ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on CNN Transformer modeling method with multi head attention in brain imaging for anxiety disorder recognition

Jing Xia[1,2], *Leilei Li*[3], *Wei Wang*[4] *and Wenjing Fu*[1,*]

[1] School of Medicine, Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China
[2] School of Pharmacy, China Medical University, Shenyang 110122, Liaoning, China
[3] Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, Shanghai, China
[4] School of Basic Medicine, Xinjiang Medical University, Urumqi 830017, Xinjiang, China

SUMMARY: *Anxiety disorder is a highly prevalent mental illness, and its current diagnosis mainly relies on symptomatology standards, which can lead to missed diagnosis and misdiagnosis. There is a lack of reliable imaging biomarkers and efficient recognition methods. To solve this problem, this paper proposes a brain image multi head attention CNN Transformer modeling method for anxiety disorder recognition. Firstly, patients with anxiety disorders (GAD and PD) and healthy individuals were selected as the research subjects, and 3.0T magnetic resonance imaging data was collected. After preprocessing, multi-modal features of ALFF, GMV, and FA were extracted and feature maps were constructed; Secondly, design a simplified multi head attention CNN Transformer model that removes the decoder and only optimizes the encoder, combined with Adam optimizer and dropout regularization to improve model performance; Finally, the effectiveness of the model was validated through baseline comparison experiments and ablation experiments. The experimental results show that the proposed model performs stably on three datasets, with accuracy rates of 86.0% and 86.9% in datasets 2 and 3, respectively, significantly better than the baseline algorithm; The ablation experiment showed that after removing CNN or multi head attention modules, the maximum decrease in accuracy reached 13.3%, and each module is crucial to the performance of the model. Research has shown that the proposed model can effectively capture anxiety disorder features in brain imaging, with good effectiveness and generalization, and can provide reliable technical support for early diagnosis of anxiety disorders.*

KEYWORDS: *Anxiety disorder; Multi head attention; Brain imaging; Magnetic resonance imaging; Transformer; CNN*

1 Introduction

Anxiety disorder is a common mental illness with a chronic and fluctuating clinical course, often recurring and prone to comorbidities with other mental illnesses such as depression, substance abuse, and dependence, leading to mental disability, social dysfunction, etc., seriously damaging the patient's life, work, and study [1]. According to epidemiological research, the prevalence of anxiety disorders worldwide is 7.28% (4.76% -10.85%); A study in 2024 shows that the prevalence of anxiety disorders in China is about 5.71%, while in European

*ccaabb202@163.com
https://doi.org/10.65102/is2026739
Published: 30 April, 2026

and American countries, the prevalence of anxiety disorders is even higher, at 13.90% and 18.06% respectively, and women are twice as likely as men. A recent global study on disease burden found that approximately 282 million people worldwide suffer from anxiety disorders, with approximately 43.5 million patients diagnosed each year, making it one of the main causes of the global disease burden. Therefore, anxiety disorders cause a heavy economic and social burden and are a serious public health problem that cannot be ignored [2].

The universal neurobiological model of anxiety disorders suggests that fear acquisition and elimination disorders are the neural basis for the onset and maintenance of pathological anxiety, typically consisting of two stages [3]: (1) fear acquisition, where neutral stimuli are repeatedly paired with fear or unpleasant stimuli. The former does not elicit an intrinsic fear response, also known as a conditioned stimulus, while the latter is called an unconditional stimulus, leading to a conditioned response. (2) Fear subsides when conditioned stimuli repeatedly appear without fear stimuli, and over time, the fear response triggered by conditioned stimuli further diminishes. Fading represents a new, inhibitory memory that competes with conditioned fear memories. When faced with fear stimuli again, which memory dominates determines whether fear will be expressed. At present, the diagnosis of anxiety disorders mainly relies on the criteria of symptomatology, namely the patient's clinical psychiatric symptoms and signs. Such a symptom-based diagnosis can sometimes lead to missed diagnosis or misdiagnosis, resulting in inadequate treatment [4]. Therefore, early diagnosis and providing effective treatment measures can greatly alleviate the serious economic and social burden caused by these diseases. Considering that anxiety disorders are easily overlooked in clinical practice, the main issue lies in the biological background of anxiety disorders, in addition to the disease nature of anxiety disorders themselves and the reluctance of some patients to seek help. Similar to other mental illnesses, we still lack a comprehensive understanding of specific biomarkers that may serve as the basis for the diagnosis and treatment of anxiety disorders. Therefore, it is necessary to continue exploring the potential neurobiological pathological mechanisms of anxiety disorders. At present, the biological and pathological mechanisms of anxiety disorders are not clear, and it is also unclear whether there are shared neuropathological changes among different subtypes of anxiety disorders and their changes with treatment. Therefore, further research is needed on the potential neurobiological pathological mechanisms between different subtypes of anxiety disorders, which is of great significance for the early diagnosis, treatment, and prognosis of anxiety disorders [5].

Evidence from structural MRI studies suggests that patients with anxiety disorders exhibit a wide range of brain structural abnormalities, mainly manifested in the prefrontal cortex, frontal lobe, and limbic system, with the nucleus accumbens and cingulate cortex being particularly prominent. This study adopted a rigorous research design (combining horizontal and vertical methods), selecting patients with different subtypes of anxiety disorders (GAD and PD) who met the inclusion criteria and did not meet the exclusion criteria as the research subjects, and selecting healthy individuals as controls. Clinical and imaging data were collected, and a multi head attention CNN Transformer modeling method for anxiety disorder recognition was designed to observe the relationship between imaging biomarkers and disease status, identify biomarkers with diagnostic significance, and improve the early diagnosis rate of anxiety disorder patients.

2 Related work

2.1 Brain imaging indicators for anxiety disorders

Among the existing biological diagnostic techniques, magnetic resonance imaging (MRI) has

the characteristics of noninvasiveness and easy reproducibility. It includes functional magnetic resonance imaging (fMRI) and structural magnetic resonance imaging (sMRI), which can detect the structural and functional signals of gray matter and white matter, and is the most likely biological indicator for the objective classification of mental disorders [6].

