Vol. 9 (2025): Proceedings of the 2025 International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability
(IC-AIMEES 2025)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.202505

Research Progress on Artificial Intelligence and Brain–Computer Interfaces in the Emotion Recognition of Patients with Depression

Zhenlin Yang*

School of Information and Communication Engineering，Hainan University, Hai Kou, 570228, China

**Corresponding author: Zhenlin Yang **

Abstract

Depression represents a significant public health concern on a global scale. Conventional diagnostic approaches rely on subjective scales, which have limitations, including low consistency and difficulty in early identification. In recent years, the integration of artificial intelligence (AI) and brain–computer interface (BCI) technologies has led to the development of new solutions for mood recognition in patients with depression. This review examines emotion recognition in depression via multimodal data, including EEG signals, eye tracking, and facial expression analysis. This highlights the application and performance of deep learning methods, such as the EEGNet and LSTM-CNN parallel models. The study also discusses multimodal fusion techniques, including graph neural networks and dynamic weighted fusion. Research indicates that AI and BCI technologies enable objective quantification, real-time monitoring, and personalised intervention. These capabilities significantly increase the accuracy and robustness of depression recognition. Nevertheless, this field continues to face challenges, including ethical controversies and the limited generalizability of models. Future work should prioritise algorithmic improvements, facilitate clinical translation, and strengthen ethical frameworks to support broader implementation.

Keywords

artificial intelligence, brain-computer interface, depression, emotion recognition, deep learning

1. Introduction

Depression is a major global public health challenge. The World Health Organisation (2023) reported that over 380 million people live with depression worldwide. Nearly half do not receive timely or effective diagnosis and treatment. In China, the lifetime prevalence of depression is 6.8%. Among adolescents, the incidence has increased by nearly 40% in the past decade. Moreover, approximately 15–20% of patients have attempted suicide. Over 50% of suicide deaths are directly related to depression. Current clinical diagnosis relies heavily on subjective scales. These include the HAMD and PHQ-9. These methods have low consistency and are prone to rater bias. This results in high misdiagnosis rates and challenges in early detection. Delays in intervention increase the risk of symptom worsening and suicide.

The integration of artificial intelligence (AI) and brain–computer interface (BCI) technology presents a novel approach to emotion recognition and depression diagnosis. BCI systems acquire and decode neural signals, such as EEGs, to directly quantify physiological traits related to emotional regulation. When combined with AI algorithms, these signals enable objective, real-time, and personalised assessment of emotional states. For example, the EEGNet model achieves an accuracy of 91.25% in detecting depression-related EEG patterns. A dynamically weighted fusion system further improves early recognition rates to 89.7%. These methods can extract and quantify key biomarkers—such as prefrontal alpha asymmetry—markedly enhancing diagnostic precision and substantially reducing the risks associated with misdiagnosis or missed cases.

Compared with traditional methods, AI-BCI technology offers multiple advantages: it provides greater objectivity by reducing reliance on subjective scales. It enables real-time and continuous monitoring. It supports multimodal fusion, integrating EEG, eye-tracking, and facial expression data to enhance model robustness and generalisation. It allows for personalisation through techniques such as transfer learning, which enables adaptation to individual characteristics.

From a historical perspective, early research focused primarily on analysing single-modal EEG signals. With advances in computational power and algorithms, the field has progressively shifted toward multimodal fusion approaches. These include novel deep learning frameworks such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, graph neural networks (GNNs), and cross-modal contrastive learning. This progress has not only improved the accuracy and early detection of depression but has also laid the foundation for closed-loop intervention and personalised treatment.

This paper systematically presents advances in AI-BCI technology for emotion recognition in depression, tracing the evolution from single-modal EEG analysis to multimodal fusion approaches. It focuses on the principles and performance of various deep learning and fusion methods and discusses opportunities and challenges in clinical translation. This study aims to provide theoretical and technical support for establishing a more objective, accurate, and generalizable diagnostic framework for depression, thereby facilitating the transition from laboratory research to real-world clinical application.

