Journal of Network Intelligence ©2025 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **10**, Number **1**, February **2025**

Deep Learning-based Artificial Intelligence-Assisted Diagnosis of Psychological Disorders

Jia Liu*

Student Mental Health Education and Counseling Center
Xi'an International Studies University, Xi 'an 710000, P. R. China
qwerty00202312@163.com

Ai-Feng He

University of Campania Luigi Vanvitelli, Naples 80138, Italy
xa2750@163.com

*Corresponding author: Jia Liu
Received March 15, 2024, revised May 30, 2024, accepted August 1, 2024.

Abstract. *Aiming at the issue of insufficient extraction of psychological feature details in the current methods of assisted diagnosis of psychological disorders, which results in low classification efficiency, this article suggests an artificial intelligence assisted diagnosis method of psychological disorders relied on deep learning. First, the attention mechanism is optimized by two one-dimensional global pooling operations, and the input features in the vertical and horizontal directions are pooled with the global average and maximum pooling respectively, so as to make full use of the original detail information. Then the original Electroencephalogram (EEG) signals are preprocessed by the first-order differential mean absolute value to obtain the normalized results, followed by the extraction of spatio-temporal features, weighting the input features in the spatio-temporal channel with the enhanced attention mechanism, and adopting the layer-jumping connection to process the weighted data of the two channels for the goal of enhancing the accuracy of the model. Next, various EEG signals were weighted with intrinsic relationships to optimize feature extraction. Finally, the classification outcome is fused with the classification results output from the Graph Convolutional Neural Network (GCNN) to realize the accurate diagnosis of psychological disorders. The experipsychological outcome implies that compared with the existing diagnostic methods, the suggested method has higher accuracy, sensitivity, specificity, F1 score and better classification performance.*
Keywords: Diagnosis of psychological disorders; Deep learning; Attention mechanisms; Graph convolutional neural networks; Feature extraction

1. **Introduction.** Recently, psychological disorders have become a widespread societal problem affecting people of all ages, genders, or industries in every country around the globe [1]. However, when the emotional threshold is higher, the likelihood of developing a psychological disorder increases, thus creating a vicious cycle. According to the United Nations Secretary General's World Health Day statement, approximately 1 billion people worldwide are suffering from psychological disorders, with depression and autism being among the most common psychiatric disorders [2]. Psychological disorders may bring about various adverse physical and psychological reactions such as autism, insomnia, fatigue, indigestion, headaches, etc. [3, 4, 5]. Prolonged exposure to the psychological stress state of severe disorders may also lead to serious health problems such as hypertension, cardiovascular diseases, and immune system disorders [6, 7]. Therefore, how to make

accurate auxiliary diagnoses of psychological disorders is of significant research value and practical importance.

1.1. **Related work.** Traditional methods focus on extracting and analyzing the characteristics of patients, which may include emotional expression, speech characteristics, behavior patterns, and so on. By observing and recording these characteristics, doctors can draw a preliminary judgment about the patient's psychological state. Smith et al. [8] extracted different features as effective predictors of depression with respect to depressive voice vocalization characteristics. Bangerter et al. [9] used the Mel Frequency Cepstrum Coefficient (MFCC) and resonance peak features, combined with the Gaussian Mixture Model (GMM) model for autism recognition. Zhang et al. [10] analyzed the frequency, rate of change, intensity, and other features of the action units of the patient and the normal person in the process of the interview, and classified them by SVM, but the classification efficiency was low. Wang et al. [11] used SVM to diagnose people with psychological disorders based on changes in eye, eyebrow, and mouth movements with an accuracy of 78.85%. Akbari et al. [12] proposed a depression detection method based on the reconstruction of phase space and geometric features from EEG signals and used particle swarm optimization algorithms and SVM classifiers to select and classify features. Stolicyn et al. [13] used pupil size and other features in eye movement signals, combined with machine learning algorithms, to build a prediction model for psychological disorders. Jiang et al. [14] proposed the use of differential entropy and genetic algorithms for the extraction and selection of EEG features, and the use of a decision tree for classification. Zhang et al. [15] proposed a multimodal depression detection method using a multi-intelligent body strategy, exploring both physiological and behavioral perspectives at the same time, incorporating EEG and sound signals. Traditional methods usually select and analyze features based on doctors' experience and professional knowledge, which may lead to subjectivity and limitations. Different doctors may have different judgment criteria, which may lead to different diagnosis results.

With significant advances in neural network models and learning algorithms, a new era of deep learning-based artificial intelligence has begun. Shoeibi et al. [16] mined psychological disorder representations from sound cues and adopted LSTM and CNN coding to identify discriminative audio representations. Niu et al. [17] designed a new framework integrating an attention module for the temporal-frequency channel of EEG signals but ignored the intrinsic feature interactions between signals. Zhu et al. [18] suggested a deep convolutional neural network-based facial appearance and motion modeling approach for diagnosis, but the classification efficiency was not high. He et al. [19] adopted Convolutional Neural Networks (CNN) to extract global and local information from video frame images [20] for the identification of psychological disorders. Malhotra and Jindal [0] offered a real-time deep learning system that fuses individual vector representations of multiple modalities from a user's social media feeds (text, images, and videos) to obtain a joint representation. Pan et al. [0] suggested a Graph CNN (GCNN) based method for the diagnosis of psychological disorders, where the connectivity matrix of the EEG is used as input and spatial features are extracted from it and thus categorized.

