Self-Supervised Federated Adaptation for Multi-Site Brain Disease Diagnosis

Qiming Yang, Qi Zhu, Mingming Wang, Wei Shao, Zheng Zhang, *Senior Member, IEEE*, and Daoqiang Zhang, *Senior Member, IEEE*

***Abstract*—The multi-site approach has attracted increasing attention in brain disease diagnosis, because it can improve the prediction performance by integrating sample information from different medical institutions. However, its training procedure requires the transmission of subject's original images or features among sites, which may cause privacy disclosure. In this article, we propose a self-supervised federated adaptation (S2FA) framework for robust multi-site prediction, which can reduce the risk of privacy disclosure. As far as we know, it is the first work to investigate the cross-site brain disease diagnosis, which trains model on source sites and tests on target site, often occurring in clinical practice. First, we implement a decentralized federated optimization strategy, by which each site communicates model parameters periodically. Second, we construct an auxiliary self-supervised model for target site through transferring knowledge from source sites with self-paced learning. Then, a hash mapping is proposed to encode the target feature, simultaneously reducing the risk of privacy information disclosure and alleviating data heterogeneity among sites. Finally, we achieve the cross-site prediction by weighted federated source model and auxiliary target model. Experimental results on multi-site datasets show that the proposed S2FA can accurately identify brain disease. Our codes are available at https://github.com/nuaayqm/S2FA.**

***Index Terms*—Brain disease identification, domain adaptation, federated learning, self-supervised learning, self-paced learning.**

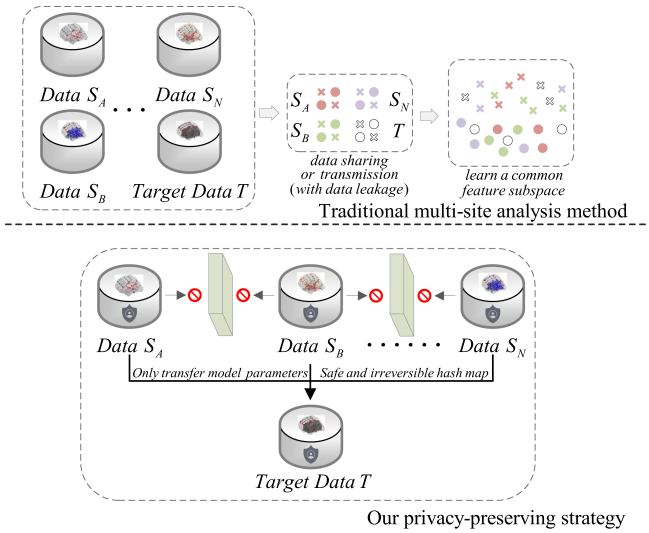


Fig. 1. The schematic diagram of traditional multi-site learning method and our federated learning strategy. The upper part of the figure shows the traditional multi-site analysis method, which mixes multi-site data and learns a common feature subspace to classify unlabeled target domain subjects. The lower part presents our federated learning strategy, in which each site trains their model locally and only transfers model parameters to the target site.

I. Introduction

NEUROIMAGING analysis technology can help reveal the underlying pathologic mechanisms of brain diseases, which has been confirmed to be effective in brain disease diagnosis [1], [2], [3], [4]. For single medical institution, it is hard to obtain sufficient subjects to train a robust model due to the high-cost data acquisition and annotations in neuroimaging. In recent years, many machine learning methods [5], [6], [7] have been proposed to fuse multi-site data to overcome the small sample size problem in single site.

The existing multi-site methods focus on integrating or fusing the samples from different sites and use these samples to train a model for disease diagnosis. Most of these methods typically assume that multi-site data have the same distribution [8], and directly train the prediction model with the samples from different sites. However, due to the inter-site variation in scanners or protocols, multi-site diagnosis is often suffered from data heterogeneity, which may severely degrade the model performance. [9]. To handle the data heterogeneity problem, some studies [10], [11] adopted multi-source domain adaptation (MSDA) to learn domain invariant representation. The MSDA can align the source and target features in a common feature space, and has been applied to multi-site brain disease diagnosis tasks [12].

As Fig. 1 shows, most of the multi-site methods only focus on improving the generalization of the diagnostic model, and our approach enables multi-site joint learning while protecting

Manuscript received 27 October 2022; revised 3 March 2023; accepted 30 March 2023. Date of publication 3 April 2023; date of current version 1 September 2023. This work was supported in part by National Natural Science Foundation of China under Grants 62076129, 62136004, 61501230, 62272226, and 61732006, in part by Key Research and Development Plan of Jiangsu Province under Grant BE2022842, and in part by the National Key R&D Program of China under Grants 2018YFC2001600 and 2018YFC2001602. Recommended for acceptance by L. Zhu. *(Qiming Yang and Qi Zhu contributed equally to this work.) (Corresponding authors: Qi Zhu; Wei Shao.)*

Qiming Yang, Qi Zhu, Mingming Wang, Wei Shao, and Daoqiang Zhang are with the MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: yangqiming@nuaa.edu.cn; zhuqinuaa@163.com; wangmingming@nuaa.edu.cn; shaowei-20022005@nuaa.edu.cn; dqzhang@nuaa.edu.cn).

Zheng Zhang is with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China, and also with the Peng Cheng Laboratory, Shenzhen, Guangdong 518055, China (e-mail: darrenzz219@gmail.com).

Digital Object Identifier 10.1109/TBDATA.2023.3264109

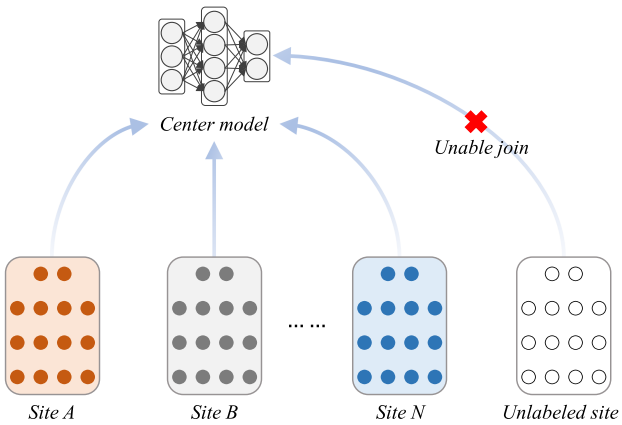


Fig. 2. The diagram of local training blocking problem in the multi-site prediction model.

privacy to some extent. Under the consensus of emphasizing the protection of privacy, the traditional multi-site approach encounters the following challenges in the real application. First, compared with local training, the multi-site method undoubtedly increases the risk of data leakage, because it requires data-sharing with other sites. Second, as privacy issues about medical data arouse more and more attention, patients might be hesitant to share their medical data because of the risk that it may be shared with employers or utilized for future health insurance decision-making [13]. Finally, from a legal point of view, many countries enforced the protection of user data via different regularizations [14].

To tackle the data-sharing problem, federated learning [15] was applied to protect privacy by using training data distributed among multiple sites. Instead of uploading data directly to a centralized data storage for building machine learning models, federated learning trains statistical models over siloed data sites, while keeping data localized. Then, it employs the statistical model of each site to learn a federated model on the central server, which can reduce the risk of data-sharing problem and encourages multi-site collaboration. Existing federated learning works mainly focus on exploring a strategy of federated aggregation to construct a central model with high generalization capability [16], [17], [18], and have been applied into learning activities of mobile phone users, adapting to pedestrian behavior in autonomous vehicles and predicting health events from wearable devices [19], [20], [21], [22]. In unsupervised multi-site diagnosis task, as shown Fig. 2, the target site only contains the unlabeled samples, which cannot join the federated training process due to the local training blocking. Besides, applying the central model directly to this site may obtain an unsatisfactory result because of the data distribution drift.

Many studies found that the communication of model parameters in federated learning also has the risk of privacy disclosure [23], [24], [25]. Differential privacy (DP) is a technique used to protect the privacy of individual data in training machine learning models. Differential privacy algorithms add noise to the data to prevent the extraction of sensitive information about individual users. However, recent studies have shown that federated learning with differential privacy algorithm requires a tradeoff between a convergence performance and privacy protection levels, i.e., higher protection level leads to a worse convergence performance [26].

In this paper, we propose a general self-supervised federated learning framework with domain adaptation for multi-site brain disease identification, called self-supervised federated adaptation (S2FA), which is designed to handle the data heterogeneity and local training blocking problems in multi-site diagnosis. Specifically, we first implement a decentralized federation optimization strategy, and employ the target federal site as a central server for receiving locally trained models from each source site. Then, we extract the self-supervised information from the unlabeled target site through the source models and use it to train an auxiliary model by a self-paced strategy. A non-linear hash mapping method is applied to extract encoded feature from target site. Third, we compute the federated weights based on the source model contribution, which is used for target federated model aggregation. Finally, we transfer the federated model and hash feature to source sites and repeat the process until convergence.

