STM JOURNALS

International Journal of
Optical Innovations & Research

Volume 3, Issue 2, 2025
July–December
DOI (Journal): 10.37591/IJOIR

https://journals.stmjournals.com/ijoir

Review IJOIR

Emotion Recognition from Electroencephalogram Signal and Eye Movement Based on Deep Learning

Rajendra Khanal[1], Raj Kumar Paneru[1], Susheel Thapa[1], Surendra Shrestha[2,*]

Abstract

Emotion recognition from electroencephalogram (EEG) signals has gained significant attention due to its potential in human–computer interaction (HCI), mental health monitoring, and personalized content delivery. This paper presents the use of convolutional neural networks (CNNs) to classify emotions such as happiness, sadness, fear, neutral, and disgust by leveraging a fusion of EEG signals and eye movements data. Compared to conventional methods of emotion detection, such as those that rely on body language, voice tone, or facial expressions, EEG-based emotion recognition offers a distinct advantage. EEG signals directly record neuronal oscillations linked to affective and cognitive processes in the brain, even if these external cues can be intentionally altered or hidden. Because of this, EEG is now a more objective and trustworthy method of detecting minute emotional changes that may not be readily apparent. Effective feature extraction and pattern detection are still difficult jobs, nevertheless, because of the high complexity and noise present in EEG data. The SEED-V dataset, a reputable benchmark database that contains synchronized EEG and eye tracking recordings from several participants exposed to emotionally charged video clips, was used for the research. Standard filtering techniques were used to preprocess each participant's data to eliminate artifacts like eye blinks and muscle noise. After being cleaned, the signals were divided into temporal windows and input into the CNN model for classification and feature learning. Multiple convolutional and pooling layers were used in the CNN architecture to collect localized spatial data. Fully connected layers were then used for the final classification of emotions. The capacity to represent intricate spatial and temporal correlations within EEG signals has greatly increased thanks to recent developments in machine learning and deep neural networks, especially CNNs. Utilizing the SEED-V dataset, our CNN model achieved a training accuracy of 97.92%, a validation accuracy of 94.44%, and a test accuracy of 93%. The integration of eye movements features with EEG signals significantly enhanced the model's performance. Experimental results validate the effectiveness of CNNs in multi-modal emotion recognition, achieving an overall weighted average F1-score of 0.93. This work highlights the promise of combining EEG and eye tracking data for enhancing interactive systems and lays a foundation for future advancements in affective computing.

***Author for Correspondence**
Surendra Shrestha
E-mail: surendra.shrestha@nou.edu.np

[1]Student, Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, TU Lalitpur, Nepal
[2]Associate Proofesor Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, TU School of Science, Health & Technology, Nepal Open University, Lalitpur, Nepal

Received Date: November 05, 2025
Accepted Date: November 10, 2025
Published Date: December 31, 2025

Citation: Rajendra Khanal, Raj Kumar Paneru, Susheel Thapa, Surendra Shrestha. Emotion Recognition from Electroencephalogram Signal and Eye Movement Based on Deep Learning. International Journal of Optical Innovations & Research. 2025; 3(2): 8–15p.

Keywords: Electroencephalogram (EEG), SEED-V dataset, human–computer interactions (HCI), convolutional neural networks (CNNs), fear, disgust, precision

INTRODUCTION

Emotions play a critical role in human cognition and influence decision making, reasoning, and interpersonal interactions. Accurately recognizing emotions is essential for developing emotionally intelligent systems in human–computer interactions (HCI). Electroencephalogram (EEG) signals, which capture brain activity with high

temporal resolution, offer a noninvasive and objective method for emotion recognition, as they are difficult to fake or suppress. With advancements in deep learning, particularly convolutional neural networks (CNNs), automated feature extraction from EEG signals has significantly improved classification accuracy. This study focused on developing a CNN-based model to classify emotions from EEG signals using SEED-V datasets. The model targets emotions such as happiness, sadness, fear, neutrality, and disgust, addressing challenges such as noise and inter-subject variability in EEG data. The objective was to design a robust emotion recognition system, validate its performance, and demonstrate its applicability in real-world scenarios, such as mental health monitoring and HCI.