1.2. **Contribution.** Psychological disorder diseases bring a great burden to patients, families, and society, so there is a crucial role for their auxiliary diagnosis. At present, some methods have achieved preliminary outcomes, but there is still the issue of insufficient extraction of psychological feature details. Focusing on the shortcomings of current research, this article suggests an artificial intelligence-assisted diagnosis method for psychological disorders based on deep learning.

Firstly, the attention mechanism is optimized by two one-dimensional global pooling operations, and the input features in vertical and horizontal directions are pooled by global average pooling and maximum pooling respectively, so as to make full use of the original detail information. The original EEG signals are then preprocessed to obtain the normalized results, followed by the extraction of spatio-temporal features, weighting the input features in the spatio-temporal channel adopting the enhanced attention mechanism, and processing the weighted data of the two channels using the hopping-layer connection to enhance the accuracy of the model. Secondly, different EEG signals are weighted with intrinsic features to optimize the extraction of EEG signal features. Finally, the classification results are fused with the classification results output from the GCNN module to obtain the final diagnosis prediction. The experimental outcome indicates that the suggested method has higher classification performance and efficiency compared with the current algorithms.

2. Theoretical analysis.

2.1. Graph convolutional neural network.

Compared with traditional CNN, GCNN [0] is able to handle irregular data, transfer and aggregate information over the whole graph to obtain more global information. In addition, GCNN adopts an adaptive filter, which can adaptively adjust the filter size according to the number of neighbors and features of different nodes to better adapt to different graph structures. As indicated in Figure 1.

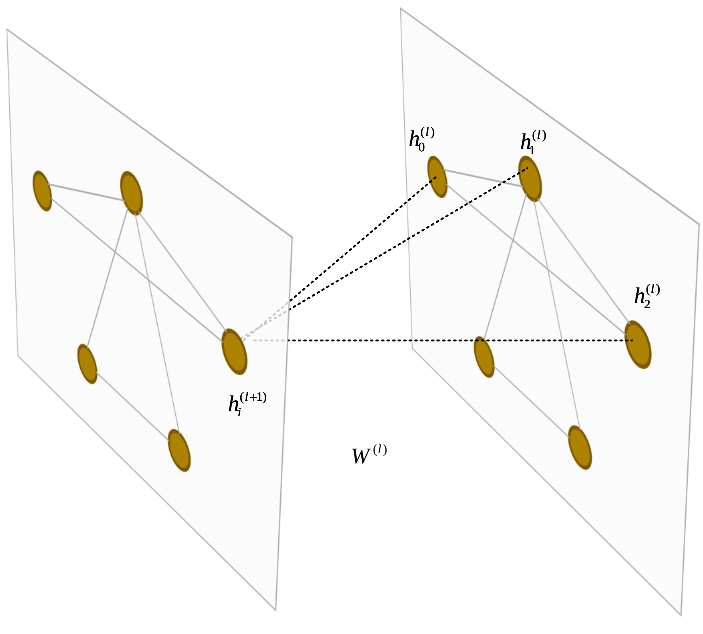

Figure 1. The computation progress of GCNN

Figure 1 is a schematic diagram of graph convolution computation, let the three neighboring points of the current node g_j be (g_0, g_1, g_2), and the degree of this node be 3.

The point g_i^{k+1} in layer $k+1$ is determined by the three reachable points in layer k at the time of feature propagation, and $V^{(k)}$ is the dashed portion, which represents the weight information of each node in this layer.

Given a graph $H = (W, E)$, where W is the set of nodes and E is the set of edges. Each node $w_j \in W$ has a feature vector $x_j \in R^D$, where D represents the dimension of the feature vector. Let M_j be the set of neighboring nodes of node w_j, $M_j = w_i|(w_j, w_i) \in E$. The graph convolution can be expressed as below:

$$g_j = \delta\left(\sum_{i=1}^{M_j} \frac{1}{de_j} V^{(k)} g_i^{(k)}\right) \tag{1}$$

where $g_i^{(k)}$ denotes the feature vector of node w_i in the k-th layer convolution, $V^{(k)}$ is the learnable weight matrix of the k-th level, δ is the activation function, de_i denotes the degree of node w_i, and $h_i^{(k)}$ denotes the feature vector extracted by node i in the k-th level convolution.

2.2. **Attention mechanism.** The attention mechanism can adaptively calibrate the weights of each channel by modeling the importance between individual features to determine the features of important attention [0], as indicated in Figure 2.