In brief, contributions can be summarized as follows.

1) A novel self-supervised federated adaptation framework (S2FA) is proposed for multi-site brain disease identification, which can transfer multi-site knowledge to the target site and reduce the risk of privacy disclosure.
2) The self-supervised contrast constraint with self-paced learning strategy module is designed to assemble source site knowledge to promote the performance of target federated model.
3) A hash mapping strategy is proposed to encode the target feature, which can alleviate data heterogeneity from source and target sites.
4) We have verified the effectiveness of our proposed model on the datasets of autism spectrum disorder (ASD) and schizophrenia with several state-of-the-art methods.

The rest of the paper is organized as follows: Section II introduces the related works; Section III describes our proposed S2FA method; Section IV shows the experimental settings and results for multi-site ASD and schizophrenia identification tasks compared with several state-of-the-art methods; Section V presents the discussion and limitations. Section VI concludes this paper.

II. Related Works

A. Multi-Site Brain Disease Diagnosis With rs-fMRI

Multi-site analysis method has drawn increasing attention because it can integrate the samples from different sites to improve the prediction accuracy. The data heterogeneity among sites has been one of the main challenges in multi-site data analysis, which may cause worse model performance. A large number of researchers have focused on leveraging domain adaptation methods to alleviate the data heterogeneity problem in multi-site datasets based on rs-fMRI [12].

To overcome this challenge, Wang et al. [27] proposed a multi-site adaptation framework via low-rank representation decomposition for ASD identification based on functional MRI (fMRI). Zhang et al. [28] propose a transport-based multi-site

joint distribution adaptation framework to reduce heterogeneity for ASD diagnosis by aligning joint feature and label distributions using optimal transport (OT). Cheng et al. [29] proposed a multi-domain transfer learning framework for early diagnosis of Alzheimer's Disease (AD) via the joint learning of tasks in multi-auxiliary domains and the target domain. Although these efforts are quite effective in dealing with data heterogeneity, the clinical practical applications are still faraway due to the data sharing problem, which leads latent risk of privacy disclosure. In this paper, we develop a federated learning based multi-site brain disease diagnosis method to achieve high classification accuracy while reducing the risk of privacy leakage.

B. Federated Learning

In general, federated learning can be accomplished in two ways: 1) training the model in each domain using private data and just transferring model parameters, and 2) employing encryption techniques to allow secure interactions between different sites. As a result, the details of the data are not shared between the domains [30].

The transferring model parameter-based methods aimed to explore a reasonable federated weight to aggregate federated parameters for the central model. The most commonly used strategy for aggregating model is Federated Averaging (FedAvg) [18], which averages local stochastic gradient descent (SGD) updates for the primal problem. However, FedAvg lacks convergence guarantees and may diverge in data heterogeneity. Chen and Chao [31] constructed local model on each source domain and then combined them via Bayesian model ensemble.

The encryption technique-based methods allowed secure interactions between different data domains. In this setting, the differential privacy [32] and k-Anonymity [33] methods are employed for data privacy protection. The methods of DP and k-Anonymity mainly include adding noise to the data or using generalization methods to obscure sensitive features until the other site cannot distinguish the source data, making the data impossible to restore to protect user privacy. Nevertheless, these methods still require the data to be transmitted elsewhere and usually involve a trade-off between privacy and accuracy.

III. Proposed Method

Suppose we have N source sites $S = \{S_1, S_2, \ldots, S_N\}$ and one target site T. In federated adaptation problem, the k-th source site S_k contains n_k labeled examples are regarded as $S_k = \{(x_i^k, y_i^k)\}_{i=1}^{n_k}$ and the target site T with n_t unlabeled examples are considered as $T = \{x_i\}_{i=1}^{n_t}$. Without loss of generality, we assume there are n_c classification labels in both source and target sites. We aim to learn a multi-site prediction model under the constraint that source site data is stored and processed locally, with only model parameter updates by communicated periodically with the target site. In particular, the goal is typically to obtain the final target model $F_T(T)$:

$$F_T(T) = \sum_{k=1}^{N} \alpha_k F_k\left(S_k\right) + \alpha_{N+1} F_{N+1}\left(S_{N+1}\right) \tag{1}$$

where $F_k(S_k)$ is the source model for the k-th site, $F_{N+1}(S_{N+1})$ is the prediction model for auxiliary dataset S_{N+1}, and α_k is the federated weight where $\alpha_k \geq 0,\ \sum_{k=1}^{N+1} \alpha_k = 1$. Then we will introduce these models and parameters in detail in the following sections. The self-supervised contrast constraint with self-paced learning strategy in Section III-B is used to learn the auxiliary model $F_{N+1}(S_{N+1})$ in (1). The federated aggregation strategy in Section III-C is used to learn the federated weight α_k and α_{N+1} in (1). The hash MMD distribution alignment module in Section III-D is used to help source site to learn a robust model $F_k(S_k)$ in (1).

A. Overview

In this study, we employ federated adaptation to achieve disease identification across different sites via weighted federated learning without raw data communication. In this case, the target site can share the model knowledge among different sites by transferring the model parameters rather than the raw samples. Similarly, the source sites use the same setup as the target site.

Fig. 3 shows the framework of the proposed model. First, each source site trains its model and transfers the model parameters to the target site. Second, the target site acquires each model contribution for federated model aggregation based on local data. Then, we extract self-supervised information from these source models through self-supervised contrastive learning on the target site. The self-supervised information is extracted to train an auxiliary model via a self-paced strategy to assist the target model aggregation. Third, we aggregate the source and auxiliary models by a weighted average strategy to obtain the target model. The target site features are encoded by a non-linear hash mapping and transferred to each source site, while the source sites use the hash code to minimize the $\mathcal{H}$-divergence to learn the invariant representation. Finally, we repeat the above process until it converges.

The overall process of S2FA is described in Algorithm 1. In next sections, we will introduce how to obtain self-supervised information with the auxiliary tasks.

B. Self-Supervised Contrast Constraint With Self-Paced Learning Strategy

The goal of self-supervised learning is to construct feature representations that are semantically meaningful via auxiliary tasks without semantic annotations [34]. One of the core steps of self-supervised learning is to obtain the valuable self-supervised information. In this section, the proposed knowledge aggregation strategy extracts the classification probability contained in the source model by the unsupervised auxiliary task. After that, the self-supervised information is subsequently applied to train the target site via self-paced learning method during each communication.

In target site, we can obtain N source site model parameters during each communication, denoted as $F_k,\ k = 1, 2, \ldots, N$. Then, as shown in Fig. 4, we can extract each source model knowledge through the target samples. Specifically, each source site model F_k can predict the classification of the target site sample x_i and obtain its probability value P_i^k. Hence, we can

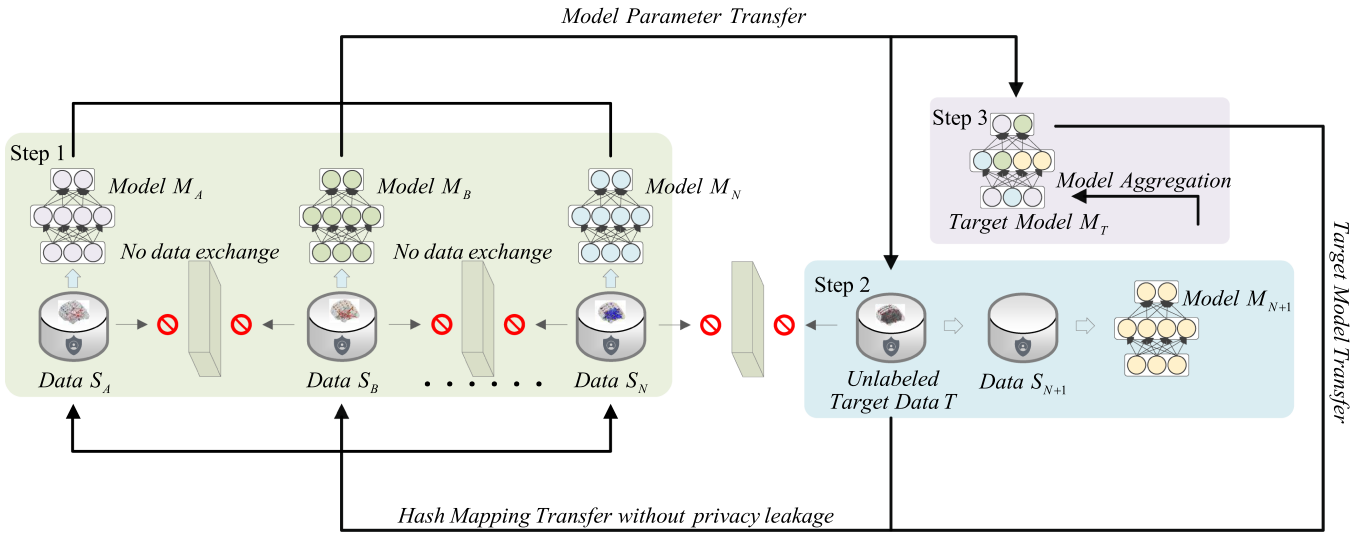


Fig. 3. The overall structure of the proposed self-supervised federated adaptation. First, each site trains model on the local data without exchanging data with other sites, and transfers the model parameter to the target site. Second, we extract the self-supervised information from unlabeled data of the target site using the source models and employ a self-paced strategy to train an auxiliary model. Meanwhile, a non-linear hash mapping strategy is employed to transfer encoded feature to source sites, which can also mitigate the impact of data heterogeneity. Finally, we obtain the target federated model by computing the federated weights based on the source model contributions, and transfer the federated model to each source site.