RELATED WORK

Benchmark Datasets and Classical Versus Deep Learning

Emotion recognition from EEG signals has become a pivotal area of research in HCI, driven by the need to endow systems with "emotional intelligence." Early contributions in this field, such as the development of the dataset for emotion analysis using physiological signals (DEAP) data, provided essential groundwork for standardized evaluation. The DEAP study recorded EEG and other physiological signals from 32 participants while they watched 40 one-minute music video clips, with participants rating each video for arousal, valence, and liking. This multi-modal dataset not only facilitated the establishment of baseline classification results using participant-specific models but also highlighted the advantages of fusing EEG with peripheral and multimedia features to enhance emotion recognition accuracy [1].

Building on these databases, researchers have conducted comprehensive evaluations of both deep and shallow traditional learning techniques. In one review, deep learning models, including CNNs and recurrent neural networks (RNN), were found to achieve higher accuracy (often exceeding 80%) compared to shallow methods like support vector machine (SVM) and k-nearest neighbors (kNN), which typically achieved accuracy in the 60–80% range. The study underscored that deep architecture is more capable of automatically extracting nonlinear Electroencephalogram (EEG) Features, particularly when larger datasets such as Database for Emotion Analysis using Physiological Signals (DEEP) or SJTU Emotion EEG Dataset (SEED) are employed [2].

Advances in Deep Learning Architectures

In addition to the unique challenges posed by EEG signals, such as nonstationarity and hierarchical temporal structures, novel architecture has emerged. For instance, the Hierarchical Bidirectional Gated Recurrent Unit (GRU) (HBGRU) model with dual attention was designed to mirror the multilevel structure of EEG data. By processing short EEG segments with a bidirectional GRU and employing attention mechanisms both within individual segments and across longer epochs, the HBGRU model significantly outperformed traditional deep models like long short-term memory (LSTM) and shallow classifiers, with improvements of over 4–11% in classification accuracy [3].

In parallel, hybrid architectures that combine CNNs with advanced modules such as transformers have shown great promise. For example, the CIT-EmotionNet model integrates a CNN branch for local spatial and spectral feature extraction with a transformer branch to capture the global temporal dependencies. An interactive fusion mechanism enables these branches to exchange information effectively, leading to state-of-the-art performance on datasets such as SEED and SEED-IV, where accuracy as high as 98.57% has been reported [4].

Feng et al. [5] proposed an innovative approach that leveraged the inherent topological structure of EEG data. Their Spatial-Temporal Graph Convolutional LSTM (ST-GCLSTM) treats EEG electrodes as nodes within a graph to capture the dynamic time-varying connectivity between different brain regions. Coupled with bidirectional LSTM and attention mechanisms, this method achieved accuracy above 95% on multiple benchmark datasets by effectively modeling both spatial and temporal features.

In addition to single-modality analysis, multi-modal emotion recognition has also garnered significant attention. Deep generalized canonical correlation analysis with attention mechanism (DGCCA-AM) is one such approach that learns joint latent representations from diverse modalities (e.g., EEG, Electrocardiogram (ECG), and facial features) while using an attention-based fusion strategy to adaptively weigh the contributions of each modality. This multi-modal fusion strategy has demonstrated superior performance over single-modality approaches, achieving mean accuracy of approximately 82% on datasets like SEED-V [6].

Review Studies and Emerging Trends

Ma et al. [7] conducted a comprehensive review mapping the evolution of deep learning in EEG-based emotion recognition and classified the models into CNN-based, RNN-based, graph-based, and hybrid approaches. Their review emphasized trends such as the adoption of 3D convolution, graph neural networks to exploit spatial electrode information, and transformer-based models to handle sequential data. It also highlights practical challenges, including inter-subject variability and the need for transfer learning and data augmentation to enhance real-world applicability.