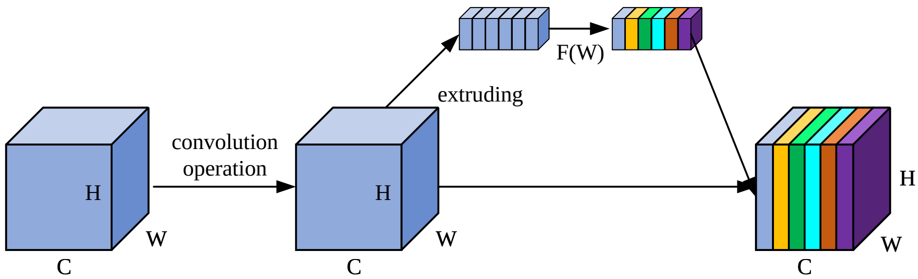

Figure 2. The attention mechanism

Attention-weighted features can be obtained by restoring the attention weights of different channels to the original features through matrix multiplication. Therefore, the channel attention extraction method can be summarized as feature dimensionality reduction, attention extraction, and weight recovery. First of all, for the feature map $x = [x_{i,j}^1, x_{i,j}^2, ..., x_{i,j}^m]$, the height and width of the feature map are represented by G and V, respectively, then the initial weights of different channels are taken as indicated in Equation (2).

$$b_l = \frac{1}{G \times V} \sum_1^m \sum_1^m x_{i,j}^l \tag{2}$$

Next, a specific method $F(x)$ is used to extract the attention weights as Equation (3).

$$a_i = F(b_i) \tag{3}$$

Finally, $\hat{x}^i = a_i x_{i,j}^l$ is adopted to weight the channel attention, and the extracted channel attention is applied to different channels for weighting, so that the weights of each channel are randomly adjusted according to the attention weights to optimize the model performance.

3. **Optimization of attention mechanisms.** The traditional attention mechanism only performs global average pooling in the calculation, without paying attention to the detailed texture information, and at the same time, in the process of feature weighting, it cannot make full use of the original feature information. Thus, on the ground of the original, the global average pooling and maximum pooling are performed on the input features, so that the original features and detail information can be more fully utilized.

Given the input x, the channels are encoded along the horizontal and vertical dimensions, respectively. The horizontal and vertical coordinates are summed up using global average pooling and global maximum pooling with dimensions $(G, 1)$ or $(1, V)$, respectively, to compute the encoding results in the horizontal and vertical planes. Therefore, the output of channel b with height G is indicated below.

$$z_b^g(\mathbf{g}) = \frac{1}{V} \sum_{0 \le j < V} x_b(\mathbf{g}, j) + \max_{0 \le j < V} x_b(\mathbf{g}, j) \tag{4}$$

The output of channel b with width v is represented as below.

$$z_b^v(\mathbf{v}) = \frac{1}{G} \sum_{0 \le j < G} x_b(j, \mathbf{v}) + \max_{0 \le j < G} x_b(j, \mathbf{v}) \tag{5}$$

After encoding, the output z^g and z^v are spliced and then transformed using the Fourier transform function F_1, as indicated in Equation (6).

$$f = L\left(F_1\left[z^g, z^v\right]\right) \tag{6}$$

where $[,]$ is the splicing operation along the spatial dimension, L is the LeakyRelu activation function, and $f \in R^{(B/s)*(G+V)}$ is the result of encoding and compressing the spliced feature information. Here, s is used to control the reduction rate of the attention mechanism module size.

Then f is decomposed into $f^g \in R^{(B/s*G)}$ and $f^v \in R^{(B/s*V)}$ along the spatial dimension, and the 1×1 convolutional transformations F_g and F_v are utilized to transform f^g and f^v to obtain $h^g \in R^{(B*G)}$ and $h^v \in R^{(B*V)}$, respectively.

$$h^g \in \mathcal{Q}\left(F_g\left(f^g\right)\right) \tag{7}$$

$$h^v \in \mathcal{Q}\left(F_v\left(f^v\right)\right) \tag{8}$$

where $\mathcal{Q}$ is the sigmoid activation function. The weights h^g and h^v are obtained after the transformation and are input into the feature map as attention weights, respectively, as indicated in Equation (9).

$$y_b(i, j) = x_b(i, j) * h_b^g(i) * h_b^v(j) \tag{9}$$

Finally, the obtained weighted feature matrix is added with the original feature matrix to get the final improved attention mechanism module output as indicated as follows:

$$o_b(i, j) = L\left(y_b(i, j) + x_b(i, j)\right) \tag{10}$$

where x_b is the original feature matrix of the model, y_b is the weighted feature matrix, and o_b is the output of the improved attention mechanism module.