Algorithm 1: Self-Supervised Federated Adaptation.

Input: Labeled source sites $S = \{S_1, S_2, \ldots, S_N\}$, unlabeled target site T
Output: Target model F_T
1: **for** S_k in S **do**
2: Train source site model M_i with classification loss and MMD loss on S_k.
3: **end for**
4: Upload $\{F_k\}_{k=1}^N$ to the target site.
5: **for** F_k in $\{F_k\}_{k=1}^N$ **do**
6: Obtain target site pseudo-label and predicted value on F_k as S^k.
7: Obtain target site ReLU feature $ReLU_k(\varphi)$ on F_k.
8: **end for**
9: Assemble $\{S^k\}_{k=1}^N$ as auxiliary dataset S_{N+1}.
10: Train auxiliary model F_{N+1} with self-paced learning strategy and contrast constraint on S_{N+1}.
11: Assess source site quality and obtain federated weights $\{\alpha_k\}_{k=1}^{N+1}$.
12: Assemble federated target model via $\{\alpha_k\}_{k=1}^{N+1}$.
13: Transfer $Hash(ReLU_k(\varphi))$ to k-th source site to learn distribution invariant feature.
14: Repeat the above procedure until the parameter convergence.

obtain N auxiliary datasets S^k as:

$$S^k = \left\{x_i, P_i^k\right\}_{i=1}^{n_t}, \quad k = 1, 2, \ldots, N \tag{2}$$

However, due to the differences in confidence of source sites, some auxiliary datasets may fail to improve the training effect of the target site. Hence, we solve this problem from three

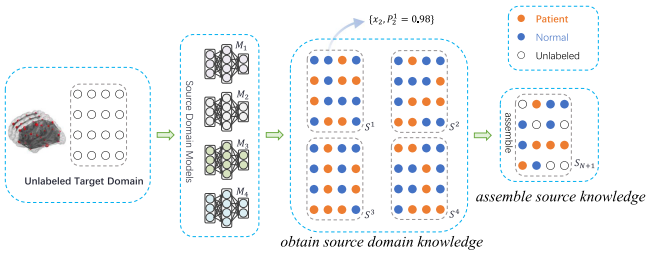


Fig. 4. The process of extracting the source model knowledge. First, we can obtain four auxiliary datasets with different predictions and pseudo-labels based on source models. Then, we assemble these datasets to acquire a more credible dataset to train the auxiliary model.

perspectives: 1) assembling the source model knowledge to improve the quality of the auxiliary dataset; 2) learning embedding representation with significant distinction by contrast constraint; 3) using a self-paced learning strategy to pick samples with high confidence to optimize target loss.

1) Assemble Source Knowledge: First, we use a confidence coefficient to filter out models that are unsure about the predicted outcomes. Specifically, for each target sample x_i, we weed out the pseudo-label with $P_i^k < \delta$ (δ is the confidence coefficient) from the auxiliary dataset. The remaining predictions are then added up to find the final pseudo-label with the high value.

For instance, the models F_1, F_2, F_3, F_4 predict the target sample x_i label $0, 0, 0, 1$ with the probability of $P_i^1 = 0.7, P_i^2 = 0.95, P_i^3 = 0.85, P_i^4 = 0.9$ respectively and the confidence coefficient $\delta = 0.8$. We first weed out the model F_1 because $P_i^1 < \delta$. We consider the remaining models F_2, F_3, F_4 are credible. Then, we discover that more models support label 0 (F_2, F_3 support label 0 and F_4 support label 1). Assuming label 0 as the pseudo-label for x_i, we can obtain the final prediction

$P_i = (P_i^2 + P_i^3)/2 = 0.9$ and the number of credible source sites $n_t^c = 2$.

After that, we can obtain a new auxiliary dataset:

$$S_{N+1} = \{(x_i, y_i), P_i, n_i^c\}_{i=1}^{n_t}, \quad P_i = \text{avg}\left\{P_i^k, P_i^k \in y_i\right\} \tag{3}$$

where y_i is the pseudo-label of target site and n_t^c is the number of credible source site. So, the auxiliary training loss can be defined as:

$$\mathcal{L}_{class} = \frac{1}{n_t}\sum_{i=1}^{n_t} L_{cla}(x_i), \tag{4}$$

$$L_{cla}(x_i) = l_c\Big(-log\left[F_{N+1}(x_i)\right], y_i\Big) \tag{5}$$

where $l_c(\cdot)$ is the cross entropy loss.

2) Self-Supervised Contrast Constraint: After obtaining the auxiliary data, we train the target site model through the contrast constraint loss. Similar to the contrastive learning methods [35], the main idea of our proposed contrast constraint is to learn an embedding representation, which preserves the distance between similar data points close and dissimilar data points far in the embedding space. To be specific, we obtain the pseudo-labels for each sample and their scores from the construction of the auxiliary dataset in the previous section. Hence, the samples with highest scores in each category are considered as the benchmark sample x_j^*. Hence, the contrast loss can be denoted as:

$$\mathcal{L}_{contrast} = \frac{1}{n_t}\sum_{i=1}^{n_t} L_{con}(x_i), \tag{6}$$

$$L_{con}(x_i) = min\left|F_{N+1}(x_i), F_{N+1}\left(x_j^*\right)\right|_{j=1}^{n_c} \tag{7}$$

where x_j^* is the benchmark sample for j-th label space and $|\cdot,\cdot|$ is the euclidean distance.

3) Self-Paced Learning Strategy: Moreover, we adopt self-paced learning strategy to optimize the target model on the auxiliary dataset to improve the model robustness. Inspired by human learning process, self-paced learning adaptively arranges the sequence of training samples to reduce the interference of noise or hard learning samples to the initial prediction model. Specifically, we definitely a score to describe the confidence for each target sample in assembling source site knowledge process, which is composed of the number of source models that choose the final pseudo-label and the average probability of the target sample pseudo-label. We select samples with high scores (e.g., $score(x_i) \geq 0.85$) to join the training. During each iteration, samples with relatively confidence are in tend to be selected from all samples, and then model parameters are updated.

$$score(x_i) = \frac{n_i^c}{N} P_i \tag{8}$$

Then we combine pseudo-label and contrast loss to optimize the final objective function:

$$\mathcal{L}_{aux} = \frac{1}{n_t}\sum_{i=1}^{n_t}\begin{cases} L_{cla}(x_i) + \beta L_{con}(x_i) & if\ score(x_i) \geq \delta \\ 0, & if\ score(x_i) < \delta \end{cases} \tag{9}$$

where β is a trade-off parameter. Here, we use the confidence coefficient δ as the score threshold.

C. Federated Aggregation

The main challenge of federated adaptation is how to determine the federated weight α in (1). The most classic algorithm in federated Learning is FedAvg, which directly averages the parameters of models coming from all participating sites without transporting local data. FedAvg has demonstrated excellent performance in many scenarios. However, due to the heterogeneity of the multi-site source data, the federated average algorithm may result in performance decrease or even divergence [36]. In addition, our task is to train the target object function F_T with multi-site source data based on federated learning. Hence, we design a personalized federated aggregation algorithm for target model through the source site contribution.