Among them, there are many analysis methods for sMRI, including voxel-based morphometry (VBM), which is one of the most commonly used methods for analyzing differences in brain structure. The analysis process is highly automated and can obtain relatively stable and highly reproducible results. FMRI can be divided into task state fMRI and resting state fMRI based on different signal acquisition methods. Resting state MRI refers to allowing subjects to collect signals in a quiet, nonmoving state, which can be done with their eyes closed or still, but cannot enter a sleep state. The variables of resting state fMRI are relatively easier to control, the acquisition of signals is easier, and variations caused by different tasks can be avoided. Therefore, more and more researchers tend to use resting state fMRI to explore the pathological and intervention mechanisms of psychological problems. The analysis methods of fMRI include Amplitude of Low Frequency Fluctuations (ALFF), Regional Homogeneity (ReHo), etc. [7]. ALFF reflects the intensity of brain activity in a certain area by calculating the BOLD signal strength of spontaneous activity in that area. ReHo describes the functional consistency within a local brain region by analyzing the Kendall harmony coefficient between voxels and surrounding voxels. More and more evidence also supports the existence of abnormalities in cerebellar brain regions in some brain function studies involving AD.

2.2 Classification algorithm based on machine learning

The application of machine learning (ML) in the field of mental illness is becoming increasingly widespread, which can distinguish AD patients from healthy samples and assist doctors in diagnosis. Therefore, the effective detection of objective brain changes using neurobiological methods has aroused great research interest. For example, Kapoor & Goel [8] used SVM to achieve a recognition rate of approximately 69.32% for GAD and Major Depressive Disorder (MDD) based on gray matter and white matter volume; Rezaei et al. [9] achieved a recognition rate of approximately 91.87% for AD patients based on electroencephalography using the Support Vector Machine (SVM) method; Wanderley-Espinola et al. [10] achieved recognition rates of approximately 84.26% and 88.19% for anxiety disorders and GAD, respectively, using SVM to classify features based on brain functional connectivity and independent component analysis. It can be seen that there are currently many studies that use MRI, electroencephalography, and other features for diagnostic classification. By processing brain neurobiological data through different signal processing methods, brain data features can be obtained from different perspectives. The performance of SVM depends on the selection of kernel function, regularization parameters, and other parameters, and choosing the wrong hyperparameters can lead to overfitting or underfitting of the model, resulting in poor performance. The existing research on classification methods such as SVM and logistic regression requires the introduction of more advanced recognition techniques. For other mental illnesses, such as attention deficit hyperactivity disorder (ADHD), classification algorithms for different subtypes of the disease have already been developed. Overall, classification algorithms consist of feature selection, feature extraction, and category decision-making. Feature selection refers to selecting relevant features that are beneficial for ADHD classification from all biological signal features. For example, the classic methods in the field of statistics, minimum absolute value convergence and selection operators, are often used to extract features, which can achieve feature selection while improving robustness through regularization methods; Support vector machine recursive feature elimination is also commonly used, which can filter features based on their importance during the iteration process.

Feature extraction is learning a low dimensional feature space based on the original data space, and mapping the original data features to the learned low dimensional feature space to achieve feature extraction, which can enhance the robustness of the model. Since the mapped feature space is not a direct selection of the original data, but an abstract representation of the original data, this process is called feature extraction. For example, Chivu et al. [11] started from EEG signals and used dynamic sparse coding algorithm to obtain a dictionary space, and then implemented classification in this space. Al-Ezzi et al. [12] used independent component analysis to extract all connectivity patterns from fMRI data to obtain differences between control groups. Robertson & Mortimer [13] greatly improved classification accuracy by using subspace clustering method. The category decision module typically serves as a classifier, using the feature data label information obtained from the feature extraction process for training, and then performing classification on the test set. As the final step, category decision-making directly affects classification accuracy and has a crucial impact on performance. Support vector machines benefit from their excellent performance, and many scholars have improved based on this classifier and applied it to the research field of mental illness classification such as anxiety disorders.

2.3 Classification algorithm based on deep learning

The inspiration for Deep Learning (DL) comes from the structure and function of the human brain, which can use deep neural networks composed of multiple layers of interconnected nodes to process large amounts of data and extract meaningful patterns or features from it. In the process of processing data, only multi-layer neural networks are used, and feature selection, extraction, and category decision-making are generally not explicitly performed. Instead, the above three processes are organically integrated, utilizing the strong fitting ability brought by the multi-level and non-linear nature of Convolutional Neural Networks (CNN). Compared to traditional simple machine learning methods, deep learning can extract and represent complex patterns in data. There are currently many popular deep learning techniques. Dalton et al. [14] used Graph Signal Processing (GSP) technology to classify children with Attention Deficit Hyperactivity Disorder (ADHD) based on resting state brain functional connectivity, with a recognition rate of about 95%; Dam et al. [15] used biomarkers of neurotrophic factors and cytokines, and employed a backpropagation neural network method to classify 117 AD patients and 145 healthy individuals, achieving an accuracy rate of 93.76%; Akbari et al. [16] achieved a recognition rate of over 90% for Alzheimer's disease using GSP technology; Nagano et al. [17] used ensemble clustering method to classify the ALFF index of 953 subjects (189 patients with schizophrenia, 168 patients with bipolar disorder, 209 patients with MDD, and 357 healthy controls) through three steps of dimensionality reduction, ensemble clustering, and optimization. They found that the functional imbalance between the frontal and posterior brain regions could be used as a discriminative feature. It can be seen that using more advanced deep learning techniques can more effectively extract features and help objectively classify anxiety disorders.

3 Dataset construction and preprocessing

3.1 Magnetic resonance data acquisition

This study first used a cross-sectional research method to evaluate the clinical symptoms and brain imaging markers of anxiety disorder (GAD and PD) patients in the baseline period without drug influence. Then, we treated anxiety disorder beneficiaries (GAD and PD) who had never

taken medication with a single SSRI drug (paroxetine) for 4 weeks to explore the changes in imaging biomarkers before and after treatment, as well as their relationship with symptom improvement.

The acquisition of magnetic resonance data was completed on a Philips 3.0T MRI scanner, and brain structural MRI examinations were performed on patients on the day of enrollment and at the end of the 4-week follow-up period. Before the start of the scan, the subjects were informed in detail of the precautions related to the MRI examination, and all metal items carried with them were removed. Soundproof headphones and sponge earplugs were provided to reduce the noise during the scan, and the subjects were required to close their eyes, lie flat and remain awake during the examination. The scanning parameters are as follows: flip angle=8, echo time (TE)=3.68ms, matrix size=240 × 240, repetition time (TR)=1980ms, slice thickness=2mm, field of view (FOV)=240 × 240mm, gap=0.5mm, number of slices=190. The magnetic resonance imaging of patients with anxiety disorders is shown in Figure 1.

