

Article

Domain Adaptation Using a Three-Way Decision Improves the Identification of Autism Patients from Multisite fMRI Data

Chunlei Shi ¹, Xianwei Xin ¹ and Jiakai Zhang ^{1,2,*}

¹ School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China; 201931210011@mail.bnu.edu.cn (C.S.); xinxianwei@mail.bnu.edu.cn (X.X.)

² Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing 100875, China

* Correspondence: jiakai.zhang@bnu.edu.cn

Abstract: Machine learning methods are widely used in autism spectrum disorder (ASD) diagnosis. Due to the lack of labelled ASD data, multisite data are often pooled together to expand the sample size. However, the heterogeneity that exists among different sites leads to the degeneration of machine learning models. Herein, the three-way decision theory was introduced into unsupervised domain adaptation in the first time, and applied to optimize the pseudolabel of the target domain/site from functional magnetic resonance imaging (fMRI) features related to ASD patients. The experimental results using multisite fMRI data show that our method not only narrows the gap of the sample distribution among domains but is also superior to the state-of-the-art domain adaptation methods in ASD recognition. Specifically, the ASD recognition accuracy of the proposed method is improved on all the six tasks, by 70.80%, 75.41%, 69.91%, 72.13%, 71.01% and 68.85%, respectively, compared with the existing methods.

Keywords: autism spectrum disorder; machine learning; three-way decision; domain adaptation



Citation: Shi, C.; Xin, X.; Zhang, J. Domain Adaptation Using a Three-Way Decision Improves the Identification of Autism Patients from Multisite fMRI Data. *Brain Sci.* **2021**, *11*, 603. <https://doi.org/10.3390/brainsci11050603>

Received: 23 March 2021

Accepted: 4 May 2021

Published: 8 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

1. Introduction

Autism spectrum disorder (ASD) is a common neurodevelopmental disease originating in infancy [1–6]. According to a recent study, one in 45 children in the world has autism, and the number of affected children has increased by 78% in the last decade [7]. Some symptoms of ASD even appear in young children by the age of two years [8]. Therefore, the early diagnosis of and intervention in ASD have received great attention in recent years [9,10]. Researchers have applied machine learning methods to identify biomarkers from resting-state functional magnetic resonance imaging (rs-fMRI) data to assist in diagnosing ASD [11–13].

Machine learning methods have demonstrated their effectiveness with the assumption that we have sufficient training data and test data drawn from the same distribution [14,15]. However, this assumption calling for enough examples is not always satisfied in practical applications and is not true in most cases, which will lead to the poor generalization ability of a model trained on one dataset when applied to another new dataset. First, clinical neural image datasets often face the problem of small dataset size due to their expensive acquisition and time-consuming labels. Therefore, multisite rs-fMRI data are often combined to expand the dataset in some research, such as ASD diagnosis, which leads to the second problem: samples from different scanners or acquisition protocols do not follow the same distribution in most cases [16,17].

The fMRI samples from different sites have also been named domains in the machine learning research community. In addition to the distribution difference of the training set (source domain) and the test set (target domain), the scarcity of labelled samples is another challenge to ASD recognition. Previous studies have investigated domain adaptation approaches to overcome site-to-site transfer [18]. Many studies have successfully applied

domain adaptation to object recognition [19], activity recognition [20], speech recognition [21], text classification [22] and autism recognition [7]. The main goal of domain adaptation is to reduce the difference in the data distribution between the source domain and the target domain and then train a robust classifier for the target domain by reusing the labelled data in the source domain.

At present, the research on domain adaptation mainly focuses on three methods, namely, instance adaptation, feature adaptation and classifier adaptation. Specifically, the instance-based domain adaptation method reuses samples from the source domain according to a certain weighting rule. Instance adaptation has achieved good results by eliminating cross domain differences [23]. However, this method needs to satisfy two strict assumptions: (1) the source domain and the target domain follow the same conditional distribution, and (2) some data in the source domain can be reused in the target domain by reweighting. The classifier-based domain adaptation method transfers knowledge from the source domain to the target domain by sharing parameters between the source domain and target domain [24,25]. Classifier transfer has performed well with labelled samples. However, regarding ASD diagnosis from fMRI, the data distributions of different sites are different, and reliable labelled data are difficult to obtain.

Therefore, the application of domain adaptations based on instances or classifiers is relatively difficult. However, the feature-based domain adaptation method can learn the subspace geometrical structure [26–29] or distribution alignment [30–32]. This method gears the marginal or conditional distributions of different domains in a principled dimensionality reduction procedure. Our work employed the feature-based domain adaptation method to eliminate the divergence of the data distribution.

Currently, most feature-based domain adaptation research is devoted to the adaptation of the marginal distribution, conditional distribution, or both. For example, Long et al. [26] found that the marginal distribution and conditional distribution between domains are different in the real world, and better performance can be achieved if the two distributions are adapted simultaneously. Subsequently, some studies based on joint distribution adaptation have been proposed successively [33–35], and these works have greatly contributed to the development of domain adaptation.

It is worth noting that in order to obtain the pseudolabels of the target domain data, the traditional methods usually directly apply the classifier trained in the source domain to the prediction of the target domain data. However, these pseudolabels might lead to some error due to the possible domain mismatch. Here, we proposed a robust method using a three-way decision model derived from triangular fuzzy similarity. The proposed model roughly classified the samples in the target domain into three domains, i.e., the positive region, the negative region and the boundary region. Then, the label propagation algorithm was used to optimize the label and make secondary decisions on the boundary region samples. The experiments demonstrate that our method can effectively improve the classification performance for automated ASD diagnosis.

The contributions of this paper is as follows:

1. A three-way decision model based on triangular fuzzy similarity is proposed to reduce the cost loss of target domain data prediction. To the best of authors' knowledge, it is the first time to combine the three-way decision model and the distribution adaptation method to reduce the distribution differences between domains. The proposed method extends the application of machine learning in the field of decision making.
2. Our method utilizes the label information from the source domain and the structural information from the target domain at the same time, which not only reduces the distribution differences between domains but also further improves the recognition ability of the target domain data.
3. Comprehensive experiments on the Autism Brain Imaging Data Exchange (ABIDE) dataset prove that our method is better than several state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 reviews the related work concisely. In Section 3, we elucidate the foundation of the proposed method. Our

proposed method is illustrated in detail in Section 4. Then, the results and discussion are presented in Sections 5 and 6, respectively. Finally, the paper is concluded in Section 7.

2. Related Work

It has been a lasting challenge to build maps between different domains in the field of machine learning. Domain adaptation has become a hot research topic in disease diagnosis with machine learning. In this paper, we proposed transfer learning based on distribution adaptation and three-way decisions. To elaborate the proposed method, we will introduce the related work from the following three aspects in this section.

2.1. Distribution Adaptation

Distribution adaptation is one of the most commonly used methods in domain adaptation. It seeks a space translation and eliminated data distribution differences between source and target domains by explicitly minimizing the predefined distance in this feature space. According to the nature of the data distribution, distribution adaptation can be divided into three categories: marginal distribution adaptation, conditional distribution adaptation and joint distribution adaptation.

Pan et al. [36] first proposed a transfer component analysis (TCA) method based on marginal distribution adaptation, which used the maximum mean discrepancy (MMD) to measure the distance between domains and achieve feature dimensionality reduction. The method assumes that there is a mapping so that the marginal distribution of the mapped source domain and target domain is similar in the new space. The disadvantage of TCA is that the algorithm only focuses on reducing the cross-domain marginal distribution difference without considering reducing the conditional distribution difference. Long et al. [37] proposed a transfer joint matching (TJM) method, which mainly combines source domain sample selection and distribution adaptation to further eliminate cross domain distribution differences.

Recently, in the work based on conditional distribution adaptation, Wang et al. [38] proposed a stratified transfer learning method (STL). Its main idea is to reduce the spatial dimension in the reproducing kernel Hilbert space (RKHS) by using the intraclass similarity so as to eliminate the distribution differences. However, in the real world, differences may exist in both marginal distributions and different conditional distributions. Adjusting only one of the distributions is insufficient to bridge domain differences. In order to solve this problem, Long et al. [26] proposed the joint distribution adaptation (JDA) method. The goal of JDA is to jointly adjust the marginal distribution and the conditional distribution using a principled dimensionality reduction process, and the representation in this common feature space reduced the domain differences significantly. Other work extended JDA by adding structural consistency [29], domain invariant clustering [30] and label propagation [31].

To provide supervised information for the target domain, JDA methods applied source domain classifiers in the target domain and took the classifier outputs as the pseudolabels of the target domain data. However, due to the different data distributions of domains, the direct use of these inaccurate pseudolabels will result in the degradation of the final model's performance.

Considering the domain gap both in labels and samples, three-way decisions provided a novel method to transmit the label information between domains and reuse the intrinsic structural information of the target domain data to further improve the performance of the model in the domain adaptation process.

2.2. Three-Way Decisions

As an effective extension of traditional rough sets, three-way decision [39] (3WD) theory has been widely applied to address uncertain, inaccurate and fuzzy problems, such as medical diagnosis [40], image processing [41], emotion analysis [42], etc. In simple terms, 3WD divides the universe of discourse into three disjoint parts, i.e., the positive region (Pos), the negative region (Neg), and the boundary region (Bnd), through a pair of upper

and lower, approximately. Acceptance and rejection decisions were made for the objects in Pos and Neg, respectively. Specifically, the objects in Bnd adopt the delay decision.

Strictly speaking, the current 3WD research can be divided according to whether it is based on decision-theoretic rough sets (DTRSs) [43]. For example, Zhang et al. [44] proposed a 3WD model for interval-valued DTRSs and gave a new decision cost function. Liu et al. [45] introduced intuitionistic fuzzy language DTRSs and 3WD models to obtain fuzzy information in uncertain languages. Agbodah [46] focused on the study of the DTRS loss function aggregation method in group decision making and utilized it to construct a 3WD model.

In addition, scholars have also conducted in-depth explorations on 3WD outside the DTRS framework. For example, Liang et al. [47] integrated the risk preference of decision makers into the decision-making process and proposed a 3WD model based on the TODIM (an acronym in Portuguese for interactive multicriteria decision making) method. Qian et al. [48] investigated three-way formal concept lattices of objects (properties) based on 3WD. Yang et al. [49] presented a 3WD model oriented to multigranularity space to adapt 3WD to intuitionistic fuzzy decisions.

