

Article

A Multi-Site Anti-Interference Neural Network for ASD Classification

Wentao Lv, Fan Li, Shijie Luo and Jie Xiang * 

College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China; lvwentao0435@link.tyut.edu.cn (W.L.); lifan0583@link.tyut.edu.cn (F.L.); luoshijie0594@link.tyut.edu.cn (S.L.)

* Correspondence: xiangjie@tyut.edu.cn

Abstract: Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that can reduce quality of life and burden families. However, there is a lack of objectivity in clinical diagnosis, so it is very important to develop a method for early and accurate diagnosis. Multi-site data increases sample size and statistical power, which is convenient for training deep learning models. However, heterogeneity between sites will affect ASD recognition. To solve this problem, we propose a multi-site anti-interference neural network for ASD classification. The resting state brain functional image data provided by the multi-site is used to train the ASD classification model. The model consists of three modules. First, the site feature extraction module is used to quantify the inter-site heterogeneity, in which the autoencoder is used to reduce the feature dimension. Secondly, the presentation learning module is used to extract classification features. Finally, the anti-interference classification module uses the output of the first two modules as labels and inputs for multi-task adversarial training to complete the representation learning that is not affected by the confounding of sites, so as to realize the adaptive anti-interference ASD classification. The results show that the average accuracy of ten-fold cross validation is 75.56%, which is better than the existing studies. The innovation of our proposed method lies in the problem that the traditional single-task deep learning ASD classification model will be affected by the heterogeneity of multi-site data and interfere with the classification. Our method eliminates the influence of multi-site factors on feature extraction through multi-task adversarial training, so that the model can better adapt to the heterogeneity of multi-site data. Meanwhile, large-scale 1DconV is introduced to extract features of brain functional network, which provides support for the interpretability of the model. This method is expected to take advantage of multiple sites and provide reference for early diagnosis and treatment of ASD.



Citation: Lv, W.; Li, F.; Luo, S.; Xiang, J. A Multi-Site Anti-Interference Neural Network for ASD Classification. *Algorithms* **2023**, *16*, 315. <https://doi.org/10.3390/a16070315>

Academic Editor: Frank Werner

Received: 7 June 2023

Revised: 19 June 2023

Accepted: 25 June 2023

Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: anti-interference neural networks; multi-site data; autoencoder; large-scale 1DconV; ASD

1. Introduction

Autism spectrum disorder (ASD) is a generalized neurodevelopmental disorder characterized by difficulty in social communication, repetitive patterns of behavior, and narrow interests [1]. Recent studies show that 23 out of every 1000 eight-year-olds in the United States have ASD [2]. Early diagnosis and intervention are key to preventing the exacerbation of symptoms in ASD patients, improving the quality of life of patients and easing the burden on families [3]. At present, the diagnosis of autism is based on symptom-based clinical criteria, which requires a large number of behavioral assessments and requires high professional knowledge of doctors. Moreover, the diagnosis results are affected by doctors' subjectivity, which may lead to misdiagnosis and delayed diagnosis [4]. Therefore, it is necessary to develop an objective, accurate and rapid diagnostic method for ASD. Neuroimaging has been shown to be useful in diagnosing brain diseases and explaining underlying pathologic mechanisms [5]. Resting-state functional magnetic resonance imaging (rs-fMRI) has become one of the commonly used imaging methods in ASD research due to its non-invasiveness, easy acquisition, low patient effort and high generalization ability [6–8].

The amount of rs-fMRI data at a single site is usually small, leading to the possibility of disputable reproducibility and universality of studies [9]; using multi-site data to compose big data integration is an effective method to solve this problem. Autism Brain Imaging Data Exchange (ABIDE) aggregates functional and structural brain imaging collected by laboratories around the world, providing a public multi-site dataset for ASD-related research. In specific ASD classification methods, the high dimension of the original image may lead to model overfitting, and the derived functional connectivity dimension is low and can reflect the specific characteristics of ASD [10], so most methods detect autism spectrum disorders from functional connectivity. For example, Eslami et al. [11] designed a joint learning program of autoencoder (AE) and single-layer perceptrons, connected these networks with low-dimension representations in a mixed way with learning functions, and completed classification on ABIDE multi-site data to reach an accuracy of 70.3%. Wang et al. [12] first defined the graph structure based on functional connections, and proposed a graph convolutional network (cGCN) based on functional connections for ASD classification. This method can extract the spatial features of the neighbor domain of the target brain region from the functional connections. The neighbors of the target brain area are calculated by group functional connections, and each convolution result is the feature of the target brain area. Therefore, the feature finally extracted is the result of all brain areas after considering the functional neighbor information, which can be consistent with the functional organization of the brain. The purpose of the design is to make the convolution operation have brain physiological significance, so as to extract features more efficiently. Finally, the accuracy rate reaches 71.6%.

However, different sites usually have differences in collection equipment, collection parameters and collection population, which results in the natural heterogeneity of multi-site datasets [9,13]. The confounding effect caused by multiple sites will affect the relationship between input data and output variables, leading to the correlation between site differences and biological predictions. Such correlations can hamper estimates of true biological changes, or extrapolate non-biological differences to biological differences, resulting in false or biased predictions from models [14]. For example, when the collection equipment is different, the fMRI images obtained by the same subject will be different, which will affect the analysis of physiological differences such as gender differences [15]. For example, when the purpose of a neuroimaging study is to distinguish between healthy individuals and patients with ASD, if the number of patients at one site is significantly higher than that at another site, the characteristics of the differences in the data at that site may be the criteria for the model to identify the disease. This heterogeneity has the result that although many methods can obtain relatively high accuracy at a single site, when applied to datasets at other sites, the trained models usually fail to achieve an acceptable performance [16,17], which is not conducive to early intervention and auxiliary diagnosis of ASD. For example, Nielsen et al. [16] achieve a maximum classification accuracy of 90% in the single-site ASD classification experiment of ABIDE dataset, but only a maximum classification accuracy of 60% in the multi-site ASD classification experiment of 17 sites.

There are two methods to solve multi-site data heterogeneity: data processing and model-based. The method based on data processing solves the problem of multiple sites by eliminating the difference of data domain distribution between sites. For example, Wang et al. [18] proposed a multi-site adaptive framework based on low-rank representation decomposition, in which the data of one site are regarded as the target domain and the data of other sites as the source domain, and the data in these domains are converted into a low-rank representation public space, so as to reduce the difference in data distribution among different sites. Model-based studies remove the influence of sites through the model itself. For example, Dinsdale et al. [19] proposed a training scheme based on deep learning, which uses iterative updating to remove scanner information and generate unchanged shared features of sites, so as to realize the prediction of a site-free model. In addition, in the study on removing the influence of confounding effects, Zhao et al. [20] proposed an end-to-end method, which quantifies the statistical dependence between the features extracted from

the model and confounding factors, so as to guide the elimination of confounding effects in the process of feature extraction. The experimental results show that the model has a high accuracy in AIDS classification after removing confounding factors.

In the above methods to solve the multi-site problem, the site data domain mapping and classifier training based on data processing are independent of each other, which may reduce the learning performance, so it is necessary to design a new framework to make the two interrelated. The model-based approach may delete information related to disease classification while removing scanner information, so it is necessary to improve the architecture balance feature extraction and scanner information removal. To solve the confounding effect problem, it is necessary to establish the statistical dependence of confounding factors and other features. However, confounders from multiple sites cannot be directly correlated with classification features.

In order to solve the confounding effect caused by data heterogeneity of multi-sites, we propose a multi-site anti-interference neural network (MS-AINN) for multi-site autism classification. First, MS-AINN extracts site features from functional connections through AE, site average aging pool and feature selection, and instantiates abstract site confounding factors into vector form to establish a mapping relationship between feature extraction and site confounding. Secondly, the representation learning module designed by convolutional neural network (CNN) uses a large-scale one-dimensional convolutional kernel, which can directly extract features from functional connections and fully consider the physiological significance of brain functional networks. The reason is that the receptive field of the convolution kernel contains a whole row in the functional connection matrix, and each convolution can extract domain information of all neighboring brain regions of the target brain region. At the same time, the convolution result can be used as the feature of the target brain region, which is convenient for visualization analysis of important brain regions. It not only has advanced performance, but also makes the model physiologically interpretable. Finally, in the adversarial training, the traditional zero-sum game method is not used to train the adversarial task separately, but the objective function is used to train the classification task and the adversarial task simultaneously, and the hyperparameter is used to control the relative importance of the two tasks, so that the training of the two tasks can achieve a certain balance. It is important to prevent adversarial task from excessively restricting ASD classification feature extraction, which degrades classification performance. This ensures that the performance of classification tasks is improved while the confounding effect of sites is removed.