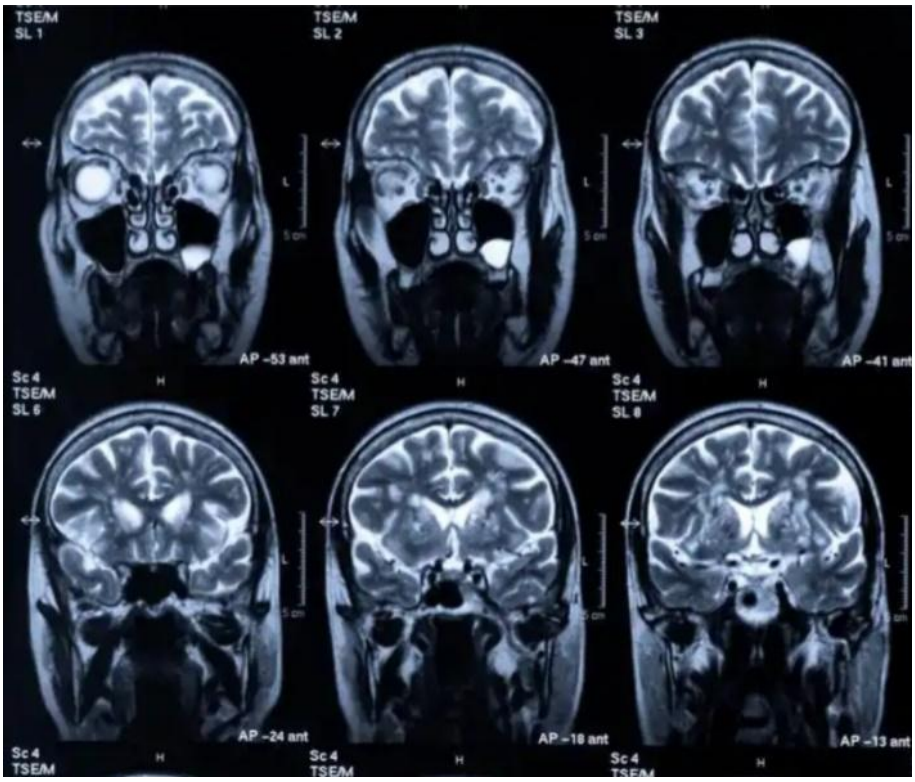

Figure 1: Magnetic Resonance Imaging of Anxiety Disorder Patients

3.2 Preprocessing of structural phase data

Preparation before analysis: (1) Export the raw image data (DICOM format) of each subject from the magnetic resonance scanner, classify them into magnetic resonance scan sequences, organize the T1DICOM folder, use mricron to convert each subject's T1DICOM data format to NIfTI format, and select the reoriented images for subsequent preprocessing. (2) Perform data quality checks on the converted images and eliminate subjects with image quality issues such as artifacts and incomplete scans.

On the Matlab 2022b platform, the Computational Anatomy Toolbox 12 toolkit under the Statistical Parametric Mapping 12 (SPM12) software package (https:/www.fil. ion. ucl. ac.uk/software/spm12/) was used for pre-processing [18]: (1) Based on tissue probability maps (TPM), the brain was divided into gray matter, white matter, and cerebrospinal fluid in individual space. Register the gray matter probability maps of all segmented subjects to the 148

standard brain template of the Montreal Neurological Institute (MNI) for spatial standardization through initial affine transformation; Using DARTEL registration to perform nonlinear transformation on gray matter probability maps, optimize registration between individuals, normalize space, and resample to voxel size images of 1.48 × 1.48 × 1.48mm. (2) Modulate the gray matter probability map using the Jacobian matrix to transform it into a gray matter volume map. After preprocessing is completed, use the CAT12 toolbox to check data quality check sample homogeneity for data quality and sample homogeneity. (3) Using a Gaussian kernel with a full width at half maximum (FWHM) of 9 × 9 × 9mm, the generated grayscale image is smoothed to improve the signal-to-noise ratio, reduce the impact of image mismatch, and facilitate statistical normality. The generated smoothed image is used for subsequent gray matter volume comparison.

For the brain regions of interest, we used the Human Brain Network Atlas, which includes detailed structures of cortical brain regions and subcortical nucleus subregions, totaling 252 fine brain regions in both hemispheres. The amygdala contains 6 subregions, hippocampus contains 5 subregions, insula contains 9 subregions, and cingulate gyrus contains 12 subregions; Based on previous animal and clinical studies on the prefrontal cortex, we focused on the role of the orbitofrontal lobe in anxiety disorders, which includes 13 subregions. We extracted gray matter volume values for each subregion of the 5 brain regions of interest for each participant and conducted subsequent statistical analysis.

3.3 Extraction of anxiety disorder feature data

Many studies have found that ALFF features, GMV features, and FA features are all related to mental disorders. Therefore, in this chapter, for anxiety disorder datasets, ALFF features, GMV features, and FA features are extracted as multimodal features. The specific details of these feature extraction processes are as follows:

(1) ALFF feature extraction. The DPARSFA toolbox was used to perform fast Fourier transform on preprocessed functional magnetic resonance imaging data, transforming the time series of each voxel into the frequency domain [19]. Afterwards, calculate the square root of the BOLD signal amplitude in the low-frequency range (0.01Hz-0.1Hz) for the power spectrum of the rs-fMRI images of the subjects. For each subject, the ALFF at voxel level can be calculated according to formula (1).

$$\mathrm{ALFF} = \sum_{k: f_k \in [0.01, 0.1]} \sqrt{\frac{a_k^2(f_k) + b_k^2(f_k)}{N}} \tag{1}$$

where, f_k represents the numerical value of frequency, while a_k and b_k represent the coefficients corresponding to different numerical frequencies. Afterwards, the Fisher transform is applied to the ALFF values to obtain the zALFF images of each subject at the whole brain level.