1) Source Site Quality Assessment: For each sample of target site, we adaptively estimate the quality of the knowledge obtained from source site S_k. In general, if a pseudo-label is supported by more source sites with higher confidence, then it will be more likely to represent the true label, which means the pseudo-label quality gets better. Therefore, for each target sample x_i, we define the related pseudo-label quality as $n_i^c \cdot P_i$ and the total pseudo-label quality Q is:

$$Q(S) = \sum_{i=1}^{n_t} n_i^c \cdot P_i \tag{10}$$

2) Target Model Aggregation: After the total pseudo-label quality defined, we obtain the marginal contribution quality of each source site as:

$$Q(S_k) = Q(S) - Q(S|S_k) \tag{11}$$

where $Q(S|S_k)$ is the total pseudo-label quality without S_k. We can learn from the (10) that if one source site is harmful, removing it will not impact the total quality Q a lot, which leads to a low contribution quality value. We can allocate appropriate weights to different source sites with the contribution value. Furthermore, when all source sites work important, Ben-David et al. [37] shows that the optimal α should be related to the amount of data in transfer learning theory. Hence, we first obtain $\alpha_{N+1} = n_t / \left(\sum_{k=1}^{N} n_k + n_t\right)$ based on the amount of data. Then we use the contribution quality value to determine the federated weight:

$$\alpha_k = \left(1 - \alpha_{N+1}\right)\frac{n_k Q(S_k)}{\sum_{k=1}^{N} n_k Q(s_k)} \tag{12}$$

Our strategy can mitigate the impact of unreliable source sites, because we assign it to a low weight.

D. Hash MMD Distribution Alignment

To alleviate data heterogeneity between source sites and target site, we employ domain adaptation method to map the samples from different sites to common space. It is worth noting that most of the existing domain adaptation methods have the risk of privacy leakage. They learn domain invariant representations by aligning the features of the source and target domains, while the features are invisible to each other in our task. Therefore, we propose a Hash-MMD, which utilizes the hash encoding with representative distributed information via minimizing maximum

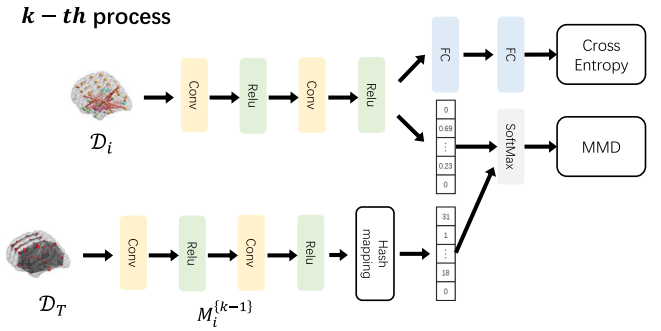


Fig. 5. The k-th training process of source domain for alleviating data heterogeneity. We perform hash mapping encoding to the extracted features using the source models on the target site. In addition, we use the feature encoding with the source feature to learn the invariant representation.

mean discrepancy (MMD) to optimize the $\mathcal{H}$-divergence. Fig. 5 shows the k-th training process of source domain for alleviating data heterogeneity.

1) Hash Encoding: Rectified Linear Unit (ReLU) is a widely-used activation function in neural network [38]. For each R-dimensional feature $\varphi = [d_1, d_2, \ldots, d_R]$, where d_i is the feature value. We can obtain $ReLU(\varphi) = max(0, \varphi)$ in training process. It means ReLU has filtered out all non-zero elements in the feature φ. Inspired by this phenomenon, we propose a hash function which can represent the distribution of non-zero elements in the feature:

$$Hash\left(ReLU\left(\varphi\right)\right) = \begin{cases} 1, & if\ d_i > 0 \\ 0, & if\ d_i = 0 \end{cases} \tag{13}$$

Then, we obtain the target non-zero distribution vector $v_{_T}$ by adding up all target feature encoding:

$$v_{_T} = \sum_{i=1}^{n_t} Hash\left(ReLU\left(\varphi_i\right)\right) \tag{14}$$

where $v_{_T}$ is a vector with R-dimension. We can transfer above model to the source sites, because it can reduce the risk of privacy leakage.

2) MMD Alignment: To alleviate the data heterogeneity between source and target site, the kernel-based MMD method has been widely used in domain adaptation tasks [39], [40]. MMD method constructs a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ endowed with a characteristic kernel k which can optimize the $\mathcal{H}$-divergence by minimizing the MMD distance on $\mathcal{H}_k$. MMD can be computed by the kernel trick:

$$MMD\left(S, v_{_T}\right) = \left\| \frac{1}{{n_k}^2} \sum_{i=1}^{n_k} k\left(f_k(x_i^k)\right) - k\left(v_{_T}\right) \right\| \tag{15}$$

where $f_k(x_i^k)$ is the feature extracter in model F_k.

Then, we optimize the MMD loss between the source sites and target site though the target non-zero distribution vector $v_{_T}$ to learn the target similarity representation. Then the source training loss can be summarized as:

$$\mathcal{L}_S = \frac{1}{n_k} \sum_{i=1}^{n_k} l_c\Big(-log\left[F_k\left(x_i^k\right)\right], y_i^k\Big) + MMD\left(S_k, v_{_T}\right) \tag{16}$$

TABLE I
DEMOGRAPHIC CHARACTERISTICS OF THE PARTICIPANTS FROM FIVE IMAGING SITES IN THE ABIDE DATABASE. THE VALUES ARE DENOTED AS MEAN ± STANDARD DEVIATION. M/F: MALE/FEMALE

Site	ASD		NC	
	Age	M/F	Age	M/F
Leuven	13.10±4.79	21/4	18.80±9.00	24/8
NYU	17.59±7.84	66/5	16.49±7.68	79/14
UCLA	16.27±6.48	28/8	14.65±4.97	31/7
UM	17.05±8.36	43/5	17.35±7.12	56/9
USM	15.77±7.21	30/8	17.34±9.53	21/1

TABLE II
DEMOGRAPHIC CHARACTERISTICS OF THE PARTICIPANTS FROM FIVE IMAGING SITES IN THE SCHIZOPHRENIA DATASETS. THE VALUES ARE DENOTED AS MEAN ± STANDARD DEVIATION. M/F: MALE/FEMALE

Site	Schizophrenia		NC	
	Age	M/F	Age	M/F
NBH	30.78±9.01	6/15	35.29±7.94	10/14
COBRE	36.75±13.68	42/11	34.82±11.28	46/21
Nottingham	33.34±9.05	27/5	33.38±8.98	95/85
Taiwan	31.59±9.60	35/34	29.87±8.62	25/37
Xiangya	23.37±7.83	49/34	27.17±6.64	35/25

IV. EXPERIMENTS AND RESULTS

A. Materials

1) Data Acquisition: To verify the effectiveness and efficiency of our proposed method, the experiment was carried out on two multi-site brain disease datasets: Autism Brain Imaging Data Exchange (ABIDE) [41] and schizophrenia datasets.

ABIDE: The Autism Brain Imaging Data Exchange (ABIDE) dataset is a qualified multi-site heterogeneous data collection. It includes resting-state fMRI and clinical data on 1,112 subjects from 17 different sites. Considering that several sites have small sample sizes, we rely on all the five sites with more than 50 subjects participated in the experiment: Leuven, USM, UCLA, UM, and NYU. These sites contain 468 subjects, including 218 patients with ASD and 250 age-matched typical control individuals (TCs). The detailed demographic information is summarized in Table I.

Schizophrenia: Five schizophrenia fMRI datasets we used are Nanjing Brain Hospital (NBH) dataset, The Center for Biomedical Research Excellence (COBRE) dataset, Nottingham(Not) dataset, Taiwan dataset, and Xiangya dataset. The subjects have the following requirements: (1) no other DSM-IV disease exists; (2) no history of drug abuse; (3) no clinically significant head trauma. The detailed demographic information is summarized in Table II.

2) Data Preprocessing: First, we perform preprocessing on resting-state fMRI, including slice-timing correction, image realignment and nuisance regression. After that, the mean

time series for a set of brain regions were extracted by the anatomical automatic labeling (AAL) [42] atlas that comprises 90 pre-defined regions-of-interest (ROIs) for ADIBE dataset and schizophrenia dataset. Finally, we used the mean time sequences of ROIs to compute the correlation matrix as functional connectivity expressed by a 90×90 symmetrical matrix for each dataset. Each element of the correlation matrix denotes the Pearson correlation coefficient between a pair of ROIs.

B. Compared Methods

In the experiments, we compare the proposed S2FA method with the following five methods, including a baseline method, three federated learning based methods and an unsupervised multi-source domain adaptation (UMDA) method.

Baseline: In this method, we use two layers of convolutional neural networks (CNN) to extract the matrix feature. The fully connected layers are utilized for classification. Specifically, we first train a model using one site until the loss converges then test the model on another site.

FedAvg[18]: In this method, we use the FedAvg, one of the most widely used aggregation methods in federated learning. To match the unsupervised task, we treat the target site as a newly federated site and directly use the trained central model for testing. To be specific, we first select one site as the target site and train the federated model with the remaining sites. Then, the FedAvg is used to accomplish the model aggregation. Finally, we test the model on the target site.

FedBN [43]: The FedBN method is a federated learning framework dealing with non-i.i.d. data which keeps the client BatchNorm (BN) [44] layer parameters updated locally and aggregating other parameters at the server. In practice, we update the non-BN layers using FedAvg and keep the client BN layers updated locally to against the data heterogeneity. In this method, we choose the model that has the best performance at all sites and apply it to the unsupervised target site.