4. Deep learning-based artificial intelligence-assisted diagnosis of psychological disorders.

4.1. Preprocessing of EEG signals. Focusing on the issue that current diagnostic methods for psychological disorders do not fully extract the features of EEG signals, which leads to inefficient classification, this article designs an artificial intelligence-assisted diagnosis method for psychological disorders based on deep learning. Firstly, the original EEG signals are preprocessed, and secondly, the spatio-temporal features of the standardized EEG signals are extracted, and the input features are weighted using an improved attention mechanism to improve the model accuracy. Then the connected data are extracted as spatio-temporal features by graph convolution. Finally, the classification results are fused with the classification results output from the GCNN to achieve accurate diagnosis of psychological disorders. The entire model of the suggested method is indicated in Figure 3.

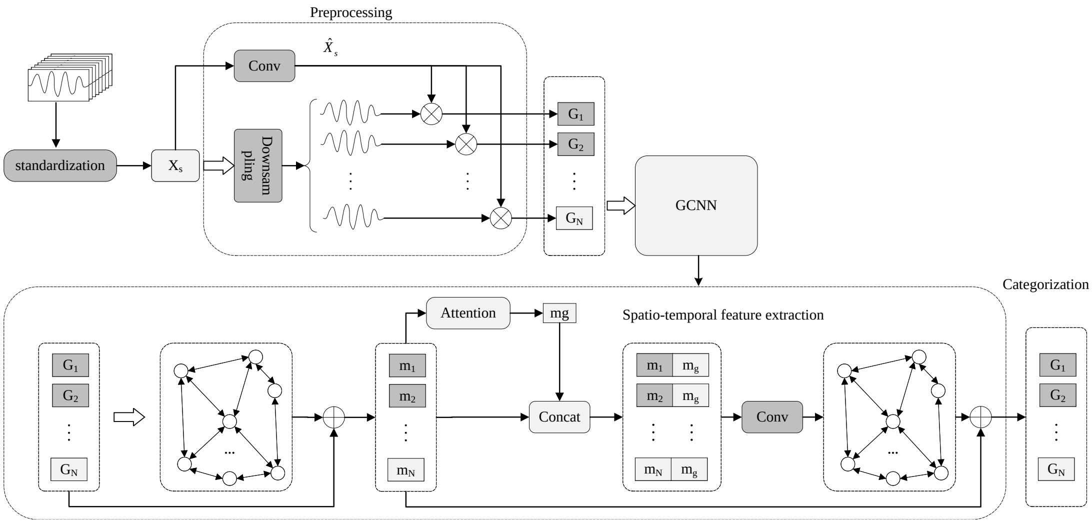

Figure 3. The overall model of the designed method

The currently available methods for the adjunctive diagnosis of psychological disorders rely mainly on biometric signals, and a commonly used biometric signal is the EEG [?]. EEG has extremely important spatio-temporal features that need to be extracted and utilized. Spatio-temporal features are the signal statistics of EEG signals in both the temporal and spatial domains. When extracting spatio-temporal features, to achieve accurate EEG data, it is necessary to reduce the sampling frequency by re-sampling the known sampling sequences at fixed intervals to gain new sequences, which is called downsampling, and the relationship is indicated in Equation (11).

$$y(n) = x(Cn) \tag{11}$$

where n denotes the extraction order and C denotes the extraction interval.

$$\mu_E = \frac{1}{M}\sum_{n=1}^{M} E(m) \tag{12}$$

where $E(m)$ is the original EEG signal, M is the number of sampling points. In this article, the original EEG signal is preprocessed by first-order differential averaging absolute value, as indicated in Equation (13).

$$\overline{\vartheta}_E = \tfrac{1}{M-1}\sum_{m=1}^{M-1} \left|\overline{E}(m+1) - \overline{E}(m)\right| = \frac{\vartheta_E}{\delta_E} \tag{13}$$

where $\overline{E}(m)$ denotes the normalized result of the original EEG signal, ϑ_E denotes the standard deviation of the original EEG signal, and δ_E denotes the mean absolute value of the first-order difference of the original EEG signal.

4.2. **Spatio-temporal feature extraction of EEG based on attention map convolutional network.** To be more effective in extracting spatio-temporal features of EEG, the input features are weighted adopting an enhanced attentional mechanism on both temporal and spatial channels to enhance the model accuracy. In this paper, feature maps are performed along the temporal channels, and the features on each spatial channel (C, V) are mapped to a value for subsequent operations, as indicated in the following equation.

$$S = L\left(\frac{1}{D * W}\sum_{d=1}^{D}\sum_{w=1}^{W}(V_1^T * X_{d,w})\right) \tag{14}$$

where $L(\cdot)$ represents the activation function LeakyRelu, X represents the model input matrix, and V_1^T is a trainable weight that represents the weight of the temporal channel in a cycle T.

The attention weighting module can effectively search the temporal-spatial correlation. The role is to weight the input data in order to emphasize key features that are meaningful for the diagnosis of psychological disorders, thus improving the accuracy of the diagnosis. The attentional weighting module is indicated in Equation (15) and Equation (16).

$$S_a = SoftMax(S^T * V_2^T) \tag{15}$$

$$X_e = L(S_a^T * X + X) \tag{16}$$

where $SoftMax(\cdot)$ denotes the SoftMax function; V_2^T is a trainable weight that represents the weight of the current input on the spatio-temporal channel.