2. Materials and Methods

2.1. Dataset

The ABIDE public dataset is adopted in this experiment. ABIDE is a multi-site dataset that collects the resting state fMRI data and corresponding phenotypic information of 17 international site subjects and contains 1112 datasets, including 539 autistic patients (ASD) and 573 typical development (TD) controls. The 1112 datasets were composed of structural and resting state fMRI data and corresponding phenotypic information. Of these 1112 subjects, 1035 were screened as eligible study candidates because these subjects had complete phenotypic information. Among the 1035 subjects, there were 505 ASD and 530 TDS, 157 women and 878 men.

Figure 1 shows the data distribution of NYU Langone Medical Center (NYU) and University of Michigan (UM) sites in ABIDE dataset. Table 1 shows the data distribution for the remaining 15 sites in the ABIDE dataset. Specifically, the first principal component of all subjects at two sites was obtained through PCA, and then the frequency distribution histogram of the first principal component of all subjects at each site was drawn. The PCA algorithm maps functional connections in a high-dimensional space to a low-dimensional space, where the first principal component is the direction of the maximum variance in the data and is the line that best illustrates the shape of the point group. Therefore, the first principal component can be used to reflect the distribution of functional connection data of

different sites to the greatest extent. It can be seen that the data distribution of the different sites is heterogeneous.

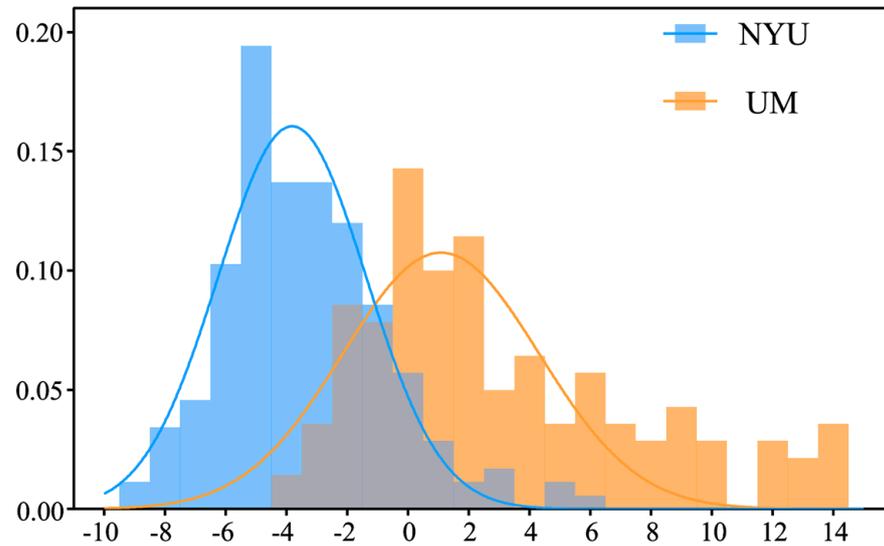


Figure 1. The first principal component is distributed at NYU site and UM site.

Table 1. Distribution of the first principal component at different sites. The range of the first principal component of the first behavior in the table. The first column in the table is the site name. Respectively are University of Pittsburgh School of Medicine (PITT), Olin, Institute of Living at Hartford Hospital (OLIN), Oregon Health and Science University (OHSU), San Diego State University (SDSU), Trinity Centre for Health Sciences (TRINITY), University of Utah School of Medicine (USM), Yale Child Study Center (YALE), Carnegie Mellon University (CMU), University of Leuven (LEUVEN), Kennedy Krieger Institute (KKI), Stanford University (STANFORD), University of California, Los Angeles (UCLA), Ludwig Maximilians University Munich (MAX_MUN), California Institute of Technology (CALTECH) and Social Brain Lab BCN NIC UMC Groningen and Netherlands Institute for Neurosciences (SBL). The percentage value in the table is the number of subjects as a percentage of the total number of subjects at the site.

	(−11,−8)	(−8,−5)	(−5,−2)	(−2,1)	(1,4)	(4,7)	(7,10)	(10,13)	(13,+∞)
PITT	0.00%	0.00%	35.71%	35.71%	19.64%	1.79%	5.36%	0.00%	1.79%
OLIN	2.94%	8.82%	17.65%	29.41%	17.65%	11.76%	8.82%	0.00%	2.94%
OHSU	3.85%	11.54%	34.62%	19.23%	15.38%	11.54%	0.00%	3.85%	0.00%
SDSU	5.56%	2.78%	27.78%	25.00%	13.89%	8.33%	8.33%	5.56%	2.78%
TRINITY	0.00%	6.38%	25.53%	48.94%	8.51%	6.38%	4.26%	0.00%	0.00%
USM	0.00%	2.82%	15.49%	26.76%	33.80%	11.27%	5.63%	4.23%	0.00%
YALE	1.79%	14.29%	33.93%	30.36%	8.93%	10.71%	0.00%	0.00%	0.00%
CMU	3.70%	25.93%	33.33%	14.81%	11.11%	3.70%	3.70%	3.70%	0.00%
LEUVEN	4.76%	28.57%	42.86%	17.46%	3.17%	0.00%	3.17%	0.00%	0.00%
KKI	0.00%	0.00%	4.17%	31.25%	41.67%	14.58%	8.33%	0.00%	0.00%
STANFORD	0.00%	0.00%	2.56%	17.95%	41.03%	25.64%	12.82%	0.00%	0.00%
UCLA	3.06%	12.24%	34.69%	16.33%	13.27%	6.12%	4.08%	8.16%	2.04%
MAX_MUN	0.00%	5.77%	32.69%	32.69%	13.46%	1.92%	7.69%	1.92%	3.85%
CALTECH	0.00%	5.41%	24.32%	32.43%	18.92%	8.11%	2.70%	2.70%	5.41%
SBL	0.00%	13.33%	33.33%	36.67%	13.33%	3.33%	0.00%	0.00%	0.00%

2.2. rs-fMRI Data Preprocessing

The ABIDE dataset is derived from the abide dataset preprocessed by the preprocessed connectome project (PCP) [21]. A configurable pipeline for analysis of connectomes (CPAC) was selected for the preprocessing pipeline. The steps include slice timing correction, head motion correction, intensity normalization, nuisance signal removals, band-pass filtered (0.01–0.1 Hz), standard spatial registration. A pre-processed fMRI scan data is a 4D time series, including three-dimensional space dimension and one-dimensional time dimension. The time series used the mean time series signal or BOLD signal of voxels within the region of interest (ROI) in the brain map. The brain mapping template we selected was Craddock 200 (CC200), with 200 ROIs defined. Pearson correlation coefficients were used to evaluate the functional connections between the mean time series of each ROI pair, resulting in a 200×200 functional connection for the CC200 brain map.

2.3. Multi-Site Anti-Interference Neural Network

We propose a multi-site autism classification method with anti-interference neural networks (MS-AINN). MS-AINN is composed of representation learning module, site feature extraction module and anti-interference classification module, and adopts multi-task adversarial design. The overall architecture is shown in Figure 2.