FedProx [45]: The FedProx method is designed to address the challenges of heterogeneity both theoretically and empirically in federated networks. It adds a proximal term to the objective to improve the stability of the model. This term provides a principled way for the server to deal with heterogeneity associated with partial information. In experiment, the best performance server model is utilized to test the unsupervised target site.

KD3A [46]: This is a decentralized multi-source domain adaptation method, aiming to find the domain invariant representation without data leakage. KD3A first obtains the federated aggregation weight by knowledge distillation from the source domains. Then, a decentralized optimization strategy based on BatchNorm is employed to alleviate the data heterogeneity.

C. Experimental Setup

We implemented S2FA in Pytorch version 1.10.1, and use SGD optimizer with an initial learning rate 0.001 for the training of our models on NVIDIA RTX 3070. Grid search strategy is utilized to set the parameter β and δ. For each site, we train a model using the baseline. In addition, we obtain the experimental result after the target federated model converges. The communication interval for each method was set to be 1 epoch. The experimental setting was consistent across all methods.

Then, we use five evaluation metrics to assess the classification performance, including classification accuracy (ACC), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic (ROC) curve (AUC). In detail, we denote TP, TN, FP and FN as True Positive, True Negative, False Positive, and False Negative, respectively. The evaluation metrics can be defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$BAC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{18}$$

$$PPV = \frac{TP}{TP + FP} \tag{19}$$

$$NPV = \frac{TN}{TN + FN} \tag{20}$$

For these metrics, higher values mean better performance.

D. Results

Tables III and IV depict the comparison results between the different methods employed in our study. First, we observed that the baseline method has unsatisfactory results in most cases, as it directly applies the source model to the target site without considering the impact of data heterogeneity. Although the baseline method may achieve high results on a few metrics for some sites, such as the case of tr(NYU) with USM in Table III, it is unstable in practice. Second, while FedAvg has proven effective in many scenarios, it yields lower accuracy than the baseline method in some sites (e.g., UM, USM, COBRE, Nottingham, and Taiwan). When data are siloed from each other, the model that resemble the training distribution achieves better results. In the experiment, we employed two additional federated learning algorithms, including FedBN and FedProx, on the same datasets. We trained a federated model on source sites and deployed it on the target site. Results show that these methods outperformed FedAvg in dealing with data heterogeneity. However, these methods only mitigated the effects of data heterogeneity and did not consider the data distribution of the target site.

In the experiment, we also adopted the multi-source domain adaptation method KD3A as the comparative method. Our experimental results demonstrated that our proposed method not only achieved better result than the baseline and FedAvg methods but also outperformed the other comparision methods. The results show that our method can better handle data heterogeneity and improve model performance.

E. Comparison With State-of-the-Art in ABIDE

In this section, we compare our S2FA method with two state-of-the-art methods in ABIDE, including federated learning via domain adaptation method and non-privacy-protected multi-site disease diagnosis method.

TABLE III
PERFORMANCE (%) OF FIVE DIFFERENT METHODS IN ASD IDENTIFICATION ON THE MULTI-SITE ABIDE DATASET. HERE, EACH SITE IS ALTERNATIVELY USED AS THE TARGET SITE, AND THE REMAINING SITES ARE TREATED AS SOURCE SITES

Target Site	index	Baseline				FedAvg	FedBN	FedProx	KD3A	Ours
		tr(NYU)	tr(UCLA)	tr(UM)	tr(USM)					
Leuven	ACC	57.89	54.39	59.65	59.65	64.91	66.67	64.91	66.67	**68.42**
	AUC	62.37	58.63	66.88	52.50	61.75	63.50	68.63	64.75	**69.75**
	BAC	55.94	54.56	55.75	57.50	62.19	63.31	62.19	63.31	**65.31**
	PPV	60.53	60.71	59.57	61.54	64.25	64.44	64.29	64.44	**65.91**
	NPV	52.63	48.28	60.00	55.56	66.67	75.00	66.67	75.00	**76.92**
		tr(Leuven)	tr(UCLA)	tr(UM)	tr(USM)					
NYU	ACC	53.66	65.24	62.80	60.37	65.24	66.46	66.46	65.85	**70.12**
	AUC	56.67	69.80	66.55	65.05	72.36	71.09	74.33	74.42	**75.07**
	BAC	54.48	64.69	62.04	61.06	65.86	64.43	65.60	65.56	**69.66**
	PPV	61.64	69.57	67.02	68.42	73.08	67.27	69.79	70.79	**73.91**
	NPV	47.25	59.72	57.14	53.41	58.14	64.81	61.67	60.00	**65.28**
		tr(Leuven)	tr(NYU)	tr(UM)	tr(USM)					
UCLA	ACC	59.46	71.62	68.92	59.46	68.92	71.62	72.97	70.27	**75.68**
	AUC	58.19	71.60	76.13	63.30	79.24	80.48	79.90	75.73	**82.53**
	BAC	59.06	71.42	68.49	59.43	68.64	71.27	72.73	69.96	**75.44**
	PPV	58.33	69.77	65.31	60.53	66.67	68.09	70.45	67.39	**72.73**
	NPV	61.54	74.19	76.00	58.33	72.41	77.78	72.73	75.00	**80.00**
		tr(Leuven)	tr(NYU)	tr(UCLA)	tr(USM)					
UM	ACC	49.56	53.98	57.52	53.98	53.98	54.87	57.52	55.75	**66.37**
	AUC	63.24	63.56	67.56	70.35	74.29	75.61	76.19	70.06	**76.63**
	BAC	52.61	58.64	61.99	58.37	58.64	59.95	61.71	59.09	**69.41**
	PPV	61.76	78.26	84.00	76.00	78.26	85.00	81.48	72.73	**86.49**
	NPV	44.30	47.79	50.00	47.73	47.78	48.39	50.00	47.75	**56.58**
		tr(Leuven)	tr(NYU)	tr(UCLA)	tr(UM)					
USM	ACC	56.67	68.33	61.67	66.67	61.67	65.00	68.33	65.00	**71.67**
	AUC	50.24	73.92	65.19	69.98	73.56	69.26	70.93	73.80	**75.84**
	BAC	51.44	67.34	60.17	66.99	65.91	66.63	68.30	67.58	**71.89**
	PPV	38.89	56.00	48.00	53.57	48.65	51.61	55.56	51.52	**59.26**
	NPV	64.29	77.14	71.43	78.12	**82.61**	79.31	78.79	81.48	81.82

In the federated learning via domain adaptation method, Li et al. [47] proposed two domain adaptation methods in federated learning formulation. One is a federated learning via mixture of experts (*Fed-MoE*) strategy, which mixes the outputs of federated model and the domain expert model for domain adaptation. The other one is a federated adversarial domain alignment (*Fed-Align*) method, which uses a domain-specific local feature extractor and a global discriminator in the classification networks for aligning the feature distribution. However, both in training a domain expert model and a domain-specific local feature extractor, the target site must include a number of labeled data. Hence, we randomly label [10%, 30%, 50%] of the target site data in this experiment.

In the non-privacy-protected multi-site disease diagnosis method, Wang et al. [27] proposed a multi-site adaption framework via low-rank representation decomposition (*maLRR*) for ASD identification based on functional MRI (fMRI). Its main idea is to determine a common low-rank representation for data from the multiple sites, aiming to reduce differences in data distributions. In this experiment, we treat one site as a target domain for testing and the remaining sites as source domains for training without privacy-protecting.

The experimental results are shown in Table V. In federated learning based methods, we find that with the increasing number of annotated samples, the accuracy of the model become higher. We consider that more annotated samples can assist the model to alleviate domain shift. Then, compared to the Fed-MoE approach, which focuses more on local domain expert model, our proposed method extracts the knowledge of source and target sites to improve model performance. Compared to the Fed-Align, which focuses more on domain-specific local feature extractor, our proposed method not only uses hash MMD distribution alignment module to train target-specific model, but also trains the auxiliary model on the target site, to further extract the target site knowledge. Hence, we can learn from the results that even though the comparison method has an advantage in experimental setting, in which 50% samples are unlabeled in Fed-MoE and Fed-Align and all samples are unlabeled in our method, our method performs better in most situations. In addition, we compared a non-privacy-protect multi-source domain adaptation method, i.e. maLRR, with our method. Our method has higher accuracies in sites Leuven and UCLA compared with maLRR. Therefore, the experiment results demonstrate the effectiveness of our approach in multi-site disease diagnosis task.

V. DISCUSSION

A. Ablation Study

In this section, we demonstrate the effect of each module in our method via the ablation study, including the self-supervised contrast constraint module, the federated aggregation module, and the hash MMD distribution alignment module.