The weighted data is convolved with the EEG signal data to compute in order to obtain the node features G_i of the EEG as indicated below.

$$G_i = \sum_{i=1}^{m}(V_i * X_e \odot B_s) \tag{17}$$

where $B_s \in \mathbb{R}^{N\times N}$ is the adjacency matrix in GCNN learning, which is updated by gradient descent during the training process. $G = [G_1, G_2, ..., G_M]$ is the node feature of GCNN input, and all the EEG feature signals share the correlation matrix B_s. This static graph models the intrinsic relationship of all the training samples, captures their strong correlations, and reduces the interference of edge noise. The procedure is as follows.

$$M = \vartheta(B_s G v_s) \tag{18}$$

where $\vartheta(\cdot)$ is the activation function and $M = [M_1, M_2, ..., M_M]$ is the updated intrinsic feature representation.

Subsequently, the intrinsic relationship is captured for each EEG signal. The updated node features M_M and G_M are superimposed to obtain new spatio-temporal features as the node features for the dynamic graph convolution input. In addition, a dynamic graph $B_D \in \mathbb{R}^{N\times N}$ is constructed for each EEG recording by connecting the correlation matrix m_M to the global features m_g through an adaptive global average pooling and convolution to obtain a feature vector containing global relationships.

$$B_D = \vartheta(f(m_M, m_g)) \tag{19}$$

where $f(\cdot)$ denotes the convolutional level. Because the structure of the graph depends on the features of each EEG signal, so the adjacency matrix of each signal is dynamically changing, so as to eliminate the interference of useless features on the fitting results and reduce the risk of overfitting. The spatio-temporal characteristics of the final EEG are indicated below.

$$G_i = \sum_{i=1}^{m}(V_i * X_e \odot B_D) \tag{20}$$

4.3. **Auxiliary diagnostic classification with loss function.** For each input spatiotemporal feature G_i, a feature mapping $G_s \in \mathbb{R}^{C\times D}$ is obtained after GCNN, a classifier is constructed based on the feature mapping, and after classifying and activating it, the activation mapping of a specific label is computed through an attention mechanism to obtain $T = [T_1, T_2, ..., T_M]$, and the obtained T is converted into a feature representation $H = [H_1, H_2, ..., H_M]$ that can represent the intrinsic relationship between EEG signals.

$$H_M = T_M^T G_i = \sum_{i=1}^{D} T_i^M \hat{G}_s \tag{21}$$

where D is the length of the signal after feature mapping, T_i^M denotes the weight of the activation mapping of the M-th EEG signal, and $\hat{G}_s$ denotes the feature vector of the feature mapping at location s.

Predictive psychological trait representations of EEG diagnostics are then generated via an activation function as follows.

$$T_M = soft\ max\left(\frac{H_M \hat{G}_s^T}{\sqrt{d_l}}\right) V \tag{22}$$

where d_l is the dimension of the input feature vector.

In this paper, the cross-entropy function [0] is chosen as the loss function for model training. It can effectively calculate the distance between the actual output and the expected output of the model and constantly correct the convergence direction of the model. It is a commonly used loss function in GCNN for classification tasks and is used to evaluate the difference between the fitted distribution of the model and the real distribution of the data. For the psychological disorder diagnosis and classification problem in this paper, the loss function is calculated as follows.

$$Loss = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{l=0}^{K-1} y_{i,k} \ln p_{i,k} \tag{23}$$

where k is the eigenvalue, N is the amount of data, $y_{i,k}$ represents the true feature of the i-th EEG as k and $p_{i,k}$ represents the probability that the i-th data is predicted to be the psychological feature k. The distance between categories is increased to some extent by fitting this loss function.

5. Performance testing and analysis.

5.1. **Comparison of classification performance.** In this article, EEG images from ABIDE were adopted as the experipsychological dataset. The ABIDE dataset [0] consists of EEG images of 621 individuals with autism and 603 normally developing controls. In this study, 1172 data (575 autism cases and 597 normal developpsychological controls) were selected, and the data were categorized as 80% for training, 10% for validation, and 10% for testing. In order to verify the performance of the method designed in this paper, comparative experiments were conducted, and for ease of analysis, the algorithm in this paper is denoted as DLPD, DRBR in literature [12], ACNN in literature [19], and MAMC in literature [22].

The models were trained and tested based on the PyTorch framework. The cross-entropy loss function and the Adam optimizer were used to optimize the model weights. To ensure the control effect, the principle of consistency of variables was followed, so that the parameter settings of DLPD, DRBR, ACNN, and MAMC were set to use the default parameter values. During the training period, the learning rate was fixed at 0.001 and the number of iterations was 100.

This article performs a classification performance comparison between autism disorder and normal developpsychological controls, and will compare classical machine learning and deep learning-based diagnostic methods for psychological disorders in five dimensions, namely Accuracy, Sensitivity, Specificity, F1 score, and AUC [0], as indicated in Table 1.