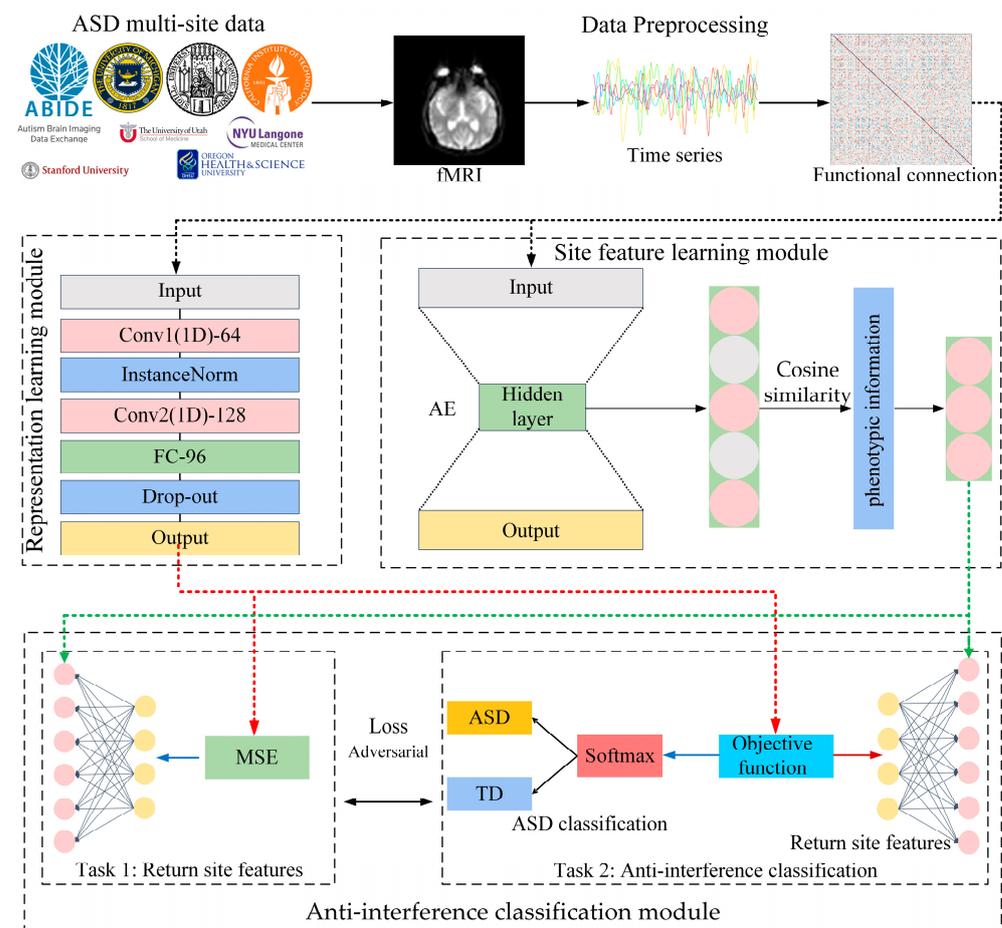


Figure 2. Multi-site anti-interference neural network classification model. The black arrows in the diagram represent the output of the previous module as the input of the next module. The red arrows represent the output features of the representation learning module as input vectors for the two tasks of the anti-interference classification module. The green arrow represents the site feature vector obtained by the site feature extraction module as the label of the two tasks of the anti-interference classification module.

Firstly, multi-site fMRI images are extracted into functional connections after data preprocessing, and functional connections are used as input data for the model. Then, the site feature extraction module will extract site features from the functional connections of the subjects to quantify the heterogeneity between sites, and the site features will be used as labels in the multi-task. The specific process is to first extract low-dimensional subject level feature vectors through AE, then use site average pooling to calculate the mean vector of all subject level feature vectors in the site as site features, and finally select low-redundancy site features through phenotypic information correlation features.

After the above two parts of the process, we have the input data functional connection required for MS-AINN multitask design, as well as two kinds of labels, disease type labels and site characteristics labels that quantify site heterogeneity. MS-AINN can then be used for ASD classification training. Firstly, ASD classification features are extracted from functional connections through the representation learning module. Specifically, large-scale one-dimensional volume nuclei are used for feature extraction, because the weight of the convolutional nuclei corresponds to brain regions, and the importance of each brain region to ASD classification can be analyzed. The convolution result will be further classified through the Fully Connected Layer (FC) for further feature extraction. Then, the obtained classification features will perform two tasks of MS-AINN through the anti-interference classification module, which are site feature regression task and anti-interference classification task. The reason is that the objective function of the anti-interference classification task is composed of ASD classification loss and site feature regression loss. However, the direction of regression loss optimization in the anti-interference classification task is opposite to that in the site feature regression task, that is, the anti-interference classification task will increase the regression loss but the site feature regression task will reduce the regression loss. The anti-interference classification module performs two kinds of tasks with ASD classification features as input through the FC layer. When it comes to site feature regression, both tasks also use the same FC layer for regression.

MS-AINN adopts the training mode of alternating iteration of two tasks, and the updated parameters of each task are independent of each other and do not overlap each other. After the adversarial training of the two tasks, the model can effectively remove the influence of multi-site confounding effect (multi-site confounding effect refers to the confounding effect caused by the heterogeneity of multi-site data) on feature extraction of the representation learning module.

2.4. Site Feature Learning

We use autoencoder (AE) and site average pools to extract site image features from functional connections. The low-dimensional vectors reflecting the individual characteristics of the subject are obtained by using the autoencoder, and the average vectors of all individual feature vectors in the site are calculated by using the site average pool. From this, we can obtain low-dimensional feature vectors that can reflect site information. Autoencoder is a kind of feedforward neural network, which is composed of encoder and decoder, where the encoder encodes the input data x into a low-dimensional representation, as shown in Formula (1).

$$h_{enc} = \phi_{enc}(x) = f(W_{enc}x + b_{enc}) \quad (1)$$

f is the activation function, W_{enc} and b_{enc} are the weight and bias of the encoder, respectively. The decoder reconstructs the output of the encoder back to the original input data, as shown in Formula (2), where W_{dec} and b_{dec} are the weight and bias of the decoder, respectively.

$$x' = \phi_{dec}(h_{enc}) = W_{dec}h_{enc} + b_{dec} \quad (2)$$

The autoencoder completes the model training by minimizing the reconstruction Error. The loss function is the Mean Squared Error (MSE) between the input data x and the

reconstruction result x' . After the encoder completes the training, the output of the encoder can be regarded as a low-dimensional feature of the input data x .

Using the characteristics of unsupervised training and nonlinear dimensionality reduction of the autoencoder, the low-dimensional individual feature vector of the subject is extracted first. Because the functional connection matrix is symmetric, the upper triangle part of the matrix is repeated with the lower triangle part. In order to reduce the parameter number of the autoencoder and speed up the training efficiency, the lower triangle part containing the main diagonal is deleted, and the upper triangle part is planar as a one-dimensional vector as the input of the autoencoder. The autoencoder adopts a single hidden-layer structure, and the encoder part and the decoder part are a single-layer fully connected layer. The autoencoder weights are $19,900 \times N$ and $N \times 19,900$, respectively, where N is the embedded dimension of the hidden layer, which is determined by hyperparameter optimization. The data of all site subjects are used as input to train the model, and the optimal model is saved in the position with the least loss value. Then, the output vector of the encoder is obtained by feeding the one-dimensional function connection vector into the autoencoder, which is a low-dimensional vector reflecting the individual characteristics of the subject. In order to further obtain the feature vector that can reflect the difference information of the sites, the mean vector of the individual feature vector of all subjects in each site is calculated by site average pooling. The reason is that we use the mean of individual characteristics of the subjects in the site to reflect the characteristics of a site as a whole. The process is shown in Figure 3.

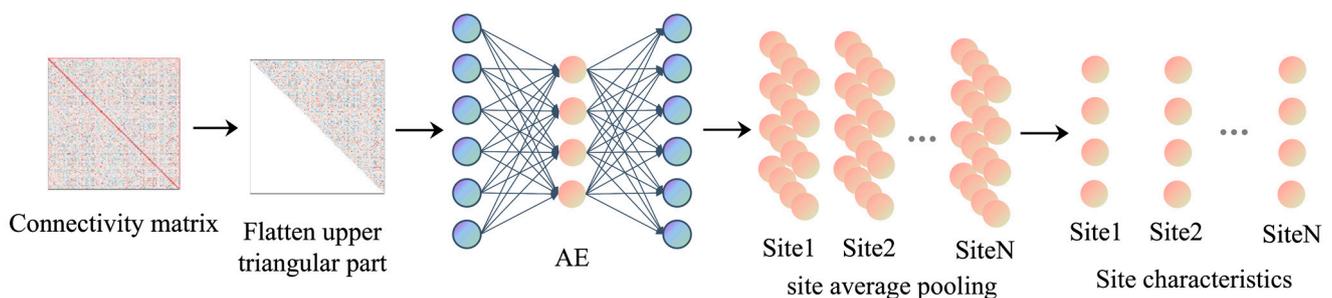


Figure 3. Site feature learning. The arrows in the figure represent the sequence of steps in the site feature learning process.

In addition to site heterogeneity in functional connections, site heterogeneity was also included in subjects' phenotypic information. However, the amount of phenotypic information, such as age and gender, is much smaller than the number of site features extracted from functional connections. The dimension of the vector composed of phenotypic information may be single digits while the dimension of the feature vector extracted from functional connections is in the hundreds. Directly connecting the two vectors together may make it difficult to reflect the site heterogeneity contained in the phenotypic information. Therefore, we select features based on the cosine similarity between the site feature vector extracted from the functional connection and the phenotype information, so that the final site features contain the phenotype information indirectly. The site feature vector needs to calculate the cosine similarity with the phenotypic information of the site, so we use the mean and variance of the phenotypic information of all subjects in a site to represent the phenotypic information of the site. The way to calculate cosine similarity is to first arrange the mean and variance of phenotypic information of all sites into mean vector and variance vector, respectively, according to the order of sites, and then some dimensional features of all site feature vectors are formed into a vector according to the same site order. The form of the vector is the same as that of the mean vector and variance vector, and the cosine similarity can be calculated with the mean vector and variance vector, respectively. Finally, some dimensions with high cosine similarity are selected to form a new site feature vector.

