An Autism Classification Method based on Deep Neural Network and Attention Mechanism

XiaoXu Ma, Xiang Guo

Autism Spectrum Disorder (ASD) is a developmental disorder whose incidence has been increasing, significantly impacting patients' daily lives. Recent research trends focus on leveraging large-scale, multi-center neuroimaging datasets to enhance clinical applicability and statistical validity. However, the lack of reliable biomarkers and data heterogeneity across datasets limit classification effectiveness. This paper proposes a resting-state functional MRI (rs-fMRI) based approach to improve classification accuracy for ASD by integrating multi-site data.

In this study, a MultiModal Deep Attention (MMDA) model is introduced for ASD identification, which effectively combines rs-fMRI features with demographic characteristics through three modules: feature extraction, feature learning, and deep perception. The feature extraction module uses autoencoders to clean rs-fMRI time series data; the feature learning module employs multi-head attention mechanisms and convolutional neural networks to uncover intrinsic data structures; and the deep perception module integrates multimodal features to produce final ASD classifications. Simulation results demonstrate that the MMDA model outperforms benchmark algorithms in ASD identification.

In summary, this research advances ASD diagnostic accuracy by integrating multimodal data with neural networks, offering a promising tool for objective auxiliary diagnosis and providing new insights into ASD pathomechanisms.

***Index Terms*—Autism Spectrum Disorder; resting-state fMRI; domain adaptation; feature representation learning; data heterogeneity; deep neural networks.**

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder characterized primarily by persistent deficits in social interaction and communication, along with restricted, repetitive patterns of behavior and interests. According to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), individuals with ASD often exhibit impairments in social-emotional reciprocity, abnormalities in nonverbal communicative behaviors used for social interaction, and difficulties in developing and maintaining relationships, imposing significant burdens on families and society[1]. Functional magnetic resonance imaging (fMRI), as a crucial brain imaging technique, offers a non-invasive way to reflect the activity of different brain regions, providing potential for early detection of ASD[2].

Despite offering objective biomarkers, traditional neuroimaging analysis methods have certain limitations. For instance, analyzing imaging data requires substantial manual intervention[3], which is time-consuming and prone to human error. Additionally, given the complexity and diversity of brain structure and function differences in ASD patients, traditional methods may struggle to effectively identify all potential abnormalities. To address this issue, machine learning techniques have been increasingly applied in the medical field. Compared to traditional clinical diagnostic methods, machine learning approaches can handle large volumes of complex imaging data and provide more objective diagnostic outcomes[4]. In the classification and prediction of ASD, machine learning methods can accurately extract important biomarkers from rs-fMRI data, offering robust support for early diagnosis and personalized treatment.

At present, ASD classification methods based on machine learning have made significant research progress[5]. Traditional machine learning algorithms, such as support vector machines (SVM), ridge regression algorithms (Ridge Regression) and random forests, have been widely used in ASD classification tasks[6]. These algorithms can identify the differences between different categories by modeling training data and make predictions in new data. However, although traditional machine learning methods can improve the accuracy of diagnosis to a certain extent, due to their limited processing capabilities for high-dimensional data, the classification effect is often affected by the selection and processing methods of data features[7]. In order to solve these problems, neural network technology, including models such as deep neural networks (DNN) and graph convolutional networks (GCN), has been increasingly widely used in ASD classification in recent years[8]. Compared with traditional machine learning methods, neural networks have stronger self-learning ability, can automatically extract more complex features from raw data, and can process larger and more complex imaging data. This makes deep learning show great potential in the classification task of ASD[9]. Studies have shown that compared with traditional machine learning methods, neural network methods can better identify the differences between ASD patients and normal controls, and have achieved superior performance in classification tasks.

Manuscript received March 04, 2025
XiaoXu Ma, School of computer science and technology, Tiangong University, Tianjin, China.
Xiang Guo, School of computer science and technology, Tiangong University, Tianjin, China.

In the study of applying neural networks to ASD identification, due to the insufficient sample size of subjects at the data collection point, the model cannot fully learn the complex feature representation related to ASD, resulting in underfitting of the neural network model[10]. In view of this, the method in this paper makes full use of demographic information and rs-fMRI structural information, optimizes feature extraction and feature learning methods, and aims to enhance the learning ability and generalization performance of the model, thereby improving the overall performance of the ASD identification model. The main contributions are as follows:

Based on the BP (Back Propagation) neural network model, this paper proposes a multimodal deep attention model (MultiModal Deep Attention, MMDA). MMDA first uses an autoencoder to reconstruct the subjects' rs-fMRI data to remove the heterogeneity between subjects. Then the model combines the attention mechanism with the CNN network, so that the model can capture a subset of features with strong correlation, thereby enhancing the understanding of key information. Finally, the learned features are input into the deep perception network to obtain the ASD classification results. In addition, MMDA adds phenotypic information features in the feature selection stage to improve the robustness of the model and improve the performance of ASD recognition.