From a broad perspective, 3WD can be classified as static or dynamic. Static 3WD includes related research based on the DTRS framework and fusion of other theories. Dynamic 3WD mainly addresses the problem of constantly changing data in time series and space, and its typical representative is the sequential 3WD model [50]. For example, Yang et al. [51] proposed a three-way calculation method for dynamic mixed data based on time and space. Zhang et al. [52] systematically investigated a new sequential 3WD model to balance autoencoder classification and reduce its misclassification cost. Liu et al. [53] combined 3WD and granular computing to construct a dynamic three-way recommendation model to reduce decision-making costs.

3WD theory has been widely used in many areas, such as emerging three-way formal concept analysis [54], three-way conflict analysis [55], three-way granular computing [56], three-way classification [57], three-way recommendation [58], and three-way clustering [59]. This paper will combine the idea of 3WD to improve the performance of heterogeneous ASD data diagnosis by reducing the difference in the data distributions between the source domain and target domain.

2.3. Application of Machine Learning in Identification of ASD Patients

In recent years, magnetic resonance imaging (MRI) has been widely used in clinical practice [60,61]. The commonly used MRI can be divided into structural MRI (sMRI) and functional MRI (fMRI). As fMRI can measure the hemodynamic changes caused by the activity of brain neurons, it has been widely used in the research of brain dysfunction diseases. For example, Li et al. [62] proposed a 4D deep learning model for ASD recognition that can utilize both temporal and spatial information of fMRI data. In the work of Riaz et al. [63], they proposed an end-to-end deep learning method called DeepfMRI for accurately identifying patients with Attention Deficit Hyperactivity Disorder (ADHD) and achieved an accuracy rate of 73.1% on open datasets. To study the relationship between mild cognitive impairment (MCI) and Small Vessel Disease (SVD), Diciotti et al. [64] applied the Stroop test to the rs-fMRI data of 67 MCI subjects and found that regional homogeneity of rs-fMRI is significantly correlated with measurements of the cognitive deficits.

As a neurodevelopmental disorder, early diagnosis of ASD is very important to improve the quality of life of patients. In recent years, researchers have attempted to extract biomarkers representing ASD from fMRI data using machine learning methods, so as to provide an auxiliary diagnosis for clinicians. For example, Lu et al. [65] proposed a multi-kernel-based subspace clustering algorithm for identifying ASD patients, which still has a good clustering effect on high-dimensional network datasets. Leming et al. [66] trained a convolutional neural network and applied it to ASD recognition, and their experiments showed that deep learning models that distinguish ASD from NC controls focus broadly on temporal and cerebellar connections. However, the problem of small size fMRI data prevented the generalization of the above research works [67].

To solve this problem, the Autism Brain Imaging Data Exchange, an international collaborative project, has collected data from over 1000 subjects and made the whole database publicly available. Based on the ABIDE database, many advanced machine learning models have been proposed for the identification of ASD patients. For example, Eslami et al. [68] used autoencoder and single-layer perceptron to diagnose ASD and proposed a deep learning framework called ASD-DiagNet, which achieved classification accuracy of 70.3%. Bi et al. [69] used randomized support vector machine (SVM) clusters to distinguish ASD patients from normal controls and identified a number of abnormal brain regions that contribute to ASD. Mladen et al. [70] selected 368 ASD patients and 449 normal controls from ABIDE database, and then used the Fisher score as the feature selection method to quantitatively analyze 817 subjects and obtained classification accuracy of 85.06%.

3. Preliminaries

We start with the definition of the problem and the terms and introduce the notation we will use below. The source domain data denoted as $X_s \in \mathbb{R}^{d \times n_s}$ are drawn from distribution $P_s(X_s)$, and the target domain data denoted as $X_t \in \mathbb{R}^{d \times n_t}$ are drawn from distribution $P_t(X_t)$, where d is the dimension of the data instance and n_s and n_t are the number of samples in the source and target domains, respectively.

Assume a labelled source domain $\mathcal{D}_s = \{(x_i, y_i)\}_{i=1}^{n_s}$, where $x_i \in \mathbb{R}^{d \times n_s}$, and an unlabeled target domain $\mathcal{D}_t = \{(x_j)\}_{j=1}^{n_t}$ and $x_j \in \mathbb{R}^{d \times n_t}$. We assume that their feature space and label space are the same, i.e., $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$, but their marginal distribution and conditional distribution are different, i.e., $P_s(X_s) \neq P_t(X_t)$ and $P_s(Y_s|X_s) \neq P_t(Y_t|X_t)$.

Domain adaptation methods often seek to reduce the distribution differences across domains by explicitly adapting both the marginal and conditional distributions between domains. To be specific, domain adaptation seeks to minimize the distance (Equation (1)):

$$\mathcal{D}(\mathcal{D}_s, \mathcal{D}_t) \approx \mathcal{D}(P_s(X_s), P_t(X_t)) + \mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t)) \quad (1)$$

where $\mathcal{D}(P_s(X_s), P_t(X_t))$ and $\mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t))$ are the marginal distribution distance and conditional distribution distance between domains, respectively.

There are many metrics that can be used to estimate the distance between distributions, such as the Kullback–Leibler (KL) divergence. However, most of these distance metrics are based on parameters, and it is difficult to calculate the distance. Therefore, Borgwardt et al. [71] proposed a nonparametric distance metric MMD using a kernel learning method to measure the distance between two distributions in RKHS. The definition of the MMD is as follows:

Definition 1. Given two random variables X_s and X_t , their MMD squared distance is calculated as follows (Equation (2)):

$$\text{Dist}(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \varnothing(x_i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \varnothing(x_j) \right\|_{\mathcal{H}}^2 \quad (2)$$

where \mathcal{H} is a universal RKHS [72], and $\varnothing : \mathcal{X} \rightarrow \mathcal{H}$.

Next, we introduce the concepts of triangular fuzzy numbers and three-way decisions.

Definition 2. [73]. Let $\tilde{t} = [t^L, t^M, t^T]$ be a triangular fuzzy number, where t^L and t^T denote the upper bound and lower bound of \tilde{t} , respectively, and t^M is the median of \tilde{t} . If $0 < t^L \leq t^M \leq t^T$ is satisfied, then \tilde{t} is called a normal triangular fuzzy number. For any two triangular fuzzy numbers $\tilde{t} = [t^L, t^M, t^T]$ and $\tilde{k} = [k^L, k^M, k^T]$, the distance between them is as follows (Equation (3)):

$$d(\tilde{t}, \tilde{k}) = \sqrt{\frac{(t^L - k^L)^2 + (t^M - k^M)^2 + (t^T - k^T)^2}{3}} \quad (3)$$

In addition, the basic operations between $\tilde{t} = [t^L, t^M, t^T]$ and $\tilde{k} = [k^L, k^M, k^T]$ are as follows (Equation (4)):

$$\begin{aligned}\tilde{t} + \tilde{k} &= [t^L, t^M, t^T] + [k^L, k^M, k^T] = [t^L + k^L, t^M + k^M, t^T + k^T] \\ \tilde{t} - \tilde{k} &= [t^L, t^M, t^T] - [k^L, k^M, k^T] = [t^L - k^L, t^M - k^M, t^T - k^T] \\ \tilde{t} \times \tilde{k} &= [t^L, t^M, t^T] \times [k^L, k^M, k^T] = [t^L \times k^L, t^M \times k^M, t^T \times k^T]\end{aligned}\quad (4)$$

Definition 3. [74]. Let U be the universe of discourse, $\forall X \in U$. If threshold $0 \leq \beta < \alpha \leq 1$ exists, then its positive region, negative region and boundary region are defined with threshold (α, β) (Equation (5)):

$$\begin{aligned}Pos_{(\alpha, \beta)}(X) &= \{x \in U | \Pr(X|[x]) \geq \alpha\} \\ Bnd_{(\alpha, \beta)}(X) &= \{x \in U | \beta < \Pr(X|[x]) < \alpha\} \\ Neg_{(\alpha, \beta)}(X) &= \{x \in U | \Pr(X|[x]) \leq \beta\}\end{aligned}\quad (5)$$

where $[x]$ is the equivalence class containing x , and $\Pr(X|[x])$ is the conditional probability.

4. Methods

4.1. Joint Distribution Adaptation

Domain adaptation seeks an invariant feature expression for the source domain and the target domain in a low-dimensional ($K < d$) space. Let $W \in \mathbb{R}^{d \times k}$ be the linear transformation matrix and $Z_s = W^T X_s$ and $Z_t = W^T X_t$ be the projected variables from the source and target data, respectively. We use the nonparametric metric MMD, which computes the distance between the sample means of the source and target data in the k -dimensional embeddings, to estimate the difference between distributions. Specifically, according to Equation (2), $\mathcal{D}(P_s(X_s), P_t(X_t))$ can be expressed as (Equation (6)):

$$\mathcal{D}(P_s(X_s), P_t(X_t)) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} W^T x_i - \frac{1}{n_t} \sum_{j=1}^{n_t} W^T x_j \right\|^2 \quad (6)$$

By further using the matrix transformation rule and regularization and then minimizing the marginal distribution distance, Equation (6) can be formalized as follows (Equation (7)):

$$\mathcal{D}(P_s(X_s), P_t(X_t)) = \text{tr}\left(A^T X M_0 X^T A\right) \quad (7)$$

where X represents the input matrix containing X_s and X_t . In addition, following [26], M_0 is the MMD matrix and can be constructed as follows (Equation (8)):

$$(M_0)_{ij} = \begin{cases} \frac{1}{n_s^2}, & x_i, x_j \in \mathcal{D}_s \\ \frac{1}{n_t^2}, & x_i, x_j \in \mathcal{D}_t \\ -\frac{1}{n_s n_t}, & \text{otherwise} \end{cases} \quad (8)$$

However, the label information of the domain data is not considered, which will lead to the lack of sufficient discriminability of the adapted features; therefore, so it is insufficient to adapt to the marginal distribution only. To solve this problem, we will next adjust the conditional distribution between domains.