2. Key problems and Principles

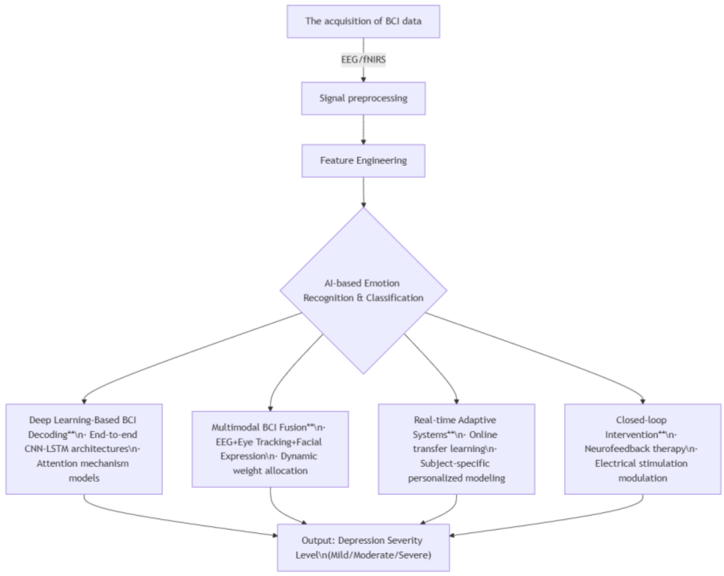
2.1 Depression-Related Issues and Manifestations

Depression is a complex mental disorder characterised by persistent low mood, diminished interest, and loss of pleasure. Its underlying mechanisms are closely linked to dysfunctions in the brain's emotion-regulation systems. Reduced activity in the prefrontal cortex impairs cognitive control, whereas hyperactivation of the amygdala amplifies negative emotions. Patients often experience significant cognitive impairment and disrupted circadian rhythms, with most suffering from sleep disturbances. Neuroelectrophysiological studies have identified specific biomarkers in depression, such as prefrontal alpha asymmetry and increased theta band power. These objective features provide a critical foundation for AI-BCI-based depression recognition (Thibodeau et al., 2023). By quantifying these abnormal physiological signals and integrating multimodal data, AI-BCI systems hold promise for overcoming the limitations of traditional subjective assessments and enabling the early and accurate detection of depression.

2.2 Technical Principles of AI and BCI in the Emotional Recognition of Patients

As illustrated in Figure 1, this approach begins by collecting EEG signals via brain monitoring devices, along with other data such as eye movement and facial expressions. After preprocessing and feature extraction, advanced AI algorithms are used to classify emotional states. Among these methods, end-to-end deep learning models can directly extract features related to depression from raw EEG signals. Multimodal fusion techniques integrate information from different sources through dynamic weighting. Studies have also developed real-time adaptive systems that use transfer learning to account for individual differences. Some systems employ closed-loop intervention strategies, which can trigger neurofeedback when depressive states are detected. These advances provide new tools for diagnosing and treating depression.

Figure 1: Technical Roadmap



3. Related Technologies

3.1 Deep Learning-Based BCI Decoding Algorithms

Deep learning algorithms analyse physiological data such as EEG signals collected by brain–computer interface devices. They extracted biomarkers related to depression to achieve objective and accurate emotion recognition, thereby assessing the severity of depressive states. Below are several core algorithms used in emotion recognition.

3.1.1 CNN and EEGNet

Convolutional neural networks (CNNs) are a type of deep learning model designed for processing grid-structured data (Figure 2). They use three main layers to extract features and perform classification. The first is the convolutional layer, which applies a sliding window to detect local patterns and spatial relationships in the input. Next, the pooling layer reduces the size of the feature maps by computing the maximum or average value within fixed regions. This helps compress the data, lowers the risk of overfitting, and reduces the computational cost. Finally, the fully connected layer converts the extracted features into a one-dimensional vector and applies a Softmax function for classification; the structure is shown in the figure. In emotion recognition, CNNs can effectively identify depression-related patterns in EEG signals and classify the severity of depression.

Figure 2: CNN Architecture

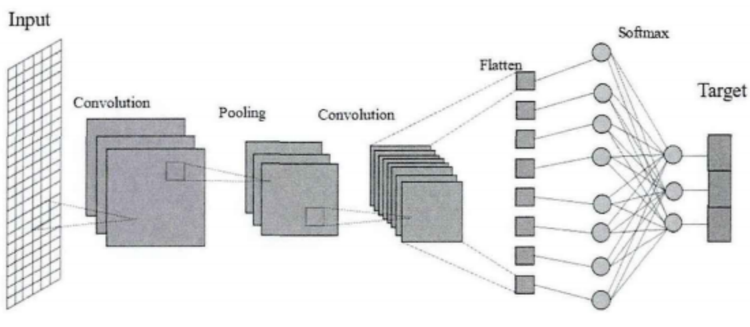
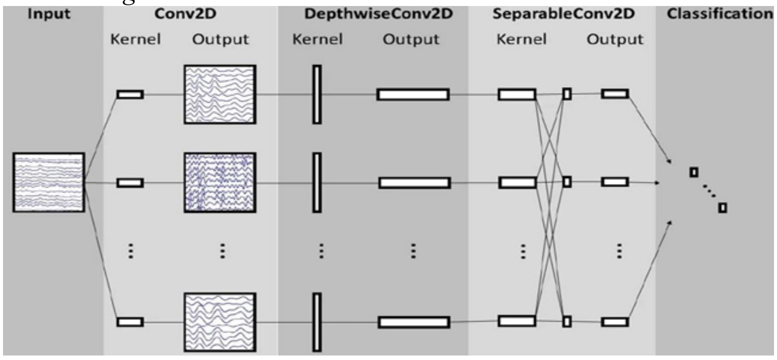