Table 1. Comparison of classification performance of different diagnostic models

Method	Accuracy	Sensitivity	Specificity	F1 Score
DRBR	0.6928	0.6712	0.7159	0.6814
ACNN	0.7411	0.7275	0.7628	0.7316
MAMC	0.7716	0.7967	0.8156	0.7598
DLPD	0.8351	0.8147	0.8712	0.8475

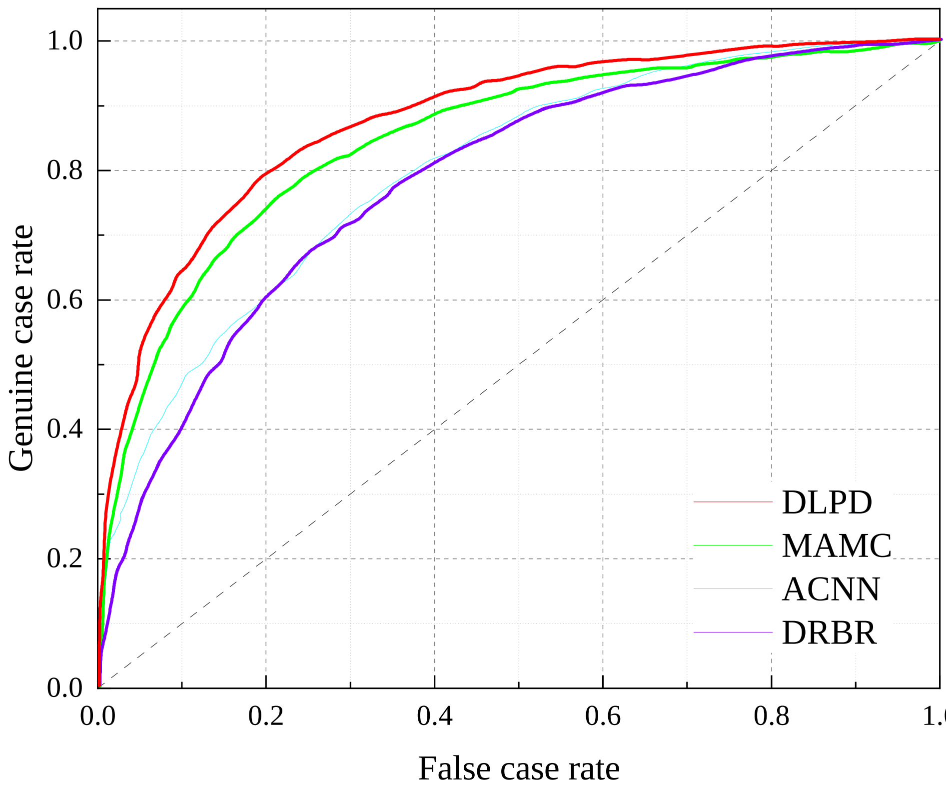

Figure 4. Comparison of ROC curves for different methods

The proposed DLPD method achieves the best performance in Accuracy, Sensitivity, Specificity, and F1 score, which is 20.54%, 21.38%, 21.69%, and 24.38% better than the DRBR method, and 12.68%, 11.99% better than the ACNN method, 14.21%, 16.61%, and 8.22%, 2.26%, 6.82%, 11.54% than the MAMC method, respectively. From this, it can be seen that the DLPD method is significantly better than the DRBR, ACNN, and MAMC methods. This is because DRBR relies on an SVM classifier to classify patients by EEG signals, and the input data is the similarity matrix of EEG, which makes it hard to find the correct hyperplane because of the high feature dimension, resulting in a much worse classification effect than the other three recommendation models based on deep learning. ACNN is based on CNN classifier for diagnosis, and the lack of information in deepening leads to a low accuracy rate. MAMC is a diagnostic method based on GCNN, but it does not fully extract the depth features of EEG, resulting in lower performance metrics than DLPD. DLPD utilizes an optimized attentional mechanism to enable the GCNN to focus on the noteworthy parts, thus improving its performance in diagnosing the important features in the disorder data, which makes the suggested method the best performer in terms of all metrics.

Figure 4 indicates the comparison of the ROC curves of different models, through which the performance is evaluated by the area metric (AUC). Comparison results. The outcome indicates that the DRBR model has the worst performance with an AUC value of 0.71, while the ACNN and MAMC models have AUC values of 0.76 and 0.8, respectively, which are slightly better than the DRBR model. However, the DLPD model proposed in this paper has the best performance with an AUC value of 0.87. This indicates that the DLPD model has stronger classification ability and robustness, and can play a better role in the diagnosis of psychological disorders. Therefore, the DLPD method proposed in this paper has greater potential in practical applications and provides strong support for research and practice in the field of psychological disorder diagnosis.