Since no label information is available in the target domain, we cannot directly estimate the conditional distribution $P_t(Y_t|X_t)$ of the target domain. Here, based on the concept of sufficient statistics, we can replace $P_t(Y_t|X_t)$ and $P_s(Y_s|X_s)$ with class conditional distributions $P_t(X_t|Y_t)$ and $P_s(X_s|Y_s)$, respectively. However, obtaining target domain label information through source domain data while reducing the distribution difference between domains is a challenging problem in unsupervised domain adaptation. In Section 4.2, we introduce how to obtain the label information of the target domain data so as to

obtain the above class conditional distribution. Thus far, we can match the class condition distribution of the two domains. Similar to the calculation of the marginal distribution, we use the modified MMD formula to estimate the conditional distribution $\mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t))$ between domains. $\mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t))$ can be represented as (Equation (9)):

$$\mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t)) = \sum_{c=1}^C \left\| \frac{1}{n_s^c} \sum_{x_i \in \mathcal{D}_s^{(c)}} W^T x_i - \frac{1}{n_t^c} \sum_{x_j \in \mathcal{D}_t^{(c)}} W^T x_j \right\|^2 \quad (9)$$

where $c \in \{1, 2, 3, \dots, C\}$ is the class label, and $\mathcal{D}_s^{(c)}$ and $\mathcal{D}_t^{(c)}$ are samples belonging to class c in the source domain and target domain, respectively. n_s^c and n_t^c are the number of samples belonging to class c in the source domain and target domain, respectively.

Similar to the marginal distribution, we formalize Equation (9) as Equation (10) by using matrix transformation rules and regularization:

$$\mathcal{D}(P_s(Y_s|X_s), P_t(Y_t|X_t)) = \text{tr}(W^T X M_c X^T W) \quad (10)$$

where the MMD matrices M_c containing class labels are constructed as follows (Equation (11)):

$$(M_c)_{ij} = \begin{cases} \frac{1}{(n_s^c)^2}, & x_i, x_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{(n_t^c)^2}, & x_i, x_j \in \mathcal{D}_t^{(c)} \\ -\frac{1}{n_s^c n_t^c}, & \begin{cases} x_i \in \mathcal{D}_s^{(c)}, x_j \in \mathcal{D}_t^{(c)} \\ x_i \in \mathcal{D}_t^{(c)}, x_j \in \mathcal{D}_s^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

In order to reduce both the marginal distribution and conditional distribution between domains, we incorporate Equations (7) and (10) into one object Function (Equation (12)):

$$\begin{aligned} \min \sum_{c=0}^C \text{tr}(W^T X M_c X^T W) + \lambda \|W\|_{2F}^2 \\ \text{s.t. } W^T X H X^T W = I \end{aligned} \quad (12)$$

where the first term considers both the adaptive marginal distribution and conditional distribution, and the second term is the regularization term. $\|\cdot\|_{2F}$ is the Frobenius norm, and λ is the regularization parameter. As noted in [29], adding the constraint in Function (12) would preserve the inner properties of the original data, which implies and introduces an additional data discrimination ability into the learned model. In addition, in function (12), X represents the input matrix containing X_s and X_t ; $I \in \mathbb{R}^{(n_s+n_t) \times (n_s+n_t)}$ denotes the identity matrix; and $H = I - \frac{1}{n_s+n_t} \mathbf{1}$ is the centering matrix, where $\mathbf{1}$ is the $(n_s + n_t) \times (n_s + n_t)$ matrix of ones.

To obtain the transformation matrix W , we obtain the Lagrange solution to function (12), which is rewritten as (Equation (13)):

$$L = \sum_{c=0}^C \text{tr}(W^T X M_c X^T W) + \lambda \|W\|_{2F}^2 + \text{tr}\left(\left(I - W^T X H X^T W\right) \Phi\right) \quad (13)$$

where $\Phi = (\varnothing_1, \varnothing_2, \dots, \varnothing_d)$ is the Lagrange multiplier. Setting $\frac{\partial L}{\partial W} = 0$, the original optimization problem is transformed into the following eigen-decomposition problem (Equation (14)):

$$\left(\sum_{c=0}^C X M_c X^T + \lambda I \right) W = X H X^T W \Phi \quad (14)$$

The transformation matrix W is the solution to Equation (14) and thus builds the bridge between the source and target domains in the new expression $Z = (Z_s, Z_t)$.

4.2. Three-Way Decision Model Based on Triangular Fuzzy Similarity

In practice, the conditional distribution cannot be obtained directly because there is no label information in the target domain. In order to solve this problem, we first give the concept of the degree of information difference and apply it to the construction of triangular fuzzy numbers and the calculation of the corresponding triangular fuzzy similarity. Then, according to the degree of association of the triangular fuzzy similarity between objects in the target domain, the target domain is divided into positive regions, negative regions and boundary regions with structural information.

For the convenience of the description, suppose that both the domain of discourse U and attribute set A are nonempty finite sets and that x_i is an object in U , a_j is an attribute in A , where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$.

4.2.1. Information Difference Degree and Triangular Fuzzy Similarity

Definition 4. Let $U = \{x_1, x_2, \dots, x_n\}$ be the domain of discourse, $A = \{a_1, a_2, \dots, a_m\}$ be the set of attributes, and the value of object x_i under attribute a_j be x_{ij} . When $\forall a_j, a_k \in A$, the degree of information difference of object x_i is as follows (Equation (15)):

$$ID_i(a_j, a_k) = \frac{1}{\exp(-\log_2 \frac{|x_{ij} - x_{ik}|}{x_{ij} + x_{ik}})} \quad (15)$$

Remark 1.

- (1) The greater the value of $ID_i(a_j, a_k)$ is, the greater the degree of information difference of object x_i under a_j and a_k . When object x_i has the same description $x_{ij} = x_{ik} = 0$ for a_j and a_k , the real part of the log function will have a denominator of 0, i.e., $x_{ij} + x_{ik} = 0$. In this case, since $|x_{ij} - x_{ik}| = 0$, we can obtain that the final degree of information deviation $ID_i(a_j, a_k)$ is independent of the value of $x_{ij} + x_{ik}$. For the reasonableness of the calculation, let $\frac{|x_{ij} - x_{ik}|}{x_{ij} + x_{ik}} = 0$.
- (2) For the convenience of the representation, we obtain the information difference matrix of object x_i , which can be expressed as follows (Equation (16)):

$$ID_i = \begin{matrix} & a_1 & a_2 & \dots & a_{m-1} & a_m \\ \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_{m-1} \\ a_m \end{matrix} & \left[\begin{array}{cccccc} ID_i^{11} & ID_i^{21} & \dots & ID_i^{(m-1)1} & ID_i^{m1} \\ ID_i^{12} & ID_i^{22} & \dots & ID_i^{(m-1)2} & ID_i^{m2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ ID_i^{1(m-1)} & ID_i^{2(m-1)} & \dots & ID_i^{(m-1)(m-1)} & ID_i^{m(m-1)} \\ ID_i^{1m} & ID_i^{2m} & \dots & ID_i^{(m-1)m} & ID_i^{mm} \end{array} \right] & (16) \end{matrix}$$

where $ID_i^{jk} = ID_i(a_j, a_k)$ represents the degree of information difference of object x_i under attributes a_j and a_k .

Theorem 1. According to definition 4, we have the following conclusions:

- (1) Boundedness: $0 \leq ID_i(a_j, a_k) \leq 1$.
- (2) Monotonicity: The degree of information difference of x_i about a_j and a_k increases monotonously as the difference increases.
- (3) Symmetry: $ID_i(a_j, a_k) = ID_i(a_k, a_j)$.

Proof. Properties (2) and (3) are easily proven by Definition 4.

- (1) According to Definition 4, $\forall a_j, a_k \in A$ and $x_i \in U$. When the description of x_i under a_j and a_k appears in two extreme cases, namely, $x_{ij} = 0$ and $x_{ik} = 1$ or $x_{ij} = 1$ and $x_{ik} = 0$, we can obtain $|x_{ij} - x_{ik}| = 1$, and the information difference reaches the maximum at this time, $ID_i(a_j, a_k) = 1$. \square

Definition 5. Let U be the domain of discourse, and the triangular fuzzy number of x_i under attribute set A is $\tilde{x}_i = [x_i^L, x_i^M, x_i^T]$, where $x_i^L = \min\{\sum ID_i^{jk}\}$, $x_i^T = \max\{\sum ID_i^{jk}\}$, $x_i^M = \max\{\sum |ID_i^{jk}|\}$ and $|ID_i^{jk}|$ denotes the number of information difference values ID_i^{jk} . Then, the degree of triangular fuzzy similarity between \tilde{x}_i and \tilde{x}_k is as follows (Equation (17)):

$$\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = 1 - d(\tilde{x}_i, \tilde{x}_k) = 1 - \sqrt{\frac{(\tilde{x}_i^L - \tilde{x}_k^L)^2 + (\tilde{x}_i^M - \tilde{x}_k^M)^2 + (\tilde{x}_i^T - \tilde{x}_k^T)^2}{3}} \quad (17)$$

Theorem 2. The degree of triangular fuzzy similarity satisfies the following properties:

- (1) $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = \tilde{S}_{TF}(\tilde{x}_k, \tilde{x}_i)$.
 (2) $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = 1$ if $\tilde{x}_i = \tilde{x}_k$, and $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = 0$ if $\tilde{x}_i = [0, 0, 0]$ and $\tilde{x}_k = [1, 1, 1]$ or $\tilde{x}_i = [1, 1, 1]$ and $\tilde{x}_k = [0, 0, 0]$.

Proof. According to Definition 5, (1) obviously holds.