Figure 3: EEGNet Neural Network Architecture

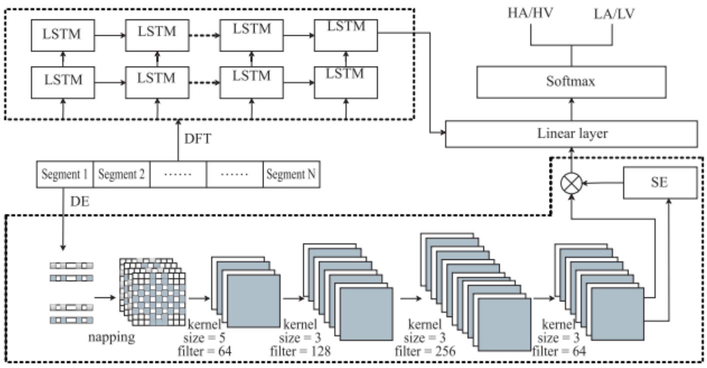


EEGNet is an efficient CNN architecture designed explicitly for EEG signal analysis. It is more suitable for emotion recognition than standard CNNs are. As shown in Figure 3, the model follows an end-to-end design that processes raw EEG signals directly via one-dimensional convolutional kernels. This enables the automatic extraction of spatial and temporal features without manual intervention, thereby reducing subjectivity and minimising information loss. A key innovation of EEGNet is its lightweight structure, which uses depthwise separable convolution to reduce the number of parameters significantly. It combines temporal and spatial convolution to capture features across time and electrode channels. This makes emotion recognition models more efficient and easier to deploy.

3.1.2 LSTM

Yang Qingran et al. proposed a parallel model that combines long short-term memory (LSTM) and a convolutional neural network (CNN) for EEG-based emotion recognition, as shown in Figure 4 (Yang et al., 2025). This model integrates the advantages of both networks to achieve efficient emotion analysis. The process begins with preprocessing EEG signals through baseline correction and bandpass filtering, followed by the extraction of differential entropy features from four key frequency bands: theta, alpha, beta, and gamma. The LSTM module captures temporal features by analysing amplitude–frequency characteristics via the discrete Fourier transform, whereas the CNN module extracts spatial features through convolutional layers. The features from both modules are fused via a weighted strategy, and emotion classification is completed through a fully connected layer and a Softmax classifier. The model achieves high accuracy rates of 90.1% in valence dimensions and 89.3% in arousal dimensions (Yang et al., 2025). Compared with using a CNN alone, this method adds a temporal dimension, leveraging the LSTM's strength in time series analysis and the CNN's ability to extract spatial features to recognise emotional states from EEG signals effectively.

Figure 4: Parallel Network Architecture Model Combining LSTM and CNN



Multimodal BCI Fusion Algorithms Multimodal fusion technology significantly enhances the robustness of depression recognition by integrating EEG signals with other physiological data, including eye movements, facial expressions, and heart rate. This approach analyses multimodal physiological signals to extract emotional features, which are then classified to determine the severity of depression. Current mainstream methods can be divided into three categories and are compared in Table 1.

Table 1: Performance Comparison of Multimodal Algorithms

Methods	Dataset	Accuracy	Specificity	Sources
GNN-based fusion	DEPRESS-LF	92.30%	89.50%	(Zhang et al., 2022)
Dynamic weighting fusion (DWF-BCI)	MODMA-Emo	88.10%	91.20%	(Chen et al., 2023)
Contrastive learning (CL-MM)	CrossMDD	85.40%	83.70%	(Liu et al., 2024)

Graph Neural Network Fusion Framework. Zhang Yuchen et al. proposed a novel GNN-based fusion model that constructs *an* EEG channel topology and eye-tracking heatmaps into a heterogeneous graph network. It employs node attention mechanisms to enable cross-modal feature interaction (Zhang et al., 2022). This model achieved an accuracy of 92.3% in depression emotion recognition, representing an 11.6% improvement *over* single-modal EEG models.