Attention-based hybrid models that combine CNNs with LSTM networks have been introduced to selectively focus on informative frequency bands and critical temporal segments within EEG data. Such models, using dual attention mechanisms, have demonstrated notable improvements in emotion classification performance on benchmark datasets, such as DEAP and SEED, thereby validating the importance of adaptive feature weighting in capturing the complex characteristics of EEG signals [8].

Finally, broader reviews of EEG signal processing provide a holistic view of the entire analytical pipeline, from signal acquisition and artifact removal (using techniques like independent component analysis and wavelet denoising) to feature extraction and classification. These reviews underscore the challenges associated with EEG's low signal-to-noise ratio and high inter-subject variability, while also noting emerging trends, such as the integration of deep learning methods and multi-modal processing, to overcome these issues [9].

Furthermore, deep learning-based multi-modal approaches, which combine EEG with other modalities, such as facial expressions and vocal intonation, consistently demonstrate enhanced performance over uni-modal systems by effectively capturing the complementary nature of different emotional cues [10].

The reviewed studies illustrate the rapid evolution and growing sophistication of deep learning methods in EEG-based emotion recognition. Collectively, the literature provides compelling evidence that leveraging advanced architecture ranging from hierarchical and attention-based networks to graph convolutional and multi-modal ModelScan significantly enhance the accuracy and robustness of emotion recognition systems. This body of work forms the backbone of the current project, which aims to further refine deep learning techniques for improved emotion classification from EEG signals.

The CNN method leverages both EEG signals and eye movement data to enhance emotion recognition. We hypothesized that combining multi-modal physiological signals can lead to improved emotional state classification, given the complementary nature of brain activity and eye movement dynamics during emotional stimuli.

The process begins with the fusion of EEG and eye movement features. The preprocessed EEG signals yielded 310 feature columns, whereas the extracted eye movement features resulted in 33 feature columns. A crucial aspect of this approach is the fusion of these multi-modal features by column-wise concatenation. This process merged 310 EEG feature columns with 33 eye movement feature columns, creating a comprehensive input feature vector of 343 columns for our model.

EXPERIMENT

Experimental Setup

Dataset

To evaluate the effectiveness of this approach of combining the EEG signal and eye movement using the SEED-V dataset curated by the brain and cognitive machine intelligence (BCMI) Laboratory at Shanghai Jiao Tong University. The dataset included EEG and eye movement recordings from 20 participants (10 males and 10 females), all right-handed students with normal vision and hearing. Prior to the experiment, all participants completed the Eysenck's personality questionnaire (EPQ) personality test to ensure a stable mental state. Emotional responses were elicited by having participants watch carefully selected film clips corresponding to five categories: happy, sad, fearful, neutral, and disgusted. EEG data were recorded from 20 participants using 62 electrodes. Eye tracking data were simultaneously captured using SensoMotoric Instruments (SMI) eye tracking glasses [11].

The BCMI laboratory preprocessed the EEG signals by downsampling them to 200 Hz and applying a bandpass filter between 1 and 75 Hz to remove noise and artifacts. Differential entropy (DE) features were then extracted from five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 H), and gamma (31–50 Hz), assuming a Gaussian distribution of the signals. For eye movement data, various features, such as pupil diameter, fixation duration, saccade amplitude, and blink frequency, were extracted based on standard methodologies from prior literature. In this study, we utilized the preprocessed version of the SEED-V dataset provided by the original authors for model training and evaluation.

Hardware and Software Environment

All model developments and training were performed using Kaggle's cloud-based infrastructure. The computational environment was provided with dual NVIDIA Tesla T4 Graphics Processing Units (GPUs) (16 GB GDDR6 VRAM each), enabling 32GB of GPU memory. This dual-GPU configuration provided significant acceleration for training deep learning models and handling parallelized operations. The system was equipped with two virtual CPUs, 13 GB RAM, and approximately 73 GB of temporary storage per session. These resources support end-to-end workflow, including data preprocessing, model training, and evaluation within the notebook environment.