5.2. **Ablation Experiment.** For the purpose of better validating the effects of the attention module and feature extraction module in the DLPD model of this paper, ablation experiments are conducted on the ABIDE dataset respectively, and two comparative models are designed for the analysis, in which the method of removing the optimized attention module is denoted as DLPD-EAM, and the method of removing the feature extraction module is denoted as DLPD-FE. The evaluation indexes are measured by three metrics, namely, Correlation coefficient R, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) [**?**]. The results of the four model evaluation metrics are given in Table 2, and the results are plotted on a visual bar comparison chart, as indicated in Figure 5.

Table 2. Comparison of classification accuracy of different methods

Method	**R**	**MAE**	**RMSE**
DLPD-EAM	0.6491	0.2451	0.3259
DLPD-FE	0.6215	0.2197	0.2941
DLPD	0.7826	0.0718	0.1128

The results of the ablation experiment indicate that the DLPD model achieves the best performance across all three evaluation metrics, demonstrating the importance of both the attention module and feature extraction module in enhancing classification accuracy. The comparison highlights that removing either module results in a decline in performance, confirming their contribution to the effectiveness of the DLPD model.

As can be seen from Table 2 and Figure 5, the accuracy evaluation indexes of DLPD are significantly better than those of DLPD-EAM and DLPD-FE. Among them, the MAE of the DLPD method is 0.0681, which is lower than that of DLPD-EAM and DLPD-FE by 0.1733 and 0.1479, respectively; the RMSE of the DLPD method is 0.1128, which is lower than that of DLPD-EAM and DLPD-FE by 0.2131 and 0.1813, respectively. Comparing the correlation coefficient R, the R-value of the DLPD method is 0.7826, which improves by 20.57% and 25.92% compared to DLPD-EAM and DLPD-FE, respectively, which suggests that the introduction of the attention mechanism to extract spatio-temporal features of the EEG in the GCN is effective, and the attention mechanism is improved so that DLPD has better fitting effect and classification accuracy.

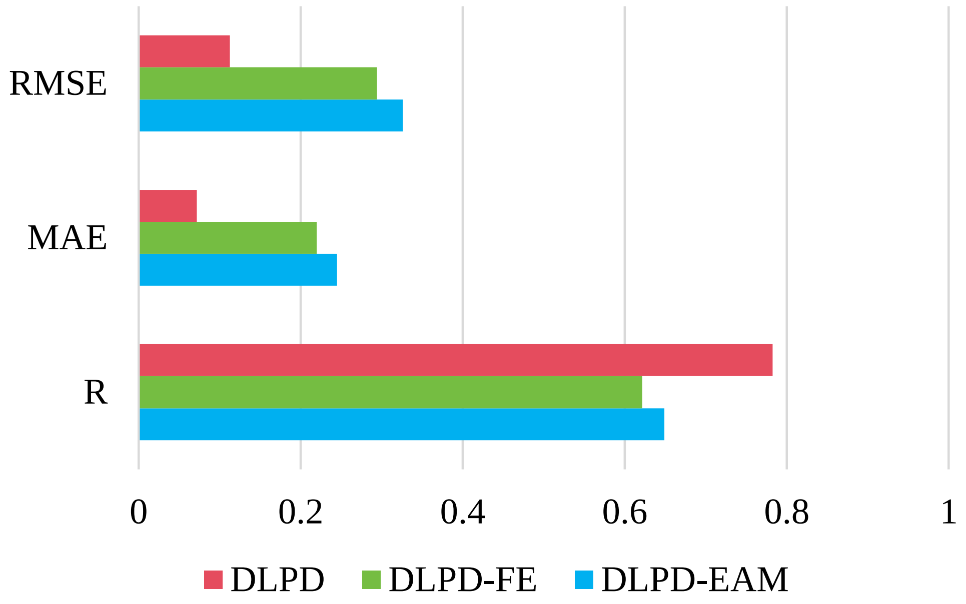

Figure 5. Comparison of ablation experiment results

6. **Conclusion.** Intending to address the issue that the classification accuracy of existing psychological disorder diagnosis methods is not high, this article designs an artificial intelligence-assisted diagnosis method for psychological disorders relying on deep learning. First, the attention mechanism is optimized, and the input features in vertical and horizontal directions are pooled by global average and maximum pooling, respectively, so that the original detail information can be more fully utilized. Then, the original EEG signals are preprocessed by the first-order differential mean absolute value to obtain normalized results, followed by the extraction of spatio-temporal features, weighting the input features in the spatio-temporal channel with the improved attention mechanism, and using the layer-jumping connection to process the weighted data of the two channels to improve the accuracy of the model.

Next, intrinsic feature weighting is performed on different EEG signals to optimize feature extraction. Finally, the classification results are fused with the classification results output from the GCNN module to obtain the final diagnosis prediction. The experimental psychological outcome indicates that the suggested method effectively enhances the accuracy, sensitivity, specificity, and F1 score of classification and can be better applied to the diagnosis of psychological disorders.