TABLE IV
PERFORMANCE (%) OF FIVE DIFFERENT METHODS IN SCHIZOPHRENIA IDENTIFICATION ON THE MULTI-SITE SCHIZOPHRENIA DATASET. HERE, EACH SITE IS ALTERNATIVELY USED AS THE TARGET SITE, AND THE REMAINING SITES ARE TREATED AS SOURCE SITES

Target Site	index	Baseline				FedAvg	FedBN	FedProx	KD3A	Ours
		tr(COBRE)	tr(Not)	tr(Taiwan)	tr(Xiangya)					
NBH	ACC	66.67	64.44	68.89	66.67	71.11	75.56	73.33	73.33	**86.67**
	AUC	75.60	74.40	77.18	82.64	85.91	86.71	86.31	85.91	**90.67**
	BAC	67.26	64.88	67.86	64.88	69.64	74.40	72.02	72.32	**86.31**
	PPV	73.68	70.00	66.67	62.86	66.67	70.97	68.75	70.00	**84.62**
	NPV	61.54	60.00	73.33	80.00	83.33	85.71	84.62	80.00	**89.47**
		tr(NBH)	tr(Not)	tr(Taiwan)	tr(Xiangya)					
COBRE	ACC	63.33	69.17	65.00	45.83	61.67	70.83	66.67	69.17	**74.17**
	AUC	72.63	76.12	77.87	46.62	75.73	76.43	76.32	77.44	**78.20**
	BAC	64.40	68.45	66.29	48.93	58.72	70.53	67.59	69.83	**73.71**
	PPV	56.52	66.00	57.75	43.48	49.48	66.67	59.70	62.50	**71.15**
	NPV	72.55	71.43	75.51	53.57	**78.26**	74.24	75.47	76.79	76.47
		tr(NBH)	tr(COBRE)	tr(Taiwan)	tr(Xiangya)					
Not	ACC	50.00	63.24	64.71	51.47	60.29	58.82	64.71	63.24	**70.59**
	AUC	67.80	73.18	66.93	54.82	76.48	74.65	75.69	71.53	**76.91**
	BAC	51.91	63.72	65.45	43.30	61.63	60.24	65.80	64.06	**71.01**
	PPV	48.21	58.97	59.52	49.09	55.10	54.00	58.70	58.14	**65.79**
	NPV	58.33	68.97	73.08	61.54	73.68	72.22	77.27	72.00	**76.67**
		tr(NBH)	tr(COBRE)	tr(Not)	tr(Xiangya)					
Taiwan	ACC	67.94	73.28	67.18	52.67	72.52	72.52	73.28	74.05	**77.10**
	AUC	74.33	78.26	74.75	53.06	81.21	78.17	78.17	79.45	**81.42**
	BAC	67.93	73.41	67.20	50.74	72.77	72.28	73.00	74.38	**77.28**
	PPV	70.15	76.56	69.70	53.10	77.05	72.60	72.97	79.66	**80.95**
	NPV	65.62	70.15	64.62	50.00	68.57	72.41	**73.68**	69.44	73.53
		tr(NBH)	tr(COBRE)	tr(Not)	tr(Taiwan)					
Xiangya	ACC	52.45	51.05	45.45	57.34	53.15	54.55	56.64	58.04	**60.14**
	AUC	59.76	46.29	41.81	57.73	57.11	55.98	58.90	57.87	**62.13**
	BAC	55.11	52.29	44.70	56.33	55.25	54.61	56.18	56.70	**59.20**
	PPV	65.31	50.66	53.25	63.41	64.81	62.50	63.64	63.53	**65.85**
	NPV	45.74	43.90	36.36	49.18	46.07	46.48	48.48	50.00	**52.46**

TABLE V
THE ACCURACY (%) OF THE STATE-OF-THE-ART METHODS IN ASD IDENTIFICATION ON THE MULTI-SITE ABIDE DATASET

Target Site	Leuven	NYU	UCLA	UM	USM
Fed-MoE(10%)	56.86	60.54	54.54	61.11	61.11
Fed-MoE(30%)	60.00	65.79	54.90	63.29	66.67
Fed-MoE(50%)	64.29	65.85	62.16	71.43	66.67
Fed-Align(10%)	56.86	57.82	65.15	59.41	61.11
Fed-Align(30%)	58.97	58.77	70.59	59.49	66.66
Fed-Align(50%)	64.29	60.98	72.97	62.50	70.00
maLRR	63.64	**71.88**	75.00	**72.73**	**74.62**
Ours	**68.42**	70.12	**75.68**	66.37	71.67

TABLE VI
THE ACCURACY (%) OF THE ABLATION EXPERIMENTS IN ASD IDENTIFICATION ON THE MULTI-SITE ABIDE DATASET

Methods	Self-supervised	Federated aggregation	Hash MMD	Target domain				
				Leuven	NYU	UCLA	UM	USM
S2FA-1	×	×	×	64.91	65.24	68.92	53.98	61.67
S2FA-2	√	×	×	64.91	65.85	70.27	56.64	65.00
S2FA-3	×	√	×	66.67	66.46	71.62	57.52	65.00
S2FA-4	×	×	√	64.91	65.85	70.27	55.75	63.33
S2FA-5	√	√	×	66.67	66.46	74.32	59.29	65.00
S2FA-6	√	×	√	66.67	67.07	70.27	62.83	66.67
S2FA-7	×	√	√	66.67	67.68	72.97	59.29	68.33
S2FA	√	√	√	**68.42**	**70.12**	**75.68**	**66.37**	**71.67**

We present the ablation experiment results in Table VI, in which S2FA-1, ..., S2FA-7 denote the variants of our proposed S2FA method. We use "×" to indicate that the corresponding module is adopted in the model, and use "√" to indicate that the corresponding module is absent. For example, S2FA-3 indicates that the variant neither uses the self-supervised contrast constraint nor introduces the hash MMD distribution alignment, but directly aggregates the source models by the weight strategy. S2FA-4 indicates that the variant uses the hash MMD distribution alignment module in the training process of the source site model, and it only uses the weight average strategy without self-supervised contrast constraint in model aggregation process. Compared with the accuracies of S2FA-5, S2FA-6, S2FA-7, and S2FA, we can find that abandoning any module of self-supervised contrast constraint, federated aggregation, and hash MMD distribution alignment will reduce the experiment accuracies. The hash MMD distribution alignment has the greatest influence on the experiment accuracies among different modules. It also can be found in Table VI that if any two modules are discarded, the accuracies will be lower than one of them is discarded, which further illustrates the important role of those three components in improving the accuracies of disease diagnosis. In addition, from the comparison of the accuracies of S2FA-2, S2FA-3, and S2FA-4, we can find that the federated aggregation module has the greatest impact on the experimental accuracies among the three components, if only one module can be preserved in the model.

Self-supervised contrast constraint module can construct an auxiliary dataset with the unlabeled samples, so the model with self-supervised contrast constraint can capture the internal knowledge embedded in the samples of the target site. Hash

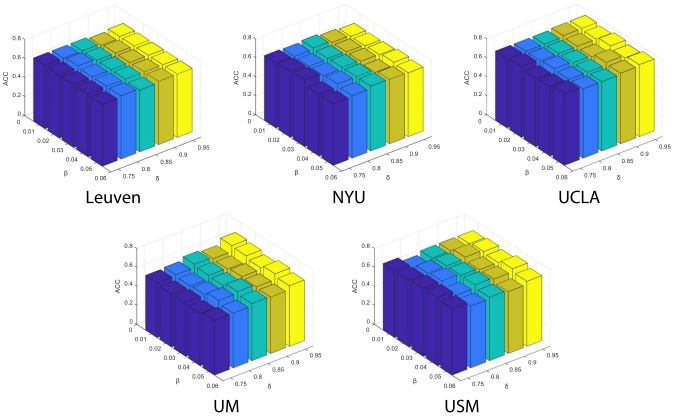


Fig. 6. Classification accuracies with different parameters of β and δ in the proposed S2FA model on five ASD sites.

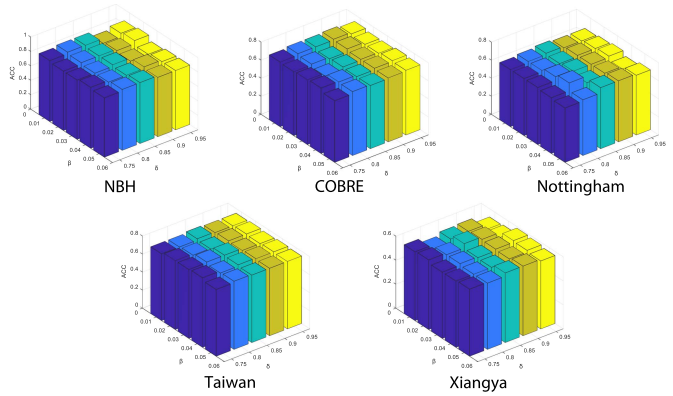


Fig. 7. Classification accuracies of different parameters of β and δ in the proposed S2FA model on five schizophrenia sites.