Dynamic weighted fusion method. Chen et al. developed a DWF-BCI system that employs a two-stage fusion strategy (Chen et al., 2023). In the first stage, primary fusion quantifies the coupling strength between *EEGs* and facial *microexpressions via* mutual information entropy. In the second stage, decision fusion dynamically adjusts the contribution of each modality through a learnable weight matrix. The system achieved a recall rate of 89.7% on a dataset of MDD patients, demonstrating strong performance in the early recognition of mild depression.

Cross-Modal contrastive learning. Liu et al. incorporated self-supervised learning into multimodal fusion. By using the InfoNCE loss function, their method maximises mutual information between EEG signals and eye movement sequences. It achieved a cross-centre validation accuracy of 85.4%*,* even with limited data (Liu et al., 2024). This approach effectively addresses the challenge of small sample sizes in depression studies.

4. Related Applications

4.1 Application and Performance of EEGNet in Depression Recognition Experiments

Song et al. showed that EEGNet can identify depression through EEG analysis (Song, 2023). In their method, resting-state EEGs were first collected from patients and healthy controls. The data come from frontal sites Fp1, Fp2, and Fz and are recorded at 1000 Hz for 5.5 minutes via an RM6280C system. They also use public MODMA data for extra samples. Next, the signals are cleaned via 7-level sym3 wavelet decomposition to remove eye artefacts. Levels 4–7 are retained, and the signal is rebuilt, after which it is downsampled to 250 Hz. The EEG is cut into 6-second segments without overlap to create more samples. The data are split 7:2:1 into training, validation, and test sets. These segments are fed into EEGNet, which uses temporal and spatial convolutions to extract features in time and across channels. The model also uses batch normalisation, pooling, and dropout before classification. It reached 91.25% accuracy on full-band RMEEG data and 87.31% accuracy on the MODMA delta band. This offers an objective option for assessing depression, which is especially useful for patients unwilling to use surveys, and helps make the diagnosis faster and more accurate.

4.2 Application of LSTM in Depression Recognition

Xie Wanqing and colleagues from the School of Biomedical Engineering at Anhui Medical University developed a multimodal system for diagnosing and screening depression and anxiety by integrating the Self-Rating Depression Scale (SDS) and the Self-Rating Anxiety Scale (SAS) with synchronised video recordings of patients during questionnaire completion (Xie et al., 2022). The system uses deep learning to analyse temporal dynamics in video sequences, including facial expressions and body movements, over time. Initially, a convolutional neural network (CNN) extracts spatial features from individual video frames, capturing static elements such as facial muscle activity and body posture. However, these frame-level features represent only isolated moments and fail to capture the progressive emotional and behavioural changes characteristic of depression. This gap is effectively addressed by incorporating a long short-term memory (LSTM) network. The LSTM model processes the sequence of CNN-extracted features to learn temporal dependencies. Its gating mechanism enables the modelling of long-range patterns, such as gradually intensifying sadness or increasing restlessness throughout the assessment. By converting raw video data into temporally structured representations, the LSTM captures clinically meaningful dynamics. The video-based temporal features are then fused with self-reported scale scores (SDSs and SASs), significantly enhancing the accuracy of depression detection. This integrated approach provides robust technical support for auxiliary diagnosis in mental health settings.

4.3 Application of AI in Analysing Multimodal Features of Depression

The application of artificial intelligence in depression recognition has evolved from single-modal approaches to multimodal fusion, aiming to improve diagnostic accuracy and comprehensiveness. As noted by Tan Fandi et al. in their review, although AI-based analysis of single modalities, such as text, voice, facial expressions, or brain MRI, has achieved notable results, the complexity of depressive symptoms makes assessing the condition using only one type of data entirely challenging. Therefore, multimodal fusion has become an important direction for future research (Tan et al., 2024). Multimodal methods integrate diverse data sources, including speech, text, facial expressions, eye movements, and neuroimaging. By employing machine learning techniques, these methods achieve information complementarity, leading to a more accurate identification of depression. For example, Zhang Yinghui proposed a deep forest-based decision-level fusion method, which significantly improved detection rates and reduced prediction errors (Zhang et al., 2018). Zhao Jiandeng et al. explored third-generation AI models that combine knowledge-driven and data-driven approaches to build more comprehensive and accurate diagnostic support systems (Zhao et al., 2023). Research shows that multimodal integration not only facilitates early screening and precise diagnosis but also has potential in predicting the risk of depression progressing to psychosis. These AI systems demonstrate assessment capabilities comparable to those of professional psychiatrists, offering a new pathway toward objective and intelligent diagnosis and treatment of depression.