To obtain ALFF features: 1) register the zALFF image onto the AAL122 brain atlas; 2) For each corresponding brain region on the AAL122 map, extract the average ALFF value of that brain region as its ALFF feature. Therefore, a 122-dimensional ALFF feature vector can be obtained.

(2) GMV feature extraction. When using the DPARSFA toolbox for T1wMRI preprocessing, a three-dimensional GMV image of the subject will be obtained during the tissue segmentation step [20]. In order to obtain GMV features, the AAL122 brain map is first registered with the three-dimensional GMV to make the image size consistent, which is conducive to subsequent feature extraction by brain region. Then, for each corresponding brain region on the AAL122 map, the average GMV value of that brain region is extracted as the GMV feature of that brain

region. Therefore, a GMV feature vector with a dimension of 122 can be obtained.

(3) FA feature extraction. To obtain the FA feature vector, first calculate the ratio of the anisotropic component of the diffusion tensor in the preprocessed DTI image to the entire diffusion tensor, which can obtain the FA value at the voxel level. The calculation of FA value is shown in formula (2).

$$FA = \sqrt{\frac{1}{2}\frac{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_3-\lambda_1)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}} \tag{2}$$

where, λ_1, λ_2, and λ_3 are the eigenvalues of the diffusion tensor matrix in DTI images. By calculating the FA values of all voxels in the brain, a three-dimensional FA value map can be obtained. Then, the JHU-WM brain map is registered onto the three-dimensional FA value map, and brain regions are divided based on the JHU-WM brain map. The FA value of any brain region is defined as the average FA value of all voxels in that region. By calculating the FA values of all brain regions, a 48-dimensional FA feature vector can be obtained.

After the feature extraction stage, the above-mentioned multimodal feature vectors are obtained, which will be used to construct multimodal graphs. These multimodal graphs will then be input into the classification model for classification.

3.4 Construction of multimodal feature maps for anxiety disorders

For any single modal feature, considering the correlation between different subjects, a subject relationship graph is constructed, denoted as $G=(X,A)$. Each node in the graph represents a subject, $X \in R^{n\times d}$ is the feature matrix, which contains the feature vector set of all nodes in the graph, *n* is the total number of nodes in the graph, and each node has a node feature vector of dimension d. The similarity between subjects is represented by the adjacency matrix $A \in R^{n\times n}$ of the graph, and A_i, A_j, A_{ij} represents the corresponding vertex elements in the *i*-th row, *i*-th column, and *i*-th row and *j*-th column of the adjacency matrix of the graph. Especially, if $A_{ij}=0$ indicates that there is no edge between node *i* and node *j*, otherwise it indicates that there is an edge between node *i* and node *j*, and the value of A_{ij} is the weight of that edge. And edge weights can affect the aggregation ability of node features, and effectively constructing edge weights can improve the classification ability of the model.

Calculate the cosine similarity between the feature vector X_i of any two subjects *i* and the feature vector X_j of subject *j* in the graph. If the cosine similarity is greater than or equal to the threshold ε, the result is retained as an edge of graph *G*, connected to nodes *i* and *j* respectively. Otherwise, it is not retained. The calculation method is shown in formula (3) [21]:

$$A_{ij} = \begin{cases} \dfrac{X_i \cdot X_j}{\|X_i\|_2 \cdot \|X_j\|_2}, & \text{if } i \neq j \text{ and } \dfrac{X_i \cdot X_j}{\|X_i\|_2 \cdot \|X_j\|_2} \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The threshold ε is related to the average number *k* of edges retained by each node in the graph. All cosine similarity values obtained through calculation in the graph are sorted from high to low, and ε is the $n \times k$ elements in the ordered sequence. Set the average retention of the first 10 edges for each node. There are $n \times k$ elements in adjacency matrix *A* that are not 0, and adjacency matrix A also represents the description of the graph structure of graph *G*, as shown in formula (4):

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{n} \end{pmatrix} \tag{4}$$

where, n is the total number of nodes in the graph, which is the number of subjects. The adjacency matrix A is a symmetric matrix, namely $A_{ij} = A_{ji}$.

For any modality, the feature matrix X under that modality and the adjacency matrix A calculated according to the above construction method are used as the two inputs for the next convolutional network.

4 Multi head attention CNN Transformer recognition model for anxiety disorders

4.1 Overall structure of identification model

As shown in Figure 2, the multi head attention CNN Transformer recognition model architecture constructed in this paper removes the decoder part of the Transformer architecture. In the task of predicting anxiety disorder subtypes using measurement curves, the data processing process is essentially a single step regression prediction, rather than a sequence generation task. Therefore, it does not have autoregressive properties and does not rely on the output of the previous step [22]. Based on this, this article has made targeted simplifications to the traditional Transformer architecture, abandoning its decoder part and only retaining and optimizing the encoder structure. The aim is to focus on utilizing the advantages of encoders in feature extraction and global dependency modeling, ultimately achieving efficient and accurate classification and prediction of anxiety disorders.

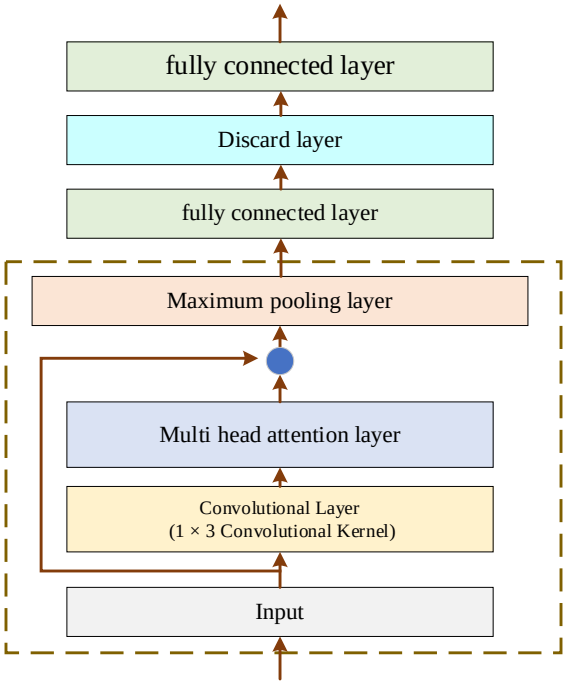

Figure 2: Multi head Attention CNN Transformer Recognition Model Architecture

For a given input data $X \in R^{B\times C\times N}$ (where C represents the number of channels, B represents the batch size, and N represents the number of features), the multi head attention CNN-Transformer recognition model performs layer by layer feature extraction and prediction, mainly including local convolution feature extraction MHAM、 Residual connection and fully connected output. The local feature extraction stage uses one-dimensional convolution to achieve sequence data feature mapping. The size of the convolution kernel is 1 × 4, gradually sliding along the dimensions of time and space of the input data, with a step size of 1, and the number of convolution kernels is 18. The output after convolution is specifically represented as [23]:

$$\boldsymbol{X}_C^{(k)} = \sigma\left(\sum_{c=1}^{c} \boldsymbol{X}_c * \boldsymbol{W}_k^{(c)} + \boldsymbol{b}_k\right) \tag{5}$$

where, $\boldsymbol{W}_k^{(c)}$ is the weight parameter matrix of the k-th convolution kernel on the c-th input channel, where $k = 1,2,\cdots,C'$ and C' is the number of convolution kernels and the number of output channels is set to 18; $\boldsymbol{b}_k$ is the bias term; * represents one-dimensional convolution operation; $\sigma(\cdot)$ represents the ReLU nonlinear activation function. The convolutional layer is mainly used to capture the temporal and spatial features of the local measurement curve data, and the output result $\boldsymbol{X}_c \in \boldsymbol{R}^{B\times C'\times N}$ is transmitted to MHAM to capture the global features.

4.2 Multi head attention mechanism

To further enhance the ability to express information, a multi head attention mechanism is introduced. For the h-th attention head, three sets of parameter matrices $\boldsymbol{W}_h^Q \in \boldsymbol{R}^{D\times d_k}$, $\boldsymbol{W}_h^K \in \boldsymbol{R}^{D\times d_k}$, and $\boldsymbol{W}_h^K \in \boldsymbol{R}^{D\times d_v}$ are used to map the node feature vectors to the query space $\boldsymbol{Q}_h$, key space $\boldsymbol{K}_h$, and value space $\boldsymbol{V}_h$, respectively. The calculation formula is as follows:

$$\boldsymbol{Q}_h = \boldsymbol{H}\boldsymbol{W}_h^Q, \boldsymbol{K}_h = \boldsymbol{H}\boldsymbol{W}_h^K, \boldsymbol{V}_h = \boldsymbol{H}\boldsymbol{W}_h^V \tag{6}$$

where, d_k and d_v respectively represent the dimensions of the key vector and value vector.

Next, the attention score is calculated, and the attention weight matrix is obtained by dot product operation combined with softmax function normalization [24]:

$$A^h = \mathrm{softmax}\left(\boldsymbol{Q}_h \boldsymbol{K}_h^T / \sqrt{d_k}\right) \tag{7}$$

$A^h \in \boldsymbol{R}^{N\times N}$ represents the attention weight matrix of the h-th head, reflecting the correlation between nodes. Use these attention weights to weight and sum the value vectors to obtain the output of each head:

$$head_h = \boldsymbol{A}^h \boldsymbol{V}_h \tag{8}$$

Concatenate the outputs of all heads and finally obtain the output of the multi head attention mechanism through linear transformation $W^O \in \boldsymbol{R}^{hd_v\times D'}$:

$$MH(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat\left(head_1, head_2, \cdots, head_h\right)\boldsymbol{W}^o \tag{9}$$

where, h represents the number of attention heads, and D' is the dimension of the final output features.

On the basis of applying different attention to anxiety disorder types through the multi head attention mechanism, the graph convolutional layer further updates the node features. For each node in the graph convolutional layer, its features are propagated and aggregated through the features of neighboring nodes to obtain an updated representation. For each node in the anxiety disorder type relationship graph, its updated feature representation is [25]:

$$H_A^{(l+1)} = \sigma\left(\frac{1}{\sqrt{\tilde{D}}}\tilde{A}\frac{1}{\sqrt{\tilde{D}}}H_A^{(l)}W^{(l)}\right) \tag{10}$$

where, $H_A^{(l+1)}$ represents the node feature matrix after l layers of graph convolution, and $\tilde{A}$ is the adjacency matrix with self-connections added. $\tilde{D}$ is the degree matrix with the self-connected adjacency matrix $\tilde{A}$ added, and $W^{(l)}$ represents the weight matrix of the *l*-th layer. σ is a nonlinear activation function. The output of the graph convolutional layer is the node feature matrix that has been propagated and updated through multiple layers, which is used for subsequent saliency evaluation and feature extraction steps.

After the multi head self-attention layer and graph convolutional layer, the saliency evaluation layer is responsible for calculating the saliency score of each anxiety disorder type node and using it in the subsequent feature extraction process to select the most representative anxiety disorder type: (1) average the embeddings of all anxiety disorder types in the set of anxiety disorder types to generate a cluster level embedding, which helps to capture the overall information of the entire anxiety disorder type family and provides reference for calculating the saliency score of each anxiety disorder type. (2) The significance score S_i of anxiety disorder type T_i is calculated using cluster embedding *C* and embedding Y_i for each anxiety disorder type. The calculation formula is [26]:

$$S_i = v^T \tanh\left(W_s Y_i + W_c C\right) \tag{11}$$

where, W_s, W_c, and v are model parameters, and tanh is the activation function. To ensure comparability of scores, significance scores are normalized using softmax operation, so that the sum of significance scores for all anxiety disorder types is 1, reflecting their relative importance in the entire anxiety disorder type.

After calculating the significance score of each anxiety disorder type node in the significance evaluation layer, the feature extraction layer is responsible for selecting the most representative anxiety disorder type based on these scores to form the final features, which includes two main steps: anxiety disorder type selection and diversity optimization.

Firstly, rank all anxiety disorder type nodes based on their significance scores, and select the anxiety disorder type corresponding to the top k nodes with the highest scores as candidate features. *k* is a pre-set parameter used to control the length of features. To ensure diversity of features, further screening of anxiety disorder types among candidate features is conducted. Using the greedy algorithm to gradually select anxiety disorder types, each time selecting the anxiety disorder type with the smallest overlap with the existing anxiety disorder types in the current features, until the predetermined feature length is reached or all candidate anxiety disorder types are considered.