The MMDA model as a whole mainly includes three core modules: Feature Extraction Module (FEM), Feature Learning Module (FLM) and Deep Perceptual Network (DPN). In the feature extraction stage, an autoencoder is used to perform deep feature mining on the preprocessed raw input data to generate a 200-dimensional feature vector representation. Then it enters the feature learning stage, which mainly includes a multi-head attention module and a convolutional neural network (CNN) module. Through the multi-head attention mechanism, the model can learn the most relevant feature subsets and enhance the understanding of key information[11], while by using the CNN network, the model can extract local information. Finally, in the classification stage, the enhanced features are connected with the phenotypic information to construct an enhanced feature set, which is then passed to the deep perceptual network. The network consists of two hidden layers and one output layer. The hidden layer uses the ReLU activation function and the output layer uses the Sigimod activation function. The deep perceptual network gradually abstracts the enhanced features and finally makes a classification decision to obtain the ASD recognition result.

II. MATERIAL AND METHODS

A. *ABIDE Database: rs-fMRI and Phenotypic Data*

This study centers on the ABIDE database, which comprises MRI and phenotypic data collected from 20 sites worldwide. Our dataset includes rs-fMRI and phenotypic data from 505 individuals with ASD and 530 typically developing (TC) controls, totaling 1035 subjects[12]. We chose a larger sample size to enhance the probability of detecting site-specific effects, despite the increased challenge of greater heterogeneity among subjects.

The ABIDE database provides comprehensive phenotypic information, including variables such as sex, age, full IQ (FIQ) scores, and handedness (left, right, or ambidextrous). Notably, the type of fMRI scanner and the duration of individual fMRI scans varied across sites, contributing to the observed heterogeneity.

To benchmark against state-of-the-art (SOTA) methods, we used the same pre-processed fMRI data from ABIDE (http://preprocessed-connectomes-project.org/abide/). This approach ensures consistency and comparability with previous studies while addressing the complexities introduced by the diverse data sources.

B. *Feature extraction module*

When dealing with complex and diverse biomedical data, data from different subjects often show significant heterogeneity. In order to effectively integrate and utilize this information, this section designs a feature reconstruction network based on an autoencoder. The network is optimized for the functional link features of rs-fMRI data. The following is a schematic diagram of the feature extraction module:

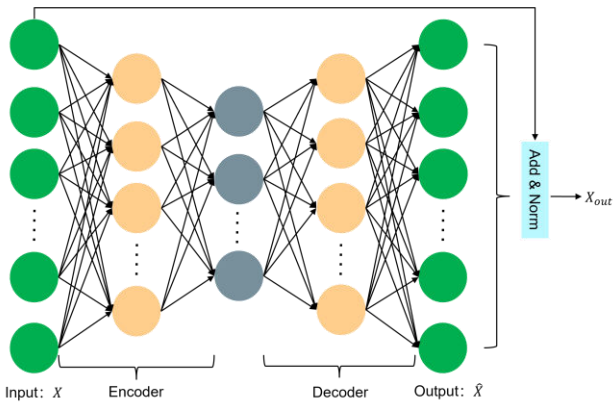

Fig.1.Feature extraction module

In the feature extraction step, due to the strong heterogeneity between data, this design uses autoencoder technology to enhance and reconstruct the 200-dimensional functional link features of rs-fMRI data. The autoencoder maps the original data to a low-dimensional space, and then restores it from the low-dimensional space to the high-dimensional space to obtain the reconstructed data. The reconstructed data not only retains the key information of the original data, but also removes noise and redundant information. The autoencoder in this design introduces a residual structure to prevent the gradient vanishing problem in deep network training and enhance the expression ability of the model. The reconstructed data is processed by RmsNorm to ensure that different features have the same dimension and distribution range.