5. Conclusion

The integration of artificial intelligence (AI) and brain-computer interface (BCI) technology presents a novel approach to emotion recognition and auxiliary diagnosis in depression, offering notable advantages. By utilising multimodal physiological data such as EEGs, eye movements, and facial expressions, AI-BCI systems enable objective, real-time, and personalised emotion assessment, effectively overcoming the subjectivity and inconsistency of traditional scale-based methods. The application of deep learning models, including EEGNet and parallel LSTM-CNN architectures, combined with multimodal fusion strategies such as graph neural networks and dynamic weighted fusion, further improves recognition accuracy and model robustness. These advancements establish a technical foundation for early screening and precise intervention in depression.

However, the field still faces multiple challenges. First, ethical concerns are prominent, particularly regarding invasive BCI technologies, which raises widespread debate over data privacy, informed consent, and safety. Second, existing models exhibit limited generalisation capability, often performing inconsistently in cross-subject and cross-centre validations, making it difficult to adapt to the high heterogeneity of depression. Finally, barriers to clinical translation remain, including high device costs, operational complexity, and a lack of large-scale clinical trial validation.

Future research should prioritize the following directions: first, developing more lightweight and highly generalisable algorithms by incorporating self-supervised learning and meta-learning to reduce reliance on annotated data; second, promoting the standardisation of multimodal datasets by establishing unified protocols for data acquisition and preprocessing; and third, accelerating the development of portable and low-cost BCI devices to facilitate the translation of the technology from laboratory research to clinical applications.

Through multidisciplinary collaboration and continuous innovation, AI and BCI technologies have become integral components of depression diagnosis and treatment systems. They offer strong support for early warning, dynamic monitoring, and personalised intervention in mental health.

References

Chen, X., Wang, Y. and Valdes-Sosa, P. A., (2023). Dynamic-weighted fusion of EEG and eye-tracking for depression detection: A bimodal deep learning approach. *IEEE Transactions on Affective Computing,* vol. 14, no. 2, pp. 1021–1034.

Liu, Z., Zhang, H. and Zhang, R., (2024). Published. Contrastive learning for cross-modal depression recognition: An EEG-eye movement study. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2024)*, Seoul. IEEE, pp. 1-5.

Song, X. W., (2023). *Research on depression detection method based on EEG signal.* Master's Thesis, Shandong University.

Tan, F. D., Ma, X. L., Wang, Y. Y., Tao, T. T., Wang, G. Y. and Huang, S. H., (2024). Application of multi-modality based on artificial intelligence feature fusion in depression recognition. *Modern Instruments & Medical Treatment,* vol. 30**,** no. 1, pp. 58-63.

Thibodeau, R., Jorgensen, R. S. and Kim, S., (2023). Frontal EEG alpha asymmetry as a diagnostic marker of depression: A meta-analytic update. *Neuroscience & Biobehavioral Reviews,* vol. 145, p. 105002.

Xie, W., Wang, C., Lin, Z., Luo, X., Chen, W., Xu, M., Liang, L., Liu, X., Wang, Y. and Luo, H., (2022). Multimodal fusion diagnosis of depression and anxiety based on CNN-LSTM model. *Computerized Medical Imaging and Graphics,* vol. 102, p. 102128.

Yang, Q. R., Fang, W. S., Xu, X. Y. and Xu, Z. Y., (2025). Emotion recognition based on LSTM-CNN parallel neural networks. *Journal of Luoyang Institute of Science and Technology(Natural Science Edition),* vol. 35**,** no. 2, pp. 69-73.

Zhang, Y. C., Li, X. F. and Wang, Z. Q., (2022). EEG-eye movement depression emotion recognition based on graph neural network. *Journal of Computer Research and Development,* vol. 59**,** no. 8, pp. 1801–1812.

Zhang, Y. H., Nie, Y. M., Sun, B., He, J. and Yang, B., (2018). A method of depression evaluation based on decision level fusion of deep forest with multimodality data. *Journal of Beijing Normal University(Natural Science),* vol. 54**,** no. 5, pp. 606-611.

Zhao, J., Zhou, L. Y. and Wu, M. Q., (2023). Artificial intelligence-based auxiliary diagnosis method for depression. *Journal of Northwest University (Natural Science Edition),* vol. 53**,** no. 3, pp. 325–335.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.