MMD distribution alignment module can constraint the spatial distribution of the source feature and target feature as similar as possible which can help improve the robustness of site aggregation process. However, the impact of these two modules on the S2FA method is not as great as federated aggregation module. The main reason is that the model aggregation strategy can adaptively fuse the models of source sites and amplifies the role of the other two modules. Therefore, although self-supervised contrast constraint module and hash MMD distribution alignment module are added to the model, the impact of abandoning the federated aggregation module on the experimental accuracies cannot be compensated.

B. Parameter Analysis

In this section, we analyze the effect of the hyper-parameters β and δ for the performance of our S2FA method. Here, we conduct experiments employing different parameters under the setting that each site is alternatively selected as the target site, and the remaining are treated ones as source sites. Specifically, we independently change the values of β and δ. β is selected from $\{0.01,\ 0.02,\ 0.03,\ 0.04,\ 0.05\}$, while δ is chosen from $\{0.75,\ 0.8,\ 0.85,\ 0.9,\ 0.95\}$. In Fig. 6, we can see that the site UM achieves good results with $\delta = 0.95$. We consider that due to the large data distribution difference between UM and other

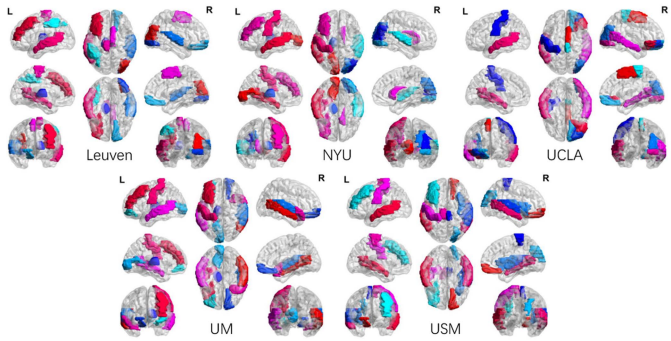


Fig. 8. Brain biomarkers associated with identifying ASD for each site.

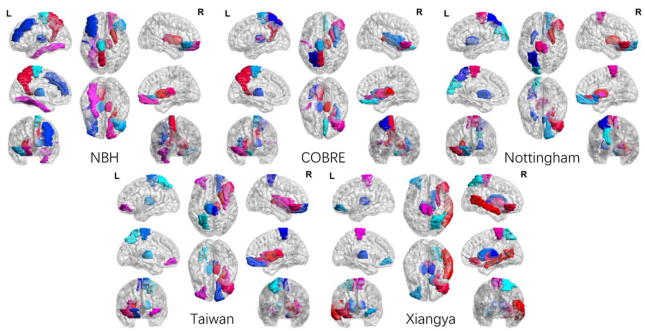


Fig. 9. Brain biomarkers associated with identifying schizophrenia for each site.

sites, the maximum confidence threshold δ is employed to obtain more credible auxiliary dataset. It is also confirmed by the results of the baseline method, with UM being the worst-performing target site. In most cases, the classification results of S2FA are stable for β and δ, indicating that our method is not sensitive to parameters.

C. Interpreting Brain Biomarkers

Apart from accuracy-related metrics, another dimension to evaluate a deep learning model is whether meaningful and robust biomarkers can be interpreted. In this subsection, we use a guided back-propagation method proposed by Li et al. [47] to interpret feature importance on ABIDE and schizophrenia dataset separately. Specifically, we applied maximizing rate of change in the model output.

$$g_k^c = ReLU\left(\frac{\partial \hat{y}^c}{\partial x_k}\right) \tag{21}$$

where c is the true label of input, and $\hat{y}^c$ is the output for label c before softmax layer, and x_k is the k-th feature of the input. g_k^c can indicate the importance of feature k for classifying as label c.

In this way, we can obtain the importance scores of 90 brain regions. Figs. 8 and 9 show the top ten brain biomarkers for each site. We can find that our method can obtain relatively stable biomarker values in different sites. Either in the ABIDE dataset or the schizophrenia dataset, we found that the Thalamus appears as a very important biomarker, which has been confirmed the effectiveness by Roy et al. [48].

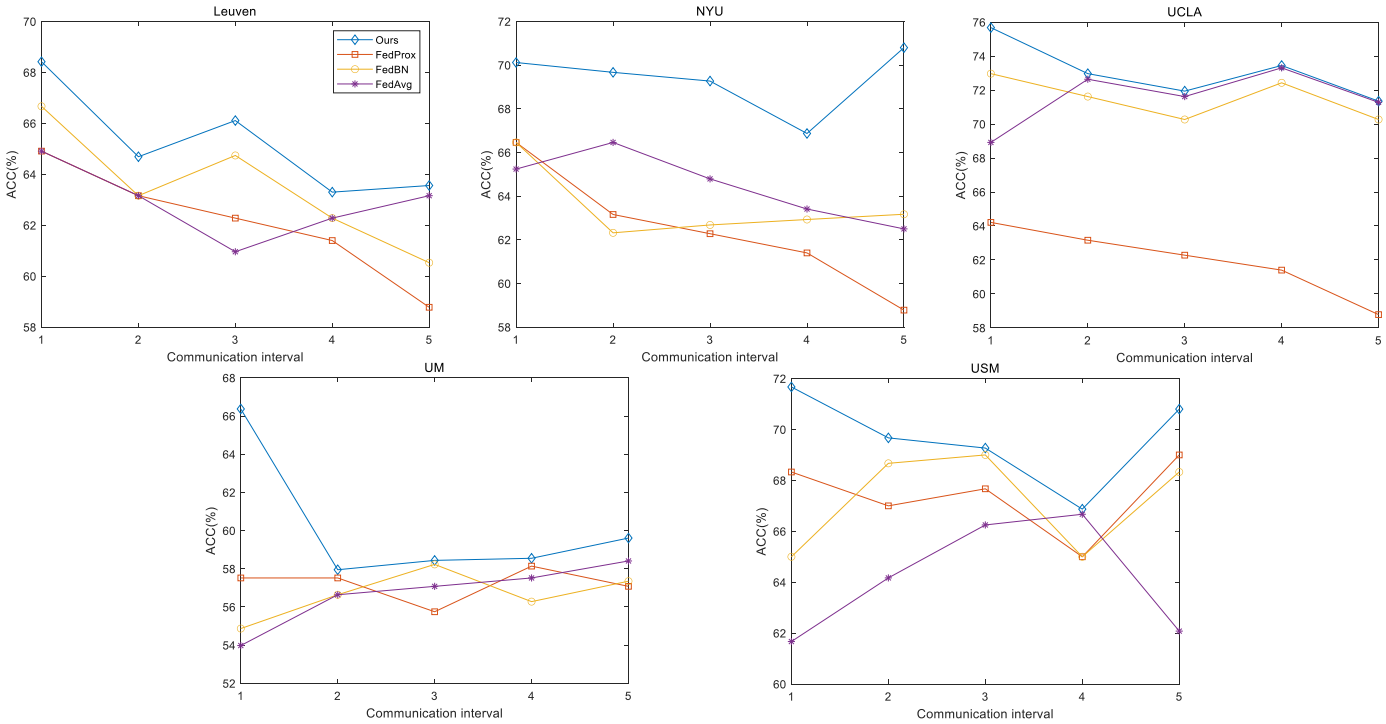


Fig. 10. The average accuracies of different communication intervals on ABIDE.

D. Communication Cost and Privacy Security

In this subsection, we conduct our S2FA method with different communication intervals and Fig. 10 shows the average accuracies on ABIDE dataset with several federated learning based methods. We regard the completion of one full training for all source sites as one communication interval. As the communication interval increases, the accuracy of our method is stable, which means that our method is robust to the communication interval. In addition, our method achieves the best classification results in each communication interval. Hence, our method can reduce the risk of privacy leakage compared to directly transferring the raw data, and it has the best classification performance compared with other state-of-the-art methods.

VI. Conclusion

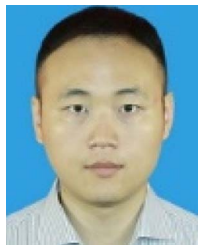
In this article, we proposed a self-supervised federated adaptation (S2FA) framework for multi-site brain disease identification. Specifically, we construct the self-supervised model for the target site by migrating prediction model knowledge from source sites with self-paced learning. Then we proposed a federated aggregation approach to assemble the models of source sites and auxiliary dataset obtained by the self-supervised learning. Finally, a hash mapping distribution alignment method is used to encode the target feature, thus both reducing the risk of privacy leakage and alleviating the data heterogeneity. Extensive experiments on both multi-site ABIDE datasets and schizophrenia datasets with fMRI demonstrate the effectiveness of the proposed method, which has promising classification results in dealing with multi-site brain disease diagnosis problem.