The software environment was based on a Linux (Debian) backend using Python3.10. Model implementation was carried out using TensorFlow2.12 and PyTorch2.0, both utilizing GPU acceleration via CUDA11.8 and cuDNN8.6. EEG signal processing was performed using the MNE-Python library, whereas data manipulation employed NumPy, pandas, and SciPy. Matplotlib and Seaborn were used for the visualization and analysis of the experimental results.

Model Architecture

The CNN model consists of two convolutional blocks, as shown in Figure 1, each with a convolutional layer, batch normalization, rectified linear unit (ReLu) activation, dropout (0.2), and a 2×2 max-pooling layer. The first block contains 16 filters (3×3). The feature maps were flattened and passed through two fully connected layers (64 neurons and num class neurons) with ReLu and SoftMax activation, respectively. The input dimension was 12,240.

Training

The model was trained for over 20 epochs using the Adam optimizer with a learning rate scheduler and early stopping to prevent overfitting. The mini-batch training balanced memory efficiency and convergence. The dataset was split into training, validation, and testing sets with 80%, 10%, and 10% of the data, respectively, resulting in train size: 576, validation size: 72, test size: 72.

Experimental Result

A CNN model trained for over 20 epochs for emotion analysis of EEG signals has provided valuable insights.

Training and Validation Loss

This study explored the performance of a machine learning model by analyzing its training and validation loss of over 20 epochs, offering key insights into its learning behavior. This evaluation helps determines the model's ability to fit the training data and generalize it to new, unseen data. The training loss, shown in blue, decreased steadily from approximately 1.4632 to 0.1008, as shown in Figure 2, reflecting effective learning and convergence on the training dataset.

The validation loss, illustrated in orange, drops from approximately 1.6044 to 0.2919, closely tracking the training loss and suggesting a strong generalization to new data, as shown in Figure 2. However, a minor divergence between the curves in the later epochs indicates potential overfitting, highlighting the need for additional regularization or hyperparameter adjustments to enhance model performance.

Training, Validation and Testing Accuracy

The training and validation accuracy curves over 20 epochs provided valuable insights into the classification performance of the model, as shown in Figure 3. The training accuracy, depicted in blue, increased from approximately 38.37% to 97.92%, indicating effective learning and robust classification