C. *Feature Learning Module*

The feature learning module integrates the Transformer's multi-head self-attention mechanism and convolutional neural network to improve the learning efficiency of multimodal feature vectors. The model is designed to integrate the information of global dependencies and local

spatial structures, thereby enhancing the understanding and representation capabilities of complex patterns.

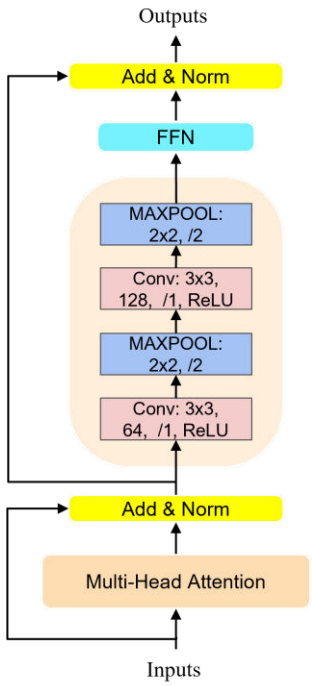

Fig.2.Feature Learning Module

First, data is processed by a Multi-Head Self-Attention (MHSA) layer to capture long-range dependencies among sequence elements. MHSA transforms the input into Query, Key, and Value matrices, computing their similarities to highlight important information across different positions. Subsequently, to extract more refined spatial features, the attention-processed data passes through a convolutional network module comprising two convolutional layers and two max-pooling layers. The first convolutional layer uses 64 filters of size 3x3, followed by a 2x2 max-pooling operation with a stride of 2. The second convolutional layer employs 128 filters of size 3x3, with an identical pooling configuration to reduce dimensions.Finally, to reconstruct and compress feature representations, the processed features are fed into a Feed-Forward Network (FFN). As part of the encoder, the FFN learns nonlinear transformations of the input features. Throughout this process, residual connections and RMS normalization are applied to stabilize training and facilitate gradient flow, ultimately yielding optimized feature representation.

In summary, the feature learning module enhances the model's understanding of complex patterns in the input data. By incorporating residual connections and normalization techniques, it addresses the vanishing gradient problem in deep networks, thereby improving overall model performance.

D. Deep Perceptual Network

This section introduces a Deep Perception Network (DPN) designed to enhance the understanding of complex patterns through multimodal data fusion. The network architecture comprises two hidden layers with 100 and 50 neurons, respectively, and an output layer with a single neuron. At the input level, the DPN integrates three primary feature sources: data from a feature learning module, phenotypic information, and augmented features (such as mean, maximum, and minimum values). Phenotypic information is One-Hot encoded, and these three features are concatenated before processing.

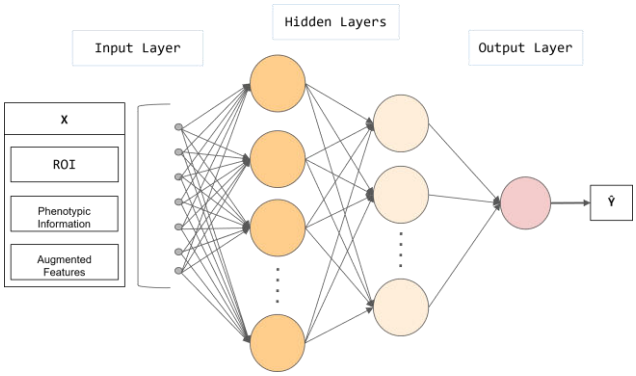

Fig.3.Deep Perceptual Network

The feature reconstruction module uses the mean absolute error (MAE) loss function as a metric to measure the difference between the model output and the original input. The classification loss uses the binary cross-entropy loss function (BCE) to evaluate the difference between the predicted probability distribution and the actual label.

III. RESULTS

A. Dataset

The Abide CC200 dataset used in the experiments in this chapter is a 200-dimensional ROI feature from rs-fMRI. In the model training phase, the optimizer uses AdamW, the initial learning rate of the experiment is 0.0001, the learning rate decay rate is 0.7, and the learning rate is adjusted once every three training loss changes below the threshold. The total number of training epochs is 300.