- (2) Since $0 \leq \tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) \leq 1$, when $\tilde{x}_i = \tilde{x}_k$, i.e., $\tilde{x}_i^L = \tilde{x}_k^L$, $\tilde{x}_i^M = \tilde{x}_k^M$, $\tilde{x}_i^T = \tilde{x}_k^T$, we have $d(\tilde{x}_i, \tilde{x}_k) = 0$, so $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = 1$. Similarly, since $\tilde{x}_i^L, \tilde{x}_i^M, \tilde{x}_i^T \in [0, 1]$ and $\tilde{x}_k^L, \tilde{x}_k^M, \tilde{x}_k^T \in [0, 1]$, $d(\tilde{x}_i, \tilde{x}_k) \in [0, 1]$. When $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k) = 0$, $d(\tilde{x}_i, \tilde{x}_k) = 1$. In this case, we can obtain $\tilde{x}_i = [0, 0, 0]$ and $\tilde{x}_k = [1, 1, 1]$ or $\tilde{x}_i = [1, 1, 1]$ and $\tilde{x}_k = [0, 0, 0]$. \square

4.2.2. Construction of the 3WD Model

Definition 6. Let U be the universe and A be the set of attributes. The triangular fuzzy similarity between any object x_i and x_k in U is $\tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k)$. If there is a threshold δ , then the δ -level classes of $x \in U$ with respect to $\tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k)$ are defined as follows (Equation (18)):

$$\begin{aligned} [\tilde{S}_{SF}^\delta]_s &= \{x \in U \mid \tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k) > \delta\} \\ [\tilde{S}_{SF}^\delta]_g &= \{x \in U \mid \tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k) \leq \delta\} \end{aligned} \quad (18)$$

where $[\tilde{S}_{SF}^\delta]_s$ and $[\tilde{S}_{SF}^\delta]_g$ are triangular fuzzy similarity classes of positive and negative fields, respectively. Specifically, the objects in $[\tilde{S}_{SF}^\delta]_s$ have the smallest degree of information difference and the largest triangular fuzzy similarity on the δ -level while the objects in $[\tilde{S}_{SF}^\delta]_g$ are the opposite to $[\tilde{S}_{SF}^\delta]_s$.

Suppose $X \subseteq U$ is a given goal concept and $\Omega = \{[\tilde{S}_{SF}^\delta]_s, [\tilde{S}_{SF}^\delta]_g\}$ is the set of states, which represents object x in the δ -level similarity domain $[\tilde{S}_{SF}^\delta]_s$ or x in the δ -level negative similarity domain $[\tilde{S}_{SF}^\delta]_g$. $\Gamma = \{a_P, a_B, a_N\}$ is the set of actions, where a_P means acceptance, a_B means delay, and a_N means rejection. According to reference [75], the losses caused by actions taken in different states are shown in Table 1.

Table 1. Cost function matrix.

Action	Cost Function	
	$\tilde{S}_{SF}^\delta _s$	$\tilde{S}_{SF}^\delta _g$
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

When the object $x \in [\tilde{S}_{SF}^\delta]_s$, λ_{PP} , λ_{BP} and λ_{NP} represent the loss of acceptance, delay and rejection decisions, respectively. Analogously, λ_{PN} , λ_{BN} and λ_{NN} represent the corresponding decision loss cost when $x \in [\tilde{S}_{SF}^\delta]_g$. Without any loss of generality, when $x \in [\tilde{S}_{SF}^\delta]_s$, we assume that the correct acceptance cost is less than the delay decision cost and less than the corresponding wrong acceptance cost, namely $\lambda_{PP} < \lambda_{BP} < \lambda_{NP}$. Similarly, when misclassified, we have $\lambda_{NN} < \lambda_{BN} < \lambda_{PN}$. Therefore, the expected losses $R(a_\bullet|x)$ ($\bullet \in \{P, B, N\}$) of object x under the above three decision actions are as follows (Equation (19)):

$$\begin{aligned} R(a_P|x) &= \lambda_{PP}\Pr([\tilde{S}_{SF}^\delta]_s|x) + \lambda_{PN}\Pr([\tilde{S}_{SF}^\delta]_g|x), \\ R(a_B|x) &= \lambda_{BP}\Pr([\tilde{S}_{SF}^\delta]_s|x) + \lambda_{BN}\Pr([\tilde{S}_{SF}^\delta]_g|x), \\ R(a_N|x) &= \lambda_{NP}\Pr([\tilde{S}_{SF}^\delta]_s|x) + \lambda_{NN}\Pr([\tilde{S}_{SF}^\delta]_g|x) \dots \end{aligned} \quad (19)$$

where $\Pr([\tilde{S}_{SF}^\delta]_s|x) = P(\tilde{S}_{SF}^\delta|x)$ and $\Pr([\tilde{S}_{SF}^\delta]_g|x) = 1 - P(\tilde{S}_{SF}^\delta|x)$ are the probabilities that object x belongs to a similar state of the δ -level positive or negative domain. By introducing Bayesian minimum risk decision theory, we have (Equation (20)):

$$\begin{aligned} (P) \quad & \lambda_{PP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{PN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{BP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{BN}(1 - P(\tilde{S}_{SF}^\delta|x)) \text{ and} \\ & \lambda_{PP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{PN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{NP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{NN}(1 - P(\tilde{S}_{SF}^\delta|x)) \\ (B) \quad & \lambda_{BP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{BN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{PP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{PN}(1 - P(\tilde{S}_{SF}^\delta|x)) \text{ and} \\ & \lambda_{BP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{BN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{NP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{NN}(1 - P(\tilde{S}_{SF}^\delta|x)) \\ (N) \quad & \lambda_{NP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{NN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{PP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{PN}(1 - P(\tilde{S}_{SF}^\delta|x)) \text{ and} \\ & \lambda_{NP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{NN}(1 - P(\tilde{S}_{SF}^\delta|x)) \leq \lambda_{BP}P(\tilde{S}_{SF}^\delta|x) + \lambda_{BN}(1 - P(\tilde{S}_{SF}^\delta|x)) \end{aligned} \quad (20)$$

Furthermore, from Equations (19) and (20), we can obtain (Equation (21)):

$$\begin{aligned} (P') \quad & \text{If } P(\tilde{S}_{SF}^\delta|x) \geq \alpha, \text{ then } x \in \text{Pos}(X), \\ (B') \quad & \text{If } \beta < P(\tilde{S}_{SF}^\delta|x) < \alpha, \text{ then } x \in \text{Bnd}(X), \\ (N') \quad & \text{If } P(\tilde{S}_{SF}^\delta|x) \leq \beta, \text{ then } x \in \text{Neg}(X) \dots \end{aligned} \quad (21)$$

where (Equation (22))

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \quad (22)$$

In the Algorithm 1, we first measure the degree of information difference for each object according to any two attributes in the target domain (line 1 and line 2). On this basis, the triangular fuzzy similarity of each object can be calculated (line 3). It is worth noting that we can obtain triangular fuzzy similarity at different levels by adjusting the threshold parameter δ . Furthermore, the triangular fuzzy similarity is regarded as the cost loss of different classification decisions, and the final decision is implemented by comparing with the decision thresholds α and β (line 4).

In addition, the higher the value of δ is, the greater the triangular fuzzy similarity between objects is. On the one hand, since $[\tilde{S}_{SF}^\delta]_s = \{x \in U \mid \tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k) > \delta\}$, $[\tilde{S}_{SF}^\delta]_g = \{x \in U \mid \tilde{S}_{SF}(\tilde{x}_i, \tilde{x}_k) \leq \delta\}$, by changing the parameter δ , we can obtain the triangular fuzzy similarity of objects in the target domain at different levels. On the other hand, the values of $[\tilde{S}_{SF}^\delta]_s$ and $[\tilde{S}_{SF}^\delta]_g$ will directly affect the values of threshold α and β . In order to visualize the impact of the final result and the threshold, we have shown it in detail in Section 6.1.

Algorithm 1 Three-way decision model based on the triangular fuzzy similarity

Input: target domain data X_t , threshold δ , α and β .

Output: positive region object set $Pos(X)$, negative region object set $Neg(X)$, boundary region object set $Bnd(X)$.

1: **BEGIN**

2: Calculate the degree of information difference $ID_i(a_j, a_k)$ of each object in the target domain under any two attributes according to Equation (15).

3: Calculate the triangular fuzzy similarity $\tilde{S}_{TF}(\tilde{x}_i, \tilde{x}_k)$ between any two objects in the target domain using Equation (17).

4: According to Equation (21), divide the target domain X_t into three domains.

5: **END BEGIN**

4.3. Adaptation Via Iterative Refinement

In this section, we integrate the methods presented in Sections 4.1 and 4.2 and finally realize unsupervised domain adaptation to the conditional distribution of cross-domain data by introducing the label propagation algorithm. Specifically, we first obtain the initial pseudolabel \hat{y}_T of the target domain according to joint distribution adaptation, then obtain the set of boundary objects of the target domain according to the three-way decision model proposed in Section 4.2 and place these objects into objects to be classified. Once the above \hat{y}_T and $Bnd(X)$ are obtained, we effectively set a semisupervised setting for the target domain data. Following [29], we use the label propagation algorithm to discriminate the boundary objects in the target domain and update \hat{y}_T . Algorithm 2 summarizes our proposed method. Algorithm 2—which in addition to the initial stage, we only adapt to the marginal distribution—and the subsequent steps consider both the marginal distribution and the conditional distribution. In addition, the accuracy of the labels in the target domain is gradually improved as the cross domain distribution differences decrease. In the following experiments, we will show that the proposed method converges to the optimal solution in a finite number of iterations and further prove the effectiveness of the proposed method.