The specific process is to first calculate the mean and variance of the phenotypic information of each site, and then arrange the mean and variance of each site according to the sequence of a site, form the mean vector and variance vector, and standardize. The final vector form is $N \times 1$, where N is the number of sites. The feature vectors of all image sites are arranged in the same order to form a site feature matrix of the form $N \times F$, where N is the number of sites and F is the feature dimension. A one-dimensional vector (form $N \times 1$) is selected along the dimensionality of the site in the site feature matrix, and the cosine similarity is calculated by means of the mean and variance vectors of the phenotypic information. The cosine similarity is calculated as shown in Equation (3), where A and B represent two vectors. According to the cosine similarity, the similarity order of phenotypic information of each site can be obtained. In cases where the mean and variance vectors of multiple phenotypic information are ordered differently, we use a voting mechanism to determine the unique ordering. The higher the similarity ranking and the stronger the phenotypic information correlation, the more likely it is to reflect the site heterogeneity. Finally, the threshold is defined according to the top percentage of similarity ranking, and the features within the threshold range are selected to form a new site feature vector. Phenotypic information we selected included sex, age, full scale IQ, verbal IQ, and operational IQ. Through experiments, the features with the top 30% cosine similarity are selected as site features.

$$\text{similarity} = \cos(\theta) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3)$$

2.5. Representation Learning

The input of the representation learning module is the functional connection matrix, and the physiological meaning is the brain functional network. There are obvious differences between the data structure and the picture, mainly reflected in the fact that the functional connection matrix is a non-Euclidean space and the picture is a Euclidean space. In the Euclidean space of the picture, the neighbors of a pixel are the surrounding pixels in space, so the convolution nuclear energy in the two-dimensional form gathers the information in the neighborhood domain. However, in the functional connection matrix, the neighbor nodes of the target node are not always adjacent in space, so the convolution kernel in two-dimensional form cannot guarantee that all the gathered information comes from the neighbor domain of the target node. This can lead to a mixture of features from different brain regions, which is not conducive to the analysis of important brain regions. The topology of the functional connection matrix is characterized by the fact that each row in the matrix represents the correlation between the brain region corresponding to that row and all the other brain regions. When the functional connection matrix is regarded as a brain functional network, a row in the matrix can also be used as the correlation between the target node and its neighbor node. Therefore, we designed the CNN (1D) representation learning module. The convolution kernel is one-dimensional and contains the same number of weights as the number of brain regions, so that the receptive field of the convolution kernel is exactly a row of the functional connection matrix, which can imitate the interaction between the target brain region and all other brain regions in the row. In this way, each convolution operation gathers the information of the neighbor domain of the target brain region, and the convolution results of each brain region are not mixed with each other, which is conducive to the final analysis of important brain regions. The CNN (1D) consists of two convolutional layers and one fully connected layer, as shown in Figure 4.

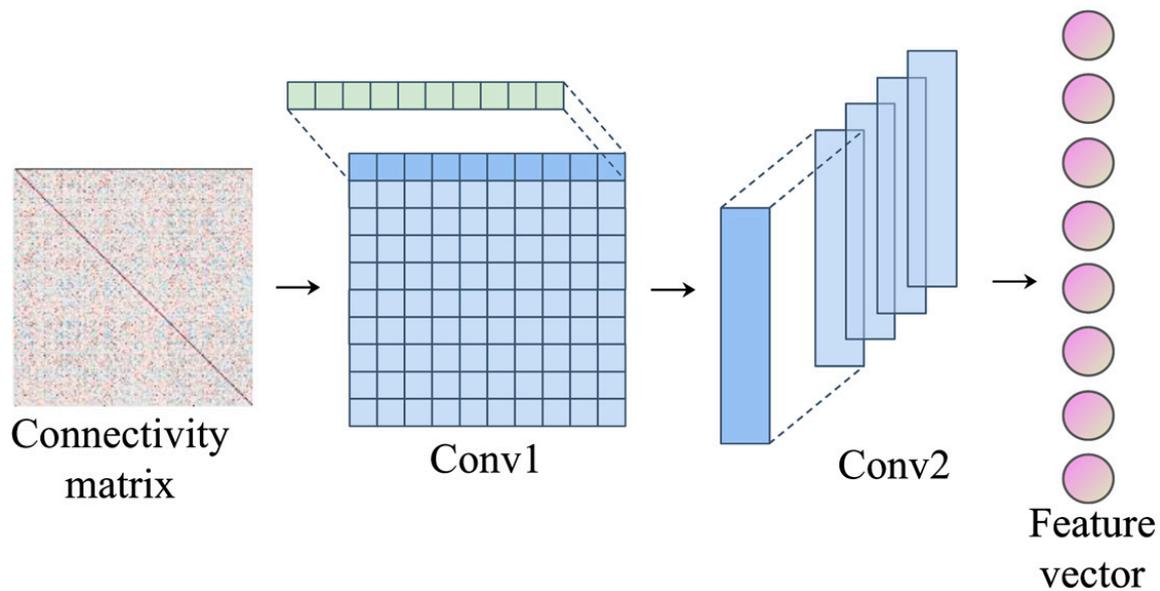


Figure 4. Representation learning.

The first layer uses the horizontal convolution kernel in the form of $64@1 \times 200$, where 1×200 represents the shape of the convolution kernel, 200 represents the number of brain regions, and 64 represents the number of channels. Since each row in the matrix represents the correlation between a certain brain region and all other brain regions, the convolution kernel in the form of 1×200 ensures that each operation is carried out between the target brain region and its adjacent brain regions, and the convolution result can be regarded as a feature of the target brain region. The feature form obtained after the first layer of convolution operation is $64 \times 200 \times 1$, meaning 64 feature maps extracted from 200 brain regions. The second layer of convolution uses vertical convolution kernel to extract whole brain features from brain region features. The convolution kernel in the form of $128@200 \times 200 \times 1$ forms the feature convolution operation of 200 brain regions to form the feature of the whole brain, and the dimension is 128. In this process, the features of brain regions are mapped to the features of the whole brain, and the weight of convolutional nuclei corresponds to the weight of brain regions. Therefore, the absolute values can be used to represent the importance degrees of different brain regions, so as to analyze the interpretability of the model. Finally, the feature is further extracted through a full-connection layer to better complete the classification task. Pooling layer is not used in the model mainly because each convolution is independent of each other based on the unit of brain region, there is no overlap in the receptive field, and the redundancy of extracted features is small. On the contrary, some key features may be lost in the downsampling of the pooling layer. In addition, the purposes of sampling under the pooling layer are to reduce the amount of computation, to prevent overfitting, and to increase the receptive field of subsequent convolution CNN (1D). In order to ensure the corresponding relationship between convolutional kernel weights and brain regions, the range of receptive fields is always the same as the number of brain regions, and pooling of the previous layer is not required to increase receptive fields. At the same time, if the calculation amount is reduced by the pooling layer, the corresponding relationship between convolutional kernel weights and brain regions will be destroyed, resulting in the mixed features of different brain regions, and the importance of specific brain regions cannot be studied. Moreover, it has been proved that downsampling of the pooling layer does not necessarily improve the performance of the CNN network [22]. The convolution operation is shown in Formula (4), where x_j^m is the output of the current layer, x_j^{m-1} is the output of the previous layer, k_{ij}^l is the parameter of the current layer, \times is the dot product operation of the corresponding receptive field, and b_j^m is the bias of the current layer.