B. Comparison of Experimental Model

In order to measure the ASD classification performance, this section selects 7 advanced autism classification models as benchmark methods for classification experiment comparison with MMDA, namely GCN[13], GAT[14], BrainGNN[15], MVS-GCN[16], PopulationGCN[17], InceptionGCN[18] and HI-GCN[19]. GCN is a semi-supervised learning method based on graph convolutional networks; GAT is a graph neural network based on attention mechanism; BrainGNN is a graph neural network containing multiple ROI graph convolutional layers; MVS-GCN is a multi-view graph convolutional neural network that includes graph structure learning and multi-task graph embedding learning; PopulationGCN is a graph convolutional network based on LSTM attention mechanism; InceptionGCN is a multi-scale graph neural network that uses kernels of different sizes to capture graph structure differences; HI-GCN is a hierarchical graph convolutional network framework that learns useful graph representations in an end-to-end manner.

The comparative experimental results are shown below:

Table 1
Comparison of experimental method performance

Method	ACC	SEN	SPE	AUC	F1
GCN	67.83	71.40	72.01	65.94	71.28
GAT	69.02	72.91	70.12	70.23	65.08
BrainGNN	70.60	61.30	71.89	64.44	65.36
MVS-GCN	68.61	68.73	64.76	67.38	69.39
PopulationGCN	71.41	77.55	64.01	71.76	74.29
InceptionGCN	70.94	**78.44**	62.37	71.52	74.63
HI-GCN	72.81	72.04	73.91	77.12	75.72
MMDA	**73.82**	76.68	**75.62**	**78.57**	**76.34**

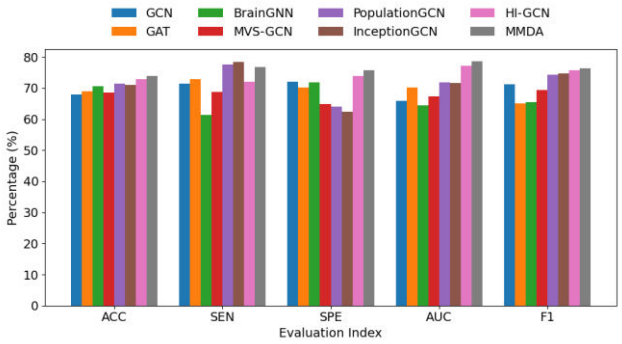

Fig.4.Experimental Results

According to the experimental results, MMDA outperforms other benchmark models across multiple evaluation metrics. Compared to HI-GCN, MMDA shows improvements of 1.01%, 4.64%, 1.71%, 1.45%, and 0.62% in ACC, SPE, AUC, F1, respectively. The comparative experiment results indicate that MMDA's ASD classification performance is generally superior to other ASD classification models. This demonstrates MMDA's capability to maintain high accuracy when dealing with imbalanced datasets and its effectiveness in handling complex data environments, thereby efficiently distinguishing between ASD and TD.

IV. DISCUSSION

In this section, a method for ASD identification based on multimodal deep neural networks is proposed and systematically discussed. Its core innovation lies in the construction of a cascaded multimodal feature learning architecture. This method consists of three main modules. The first is the feature reconstruction network, which mainly contains an autoencoder. It reconstructs the input data through unsupervised learning, effectively suppresses the intra-modal noise interference while retaining key biomarkers, and thus extracts robust and meaningful feature representations from the original data. Next is the feature learning network, which consists of an attention module and a convolution module. On the one hand, the long-range dependency of cross-brain functional connections is modeled through a multi-head self-attention mechanism to extract dynamic spatiotemporal features with global semantic associations. On the other hand, the convolutional network is used to adaptively capture the topological characteristics of local brain structures. Finally, through the deep perception network module, which consists of multiple linear layers, the features processed by the convolution layer and the attention mechanism are remapped back to the same data dimension as the input, and finally fused through residual connections and RMS Norm to obtain the final classification results.

In summary, this chapter proposes a multimodal deep attention model (MMDA). This model not only makes full use of the advantages of autoencoders, multi-head attention mechanisms and convolutional neural networks, but also further enhances the robustness and generalization ability of the model by introducing regularization and MSE loss functions. Experimental results show that the MMDA model has achieved excellent performance in multiple evaluation indicators (such as ACC, SEN, SPE, AUC and F1 score), verifying its effectiveness in ASD classification tasks.

V. CONCLUSION

This paper proposes a method for autism recognition based on deep neural network. The method consists of three main modules: feature extraction module, feature learning module and deep perception module. In the feature extraction module, the unique structural characteristics contained in rs-fMRI data are deeply mined, and the rs-fMRI time series information data is reconstructed by autoencoder to extract the original feature information with potential value. In the feature learning module, a multi-head attention mechanism is combined with a convolutional neural network to learn a more representative and discriminative latent feature representation. The multi-head attention mechanism can capture the complex relationship between different time points, while CNN extracts local features from high-dimensional data. The combination of the two greatly improves the model's ability to learn nonlinear features. Finally, the learned high-level features are further integrated through the deep perception module, and the ASD classification results are output. The experiment uses a 20-fold cross-validation strategy to evaluate the model. The results show that the proposed method achieves a recognition accuracy of 73.82%, and the recognition performance is better than the comparison method.