The model uses Cross Entropy Loss to measure the difference between the predicted probability distribution and the true distribution. For each anxiety disorder type cluster, the model predicts a significance score that represents the probability of the anxiety disorder type being selected as a feature. By applying the softmax function, these scores are normalized into a probability distribution. Subsequently, the cross-entropy loss function is used to compare the

probability distribution predicted by the model with the distribution corresponding to the manually annotated features. The mathematical expression of the loss function is as follows [27]:

$$\text{Loss} = -\sum_{d \in D} \sum_{v_i \in d} \text{Rouge}(v_i) \log S_i \tag{12}$$

$$\text{Rouge}(v_i) = \frac{R_1(v_i) + R_2(v_i)}{2} \tag{13}$$

In the above formula, *D* represents the set of all anxiety disorder type clusters, *d* is a group of anxiety disorder types in a certain anxiety disorder type cluster, and v_i represents a certain anxiety disorder type belonging to anxiety disorder type cluster *d*. R_1 and R_2 represent the ROUGE-1 and ROUGE-2 scores of manually annotated reference features, respectively, while S_i is the significance score predicted by the model after softmax normalization. $\text{Rouge}(v_i)$ is the average of ROUGE-1 and ROUGE-2 scores between v_i and manually annotated reference features, used to measure the quality of this type of anxiety disorder.

4.3 Weight optimization plan

The goal of model training is to optimize model parameters by minimizing the loss function, thereby improving the accuracy of predicting significant scores for anxiety disorder types. This article uses the Adam optimizer to update the model parameters and designs an adaptive learning rate adjustment strategy to dynamically adjust the learning rate based on the model's performance on the validation set [28]. When the model does not show significant performance improvement over multiple consecutive training cycles, the learning rate will be appropriately reduced to avoid stagnation at the local minimum of the loss function, thereby improving the convergence effect of training. To alleviate the overfitting problem, we introduced dropout regularization technique during the training process. In actual training, the model optimizes parameters through multiple iterations. Each iteration calculates the gradient of the current loss function and updates the model parameters based on the gradient information. The training process will continue until the loss function converges or reaches the preset maximum number of iterations.

This article uses the Adam (Adaptive Moment Estimation) optimization algorithm to update the model training weights in small batches. This algorithm combines the core advantages of Momentum and Adaptive Learning Rate (RMSprop), and can dynamically adjust the model weights by smoothing the speed and magnitude of weight and deviation oscillations through exponential weighted averaging and deviation optimization strategies. This is beneficial for the rapid convergence of the model and avoids problems such as training process stagnation and model generalization ability damage caused by gradient disappearance or explosion.

5 Experimental analysis

5.1 Experimental setup

In this section, in order to evaluate the impact of experimental paradigms on the effectiveness of predicting anxiety disorder types and assess the performance of the multi head attention CNN Transformer model in predicting anxiety disorder types, we conducted experiments on the dataset proposed in this paper. Three datasets each contain EEG data from 27 patients with anxiety disorders, as well as LSAS scores before and after intervention. The anxious subjects

underwent an 8-week cognitive-behavioral therapy intervention after the ERP experiment, and we collected the LSAS scores of the patients before and after the intervention treatment.

Firstly, to evaluate the performance of the multi head attention CNN Transformer model, we compared it with CNN, BiLSTM, CNN-BiLSTM-attention (CBATT), CNN-BiLSTM-Transformer (CBTransformer) A comparative experiment was conducted. This study used four indicators to evaluate the performance of the model: accuracy (Acc), precision (Prec), recall (Recall), and F1 score. Accuracy reflects the proportion of correctly classified samples in the model; The accuracy reflects the proportion of positive samples predicted as anxiety disorder types that are actually positive; The recall rate measures the proportion of positive samples of anxiety disorder types correctly identified by the model among all positive samples; The F1 score is the harmonic mean of accuracy and recall, used to comprehensively evaluate the performance of anxiety disorder type prediction models. These evaluation indicators are defined as follows:

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{14}$$

$$Prec = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = \frac{2 \times (\mathrm{Pre} \times \mathrm{Recall})}{\mathrm{Prec} + \mathrm{Recall}} \tag{17}$$

In anxiety disorder classification tasks, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are key indicators for evaluating the performance of anxiety disorder classification models. Where, TP+FP+TN+FN equals the total sample size of anxiety disorder patients. TP+TN represents the number of correctly predicted anxiety disorder patient samples by the model, TP+FN represents the actual number of positive anxiety disorder patient samples, TN+FP represents the actual number of negative anxiety disorder patient samples, and TP+FP represents the total number of predicted positive anxiety disorder patient samples. These indicators together form the basis for evaluating classification performance.

5.2 Result analysis

In order to verify the effectiveness of the CNN Transformer modeling method with multi head attention in brain imaging for anxiety disorder recognition, the experimental results of the baseline comparison algorithm and the proposed method on the three selected datasets are shown in Table 1.

Table 1: Comparison of Algorithm Performance Results

Dataset	Model	Acc	prec	recall	F1
Dataset 1	CNN	0.738	0.780	0.742	0.731
	BiLSTM	0.774	0.843	0.725	0.715
	CBATT	0.842	0.831	0.854	0.836
	CBTransformer	0.831	0.826	0.806	0.796
	BiHDM	0.847	0.856	0.853	0.880
	EERN	0.858	0.854	0.826	0.848
	ECA-CRNN	0.882	0.876	0.892	0.895
	Ours	0.815	0.824	0.835	0.815
Dataset 2	CNN	0.710	0.728	0.731	0.716
	BiLSTM	0.726	0.752	0.727	0.729
	CBATT	0.773	0.764	0.790	0.783
	CBTransformer	0.791	0.801	0.802	0.836
	BiHDM	0.814	0.823	0.798	0.830
	EERN	0.839	0.837	0.857	0.853
	ECA-CRNN	0.827	0.842	0.806	0.847
	Ours	0.860	0.869	0.876	0.870
Dataset 3	CNN	0.732	0.751	0.742	0.756
	BiLSTM	0.761	0.786	0.770	0.784
	CBATT	0.792	0.798	0.825	0.817
	CBTransformer	0.814	0.834	0.827	0.809
	BiHDM	0.838	0.848	0.825	0.831
	EERN	0.831	0.856	0.820	0.837
	ECA-CRNN	0.845	0.837	0.847	0.809
	Ours	0.869	0.874	0.873	0.886

