817622
40. Li J, Qiu S, Du C, Wang Y, He H. Domain adaptation for EEG emotion recognition based on latent representation similarity. IEEE Trans Cogn Develop Syst. (2019) 12:344–353. 10.1109/TCDS.2019.2949306
41. Li Y, Zheng W, Zong Y, Cui Z, Zhang T, Zhou X. A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. IEEE Trans Affect Comput. (2018) 12: 494–504. 10.1109/TAFFC.2018.2885474
42. Lan Z, Sourina O, Wang L, Scherer R, Müller-Putz GR. Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. IEEE Trans Cogn Develop Syst. (2018) 11:85–94. 10.1109/TCDS.2018.2826840
43. Duan RN, Zhu JY, Lu BL. Differential entropy feature for EEG-based emotion classification. In: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). San Diego, CA: IEEE; (2013). p. 81–4.
44. Teplan M. Fundamentals of EEG measurement. Meas Sci Rev. (2002) 2:1–11. Available online at: http://www.edumed.org.br/cursos/neurociencia/MethodsEEGMeasurement.pdf
45. Zhang W, Wu D. Manifold Embedded Knowledge Transfer for Brain-Computer Interfaces. IEEE Trans Neural Syst Rehabilitation Eng. (2020) 28:1117–27. 10.1109/TNSRE.2020.2985996
46. Boashash B. Time-Frequency Signal Analysis and Processing: A Comprehensive Reference. Cambridge, MA: Academic Press; (2015).
47. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. (2004) 134:9–21. 10.1016/j.jneumeth.2003.10.009
48. Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, et al. EEG-based emotion recognition in music listening. IEEE Trans Biomed Eng. (2010) 57:1798–806. 10.1109/TBME.2010.2048568
49. Sammler D, Grigutsch M, Fritz T, Koelsch S. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. Psychophysiology. (2007) 44:293–304. 10.1111/j.1469-8986.2007.00497.x
50. Huang NE. Hilbert-Huang transform and Its Applications, Vol. 16. Singapore: World Scientific; (2014).
51. Huang NE, Wu Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies. Rev Geophys. (2008) 46:1–23. 10.1029/2007RG000228
52. Zhang D, Yao L, Zhang X, Wang S, Chen W, Boots R, et al. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface. In: AAAI, New Orleans. (2018). p. 1703–10.31021810
53. Cai H, Gao Y, Sun S, Li N, Tian F, Xiao H, et al. MODMA dataset: a multi-modal open dataset for mental-disorder analysis. arXiv preprint arXiv:200209283. (2020).34512301
54. Hu B, Rao J, Li X, Cao T, Li J, Majoe D, et al. Emotion regulating attentional control abnormalities in major depressive disorder: an event-related potential study. Sci Rep. (2017) 7:1–21. 10.1038/s41598-017-13626-
55. Li X, Li J, Hu B, Zhu J, Zhang X, Wei L, et al. Attentional bias in MDD: ERP components analysis

and classification using a dot-probe task. Comput Methods Programs Biomed. (2018) 164:169–79. 10.1016/j.cmpb.2018.07.003

56. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett. (1999) 9:293–300. 10.1023/A:1018628609742119729 10
57. Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. (1998) 10:1299–319. 10.1162/089976698300017467118606 87
58. Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. IEEE Trans Neural Netw. (2010) 22:199–210. 10.1109/TNN.2010.2091281
59. Sangineto E, Zen G, Ricci E, Sebe N. We are not all equal: personalizing models for facial expression analysis with transductive parameter transfer. In: Proceedings of the 22nd ACM International Conference on Multimedia, Lisboa. (2014). p. 357–66. 10.1145/2647868.2654916
60. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. J Mach Learn Res. (2016) 17:1–35. Available online at: https://www.jmlr.org/papers/volume17/15-239/15-239.pdf
61. Li Y, Zheng W, Cui Z, Zhang T, Zong Y. A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm. (2018). p. 1561–7.
62. Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. IEEE Trans Affect Comput. (2019) 1. 10.1109/TAFFC.2019.2922912
63. Song T, Liu S, Zheng W, Zong Y, Cui Z. Instance-adaptive graph for EEG emotion recognition. In: AAAI. (2020). p. 2701–8.