Algorithm 2 Our Proposed Model

Input: source domain data X_s , target domain data X_t , labels y_s of source domain data, threshold δ , α and β

Output: y_T as labels of target domain data

1: **BEGIN**

2: Initialize $\mathcal{D}_t^{(c)}$ as Null

3: **while** not converged **do**

4: (1) $W \leftarrow$ Distribution adaptation $(\mathcal{D}_t^{(c)} \dots \hat{y}_T)$ in Equation (14) and let $Z_s = W^T X_s$ and $Z_t = W^T X_t$

5: (2) Assign \hat{y}_T using classifiers trained by Z_s

6: (3) Obtain $Bnd(X)$ in Algorithm 1

7: (4) $(\mathcal{D}_t^{(c)} \dots \hat{y}_T) \leftarrow$ execute label propagation algorithm

8: **End while**

9: $y_T \leftarrow \hat{y}_T$

10: **END BEGIN**

5. Experiments

5.1. Materials

5.1.1. Data Acquisition

In order to verify the effectiveness of our proposed method and compare this method with the existing research, our experimental data are obtained from the publicly accessible ABIDE dataset. ABIDE is a multisite platform that has aggregated functional and structural brain imaging data collected from 17 different laboratories around the world, which including 539 ASD patients and 573 neurotypical controls. All subjects had corresponding resting-state fMRI images and phenotypic information such as age and gender. More details on the data collection, exclusion criteria, and scan parameters are available on the ABIDE website, namely, http://fcon_1000.projects.nitrc.org/indi/abide/, (accessed on 8 October 2020). As different sites have different numbers of limited samples, we use the data from three different sites, including NYU, UM and USM, each with more than 50 subjects and using different fMRI protocols. Specifically, there were 343 subjects, including 159 ASD patients and 184 neurotypical controls. Detailed demographic information of the subjects is listed in Table 2. In Table 2, $m \pm \text{std}$ and M/F are short for mean \pm standard deviation and male/female, respectively. In each site, we used the two-sample t-test to evaluate the differences in age between the two groups and no significant differences was observed between the control group and the ASD group, i.e., $p = 0.42$ (NYU), $p = 0.31$ (USM), $p = 0.34$ (UM). Since the subjects across different sites follow different distributions, it is necessary to perform domain adaptation. In the experiments, we use $A \rightarrow B$ to denote the knowledge transfer from source domain A to target domain B. We construct a total of six tasks: NYU \rightarrow USM, NYU \rightarrow UM, USM \rightarrow NYU, USM \rightarrow UM, UM \rightarrow NYU, and UM \rightarrow USM.

Table 2. Demographic information of the studied subjects from three imaging sites in the ABIDE database. The age values are denoted as the mean \pm standard deviation. M/F: male/female.

Site	ASD		Normal Control	
	Age (m \pm std)	Gender (M/F)	Age (m \pm std)	Gender (M/F)
NYU	14.92 \pm 7.04	64/9	15.75 \pm 6.23	70/36
USM	24.59 \pm 8.46	38/0	22.33 \pm 7.69	23/0
UM	13.85 \pm 2.29	39/9	15.03 \pm 3.64	49/16

5.1.2. Data Pre-Processing

To ensure replicability, each rs-fMRI datapoint used in this research was provided by the Preprocessed Connectome Project initiative and preprocessed by using the Data Processing Assistant for Resting-State fMRI (DPARSF) software [76]. The image preprocessing steps are listed as follows. (1) Remove the first 10 time points, (2) conduct slice timing correction, and (3) conduct head motion realignment. (4) Next, image standardization was performed by normalizing the functional images into the echo planar imaging (EPI) template, followed by (5) spatial smoothing, (6) removing the linear trend, (7) temporal filtering, and (8) removing covariates. Subsequently, the brain was divided into 90 regions of interest (ROIs) based on the Automatic Anatomical Labelling (AAL) [77] atlas, and the average time series of each ROI was extracted. Then, for each subject, we obtained a 90×90 functional connectivity symmetric matrix, where each element represents the Pearson correlation coefficient between a pair of ROIs. Finally, we convert the upper triangle into a 4005 ($90 \times 89/2$)-dimensional feature vector to represent each subject.

5.2. Competing Methods

We compared the performance of our method with the following state-of-the-art machine learning models, including one baseline method and three representation-based methods.

Baseline: In this study, we use a support vector machine (SVM) as the base classifier, which is widely used in the field of neuroimaging [11]. Specifically, we specify site data as the source domain, directly train an SVM model using the original features on it, and then

use the rest of the site data as the target domain to test the classifier we have trained. In the SVM classifier, we applied a linear kernel and searched the margin penalty using the grid-search strategy from the range of $[2^{-5}, 2^{-4} \dots, 2^4, 2^5]$ via cross-validation.

Transfer component analysis (TCA) [36]: This is a general feature transformation method that reduces the difference in the marginal distribution between different domains by learning the transfer components between domains in RKHS.

Joint distribution adaptation (JDA) [26]: The JDA approach reduces both the marginal distribution and conditional distribution between different domains.

Domain adaptation with label and structural consistency (DALSC) [29]: DALSC is an unsupervised domain adaptation method that uses the structural information of the target domain to improve the performance of the model while adjusting the marginal distribution and conditional distribution between domains.

5.3. Experimental Setup

In this work, we use 5-fold cross-validation to evaluate the performance of each method. For our method, we set $\delta = 0.3$, β is searched in $\{0.5, 0.55, \dots, 0.85, 0.9\}$, α is searched in $\{0.55, 0.6, \dots, 0.9, 0.95\}$, and $\alpha > \beta$. In addition, to evaluate the classification performance, we calculated the true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) for the classification by comparing the classified labels and gold-standard labels. Then, six evaluation metrics on test data, including the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), balanced accuracy (BAC), positive predictive value (PPV) and negative predictive value (NPV), are utilized. These metrics can be computed as follows (Equation (23)):

$$\begin{aligned} ACC &= (TP + TN) / (TP + FN + TN + FP), SEN = TP / (TP + FN) \\ SPE &= TN / (TN + FP), BAC = (SEN + SPE) / 2 \\ PPV &= TP / (TP + FP), NPV = TN / (TN + FN) \end{aligned} \quad (23)$$

For these metrics, higher values indicate better classification performance.

5.4. Results on ABIDE with Multisite fMRI Data

In this section, we present the experimental results of the proposed method and several other comparative methods on six tasks. Note that data from each site can be used as the source domain while the data from other sites can be used as the target domain. For the three domain adaptation methods (i.e., TCA, JDA, and DALSC) and our proposed method, an unsupervised adaptive experimental setup is adopted, which has no label information of the target domain to be utilized in the prediction process. The classification performance results of various methods are shown in Table 3. From Table 3, we can make the following three observations.

Table 3. Performance of five different methods in ASD classification on the multisite ABIDE database. The number in bold indicates the best result achieved under a certain metric.

Task	Method	ACC (%)	SEN (%)	SPE (%)	BAC (%)	PPV (%)	NPV (%)
NYU→UM	Baseline	54.87	49.23	62.5	55.87	64	47.62
	TCA	62.83	58.46	68.75	63.61	71.69	55.00
	JDA	64.50	66.67	61.64	64.16	69.57	58.44
	DALSC	64.60	56.92	75.00	65.96	75.51	56.25
	Ours	70.80	72.31	68.75	70.53	75.81	64.71
NYU→USM	Baseline	67.21	78.26	60.53	69.39	54.55	82.14
	TCA	68.85	82.61	60.53	71.57	55.88	85.19
	JDA	70.49	86.96	60.53	73.74	57.14	88.46
	DALSC	72.13	73.91	71.05	72.48	60.71	81.81
	Ours	75.41	91.30	65.79	78.55	61.76	92.59
USM→UM	Baseline	57.52	35.38	87.50	61.44	79.31	50.00
	TCA	58.41	38.46	85.42	61.94	78.13	50.62
	JDA	61.06	61.54	60.42	60.98	67.80	53.70
	DALSC	64.60	73.85	52.08	62.96	67.61	59.52
	Ours	69.91	76.92	60.42	68.67	72.46	65.91

Table 3. Cont.

Task	Method	ACC (%)	SEN (%)	SPE (%)	BAC (%)	PPV (%)	NPV (%)
USM→NYU	Baseline	53.25	35.42	76.71	56.06	66.67	47.46
	TCA	57.39	40.63	79.45	60.04	72.22	50.43
	JDA	60.36	64.58	54.79	59.69	65.26	54.05
	DALSC	63.91	65.63	61.64	63.63	69.23	57.69
	Ours	72.13	78.26	68.42	73.34	60.00	83.87
UM→NYU	Baseline	58.58	83.33	26.03	54.68	59.70	54.29
	TCA	61.54	82.29	34.25	58.27	62.20	59.50
	JDA	63.31	82.29	38.35	60.32	63.71	62.22
	DALSC	64.49	92.70	27.39	60.05	62.68	74.07
	Ours	71.01	90.63	45.21	67.92	68.50	78.57
UM→USM	Baseline	54.09	78.26	39.47	58.87	43.90	75.00
	TCA	60.66	73.91	52.63	63.27	48.57	76.92
	JDA	60.66	78.26	50.00	64.13	48.65	79.17
	DALSC	57.38	73.91	47.37	60.64	45.95	75.00
	Ours	68.85	82.61	60.53	71.57	55.88	85.19

First, in terms of accuracy, the domain adaptive method based on feature representation is better than the direct use of the SVM classifier to predict the target domain.

Second, the TCA method in the domain adaptation method has the worst classification result because it only considers the marginal distribution.

Finally, the experimental results show that the classification accuracy of the proposed method is better than the existing domain adaptive methods (such as TCA, JDA and DALSC) in six tasks, and it also has good performance in SEN, SPE, BAC and other indicators.

6. Discussion

In this section, we first analyze the influence of the parameters in the proposed method on the algorithm performance and then compare the proposed method with other state-of-the-art methods.

6.1. Parameter Analysis

We first analyze the impact of the number of iterations on the performance of the proposed method. As mentioned in Section 4.3, for domain adaptation, we solve the proposed model iteratively. In order to evaluate its convergence, Figure 1 shows the change in algorithm accuracy as the number of iterations increases on the six tasks. It can be seen from Figure 1 that the classification accuracy of each task is gradually improved with the increase in the number of iterations. This indicates that our model learned an invariant data distribution among domains/sites after multiple iterations. The figure shows that the accuracy rate converges in 10–15 iterations.

In addition, the values of α and β involved in the experiment represent different decision risk cost levels, and their slight differences may induce different decision results. Without any loss of generality, in order to obtain more suitable parameters, we analyze the influence of different threshold parameters on the performance of the proposed method. Specifically, in order to evaluate the method's convergence, we conducted comparison experiments at different levels on the six tasks, and the final results are shown in Figure 2. The figures show that the accuracy of the algorithm changes as the threshold changes; and although the degree of fluctuation of the accuracy is different under different (α, β) , it will eventually converge. It can be seen from Figure 2 that the optimal values of (α, β) under six tasks NYU→UM, NYU→USM, USM→UM, USM→NYU, UM→NYU and UM→USM are (0.8, 0.7), (0.75, 0.65), (0.7, 0.6), (0.8, 0.7), (0.75, 0.55), (0.9, 0.6), respectively. Furthermore, it can be observed from Figure 2 that when given smaller β and larger α , the classification accuracy of the six tasks is relatively low. This shows that smaller β and larger α result in more samples from the target domain being divided into the boundary region. More boundary objects increase the uncertainty information when implementing the label propagation algorithm, which leads to the decline of classification performance.