Considering that the range of the functional connection matrix is -1 to 1 , the representation learning module selects the activation function \tanh with the same range. The classification layer uses Softmax function to calculate the probability of each category, as shown in Formula (5). The Dropout layer and L2 regularization are added to prevent overfitting of the model, while the InstanceNorm layer is used to normalize the features in each channel.

$$x_j^m = f \left(\sum_{i \in M_j} x_i^{m-1} \times k_{ij}^m + b_j^m \right) \tag{4}$$

$$p(y = j) = \frac{e^{x^T W_j}}{\sum_{k=1}^K e^{x^T W_k}} \tag{5}$$

2.6. Objective Function and Model Training

In the existing methods, the model is trained in a zero-sum game way when adopting adversarial design, that is, the adversarial task that removes confounding effects is trained separately and its loss value is minimized [19,20]. This method is feasible when there is only a single task, such as the generation of adversarial network. However, when the target task and the adversarial task are inconsistent, the lowest loss value of the adversarial task may not make the target task achieve the optimal result, and may even have the opposite effect. Dinsdale et al. [19] have found in the experiment that the adversarial task that removes scanner information can also cause adverse effects on the target task and reduce the final prediction performance. Therefore, MS-AINN does not train the adversarial task alone, but combines the adversarial task and the target task into an objective function, and completes the training of the two tasks simultaneously through the loss optimization of the objective function, so that the two can reach a balance in the training and ensure the performance improvement of the target task.

The objective function of anti-interference classification task is composed of loss functions of ASD classification and site feature regression. The purpose is to make the loss of ASD classification decrease while the loss of site feature regression increases, so as to complete the disease classification and site feature regression adversarials at the same time. Firstly, the symbols used in the function are introduced. $X = \{X_1, \dots, X_N\}$ represents the functional connection of N subjects. $Y = \{Y_1, \dots, Y_N\}$ represents the labels of N subjects; $C = \{C_1, \dots, C_N\}$ represents the site feature vector of N subjects; θ_E is the parameter representing the learning module; θ_p is the classifier parameter; θ_R is the regression component parameter. The loss function of site feature regression task is the mean square error, as shown in Formula (6). The classification task loss function is cross entropy loss, as shown in Formula (7).

$$L_R(X_i, C_i; \theta_R) = \sum_{i=1}^N \frac{1}{m} \sum_{j=1}^m \left(C_{j,i} - \widehat{C}_{j,i} \right)^2 \tag{6}$$

$$L_p(X_i, Y_i; \theta_E, \theta_p) = - \sum_{i=1}^N [Y_i \ln(P_i) + (1 - Y_i) \ln(1 - P_i)] \tag{7}$$

The model adopts the training mode of alternating iteration. In each epoch, the regression task loss L_R is optimized first, and the regression component parameter θ_R is trained. After optimization, the objective function loses L_t , and the representation learning module parameter θ_E and classifier parameter θ_p are trained. In order to realize the confrontation between two trainings, the objective function should minimize the classification loss L_p and also maximize the regression task loss L_R . Therefore, the regression task loss function is selected as the denominator of a fraction and the whole fraction is added to the classification task loss function. In this way, when the optimizer reduces the objective function L_t , the

classification task loss L_p will also decrease and the regression task loss L_R will increase, as shown in Formula (8). α is a hyperparameter, which is used to balance the loss optimization of the two tasks, so as to prevent the excessive increase of L_R in the training of the objective function from resulting in too-strong restriction on the representation learning module and the degradation of the classification performance of ASD.

$$L_t(X_i, Y_i, C_i; \theta_E, \theta_P) = L_p(X_i, Y_i) + \frac{\alpha}{L_R(X_i, C_i)} \quad (8)$$

In the two trainings of the model, the first regression task trained a regression component to identify the site difference information contained in the features extracted by the representation learning module. The second anti-interference classification task training optimizes in the opposite direction through two kinds of losses, namely, classification and regression, so that the representation learning module can extract classification features while containing fewer features that can reflect the difference information of the sites. In this way, after two training and alternating iteration optimizations, the ability of the regression component to identify the site difference information is gradually enhanced, and the site difference information contained in the classification features extracted by the representation learning module is gradually reduced. In this way, the confounding effect brought by site heterogeneity will be weakened in ASD classification, and anti-interference classification will finally be realized.

3. Results

3.1. Anti-Interference Classification Results of Multi-Site Data

CNN (1D) was used as the representation learning module under the ten-fold cross-verification of the whole site to compare the classification performance of norm neural networks and MS-AINN. Evaluation indicators include classification accuracy, AUC, precision, recall, f1 and site classification accuracy, and the results are shown in Table 2. The first five indexes in the table evaluated the classification performance of ASD diseases of the model. In addition, the classification accuracy of sites was obtained by training the features learned from representations separately, which was used to evaluate the degree of confounding of features by sites. CNN (1D) in the table refers to the classification training conducted by the representation learning module directly after extracting features.

Table 2. CNN (1D) anti-interference classification results.

Method	ACC%	AUC%	Precision %	Recall %	f1%	Site Classification ACC %
CNN (1D)	72.77	77.85	73.59	72.49	72.32	21.39
MS-AINN-CNN (1D)	75.56	78.99	76.60	75.36	75.20	15.76

It can be seen from the table that the classification performance of MS-AINN-CNN (1D) is significantly improved compared with CNN (1D). Accuracy increased by 2.79%, AUC increased by 1.14%, precision increased by 3.01%, recall increased by 2.87% and f1 increased by 2.88%. Meanwhile, the accuracy of site classification decreased by 5.63%.

In order to prove that the classification performance improvement of MS-AINN is due to the neural network structure rather than the representation learning part, this experiment also replaced the representation learning module with Multilayer perceptron (MLP) and CNN (2D) for testing, and the results are shown in Tables 3 and 4. The input of MLP is the same as that of AE, which is a flat one-dimensional feature vector of functional connection. Specifically, a double hidden layer design is adopted. The weight form of the first layer is $19,900 \times 512$, and the weight form of the second layer is 512×256 . CNN (2D) replaces the convolution kernel of CNN (1D) with the traditional two-dimensional convolution, with the specific forms of $32@5 \times 5$ and $64@2 \times 2$, and the input is the same as CNN (1D).

Table 3. MLP anti-interference classification results.

Method	ACC%	AUC%	Precision %	Recall %	f1%	Site Classification ACC %
MLP	72.23	75.70	73.90	71.88	71.37	37.04
MS-AINN-MLP	74.20	76.51	75.55	73.79	73.52	18.73

Table 4. CNN (2D) anti-interference classification results.

Method	ACC%	AUC%	Precision %	Recall %	f1%	Site Classification ACC %
CNN (1D)	71.37	76.17	73.47	71.01	70.41	37.54
MS-AINN-CNN (2D)	74.12	77.10	74.79	73.97	73.78	18.58

Compared with MLP, the accuracy of MS-AINN-MLP increased by 1.97%, AUC increased by 0.81%, precision increased by 1.65%, recall increased by 1.91% and f1 increased by 2.15%. Meanwhile, the accuracy of site classification decreased by 18.31%. Compared with CNN (2D), the accuracy of MS-AINN-CNN (2D) increased by 2.75%, AUC increased by 0.93%, precision increased by 1.32%, recall increased by 2.96% and f1 increased by 3.37%. Meanwhile, the accuracy of site classification decreased by 18.96%.

We also made comparisons with previous ASD classification methods that used data from the same 17 sites as ours for ASD classification in the ABIDE multi-site dataset. The results are shown in Table 5, and it can be seen that the proposed method achieves the optimal classification accuracy.

Table 5. Comparison of ASD classification methods in multi-sites.

Method	Year of Publication	ACC%
GCN [23]	2018	70.40
ASD-DiagNet [11]	2019	70.30
Attention selection based on Extra-Trees [24]	2020	72.20
Self-attention deep learning framework [25]	2021	72.48
cGCN [12]	2021	71.60
AE combined with feature selection [26]	2022	70.90
MS-AINN-CNN (1D)	-	75.56

3.2. Hyperparameter Optimization Results

We use an Adam optimizer to optimize the model and search parameters using grid search. The final learning rate is 1×10^{-4} , dropout rate is 0.5, L2 penalty factor is 1×10^{-4} , epoch is set to 100, batch size is 10, and the balance parameter in the objective function is 0.006.

Site features, as vectors quantifying site differences, are the basis for constructing the mapping relationship between representation learning features and site confounding factors. Only on this basis can MS-AINN be used to anti-interference classify the ASD. Therefore, in order to explore the impact of site feature dimensions on classification performance, this experiment tested the classification accuracy of site features extracted by MS-AINN with no AE and AE with different hidden dimensions under the ten-fold cross-verification of the whole site. Meanwhile, the results were compared with the results of the training model of norm neural networks, as shown in Figure 5. No AE means no dimensionality reduction with AE, and the mean vector of the one-dimensional functional connection feature vector of all the subjects in each site after flattening is calculated directly through site average pooling, which is taken as the site feature. The dimension is 19,900. AE Dimension represents the dimension of hidden layer when AE extracts site features, that is, the dimension of site features. It can be seen from Figure 5 that MLP, CNN (2D) and CNN (1D) have improved the accuracy of MS-AINN from 0.5% to 2.75% compared with norm neural networks under different site feature vector dimensions. When AE dimension 512 is used, the MLP accuracy rate increases by 1.97%, CNN (1D) by 2.34%, and CNN (2D)

by 2.75% to achieve the optimal performance of the corresponding model. The overall accuracy rate increases first and then decreases with the decrease of site feature dimension. The reason is that different parameters can obtain site features of different dimensions. When the dimension is large, there are more redundant features, and when the dimension is small, site differences cannot be fully expressed.