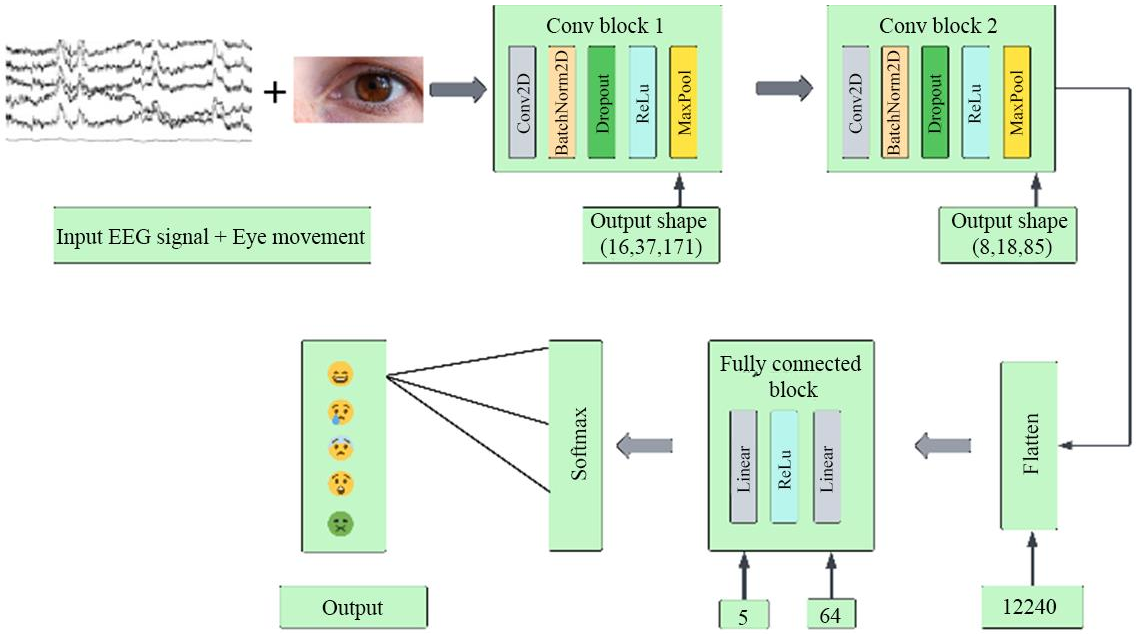

Figure 1. Emotion classification model.

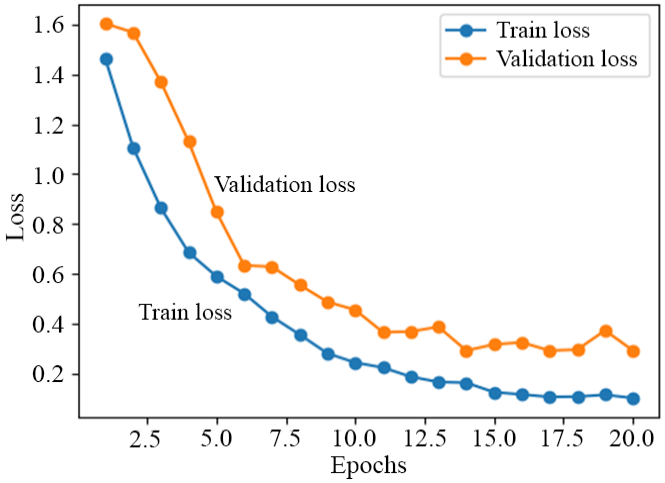

Figure 2. Training and validation loss across epochs.

of the training dataset. The validation accuracy, illustrated in orange, increased from approximately 20.83% to 94.44%, closely tracking the training accuracy and demonstrating strong generalization to unseen EEG signals. A testing accuracy of 93% further confirmed the model's consistent performance on new data, although minor fluctuations in the validation curve suggested potential areas for improvement in model stability.

Confusion Matrix

The confusion matrix, depicted in Figure 4, evaluates the performance of the CNN model across five emotions: disgust, fear, sadness, neutral, and happiness. The vertical axis represents true emotions, whereas the horizontal axis denotes predicted emotions, with cell values expressed as percentages. Notably, the top-left cell reflects a 100% accuracy for the corresponding emotions.

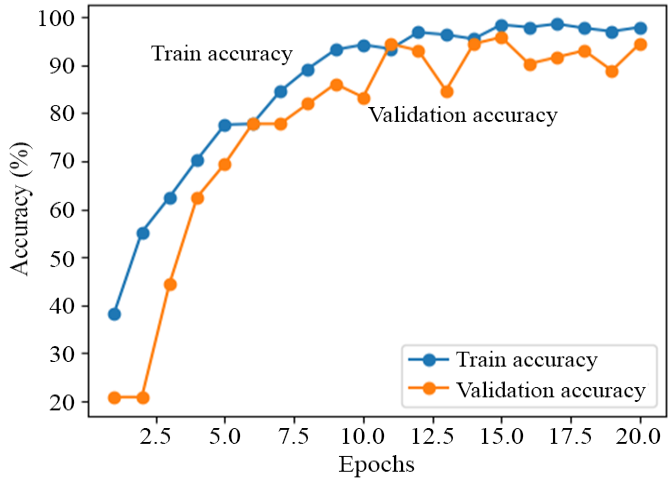

Figure 3. Training and validation accuracy across epochs.