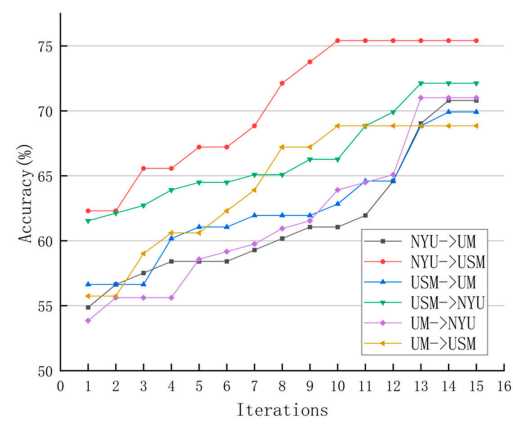
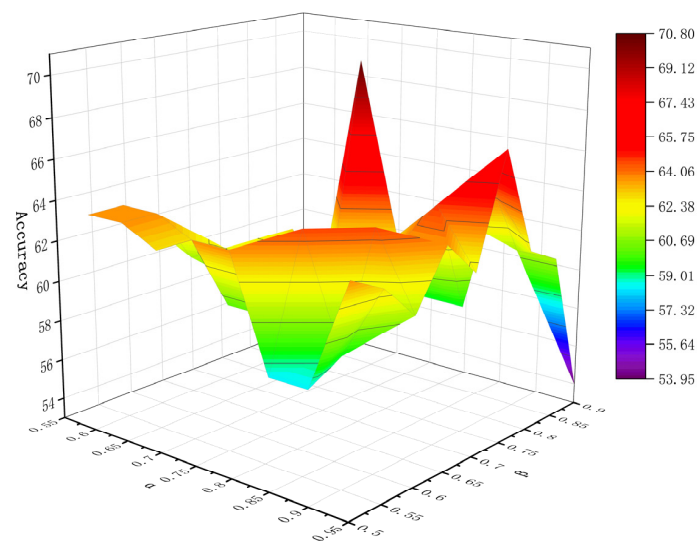
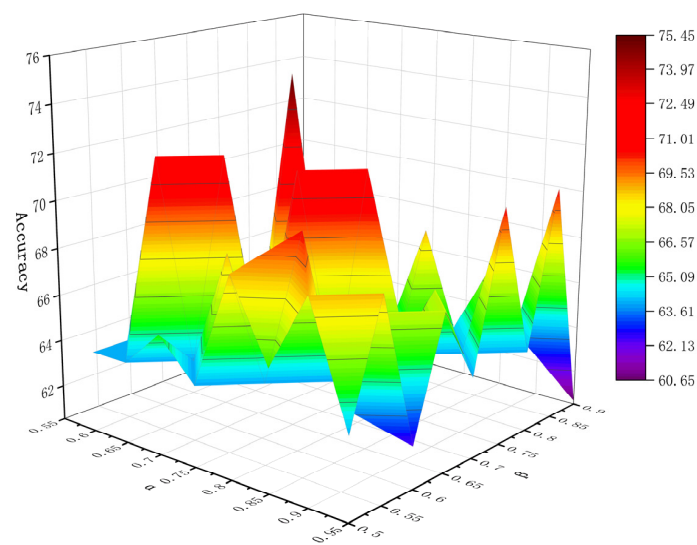


Figure 1. Classification accuracy versus the number of iterations on six domain pairs.



(a) NYU \rightarrow UM



(b) NYU \rightarrow USM

Figure 2. Cont.

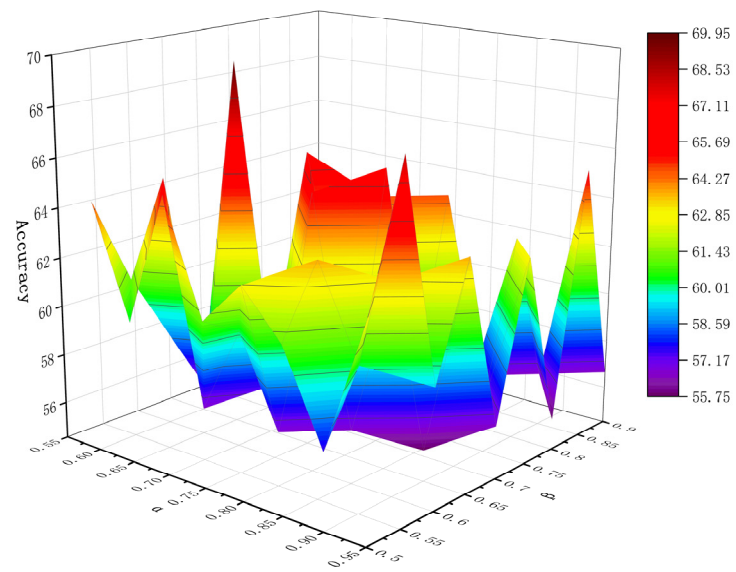
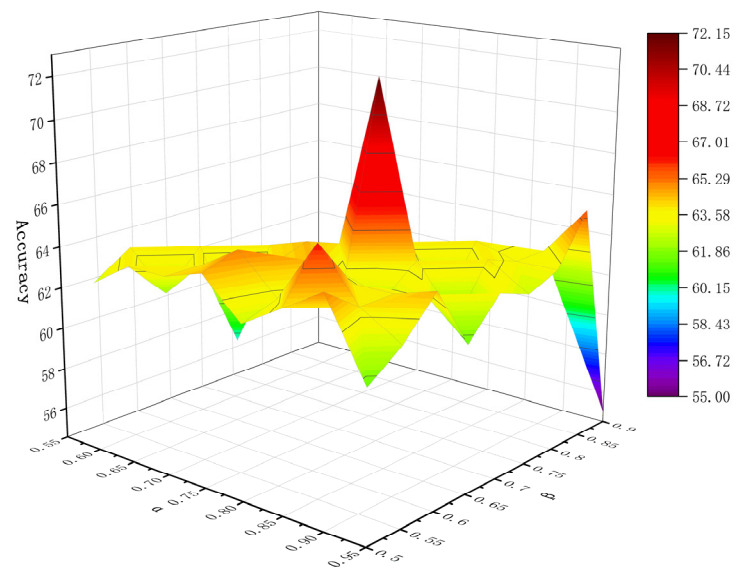
(c) USM \rightarrow UM(d) USM \rightarrow NYU

Figure 2. Cont.

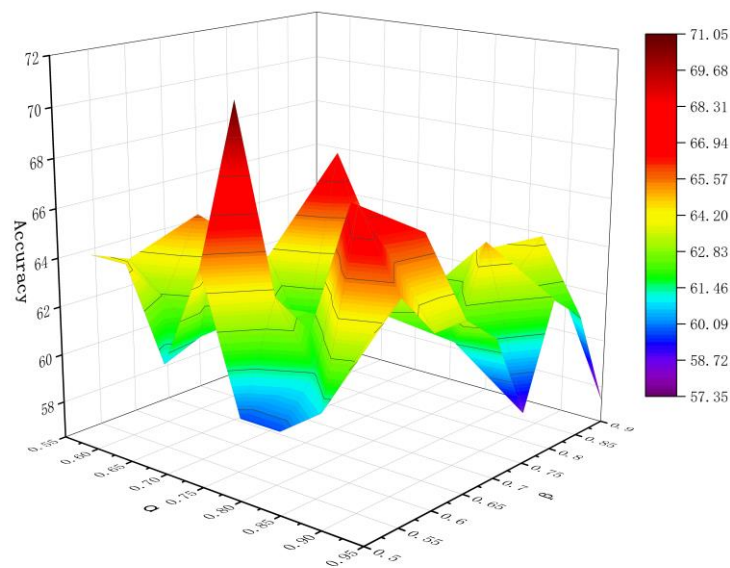
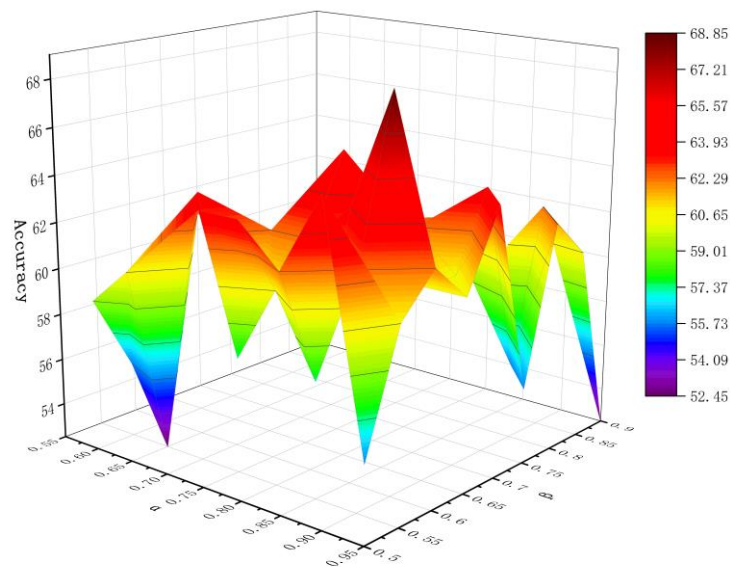
(e) UM \rightarrow NYU(f) UM \rightarrow USM

Figure 2. Classification accuracies with respect to different parameter values of α and β on six domain pairs (a) NYU \rightarrow UM; (b) NYU \rightarrow USM; (c) USM \rightarrow UM; (d) USM \rightarrow NYU; (e) UM \rightarrow NYU; (f) UM \rightarrow USM.

6.2. Comparison with State-of-the-Art Methods

To further verify the effectiveness of our proposed method, we also compare it with six other advanced methods (including the deep learning method) using the rs-fMRI data in the ABIDE database. Since only a few research papers have reported their average classification results among different sites, we only list the classification results on the NYU site in Table 4. In addition, we list the details of each method in Table 4, including the classifier and the type of feature. It is worth noting that in the research of [14,17], they selected a part of the samples from each site in proportion as the training set and then used the trained deep learning model to predict the NYU site directly.