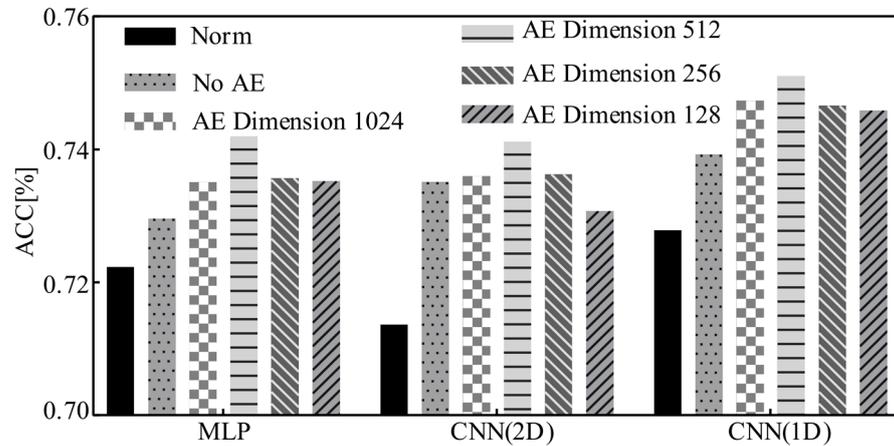


Figure 5. Classification results of site features in different dimensions.

In addition, the phenotypic information correlation features were selected for the site features extracted at AE dimension 512, and the top percentage features with the highest correlation ranking were selected to perform the site feature regression task. A threshold is defined based on the percentage that ranks high in correlation, and the threshold is used as a hyperparameter for optimization. The classification accuracy of MS-AINN-CNN (1D) under different thresholds is shown in Figure 6. The results show that as the feature selection threshold changes from low to high, the accuracy increases first, then decreases and then increases, and reaches the highest accuracy of 75.56% when the threshold is 0.3. With the change of feature selection threshold from low to high, AUC first increased, then decreased and then increased. When the threshold was 0.3, AUC reached the highest value of 78.99%.

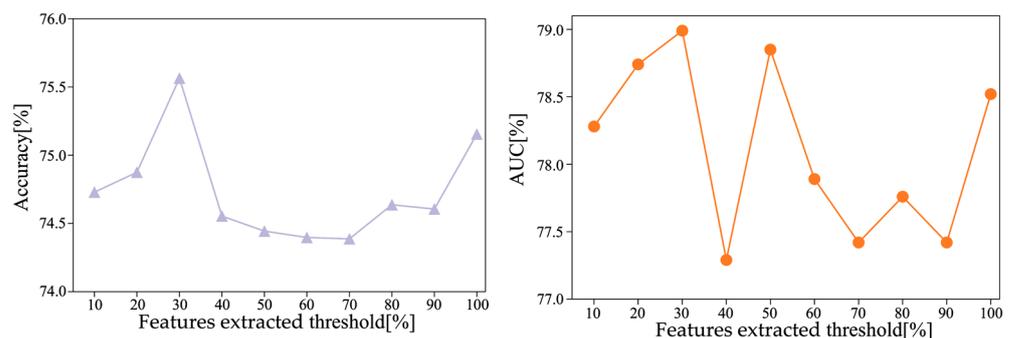


Figure 6. Site feature selection results.

3.3. Visual Analysis

Since CNN (1D) adopts a large-scale convolution kernel considering the topology of brain functional network, and there is a one-to-one correspondence between the weight of the convolution kernel and brain region, we use the results of the normalization of the absolute value of the model weight to represent the classification contribution of brain region, and visualized the CNN (1D) and MS-AINN-CNN (1D), respectively; the results are shown in Figure 7, where A is the result of CNN (1D), and B is the result of MS-AINN-CNN (1D). The redder the color, the higher the relative importance of the brain regions, and the bluer the lower. Brain regions of great relative importance in the two structures include

vermis, frontal lobe, posterior cingulate gyrus, orbital part, fusiform gyrus, thalamus, temporal transverse gyrus, precuneus, horny gyrus, caudate nucleus, and hippocampus. These brain regions are the same as those reported in the literature for differences between ASD and TD [7,27–32]. This indicates that the representation learning module can actively identify the brain regions that change due to disease, and extract the difference features for classification. Moreover, the high-contribution brain areas in the MS-AINN-CNN (1D) were the same as CNN (1D), indicating that although the MS-AINN-CNN (1D) limited the feature extraction of the representation learning module to remove the confusion of sites, it did not affect the recognition of important brain areas by the model. At the same time, parts of the brain that were bluer became bluer and less important. It shows that the MS-AINN also makes the representation learning module pay more attention to the important brain regions that may have physiological differences and ignore the non-important brain regions. After reducing the weight value of non-critical brain areas, the features extracted contain the difference information of normal patients with important brain areas, and reduce the interference of confounding factors of non-important brain areas, and finally improve the classification performance.

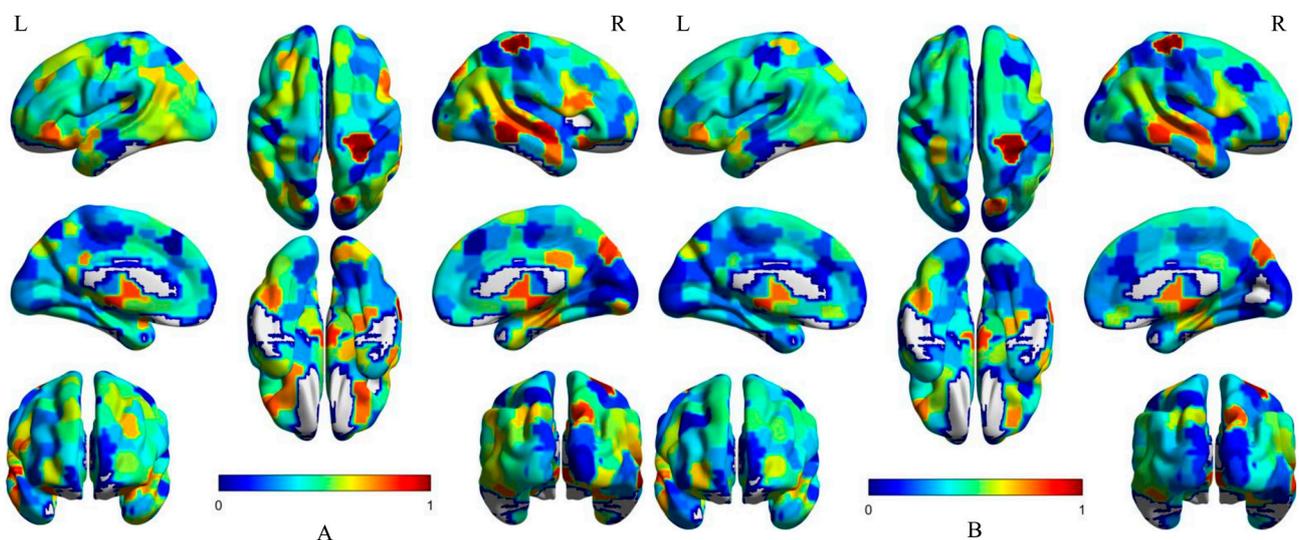


Figure 7. Visualization of brain region contribution in CNN (1D). (A) shows the relative importance of each brain region after CNN (1D) visualization. (B) shows the relative importance of each brain region after visualization by MS-AINN-CNN (1D). In (A,B), the bluer the color, the less important the brain area, and the redder the color, the more important the brain area.

In order to more intuitively reflect the impact of multi-site data heterogeneity on feature extraction, t-SNE [29] is used to visualize the spatial distribution of feature vectors extracted by CNN (1D) representation learning module, as shown in Figure 8A,B. Each color in the figure represents data for one site. In order to display results more clearly, we choose four sites with the largest number of subjects in the ABIDE dataset for visualization, namely, UM, USM, NYU and UCLA.