REFERENCES

[1] W. W. Eaton, S. S. Martins, G. Nestadt, O. J. Bienvenu, D. Clarke, and P. Alexandre, "The burden of mental disorders," *Epidemiologic Reviews*, vol. 30, no. 1, pp. 1-14, 2008.
[2] S. E. Hyman, "The diagnosis of mental disorders: the problem of reification," *Annual Review of Clinical Psychology*, vol. 6, pp. 155-179, 2010.
[3] L. Gamm, S. Stone, and S. Pittman, "Mental health and mental disorders—A rural challenge: A literature review," *Rural Healthy People*, vol. 2, no. 1, pp. 97-114, 2010.
[4] F. Edition, "Diagnostic and statistical manual of mental disorders," *Am Psychiatric Assoc*, vol. 21, no. 21, pp. 591-643, 2013.
[5] D. Borsboom, "A network theory of mental disorders," *World Psychiatry*, vol. 16, no. 1, pp. 5-13, 2017.
[6] M. Joukamaa, M. Heliövaara, P. Knekt, A. Aromaa, R. Raitasalo, and V. Lehtinen, "Mental disorders and cause-specific mortality," *The British Journal of Psychiatry*, vol. 179, no. 6, pp. 498-502, 2001.

[7] N. M. Petry, K. Zajac, and M. K. Ginley, "Behavioral addictions as mental disorders: to be or not to be?," *Annual Review of Clinical Psychology*, vol. 14, pp. 399-423, 2018.

[8] K. M. Smith, P. F. Renshaw, and J. Bilello, "The diagnosis of depression: current and emerging methods," *Comprehensive Psychiatry*, vol. 54, no. 1, pp. 1-6, 2013.

[9] A. Bangerter et al., "Automated recognition of spontaneous facial expression in individuals with autism spectrum disorder: parsing response variability," *Molecular Autism*, vol. 11, pp. 1-15, 2020.

[10] W.-R. Zhang, A. K. Pandurangi, K. E. Peace, Y.-Q. Zhang, and Z. Zhao, "MentalSquares: a generic bipolar support vector machine for psychiatric disorder classification, diagnostic analysis and neurobiological data mining," *International Journal of Data Mining and Bioinformatics*, vol. 5, no. 5, pp. 532-557, 2011.

[11] Q. Wang, H. Yang, and Y. Yu, "Facial expression video analysis for depression detection in Chinese patients," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 228-233, 2018.

[12] H. Akbari et al., "Depression recognition based on the reconstruction of phase space of EEG signals and geometrical features," *Applied Acoustics*, vol. 179, 108078, 2021.

[13] A. Stolicyn, J. D. Steele, and P. Seriès, "Prediction of depression symptoms in individual subjects with face and eye movement tracking," *Psychological Medicine*, vol. 52, no. 9, pp. 1784-1792, 2022.

[14] C. Jiang, Y. Li, Y. Tang, and C. Guan, "Enhancing EEG-based classification of depression patients using spatial information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 566-575, 2021.

[15] X. Zhang et al., "Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2265-2275, 2019.

[16] A. Shoeibi et al., "Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models," *Frontiers in Neuroinformatics*, vol. 15, 777977, 2021.

[17] M. Niu, B. Liu, J. Tao, and Q. Li, "A time-frequency channel attention and vectorization network for automatic depression level prediction," *Neurocomputing*, vol. 450, pp. 208-218, 2021.

[18] Z. Fei et al., "Deep convolution network based emotion analysis towards mental health care," *Neurocomputing*, vol. 388, pp. 212-227, 2020.

[19] L. He, J. C.-W. Chan, and Z. Wang, "Automatic depression recognition using CNN with attention mechanism from videos," *Neurocomputing*, vol. 422, pp. 165-175, 2021.

[20] D. S. B. A. Hamid, S. Goyal, and P. Bedi, "Integration of deep learning for improved diagnosis of depression using EEG and facial features," *Materials Today: Proceedings*, vol. 80, pp. 1965-1969, 2023.

[21] A. Malhotra, and R. Jindal, "Multimodal deep learning based framework for detecting depression and suicidal behaviour by affective analysis of social media posts," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 21, 164259, 2020.

[22] J. Pan, H. Lin, Y. Dong, Y. Wang, and Y. Ji, "MAMF-GCN: Multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder," *Computers in Biology and Medicine*, vol. 148, pp. 105823, 2022.

[23] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19-46, 2024.

[24] Y. Ma, Y. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121-2133, 2022.

[25] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, 40, 2019.

[26] L. Li, M. Doroslovački, and M. H. Loew, "Approximating the gradient of cross-entropy loss function," *IEEE Access*, vol. 8, pp. 111626-111635, 2020.

[27] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage: Clinical*, vol. 17, pp. 16-23, 2018.

[28] M. Ingalhalikar, S. Shinde, A. Karmarkar, A. Rajan, D. Rangaprakash, and G. Deshpande, "Functional connectivity-based prediction of Autism on site harmonized ABIDE dataset," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 12, pp. 3628-3637, 2021.

[29] T. Chai, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)," *Geoscientific Model Development Discussions*, vol. 7, no. 1, pp. 1525-1534, 2014.