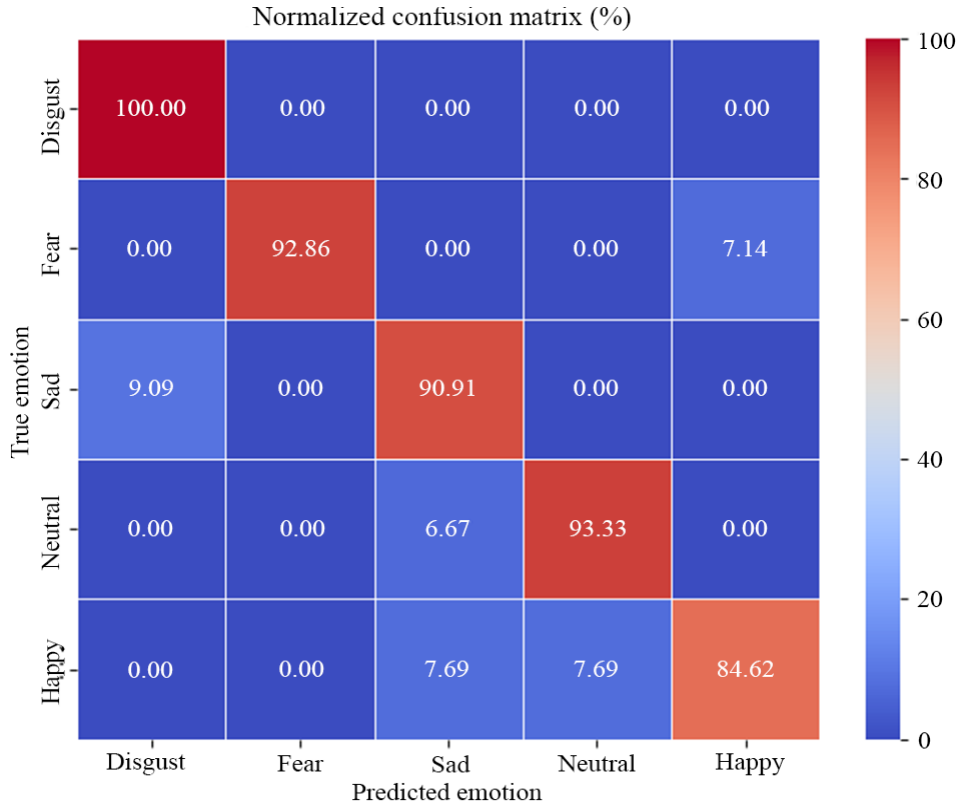

Figure 4. Confusion matrix of the model.

Table 1 compares the various EEG-based emotion recognition methods with our approach, which combines EEG signals. and eye movement data processed through a CNN, achieving 93% accuracy on the SEED-V dataset. In contrast, CTF-EmotionNet with CNN and interactive transformer scores 92.09% on SEED-IV, DGCCA-AM with deep generalized CCA reached 82.00% on SEED-V, hybrid attention model with LSTM and time attention attained 92.47% on SEED, and TL+CNN with pre-trained EEG spectrogram mapping scores 89.81% on SEED-IV and 88.23% on SEED-V. This approach outperforms these methods, particularly on SEED-V, demonstrating the effectiveness of integrating EEG and eye movement data with CNN.

DISCUSSION AND CONCLUSION

The integration of electroencephalogram (EEG) signals with eye movement data, utilizing the SEED-V dataset, marks a distinctive advancement in multi-modal emotion recognition in addition to existing methodologies. By concatenating 310 EEG feature columns with 33 eye movement feature columns, we have achieved a test accuracy of 93%, outperforming state-of-the-art methods, such as CIT-EmotionNet (92.09% on SEED-IV), DGCCA-AM (82% on SEED-V), and TL+CNN (88.23 % on SEED-V). These fusion leverages are the complementary strengths of EEG, which captures high-resolution brain activity reflective of internal emotional states, and eye movement features such as pupil diameter and fixation duration, which provide insights into attentional and arousal dynamics. The resulting model effectively classified complex emotions, happiness, sadness, fear, neutral, and disgust with a weighted average F1-score of 0.93 and a robust confusion matrix, demonstrating its ability to handle challenges such as inter-subject variability and noise inherent in EEG data.