Figure 8A shows that the feature vectors extracted by CNN (1D) are relatively dispersed in spatial distribution, and the feature vectors between different stations have a certain degree of domain separation. From Figure 8B, It can be seen that the feature vectors extracted by MS-AINN-CNN (1D) are relatively dense in spatial distribution, and the domains of feature vectors at different sites overlap each other. The reason for the two different spatial distributions lies in the heterogeneity of the multi-site data. The feature extraction directly through the representation learning module will be affected by this and result in the separation of the feature vector domain between the sites. However, in order to extract the feature vectors affected by the hybridization to the sites, the MS-AINN restricts the feature extraction process of the representation learning module by means of

adversarial training, so that the solution space of the model is concentrated in the area that does not reflect the differences between the sites, and finally the feature vector of intersite aggregation is obtained.

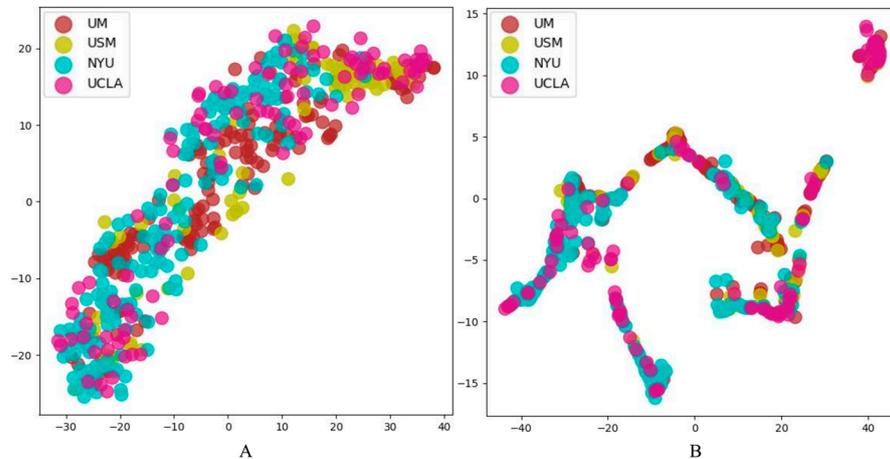


Figure 8. CNN (1D) Feature vector visualization. (A) is the result of CNN (1D), and (B) is the result of MS-AINN-CNN (1D).

4. Discussion

In this study, we propose an MS-AINN-CNN (1D) deep learning model, to classify ASD and TDS on large multi-site rs-fMRI data based on whole-brain functional connections. MS-AINN is composed of two key parts: (1) Site feature learning, which extracts site features from the functional connections of subjects by using AE, site average pooling and phenotypic information feature selection, so as to reflect the heterogeneity information among sites, and then uses it to build a mapping relationship with the representation learning features to provide necessary conditions for adversarial training against the impact of site confounding. (2) Anti-interference model training, using two tasks of site feature regression and anti-interference classification to reduce the site heterogeneity information contained in the representation learning features, so that the model classification is less affected by the multi-site confounding effect. The results show that when the representation learning module is CNN (1D), MS-AINN can reduce the accuracy of site classification of representation learning features, indicating that the impact of the representation learning module on the site confounders is weakened, which proves that MS-AINN can realize the representation learning to the site confounders. Moreover, all classification indexes of MS-AINN-CNN (1D) were improved, which proved that MS-AINN's confounder remove representation learning successfully improved the model classification performance. In addition, the results of the representation learning module replacement experiment show that MS-AINN can also bring about the same disaggregation effect and obvious classification performance improvement when using MLP and CNN (2D) representation learning modules, indicating that anti-interference ideas play a key role in the entire, while the representation learning module provides a baseline performance for classification. Among them, MLP and CNN (2D) have lower classification effect than CNN (1D), which proves that large-scale convolutional kernel design considering physiological significance has superior performance in ASD classification research.

In the optimization results of AE hidden layer, the overall accuracy rate increases first and then decreases with the decrease in site feature dimension, because different parameters will obtain site features of different dimensions. When the dimension is large, there are more redundant features, and when the dimension is small, site differences cannot be fully expressed, and the optimal results cannot be achieved in both cases. In the site feature selection threshold experiment, the results show that with the change of feature selection threshold from low to high, the accuracy and auc increase first, then decrease and

then increase, and the highest accuracy is 75.56%, auc is 78.99% when the threshold is 0.3. The reason is that when the threshold is low, the features selected are high-scale related features, which are composed of site image features and potentially contain phenotypic information, and have a strong ability to reflect the difference between sites. Therefore, the increase in such features can enable MS-AINN to more accurately realize the impact of site confounding, and the classification accuracy will therefore increase. The threshold value of 0.3 to 0.7 is the decline range of accuracy. At this time, there are fewer features related to the high scale in the site features. With the increase in the threshold value, the redundant features begin to increase, and the effect of the confounding remove begins to decline, and the accuracy also declines. The threshold value of 0.7 to 1.0 is the interval where the accuracy rate rises again. The reason is that among the site image features extracted from AE and site average pooling, there are still features that can highly reflect the difference between sites except the features with high phenotypic information correlation. The distribution of these features has nothing to do with phenotypic information correlation, but only increases with the increase of the total number of features. When a certain threshold is exceeded, the number of features in this part is enough to enhance the ability of site features to reflect site differences again, and ultimately improve the accuracy.

The experimental results of hyperparameter optimization show that different parameters will affect the classification accuracy, and there is an optimal parameter to make the accuracy highest. However, it can be seen from the experimental results that the accuracy difference between different parameters is small, and the accuracy difference between whether MS-AINN is used or not is large. Even when AE reduction and feature selection are not used, MS-AINN still has performance improvement. Therefore, it can be concluded that the reason for the effectiveness of the proposed method lies in the use of MS-AINN, and the purpose of the method is not to find the optimal parameter set, but to achieve anti-interference classification by means of the method of multi-site confounding removal.

In the interpretability analysis of the model visualization, the results of the visualization of important brain regions show that the representation learning module can actively identify the brain regions that have changes due to disease, and extract the difference features for classification. Moreover, the high-contribution brain areas after using MS-AINN were the same as before, indicating that although MS-AINN limited the feature extraction of the representation learning module to remove site confusions, it did not affect the model's recognition of important brain areas. At the same time, parts of the brain that were bluer became bluer and less important. MS-AINN also makes the representational learning module focus more on important brain regions that may have physiological differences, while ignoring non-important brain regions. After reducing the weight value of non-critical brain regions, the features extracted contain the difference information of normal patients in important brain regions, and reduce the interference of confounding factors in non-important brain regions, and finally improve the classification performance. The T-SNE visualization of the representation learning features shows that MS-AINN can make the features separated from each other gather together. The reason for the change in spatial distribution is the heterogeneity of the data of multiple sites. The feature extraction directly through the representation learning module will be affected by this, resulting in the separation of the feature vector domain between sites. In order to extract the feature vectors of the mixed influence of the sites, MS-AINN limited the feature extraction process of the representation learning module by means of adversarial training, so that the solution space of the model was concentrated in the region that did not reflect the differences of the sites, and finally obtained the feature vectors of the inter-site domain aggregation.

5. Conclusions

We proposed a multi-site anti-interference neural network ASD classification method that aims to make the representation learning results reflect the disease differences while being less affected by site confounding factors, and finally realize anti-interference classification. The experimental results show that the MS-AINN achieves the purpose of

confounding the site. At the same time, the improvement of various classification indexes shows that reducing the impact of site confounders on representation learning can effectively improve the classification performance of models. Moreover, the MS-AINN can still improve the classification performance after replacing the representation learning module, which proves the universality of the architecture. Therefore, the MS-AINN that takes site confounders into account is of great significance for the design and research of deep learning ASD classification model under multi-site datasets.

The purpose of our method is to remove the multi-site confounding effect on ASD classification, which is realized by alternating iterative training of site feature regression task and anti-interference classification task under MS-AINN. This training method can limit the feature extraction of the representation learning module, so that the extracted features are less confounded by the site. However, it is found in the experiment that the restriction representation learning module will not only weaken the confounding effect of the sites, but also affect the classification. Neither too large nor too small restriction can achieve the optimal performance. Therefore, hyperparameters are required in the objective function to balance the losses of the two tasks, which also increases the difficulty of parameter adjustment. In addition, we only test classification tasks on ASD datasets, and the next step will be to test other tasks on different multi-site datasets.