References

[1] A. Phinyomark, E. Ibanez-Marcelo, and G. Petri, "Resting-state fMRI functional connectivity: Big data preprocessing pipelines and topological data analysis," *IEEE Trans. Big Data*, vol. 3, no. 4, pp. 415–428, Dec. 2017.
[2] M. Wang, J. Huang, M. Liu, and D. Zhang, "Modeling dynamic characteristics of brain functional connectivity networks using resting-state functional MRI," *Med. Image Anal.*, vol. 71, 2021, Art. no. 102063.
[3] J. Huang, M. Wang, X. Xu, B. Jie, and D. Zhang, "A novel node-level structure embedding and alignment representation of structural networks for brain disease analysis," *Med. Image Anal.*, vol. 65, 2020, Art. no. 101755.
[4] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.
[5] M. Leming, J. M. Górriz, and J. Suckling, "Ensemble deep learning on large, mixed-site fMRI datasets in autism and other tasks," *Int. J. Neural Syst.*, vol. 30, no. 07, 2020, Art. no. 2050012.
[6] P. Dluhoš et al., "Multi-center machine learning in imaging psychiatry: A meta-model approach," *Neuroimage*, vol. 155, pp. 10–24, 2017.
[7] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *Neuroimage*, vol. 169, pp. 431–442, 2018.
[8] A. Abraham et al., "Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example," *Neuroimage*, vol. 147, pp. 736–745, 2017.
[9] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *Neuroimage Clin.*, vol. 17, pp. 16–23, 2018.
[10] J. Wang et al., "Multi-class ASD classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3137–3147, Oct. 2020.
[11] H. Guan, L. Wang, and M. Liu, "Multi-source domain adaptation via optimal transport for brain dementia identification," in *Proc. IEEE 18th Int. Symp. Biomed. Imag.*, 2021, pp. 1514–1517.
[12] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," 2021, *arXiv:2102.09508*.
[13] J. Roski, G. W. Bo-Linn, and T. A. Andrews, "Creating value in health care through Big Data: Opportunities and policy implications," *Health Affairs*, vol. 33, no. 7, pp. 1115–1122, 2014.
[14] N. Inkster, *China's Cyber Power*. Evanston, IL, USA: Routledge, 2018.

[15] Y. Kang, Y. He, J. Luo, T. Fan, Y. Liu, and Q. Yang, "Privacy-preserving federated adversarial domain adaptation over feature groups for interpretability," *IEEE Trans. Big Data*, to be published, doi: 10.1109/TBDATA.2022.3188292.
[16] M. Al-Shedivat, J. Gillenwater, E. Xing, and A. Rostamizadeh, "Federated learning via posterior averaging: A new perspective and practical algorithms," 2020, *arXiv:2010.05273*.
[17] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," 2021, *arXiv:2111.04263*.
[18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, PMLR, 2017, pp. 1273–1282.
[19] B. Liu, Y. Guo, and X. Chen, "PFA: Privacy-preserving federated adaptation for effective model personalization," in *Proc. Web Conf.*, 2021, pp. 923–934.
[20] D. Anguita et al., "A public domain dataset for human activity recognition using smartphones," *Esann*, vol. 3, 2013, Art. no. 3.
[21] L. Huang, Y. Yin, Z. Fu, S. Zhang, H. Deng, and D. Liu, "LoAdaBoost: Loss-based AdaBoost federated machine learning on medical data," 2018, *arXiv:1811.12629*.
[22] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.
[23] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the GAN: Information leakage from collaborative deep learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 603–618.
[24] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *Proc. IEEE Symp. Secur. Privacy*, 2019, pp. 691–706.
[25] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 14774–14784.
[26] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3454–3469, 2020.
[27] M. Wang, D. Zhang, J. Huang, P.-T. Yap, D. Shen, and M. Liu, "Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 644–655, Mar. 2020.
[28] J. Zhang, P. Wan, and D. Zhang, "Transport-based joint distribution alignment for multi-site autism spectrum disorder diagnosis using resting-state fMRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Springer, 2020, pp. 444–453.
[29] B. Cheng, M. Liu, D. Shen, Z. Li, and D. Zhang, "Multi-domain transfer learning for early diagnosis of alzheimer's disease," *Neuroinformatics*, vol. 15, no. 2, pp. 115–132, 2017.
[30] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.
[31] H.-Y. Chen and W.-L. Chao, "FedBE: Making Bayesian model ensemble applicable to federated learning," 2020, *arXiv:2009.01974*.
[32] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, Springer, 2008, pp. 1–19.
[33] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertainty Fuzziness Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, 2002.
[34] I. Misra and L. V. D. Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6707–6717.
[35] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2020, pp. 4182–4192.
[36] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on Non-IID data," 2019, *arXiv:1907.02189*.
[37] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, no. 1, pp. 151–175, 2010.
[38] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[39] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 97–105.
[40] P. Wei, Y. Ke, and C. K. Goh, "A general domain specific feature transfer framework for hybrid domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1440–1451, Aug. 2019.
[41] A. Di Martino et al., "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
[42] Y. Zhang, Y. Zhang, Y. Wang, and Q. Tian, "Domain-invariant adversarial learning for unsupervised domain adaption," 2018, *arXiv:1811.12751*.
[43] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-iid features via local batch normalization," 2021, *arXiv:2102.07623*.
[44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 448–456.
[45] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proc. Mach. Learn. Syst.*, vol. 2, pp. 429–450, 2020.
[46] H.-Z. Feng et al., "KD3A: Unsupervised multi-source decentralized domain adaptation via knowledge distillation," 2020, *arXiv:2011.09757*.
[47] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Med. Image Anal.*, vol. 65, 2020, Art. no. 101765.
[48] D. S. Roy et al., "Anterior thalamic dysfunction underlies cognitive deficits in a subset of neuropsychiatric disease models," *Neuron*, vol. 109, no. 16, pp. 2590–2603, 2021.

Qiming Yang received the BS degree in civil aviation electronics and electrical from the Nanjing University of Aeronautics and Astronautics, in Jul. 2018. He is currently working toward the MS degree with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. His current research interests include federated learning and domain adaptation.



Qi Zhu received the BS, MS, and PhD degrees from the Harbin Institute of Technology in 2007, 2010, and 2014, respectively. He is an associate professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. He has published more than 80 scientific articles in refereed international journal and conference proceedings. His interests include pattern recognition, feature extraction, and medical image analysis.

Mingming Wang received the BS degree from the Nanjing University of Aeronautics and Astronautics, in Jul. 2018. He is currently working toward the MS degree with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. His current research interests include federated learning and multi-Site disease diagnosis.

Wei Shao received the BS and MS degrees from the Nanjng University of Technology, Jiangsu, China, in 2009 and 2012, respectively, and the PhD degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018. He is an associate professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. His interests include machine learning and bioinformatics.

Zheng Zhang (Senior Member, IEEE) received the MS and PhD degrees from the Harbin Institute of Technology, China, in 2014 and 2018, respectively. He was a postdoctoral research fellow with The University of Queensland, Australia. He is currently with the Harbin Institute of Technology, Shenzhen, China. He has published more than 100 technical papers at prestigious journals and conferences. He is an Editorial Board member of the Information Processing & Management journal, and also serves/served as the AC/SPC/PC member of several top conferences. His research interests include machine learning, computer vision and multimedia analytics.

Daoqiang Zhang (Senior Member, IEEE) received the BS and PhD degree in computer science from the Nanjing University of Aeronautics and Astronautics (NUAA), China, in 1999, and 2004, respectively. He joined the Department of Computer Science and Engineering of NUAA as a lecturer in 2004, and is a professor with present. His research interests include machine learning, pattern recognition, data mining, and medical image analysis. In these areas, he has published more than 200 scientific articles in refereed international journals such as *IEEE Transactions Pattern Analysis and Machine Intelligence*, *IEEE Transactions Medical Imaging*, *IEEE Transactions Image Processing*, *Neuroimage*, *Human Brain Mapping*, *Medical Image Analysis*; and conference proceedings such as IJCAI, AAAI, NIPS, CVPR, MICCAI, KDD, with 12,000+ citations by Google Scholar. He was nominated for the National Excellent Doctoral Dissertation Award of China in 2006, won the best paper award or best student award of several international conferences such as PRICAI'06, STMI'12 and BICS'16, etc. He has served as a program committee member for several international conferences such as IJCAI, AAAI, NIPS, MICCAI, SDM, PRICAI, and ACML, etc. He is a member of the Machine Learning Society of the Chinese Association of Artificial Intelligence (CAAI), and the Artificial Intelligence & Pattern Recognition Society of the China Computer Federation (CCF).