This study highlights the transformative potential of combining EEG and eye movement data for emotion recognition, offering a robust framework for applications in human–computer interaction (HCI), mental health monitoring, and personalized content delivery. The superior performance on the SEED-V dataset underscores the value of multi-modal approaches in capturing the nuanced interplay between cognitive and behavioral emotional cues, surpassing single-modality systems. This work lays a foundation for future advancements in affective computing, with opportunities to explore additional modalities, such as facial expressions or heart rate variability, and to incorporate advanced architectures, such as graph neural networks or transformers. By addressing the limitations of traditional EEG-based systems, we paved the way for more accurate and generalizable emotion recognition systems, fostering emotionally intelligent technologies that can adapt to diverse populations and real-world contexts.

Table 1. Comparison of EEG-based emotion recognition methods.

Paper	Method description	Dataset	Accuracy
CIT-EmotionNet	CNN with interactive transformer for spatial-temporal EEG feature extraction	SEED-IV	92.09%
DGCCA-AM	Deep generalized Canonical Correlation Analysis (CCA) with attention for multi-modal emotion recognition	SEED-V	82%
Hybrid Attn. Model	DE features, CNN encoder, band attention, LSTM, and time attention for key features	SEED	92.47
Transfer Learning(TL) + CNN	EEG spectrogram passed to pre-trained model (TL), spatial map fed into CNN	SEED-IV	89.81%
		SEED-V	88.23%
Our Approach	EEG signal combined with eye movement passed to convolutional neural network	SEED-V	93%

REFERENCES

1. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, et al. DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput. 2012;3(1):18–31. doi:10.1109/T-AFFC.2011.15.

2. Islam MR, Moni MA, Islam MM, Rashed-Al-Mahfuz M, Islam MS, Hasan MK, et al. Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. IEEE Access. 2021;9:94601–94624. doi:10.1109/ACCESS.2021.3091487.
3. Chen JX, Jiang DM, Zhang YN. A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. IEEE Access. 2019;7:118530–118540. doi:10.1109/ACCESS.2019.2936817.
4. Lu W, Xia L, Tan TP, Ma H. CIT-EmotionNet: convolution interactive transformer network for EEG emotion recognition. PeerJ Comput Sci. 2024;10:e2610. doi:10.7717/peerj-cs.2610.
5. Feng L, Cheng C, Zhao M, Deng H, Zhang Y. EEG-based emotion recognition using spatial–temporal graph convolutional LSTM with attention mechanism. IEEE J Biomed Health Inform. 2022;26(11):5406–5417. doi:10.1109/JBHI.2022.3198688.
6. Lan YT, Liu W, Lu BL. Multimodal emotion recognition using deep generalized canonical correlation analysis with an attention mechanism. 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK. 2020. p. 1–6. doi:10.1109/IJCNN48605.2020.9207625.
7. Ma W, Zheng Y, Li T, Li Z, Li Y, Wang L. A comprehensive review of deep learning in EEG-based emotion recognition: classifications, trends, and practical implications. PeerJ Comput Sci. 2024;10:e2065. doi:10.7717/peerj-cs.2065.
8. Zhang Y, Zhang Y, Wang S. An attention-based hybrid deep learning model for EEG emotion recognition. Signal Image Video Process. 2023;17(5):2305–2313. doi:10.1007/s11760-022-02447-1.
9. Chaddad A, Wu Y, Kateb R, Bouridane A. Electroencephalography signal processing: a comprehensive review and analysis of methods and techniques. Sensors (Basel). 2023;23(14):6434. doi:10.3390/s23146434.
10. Ranganathan H, Chakraborty S, Panchanathan S. Multimodal emotion recognition using deep learning architectures. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA. 2016. p. 1–9. doi:10.1109/WACV.2016.7477679.
11. Liu W, Qiu JL, Zheng WL, Lu BL. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. IEEE Trans Cogn Dev Syst. 2022;14(2):715–729. doi:10.1109/TCDS.2021.3071170.