Author Contributions: Conceptualization, W.L. and J.X.; methodology, W.L.; software, W.L. and S.L.; validation, W.L., F.L. and S.L.; formal analysis, S.L. and J.X.; writing—original draft preparation, W.L.; writing—review and editing, W.L.; supervision, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: This study use of publicly available, previously published data from ABIDE.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anagnostou, E.; Taylor, M.J. Review of neuroimaging in autism spectrum disorders: What have we learned and where we go from here. *Mol. Autism* **2011**, *2*, 4. [[CrossRef](#)] [[PubMed](#)]
2. Maenner, M.J.; Shaw, K.A.; Bakian, A.V.; Bilder, D.A.; Durkin, M.S.; Esler, A.; Furnier, S.M.; Hallas, L.; Hall-Lande, J.; Hudson, A.; et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveill. Summ.* **2021**, *70*, 1–16. [[CrossRef](#)] [[PubMed](#)]
3. Fernell, E.; Eriksson, M.A.; Gillberg, C. Early diagnosis of autism and impact on prognosis: A narrative review. *Clin. Epidemiol.* **2013**, *5*, 33–43. [[CrossRef](#)]
4. Lord, C.; Rutter, M.; Le Couteur, A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **1994**, *24*, 659–685. [[CrossRef](#)]
5. Jie, B.; Liu, M.; Zhang, D.; Shen, D. Sub-Network Kernels for Measuring Similarity of Brain Connectivity Networks in Disease Diagnosis. *IEEE Trans. Image Process.* **2018**, *27*, 2340–2353. [[CrossRef](#)] [[PubMed](#)]
6. Yao, Z.; Hu, B.; Xie, Y.; Zheng, F.; Liu, G.; Chen, X.; Zheng, W. Resting-State Time-Varying Analysis Reveals Aberrant Variations of Functional Connectivity in Autism. *Front. Hum. Neurosci.* **2016**, *10*, 463. [[CrossRef](#)]
7. Starck, T.; Nikkinen, J.; Rahko, J.; Remes, J.; Hurtig, T.; Haapsamo, H.; Jussila, K.; Kuusikko-Gauffin, S.; Mattila, M.L.; Jansson-Verkasalo, E.; et al. Resting state fMRI reveals a default mode dissociation between retrosplenial and medial prefrontal subnetworks in ASD despite motion scrubbing. *Front. Hum. Neurosci.* **2013**, *7*, 802. [[CrossRef](#)] [[PubMed](#)]
8. Smitha, K.A.; Akhil Raja, K.; Arun, K.M.; Rajesh, P.G.; Thomas, B.; Kapilamoorthy, T.R.; Kesavadas, C. Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks. *Neuroradiol. J.* **2017**, *30*, 305–317. [[CrossRef](#)]
9. Button, K.S.; Ioannidis, J.P.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.; Munafò, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376. [[CrossRef](#)]
10. Liu, M.; Li, B.; Hu, D. Autism spectrum disorder studies using fMRI data and machine learning: A review. *Front. Neurosci.* **2021**, *15*, 697870. [[CrossRef](#)]
11. Eslami, T.; Mirjalili, V.; Fong, A.; Laird, A.R.; Saeed, F. ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data. *Front. Neuroinform.* **2019**, *13*, 70. [[CrossRef](#)]

12. Wang, L.; Li, K.; Hu, X.P. Graph convolutional network for fMRI analysis based on connectivity neighborhood. *Netw. Neurosci.* **2021**, *5*, 83–95. [[CrossRef](#)] [[PubMed](#)]
13. Heinsfeld, A.S.; Franco, A.R.; Craddock, R.C.; Buchweitz, A.; Meneguzzi, F.J.N.C. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin.* **2018**, *17*, 16–23. [[CrossRef](#)] [[PubMed](#)]
14. Shinohara, R.T.; Oh, J.; Nair, G.; Calabresi, P.A.; Davatzikos, C.; Doshi, J.; Henry, R.G.; Kim, G.; Linn, K.A.; Papinutto, N.; et al. Volumetric Analysis from a Harmonized Multisite Brain MRI Study of a Single Subject with Multiple Sclerosis. *AJNR Am. J. Neuroradiol.* **2017**, *38*, 1501–1509. [[CrossRef](#)]
15. Takao, H.; Hayashi, N.; Ohtomo, K. Effects of study design in multi-scanner voxel-based morphometry studies. *Neuroimage* **2014**, *84*, 133–140. [[CrossRef](#)] [[PubMed](#)]
16. Nielsen, J.A.; Zielinski, B.A.; Fletcher, P.T.; Alexander, A.L.; Lange, N.; Bigler, E.D.; Lainhart, J.E.; Anderson, J.S. Multisite functional connectivity MRI classification of autism: ABIDE results. *Front. Hum. Neurosci.* **2013**, *7*, 599. [[CrossRef](#)] [[PubMed](#)]
17. Abraham, A.; Milham, M.P.; Di Martino, A.; Craddock, R.C.; Samaras, D.; Thirion, B.; Varoquaux, G.J.N. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage* **2017**, *147*, 736–745. [[CrossRef](#)] [[PubMed](#)]
18. Wang, M.; Zhang, D.; Huang, J.; Yap, P.T.; Shen, D.; Liu, M. Identifying Autism Spectrum Disorder With Multi-Site fMRI via Low-Rank Domain Adaptation. *IEEE Trans. Med. Imaging* **2020**, *39*, 644–655. [[CrossRef](#)]
19. Dinsdale, N.K.; Jenkinson, M.; Namburete, A.I.L. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *Neuroimage* **2021**, *228*, 117689. [[CrossRef](#)]
20. Zhao, Q.; Adeli, E.; Pohl, K.M. Training confounder-free deep learning models for medical applications. *Nat. Commun.* **2020**, *11*, 6010. [[CrossRef](#)]
21. Craddock, C.; Benhajali, Y.; Chu, C.; Chouinard, F.; Evans, A.; Jakab, A.; Khundrakpam, B.S.; Lewis, J.D.; Li, Q.; Milham, M.; et al. The neuro bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives. *Front. Neuroinformatics* **2013**, *7*, 27.
22. Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv* **2014**, arXiv:1412.6806.
23. Parisot, S.; Ktena, S.I.; Ferrante, E.; Lee, M.; Guerrero, R.; Glocker, B.; Rueckert, D. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer’s disease. *Med. Image Anal.* **2018**, *48*, 117–130. [[CrossRef](#)] [[PubMed](#)]
24. Liu, Y.; Xu, L.; Li, J.; Yu, J.; Yu, X. Attentional Connectivity-based Prediction of Autism Using Heterogeneous rs-fMRI Data from CC200 Atlas. *Exp. Neurobiol.* **2020**, *29*, 27–37. [[CrossRef](#)] [[PubMed](#)]
25. Wang, Z.; Peng, D.; Shang, Y.; Gao, J. Autistic Spectrum Disorder Detection and Structural Biomarker Identification Using Self-Attention Model and Individual-Level Morphological Covariance Brain Networks. *Front. Neurosci.* **2021**, *15*, 756868. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, J.; Feng, F.; Han, T.; Gong, X.; Duan, F. Detection of Autism Spectrum Disorder using fMRI Functional Connectivity with Feature Selection and Deep Learning. *Cogn. Comput.* **2022**, 1–12. [[CrossRef](#)]
27. Tsai, P.T. Autism and cerebellar dysfunction: Evidence from animal models. *Semin. Fetal. Neonatal. Med.* **2016**, *21*, 349–355. [[CrossRef](#)] [[PubMed](#)]
28. Ecker, C. The neuroanatomy of autism spectrum disorder: An overview of structural neuroimaging findings and their translatability to the clinical setting. *Autism* **2017**, *21*, 18–28. [[CrossRef](#)] [[PubMed](#)]
29. Maximo, J.O.; Kana, R.K. Aberrant “deep connectivity” in autism: A cortico-subcortical functional connectivity magnetic resonance imaging study. *Autism Res.* **2019**, *12*, 384–400. [[CrossRef](#)]
30. Deng, Z.; Wang, S. Sex differentiation of brain structures in autism: Findings from a gray matter asymmetry study. *Autism Res.* **2021**, *14*, 1115–1126. [[CrossRef](#)]
31. Assaf, M.; Jagannathan, K.; Calhoun, V.D.; Miller, L.; Stevens, M.C.; Sahl, R.; O’Boyle, J.G.; Schultz, R.T.; Pearlson, G.D. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage* **2010**, *53*, 247–256. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, L.; Liu, L.; Wen, Y.; Ma, M.; Cheng, S.; Yang, J.; Li, P.; Cheng, B.; Du, Y.; Liang, X.; et al. Genome-wide association study and identification of chromosomal enhancer maps in multiple brain regions related to autism spectrum disorder. *Autism Res.* **2019**, *12*, 26–32. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.