Table 4. Comparison with state-of-the-art methods for ASD identification using rs-fMRI ABIDE data on the NYU site. HOA: Harvard Oxford Atlas. GMR: grey matter ROIs, and AAL: anatomical automatic labelling. CC200: Craddock 200. sGCN: siamese graph convolutional neural network. FCA: functional connectivity analysis. DAE: denoising autoencoder. DANN: deep attention neural networks.

Method	Feature Type	Feature Dimension	Classifier	ACC (%)
sGCN + Hing Loss [14]	HOA	111 × 111	K-Nearest Neighbor (KNN)	60.50
sGCN + Global Loss [14]	HOA	111 × 111	KNN	63.50
sGCN + Constrained Variance Loss [14]	HOA	111 × 111	KNN	68.00
FCA [17]	GMR	7266 × 7266	<i>t</i> -test	63.00
DAE [16]	CC200 Atlas	19,900	Softmax Regression	66.00
DANN [78]	AAL	6670	Deep neural network	70.90
Ours	AAL	4005	SVM	72.13/71.01

As Table 4 shows, the proposed method achieves 72.13% and 71.01% classification accuracy, respectively in the two tasks with NYU as the target domain, which is better than the models proposed in other research papers. In terms of feature type and feature dimension, this paper uses AAL atlas to divide brain regions, and obtains the original feature vector with the smallest dimension. In addition, although the sGCN, DAE and DANN are three deep learning methods, our proposed method still has a better classification effect. There may be two reasons for this. (1) Training a robust deep learning model usually requires a large number of samples. However, for multisite ASD recognition, although the data from each site can be fused together to generate a larger data set, these samples are still insufficient to train a reliable deep neural network. (2) The overfitting problem usually occurs when a deep neural network processes data with noise. In fact, fMRI data usually contain a large amount of noise information, which limits the generalization ability of the trained neural network.

7. Conclusions

In this paper, we propose a novel domain adaptation method for ASD identification with rs-fMRI data. Specifically, we introduce a three-way decision model based on triangular fuzzy similarity and divide the objects in the target domain with coarse granularity. Then, a label propagation algorithm is used to make secondary decisions on boundary region objects so as to improve the performance of ASD diagnosis based on cross-site rs-fMRI data. We conduct extensive experiments on the ABIDE dataset based on multisite data to verify the convergence and robustness of the proposed algorithm. Compared with several state-of-the-art methods, the experimental results show that the proposed method has better classification performance.

Although the classification results of our proposed method based on cross-site ASD diagnosis are significantly improved compared with the existing domain adaptation methods based on feature distribution, the following technical problems need to be considered in the future. First, although the proposed method can alleviate data heterogeneity between source and target domains, the input fMRI features are still unfiltered original high-dimensional features. However, the original high-dimensional features may have redundant features, which will reduce the performance of the model. Therefore, in the future, we will study how to combine feature selection with our methods for ASD diagnosis. Second, in this paper, we only take the functional connectivity matrix of rs-fMRI data as the feature representation of each subject without considering the network topology information. In future research, we will consider the fusion of functional brain network topology data to provide more valuable discriminant information for ASD diagnosis. Finally, in order to obtain more valuable structured information of the target domain, we will consider combining multigranularity rough sets to further improve the model performance in the future.

Author Contributions: C.S. initiated the research and wrote the paper. C.S. and X.X. performed the experiments; J.Z. supervised the research work and provided helpful suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the General Program (61977010), Key Program (61731003) of Nature Science Foundation of China and the Beijing Normal University Interdisciplinary Research Foundation for the First-Year Doctoral Candidates (BNUXKJC1925).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank all individuals who participated in the initial experiments from which raw data were collected.

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

References

1. Khan, N.A.; Waheeb, S.A.; Riaz, A.; Shang, X. A Three-Stage Teacher, Student Neural Networks and Sequential Feed Forward Selection-Based Feature Selection Approach for the Classification of Autism Spectrum Disorder. *Brain Sci.* **2020**, *10*, 754. [[CrossRef](#)] [[PubMed](#)]
2. Kong, Y.; Gao, J.; Xu, Y.; Pan, Y.; Wang, J.; Liu, J. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing* **2019**, *324*, 63–68. [[CrossRef](#)]
3. Zecavati, N.; Spence, S.J. Neurometabolic disorders and dysfunction in autism spectrum disorders. *Curr. Neurol. Neurosci.* **2009**, *9*, 129–136. [[CrossRef](#)] [[PubMed](#)]
4. Amaral, D.G.; Schumann, C.M.; Nordahl, C.W. Neuroanatomy of autism. *Trends Neurosci.* **2008**, *31*, 137–145. [[CrossRef](#)] [[PubMed](#)]
5. Khundrakpam, B.S.; Lewis, J.D.; Kostopoulos, P.; Carbonell, F.; Evans, A.C. Cortical Thickness Abnormalities in Autism Spectrum Disorders Through Late Childhood, Adolescence, and Adulthood: A Large-Scale MRI Study. *Cereb. Cortex* **2017**, *27*, 1721–1731. [[CrossRef](#)]
6. Zablotzky, B.; Black, L.I.; Maenner, M.J.; Schieve, L.A.; Blumberg, S.J. Estimated Prevalence of Autism and Other Developmental Disabilities Following Questionnaire Changes in the 2014 National Health Interview Survey. *Natl. Health Stat. Rep.* **2015**, *87*, 1–20.
7. Wang, M.; Zhang, D.; Huang, J.; Yap, P.; 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](#)]
8. 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](#)] [[PubMed](#)]
9. Mwangi, B.; Ebmeier, K.P.; Matthews, K.; Steele, J.D. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain J. Neurol.* **2012**, *135*, 1508–1521. [[CrossRef](#)]
10. Plitt, M.; Barnes, K.A.; Martin, A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage Clin.* **2015**, *7*, 359–366. [[CrossRef](#)]
11. Shi, C.; Zhang, J.; Wu, X. An fMRI Feature Selection Method Based on a Minimum Spanning Tree for Identifying Patients with Autism. *Symmetry* **2020**, *12*, 1995. [[CrossRef](#)]
12. Van den Heuvel, M.P.; Hulshoff Pol, H.E. Exploring the brain network: A review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* **2010**, *20*, 519–534. [[CrossRef](#)] [[PubMed](#)]
13. Song, J.; Yoon, N.; Jang, S.; Lee, G.; Kim, B. Neuroimaging-Based Deep Learning in Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *J. Child Adolesc. Psychiatry* **2020**, *31*, 97–104. [[CrossRef](#)]
14. Ktena, S.I.; Parisot, S.; Ferrante, E.; Rajchl, M.; Lee, M.; Glocker, B.; Rueckert, D. Metric learning with spectral graph convolutions on brain connectivity networks. *Neuroimage* **2018**, *169*, 431–442. [[CrossRef](#)] [[PubMed](#)]
15. Abraham, A.; Millham, M.P.; Di Martino, A.; Craddock, R.C.; Samaras, D.; Thirion, B.; Varoquaux, G. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage* **2017**, *147*, 736–745. [[CrossRef](#)] [[PubMed](#)]
16. Heinsfeld, A.S.; Franco, A.R.; Craddock, R.C.; Buchweitz, A.; Meneguzzi, F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin.* **2017**, *17*, 16–23. [[CrossRef](#)]
17. 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*. [[CrossRef](#)]
18. Dai, W.; Yang, Q.; Xue, G.R.; Yu, Y. Boosting for transfer learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 193–200.
19. Ren, C.X.; Dai, D.Q.; Huang, K.K.; Lai, Z.R. Transfer learning of structured representation for face recognition. *IEEE Trans. Image Process.* **2014**, *23*, 5440–5454. [[CrossRef](#)]
20. Chen, Y.; Wang, J.; Huang, M.; Yu, H. Cross-position activity recognition with stratified transfer learning. *Perov. Mob. Comput.* **2019**, *57*, 1–13. [[CrossRef](#)]