According to the experimental results in Table 1, it can be seen that: (1) in dataset 1, the experimental result of the Acc index of the proposed method is 0.815, which is lower than the 0.882 of ECA-CRNN, but is 7.7 and 4.1 percentage points higher than the basic algorithms CNN (0.738) and BiLSTM (0.774), respectively, reflecting the advantages of multi head attention fusion structure; The experimental results of the prec index of the proposed method reached 0.824, which is higher than CNN (0.780) and BiLSTM (except 0.843); The experimental result of the recall index of the proposed method is 0.835, which is 2.9 percentage points higher than CBTransformer (0.806); The experimental result of the F1 value index of the proposed method is 0.815, which is lower than BiHDM (0.880) and ECA-CRNN (0.895), but the overall performance is better than traditional deep learning models, indicating that the proposed method can preliminarily capture anxiety disorder features in brain images. (2) In dataset 2, the experimental result of the Acc index of the proposed method is 0.860, which is 2.1 percentage points higher than the optimal baseline EERN (0.839) and 15 percentage points higher than the basic model CNN (0.710); The experimental results of the prec index of the proposed method reached 0.869, which is higher than ECA-CRNN (0.842) and BiHDM (0.823); The experimental result of the recall index of the proposed method is 0.876, which is 1.9 percentage points higher than EERN (0.857); The experimental result of the F1 value index of the proposed method is 0.870, second only to CBTransformer (0.836) and BiHDM (0.830), significantly better than other algorithms, verifying the adaptability of the proposed method on this dataset. (3) In dataset 3, the experimental result of the Acc index of the proposed method is 0.869, which is 2.4 percentage points higher than ECA-CRNN (0.845); The experimental

result of the prec index of the proposed method reached 0.874, which is higher than all baseline algorithms; The experimental result of the recall index of the proposed method is 0.873, which is 4.8 percentage points higher than BiHDM (0.825); The experimental result of the F1 value index of the proposed method is 0.886, far exceeding other algorithms and improving by 4.9 percentage points compared to the suboptimal EERN (0.837).

Overall, the proposed method shows stable performance on three datasets, particularly outstanding in datasets 2 and 3. Although slightly inferior to ECA-CRNN in dataset 1, it still shows significant improvement compared to traditional models, demonstrating its good effectiveness and generalization in identifying anxiety disorders in brain imaging.

5.3 Ablation experiment

To verify the effectiveness of each module of the model, this study designed a systematic ablation experiment. The experiment evaluated the impact of gradually removing various modules from the multi head attention CNN Transformer model on classification performance, as shown in Table 2.

Table 2: Classification results of ablation experiments

Removed modules	Dataset 1	Dataset 2	Dataset 3
Proposed method	81.5%	86.0%	86.9%
Multi-head attention	78.3%	74.5%	74.3%
CNN	77.6%	73.7%	73.6%
Preprocessing of structural phase data	79.1%	74.2%	74.2%
Multi modal feature map construction	80.3%	74.7%	74.7%

According to the experimental results in Table 2, it can be seen that the proposed method performs stably on three datasets: (1) In dataset 1, the experimental result of the Acc index of the proposed method is 0.815, which is 7.7 and 4.1 percentage points higher than CNN and BiLSTM, respectively; In dataset 2, the experimental result of the Acc index of the proposed method is 0.860, which is 2.1 percentage points higher than the optimal baseline EERN; In dataset 3, the experimental result of the Acc index of the proposed method is 0.869, which is 2.4 percentage points higher than ECA-CRNN and overall better than traditional models.

The ablation experiment results showed that each module had a significant impact on the performance of the model. Compared with the complete model (Proposed method), (1) after removing multi head attention, the accuracy of the three datasets decreased by 3.2%, 11.5%, and 12.6%, respectively, with the most significant decrease, indicating that multi head attention can effectively capture key features of brain images; (2) After removing the CNN module, the accuracy decreased by 3.9%, 12.3%, and 13.3% respectively, reflecting the core role of CNN in feature extraction; (3) After removing the preprocessing of structural phase data and the construction of multimodal feature maps, the accuracy decreased by 2.4%, 11.8%, 12.7%, and 1.2%, 11.3%, 12.2%, respectively, indicating that preprocessing and feature construction can improve feature quality.

Overall, the proposed method demonstrates good effectiveness and generalization in baseline comparison, and ablation experiments further validate the necessity of various modules such as multi head attention and CNN.

6 Conclusion

This study proposes a multi head attention CNN Transformer modeling method for anxiety

disorder recognition, aiming to explore the potential neurobiological pathological mechanisms of anxiety disorders, construct an efficient brain image recognition model, and provide technical support and theoretical reference for the early objective diagnosis of anxiety disorders, in response to the problems of missed diagnosis and misdiagnosis, lack of reliable imaging biomarkers, and efficient recognition methods that rely on symptomatology standards for clinical diagnosis of anxiety disorders. This study achieved efficient identification of anxiety disorders through the fusion of multimodal features from brain imaging and the construction of deep learning models, providing a new technological path for objective diagnosis of anxiety disorders and offering imaging biomarkers as a reference for exploring the neurobiological pathological mechanisms of anxiety disorders.

However, this study still has certain limitations: (1) The sample size selected for the study is relatively limited, and only two subtypes of anxiety disorders, GAD and PD, were studied. In the future, the sample size can be expanded to include more subtypes such as social anxiety disorder and specific phobias, further verifying the generalization ability and universality of the model. (2) This study is mainly based on structural phase brain imaging data for feature extraction and model training. Subsequently, multi-source data such as functional phase brain imaging and electroencephalogram can be fused to construct a multimodal data fusion recognition model, mining richer disease feature information and improving the recognition performance of the model. (3) Thirdly, this study only completed the classification and recognition of anxiety disorders, without delving into the correlation between key brain regions and pathological mechanisms that the model focuses on during the recognition process. In the future, interpretable deep learning techniques can be used to analyze the decision-making process of the model, clarify the characteristics of brain regions that play a key role in the recognition of anxiety disorders, and provide more accurate basis for the study of the pathological mechanisms of anxiety disorders. (4) This study did not combine the model with the clinical diagnostic process for validation. Subsequent clinical controlled experiments can be conducted to apply the proposed model to clinical auxiliary diagnostic scenarios, optimize its practicality and ease of use, and promote its clinical translation.