Although the method proposed in this paper has achieved a relatively high accuracy in the ASD recognition task, there are still some shortcomings, mainly including the following aspects:

A: When performing feature engineering, this study mainly focused on the structural information features of rs-fMRI, and did not fully consider the temporal information features. Future research can further explore the temporal information features of rs-fMRI and organically integrate them with the structural information features to design a more comprehensive and accurate multimodal feature system.

B: Although the MMDA model proposed in this paper has made progress in improving recognition performance, it cannot locate specific neural markers and it is difficult to explain the decision-making process of the model, which brings significant challenges to diagnosis and treatment in clinical applications. In future research, the interpretability of the model can be enhanced so that it can more clearly reveal the relationship between features and disease states.

REFERENCES

[1] [1] Hirota T, King B H. Autism spectrum disorder: a review[J]. Jama, 2023, 329(2): 157-168.
[2] Santana C P, de Carvalho E A, Rodrigues I D, et al. rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis[J]. Scientific reports, 2022, 12(1): 6030.
[3] Santana C P, de Carvalho E A, Rodrigues I D, et al. rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis[J]. Scientific reports, 2022, 12(1): 6030.
[4] Akhavan Aghdam M, Sharifi A, Pedram M M. Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network[J]. Journal of digital imaging, 2018, 31(6): 895-903.
[5] Lancaster J L, Woldorff M G, Parsons L M, et al. Automated Talairach atlas labels for functional brain mapping[J]. Human brain mapping, 2000, 10(3): 120-131.
[6] Battineni G, Chintalapudi N, Amenta F. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM)[J]. Informatics in Medicine Unlocked, 2019, 16: 100200.
[7] Chen C P, Keown C L, Jahedi A, et al. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism[J]. NeuroImage: Clinical, 2015, 8: 238-245.
[8] Katuwal G J, Baum S A, Cahill N D, et al. Divide and conquer: sub-grouping of ASD improves ASD detection based on brain morphometry[J]. PloS one, 2016, 11(4): e0153331.
[9] Feczko E, Balba N M, Miranda-Dominguez O, et al. Subtyping cognitive profiles in autism spectrum disorder using a functional random forest algorithm[J]. Neuroimage, 2018, 172: 674-688.
[10] Kumar C J, Das P R. The diagnosis of ASD using multiple machine learning techniques[J]. International journal of developmental disabilities, 2022, 68(6): 973-983.
[11] Shao L, Fu C, You Y, et al. Classification of ASD based on fMRI data with deep learning[J]. Cognitive Neurodynamics, 2021, 15(6): 961-974.
[12] Sadiq S, Castellanos M, Moffitt J, et al. Deep learning based multimedia data mining for autism spectrum disorder (ASD) diagnosis[C]//2019 international conference on data mining workshops (ICDMW). IEEE, 2019: 847-854.
[13] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, arXiv:1609.02907.
[14] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” Stat, vol. 1050, p. 20, Oct. 2017.
[15] X. Li et al., “BrainGNN: Interpretable brain graph neural network for fMRI analysis,” Med. Image Anal., vol. 74, Dec. 2021, Art. no. 102233.
[16] G. Wen, P. Cao, H. Bao, W. Yang, T. Zheng, and O. Zaiane, “MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis,” Comput. Biol. Med., vol. 142, Mar. 2022, Art. no. 105239.
[17] S. Parisot et al., “Spectral graph convolutions for population-based disease prediction,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Cham, Switzerland: Springer, 2017, pp. 177–185.
[18] A. Kazi et al., “InceptionGCN: Receptive field aware graph convolutional network for disease prediction,” in Proc. Int. Conf. Inf. Process. Med. Imag. Cham, Switzerland: Springer, 2019, pp. 73–85.
[19] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, “Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction,” Comput. Biol. Med., vol. 127, Dec. 2020, Art. no. 104096.

XiaoXu Ma, School of computer science and technology, Tiangong University, Tianjin, China.
Xiang Guo, School of computer science and technology, Tiangong University, Tianjin, China.