21. Yi, J.; Tao, J.; Wen, Z.; Bai, Y. Language-adversarial transfer learning for low-resource speech recognition. *IEEE-ACM Trans. Audio Speech Lang.* **2019**, *27*, 621–630. [[CrossRef](#)]
22. Do, C.B.; Ng, A.Y. Transfer learning for text classification. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 299–306.
23. Xu, Y.; Pan, S.J.; Xiong, H.; Wu, Q.; Luo, R.; Min, H.; Song, H. A Unified Framework for Metric Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 1158–1171. [[CrossRef](#)]
24. Duan, L.; Tsang, I.W.; Xu, D.; Chua, T.S. Domain adaptation from multiple sources via auxiliary classifiers. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 289–296.
25. Long, M.; Wang, J.; Ding, G.; Pan, S.J.; Yu, P.S. Adaptation Regularization: A General Framework for Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 1076–1089. [[CrossRef](#)]
26. Fernando, B.; Habrard, A.; Sebban, M.; Tuytelaars, T. Unsupervised visual domain adaptation using subspace alignment. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2960–2967.
27. Gong, B.; Shi, Y.; Sha, F.; Grauman, K. Geodesic flow kernel for unsupervised domain adaptation. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 16–21 June 2012; pp. 2066–2073.
28. Sun, B.; Feng, J.; Saenko, K. Return of frustratingly easy domain adaptation. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AR, USA, 12–17 February 2016.
29. Wang, J.; Chen, Y.; Yu, H.; Huang, M.; Yang, Q. Easy transfer learning by exploiting intra-domain structures. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019; pp. 1210–1215.
30. Long, M.; Wang, J.; Ding, G.; Sun, J.; Yu, P.S. Transfer feature learning with joint distribution adaptation. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2200–2207.
31. Wang, J.; Chen, Y.; Hao, S.; Feng, W.; Shen, Z. Balanced distribution adaptation for transfer learning. In Proceedings of the 2017 IEEE International Conference on Data Mining, New Orleans, LA, USA, 18–21 November 2017; pp. 1129–1134.
32. Zhang, J.; Li, W.; Ogunbona, P. Joint geometrical and statistical alignment for visual domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1859–1867.
33. Hou, C.; Tsai, Y.H.; Yeh, Y.; Wang, Y.F. Unsupervised Domain Adaptation with Label and Structural Consistency. *IEEE T Image Process.* **2016**, *25*, 5552–5562. [[CrossRef](#)]
34. Tahmoresnezhad, J.; Hashemi, S. Visual domain adaptation via transfer feature learning. *Knowl. Inf. Syst.* **2017**, *50*, 586–605. [[CrossRef](#)]
35. Zhang, Y.; Deng, B.; Jia, K.; Zhang, L. Label propagation with augmented anchors: A simple semi-supervised learning baseline for unsupervised domain adaptation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 781–797.
36. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain Adaptation via Transfer Component Analysis. *IEEE Trans. Neural Netw.* **2011**, *22*, 199–210. [[CrossRef](#)] [[PubMed](#)]
37. Long, M.; Wang, J.; Ding, G.; Sun, J.; Yu, P.S. Transfer joint matching for unsupervised domain adaptation. In Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1410–1417.
38. Wang, J.; Chen, Y.; Hu, L.; Peng, X.; Philip, S.Y. Stratified transfer learning for cross-domain activity recognition. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications, Athens, Greece, 19–23 March 2018; pp. 1–10.
39. Yao, Y. Three-way decisions with probabilistic rough sets. *Inf. Sci.* **2010**, *180*, 341–353. [[CrossRef](#)]
40. Chu, X.; Sun, B.; Huang, Q.; Zhang, Y. Preference degree-based multi-granularity sequential three-way group conflict decisions approach to the integration of TCM and Western medicine. *Comput. Ind. Eng.* **2020**, *143*. [[CrossRef](#)]
41. Almasvandi, Z.; Vahidinia, A.; Heshmati, A.; Zangeneh, M.M.; Goicoechea, H.C.; Jalalvand, A.R. Coupling of digital image processing and three-way calibration to assist a paper-based sensor for determination of nitrite in food samples. *RSC Adv.* **2020**, *10*, 14422–14430. [[CrossRef](#)]
42. Ren, F.; Wang, L. Sentiment analysis of text based on three-way decisions. *J. Intell. Fuzzy Syst.* **2017**, *33*, 245–254. [[CrossRef](#)]
43. Yao, Y. Decision-theoretic rough set models. In *International Conference on Rough Sets and Knowledge Technology*; Springer: Berlin/Heidelberg, Germany, 14–16 May 2007; pp. 1–12.
44. Zhang, H.; Yang, S. Three-way group decisions with interval-valued decision-theoretic rough sets based on aggregating inclusion measures. *Int. J. Approx. Reason.* **2019**, *110*, 31–45. [[CrossRef](#)]
45. Liu, P.; Yang, H. Three-way decisions with intuitionistic uncertain linguistic decision-theoretic rough sets based on generalized Maclaurin symmetric mean operators. *Int. J. Fuzzy Syst.* **2020**, *22*, 653–667. [[CrossRef](#)]
46. Agbodah, K. The determination of three-way decisions with decision-theoretic rough sets considering the loss function evaluated by multiple experts. *Granul. Comput.* **2019**, *4*, 285–297. [[CrossRef](#)]
47. Liang, D.; Wang, M.; Xu, Z.; Liu, D. Risk appetite dual hesitant fuzzy three-way decisions with TODIM. *Inf. Sci.* **2020**, *507*, 585–605. [[CrossRef](#)]
48. Qian, T.; Wei, L.; Qi, J. A theoretical study on the object (property) oriented concept lattices based on three-way decisions. *Soft Comput.* **2019**, *23*, 9477–9489. [[CrossRef](#)]
49. Yang, C.; Zhang, Q.; Zhao, F. Hierarchical Three-Way Decisions with Intuitionistic Fuzzy Numbers in Multi-Granularity Spaces. *IEEE Access* **2019**, *7*, 24362–24375. [[CrossRef](#)]
50. Yao, Y.; Deng, X. Sequential three-way decisions with probabilistic rough sets. In Proceedings of the IEEE 10th International Conference on Cognitive Informatics and Cognitive Computing, Banff, AB, Canada, 18–20 August 2011; pp. 120–125.

51. Yang, X.; Li, T.; Liu, D.; Fujita, H. A temporal-spatial composite sequential approach of three-way granular computing. *Inf. Sci.* **2019**, *486*, 171–189. [[CrossRef](#)]
52. Zhang, L.; Li, H.; Zhou, X.; Huang, B. Sequential three-way decision based on multi-granular autoencoder features. *Inf. Sci.* **2020**, *507*, 630–643. [[CrossRef](#)]
53. Liu, D.; Ye, X. A matrix factorization based dynamic granularity recommendation with three-way decisions. *Knowl. Based Syst.* **2020**, *191*. [[CrossRef](#)]
54. Yao, Y. Three-way granular computing, rough sets, and formal concept analysis. *Int. J. Approx. Reason.* **2020**, *116*, 106–125. [[CrossRef](#)]
55. Lang, G.; Luo, J.; Yao, Y. Three-way conflict analysis: A unification of models based on rough sets and formal concept analysis. *Knowl. Based Syst.* **2020**, *194*. [[CrossRef](#)]
56. Xin, X.; Song, J.; Peng, W. Intuitionistic Fuzzy Three-Way Decision Model Based on the Three-Way Granular Computing Method. *Symmetry* **2020**, *12*, 1068. [[CrossRef](#)]
57. Yue, X.D.; Chen, Y.F.; Miao, D.Q.; Fujita, H. Fuzzy neighborhood covering for three-way classification. *Inf. Sci.* **2020**, *507*, 795–808. [[CrossRef](#)]
58. Ma, Y.Y.; Zhang, H.R.; Xu, Y.Y.; Min, F.; Gao, L. Three-way recommendation integrating global and local information. *J. Eng.* **2018**, *16*, 1397–1401. [[CrossRef](#)]
59. Yu, H.; Wang, X.; Wang, G.; Zeng, X. An active three-way clustering method via low-rank matrices for multi-view data. *Inf. Sci.* **2020**, *507*, 823–839. [[CrossRef](#)]
60. Taneja, S.S. Re: Variability of the Positive Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the Society of Abdominal Radiology Prostate Cancer Disease-Focused Panel. *J. Urol.* **2020**, *204*, 1380–1381. [[CrossRef](#)] [[PubMed](#)]
61. Salama, S.; Khan, M.; Shanchei, A.; Levy, M.; Izbudak, I. MRI differences between MOG antibody disease and AQP4 NMOSD. *Mult. Scler.* **2020**, *26*, 1854–1865. [[CrossRef](#)] [[PubMed](#)]
62. Li, W.; Lin, X.; Chen, X. Detecting Alzheimer’s disease Based on 4D fMRI: An exploration under deep learning framework. *Neurocomputing* **2020**, *388*, 280–287. [[CrossRef](#)]
63. Riaz, A.; Asad, M.; Alonso, E.; Slabaugh, G. DeepfMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI. *J. Neurosci. Meth.* **2020**, *335*. [[CrossRef](#)]
64. Diciotti, S.; Orsolini, S.; Salvadori, E.; Giorgio, A.; Toschi, N.; Ciulli, S.; Ginestroni, A.; Poggesi, A.; De Stefano, N.; Pantoni, L.; et al. Resting state fMRI regional homogeneity correlates with cognition measures in subcortical vascular cognitive impairment. *J. Neurol. Sci.* **2017**, *373*, 1–6. [[CrossRef](#)]
65. Lu, H.; Liu, S.; Wei, H.; Tu, J. Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network. *Exp. Syst. Appl.* **2020**, *159*. [[CrossRef](#)]
66. Leming, M.; Górriz, J.M.; Suckling, J. Ensemble Deep Learning on Large, Mixed-Site fMRI Datasets in Autism and Other Tasks. *Int. J. Neural Syst.* **2020**, *30*. [[CrossRef](#)]
67. Benabdallah, F.Z.; Maliani, A.D.E.; Lotfi, D.; Hassouni, M.E. Analysis of the Over-Connectivity in Autistic Brains Using the Maximum Spanning Tree: Application on the Multi-Site and Heterogeneous ABIDE Dataset. In Proceedings of the 2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM), Reims, France, 27–29 October 2020; pp. 1–7.
68. 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. Neuroinf.* **2019**, *13*. [[CrossRef](#)] [[PubMed](#)]
69. Bi, X.; Wang, Y.; Shu, Q.; Sun, Q.; Xu, Q. Classification of Autism Spectrum Disorder Using Random Support Vector Machine Cluster. *Front. Genet.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
70. Rakić, M.; Cabezas, M.; Kushibar, K.; Oliver, A.; Lladó, X. Improving the detection of autism spectrum disorder by combining structural and functional MRI information. *NeuroImage Clin.* **2020**, *25*. [[CrossRef](#)] [[PubMed](#)]
71. Borgwardt, K.M.; Gretton, A.; Rasch, M.J.; Kriegel, H.; Schölkopf, B.; Smola, A.J. Integrating structured biological data by Kernel Maximum Mean Discrepancy. *Bioinformatics* **2006**, *22*, e49–e57. [[CrossRef](#)]
72. Steinwart, I. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.* **2001**, *2*, 67–93.
73. Xu, Z.S. Study on method for triangular fuzzy number-based multi-attribute decision making with preference information on alternatives. *Syst. Eng. Electron.* **2002**, *24*, 9–12.
74. Yao, Y. The superiority of three-way decisions in probabilistic rough set models. *Inf. Sci.* **2011**, *180*, 1080–1096. [[CrossRef](#)]
75. Yao, Y. An outline of a theory of three-way decisions. In Proceedings of the International Conference on Rough Sets and Current Trends in Computing, Chengdu, China, 17–20 August 2012; pp. 1–17.
76. Chao-Gan, Y.; Yu-Feng, Z. DPARSF: A MATLAB Toolbox for “Pipeline” Data Analysis of Resting-State fMRI. *Front. Syst. Neurosci.* **2010**, *4*. [[CrossRef](#)]
77. Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; Joliot, M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **2002**, *15*, 273–289. [[CrossRef](#)] [[PubMed](#)]
78. Niu, K.; Guo, J.; Pan, Y.; Gao, X.; Peng, X.; Li, N.; Li, H. Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data. *Complexity* **2020**. [[CrossRef](#)]