In the future, the research team will continue to explore in depth around the above directions, continuously improve the brain image recognition system for anxiety disorders, and provide more comprehensive technical support for early diagnosis, intervention, and prognosis evaluation of anxiety disorders.

References

[1] Szuhany K L, Simon N M. Anxiety disorders: a review[J]. Jama, 2022, 328(24): 2431-2445.

[2] Salari N, Heidarian P, Hassanabadi M, et al. Global prevalence of social anxiety disorder in children, adolescents and youth: A systematic review and meta-analysis[J]. Journal of Prevention, 2024, 45(5): 795-813.

[3] Javaid S F, Hashim I J, Hashim M J, et al. Epidemiology of anxiety disorders: global burden and sociodemographic associations[J]. Middle East Current Psychiatry, 2023, 30(1): 44.

[4] Rapee R M, Creswell C, Kendall P C, et al. Anxiety disorders in children and adolescents: A summary and overview of the literature[J]. Behaviour research and therapy, 2023, 168: 104376.

[5] Butler M I, Bastiaanssen T F S, Long-Smith C, et al. The gut microbiome in social anxiety disorder: evidence of altered composition and function[J]. Translational Psychiatry, 2023, 13(1): 95.

[6] Das K P, Gavade P. A review on the efficacy of artificial intelligence for managing anxiety disorders[J]. Frontiers in Artificial Intelligence, 2024, 7: 1435895.

[7] DeGeorge K C, Grover M, Streeter G S. Generalized anxiety disorder and panic disorder in adults[J]. American family physician, 2022, 106(2): 157-164.

[8] Kapoor A, Goel S. Prediction of Anxiety Disorders using Machine Learning Techniques[C]//2022 IEEE Bombay Section Signature Conference (IBSSC). IEEE, 2022: 1-6.

[9] Rezaei S, Gharepapagh E, Rashidi F, et al. Machine learning applied to functional magnetic resonance imaging in anxiety disorders[J]. Journal of Affective Disorders, 2023, 342: 54-62.

[10] Wanderley Espinola C, Gomes J C, Mônica Silva Pereira J, et al. Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study[J]. Research on Biomedical Engineering, 2022, 38(3): 813-829.

[11] Chivu A, Pascal S A, Damborská A, et al. EEG microstates in mood and anxiety disorders: a meta-analysis[J]. Brain topography, 2024, 37(3): 357-368.

[12] Al-Ezzi A, Al-Shargabi A A, Al-Shargie F, et al. Complexity analysis of EEG in patients with social anxiety disorder using fuzzy entropy and machine learning techniques[J]. IEEE Access, 2022, 10: 39926-39938.

[13] Robertson C, Mortimer A. Quantitative EEG (qEEG) guided transcranial magnetic stimulation (TMS) treatment for depression and anxiety disorders: An open, observational cohort study of 210 patients[J]. Journal of Affective Disorders, 2022, 308: 322-327.

[14] Dalton S D P, Cooper H, Jennings B, et al. The empirical status of implicit emotion regulation in mood and anxiety disorders: A meta-analytic review[J]. Journal of Affective Disorders, 2025, 380: 256-269.

[15] Dam S, Maurel P, Coloigner J. Graph wavelet packets for the classification of brain data in anxiety and depression[C]//2024 32nd European Signal Processing Conference (EUSIPCO). IEEE, 2024: 1436-1440.

[16] Akbari M, Seydavi M, Rahmati S, et al. Psychometric validation of the Persian versions of the Generic Scale of Phubbing (GSP) and the Generic Scale of Being Phubbed (GSBP): factor structure, reliability, and construct validity in an Iranian community sample[J]. Current Psychology, 2026, 45(1): 83.

[17] Nagano T, Kurita K, Yoshida T, et al. Comparison of Resting-State Functional Connectivity Between Generalized Anxiety Disorder and Social Anxiety Disorder: Differences in the Nucleus Accumbens and Thalamus Network[J]. Brain Connectivity,

2024, 14(8): 445-456.

[18] Wlad M, Frick A, Engman J, et al. Dorsal anterior cingulate cortex activity during cognitive challenge in social anxiety disorder[J]. Behavioural Brain Research, 2023, 442: 114304.

[19] Thomas P J, Leow A, Klumpp H, et al. Default mode network hypoalignment of function to structure correlates with depression and rumination[J]. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2024, 9(1): 101-111.

[20] Thomas P J, Leow A, Klumpp H, et al. Default mode network hypoalignment of function to structure correlates with depression and rumination[J]. medRxiv, 2022: 2022.09. 02.22279551.

[21] Yogee S Y T, Savitha H P, Shankar M V, et al. Ayurveda Management of Generalized Anxiety Disorder[J]. Journal of Ayurveda, 2025, 19(3): 296-300.

[22] Hackenberg B, Döge J, O'Brien K, et al. Tinnitus and its relation to depression, anxiety, and stress—a population-based cohort study[J]. Journal of clinical medicine, 2023, 12(3): 1169.

[23] Ellwardt E, Muthuraman M, Gonzalez-Escamilla G, et al. Network alterations underlying anxiety symptoms in early multiple sclerosis[J]. Journal of neuroinflammation, 2022, 19(1): 119.

[24] İmamoğlu G, Keyvan A. Investigation of the relationships between phubbing, attachment styles and social anxiety variables in adults[J]. Am J Humanit Soc Sci Res, 2022, 6(5): 131-143.

[25] Rodriguez R L, van Zuilen M H, Gould C E, et al. The Geriatric Scholars Program: expanding the workforce equipped to care for the mental health needs of older veterans[J]. Academic psychiatry, 2025, 49(5): 474-478.

[26] Sahu N K, Gupta S, Lone H. Wearable technology insights: Unveiling physiological responses during three different socially anxious activities[J]. ACM Journal on Computing and Sustainable Societies, 2024, 2(2): 1-23.

[27] Sharma S, Ghantasala G S P, Gupta V, et al. Analyzing Employees' Mental Health Using Machine Learning Algorithms[C]//2025 12th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2025: 1-5.

[28] Mofrad L, Hall D, Tiplady A, et al. Making friends with uncertainty: evaluation of a group intervention targeting intolerance of uncertainty in a Talking Therapies service[J]. the Cognitive Behaviour Therapist, 2025, 18: e43.