

Article

JDC-DA: An Unsupervised Target Domain Algorithm for Alzheimer's Disease Diagnosis with Structural MRI Using Joint Domain and Category Dual Adaptation

Yuan Sui ¹, Yujie Zhang ^{1,*}, Ying Wei ² and Gang Yang ²¹ School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China; 2110689@stu.neu.edu.cn² College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

* Correspondence: gloriashangy97@gmail.com

Abstract

Domain shift in multi-source MRI imaging data significantly degrades the performance of Alzheimer's disease diagnostic models. This study aims to develop an effective unsupervised domain adaptation method to enhance diagnostic accuracy across different clinical datasets. We propose a Joint Domain and Category Dual Adaptation framework (JDC-DA) that integrates metric learning and adversarial learning. The method employs multi-scale feature aggregation to capture diverse lesion characteristics, generates dynamic prototype features through category clustering, and implements a novel metric learning approach that simultaneously aligns both domain-level and category-level feature distributions. Additionally, we introduce a classification certainty maximization strategy that establishes a dual adversarial mechanism between domain discriminator and classification discrepancy discriminator. The framework was evaluated on four public datasets (ADNI-1, ADNI-2, ADNI-3, AIBL) containing 1230 baseline sMRI scans for four classification tasks: AD vs. NC, MCI vs. NC, AD vs. MCI, and AD vs. MCI vs. NC. The proposed JDC-DA method achieved superior performance with accuracies of 92.16%, 83.56%, 81.96%, and 79.12% for the four classification tasks respectively, significantly outperforming existing state-of-the-art domain adaptation methods across all evaluation metrics. The JDC-DA framework effectively addresses domain shift challenges in Alzheimer's disease diagnosis through its integrated approach to feature alignment and adversarial learning. The method demonstrates strong potential for clinical application in automated diagnosis systems, particularly for handling multi-center neuroimaging data with distribution discrepancies.

Keywords: Alzheimer's disease; structural MRI; multi-scale feature aggregation; metric learning; adversarial learning; classification certainty maximization strategy



Academic Editors: Xiaopeng Yang and Danila Germanese

Received: 14 February 2026

Revised: 13 March 2026

Accepted: 17 March 2026

Published: 21 March 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

MSC: 68U10; 68T45; 68T07

1. Introduction

Population aging has made Alzheimer's disease (AD) a major elderly health threat. Traditional diagnostics relying on clinical assessments and cognitive tests suffer from subjectivity and delayed detection. Deep learning advancements have enabled automated AD diagnosis through end-to-end learning from medical images, extracting high-dimensional features directly from complex imaging data.

AD progression strongly correlates with cerebral atrophy. Structural MRI (sMRI) serves as a critical biomarker in AD research due to its sensitivity to brain morphological changes [1]. sMRI quantifies tissue changes and reveals characteristic alterations like hippocampal atrophy, demonstrating significant computer-aided diagnostic value. Although widely used, MRI interpretation remains expertise-dependent, being time-consuming and subjective. Early-stage subtle brain changes particularly challenge manual detection, increasing misdiagnosis risks.

Deep learning advances in medical image analysis [2–7] enable automatic brain feature extraction and precise pathological quantification, facilitating accurate AD staging. While enhancing diagnostic efficiency and enabling early intervention, AD imaging data face challenges including high dimensionality, privacy concerns, and limited availability.

sMRI captures tissue-specific changes underlying AD-related atrophy. However, scanner variations in manufacturers, protocols, and software introduce biases that compromise data stability. These confounders introduce scanner-dependent features, obscure disease-related information, reduce model generalization, and cause domain shift. Real-world data from different sources exhibit substantial distribution discrepancies due to varying imaging conditions, device parameters, processing pipelines, and population characteristics. Direct application of source-trained models to target domains typically causes performance degradation, as features become source-domain specific with limited target generalization. Domain adaptation research consequently focuses on eliminating cross-domain distribution differences.

To address AD diagnostic performance degradation from MRI domain shifts, we integrate metric learning and adversarial learning within a joint adversarial domain adaptation framework. Our approach employs multi-scale feature aggregation to enhance multi-scale lesion and structural variation perception, generating dynamic category-clustered prototypes. We compute fused domain and category prototypes across source and target domains, proposing a metric learning method combining both feature differences for cross-domain alignment. A classification certainty maximization strategy establishes a joint adversarial mechanism with domain and classification discrepancy discriminators, creating a dual adversarial learning method. Unlike conventional approaches relying solely on classifier consistency, our method simultaneously optimizes output consistency and certainty, preventing ambiguous decisions while enhancing discriminative capability, stability, and accuracy.

For cross-domain variations in resolution, lesion size, and structural complexity, we introduce a multi-scale feature aggregation module that integrates hierarchical features, preserving semantic information while capturing spatial details. This enables simultaneous perception of fine small-lesion structures and large-lesion contours, mitigating scale-related information loss while improving generalization and robustness.

In summary, we propose a joint adversarial domain adaptation method for AD diagnosis incorporating multi-scale feature aggregation, combining metric and adversarial learning advantages. Our integrated feature alignment method and dual adversarial learning with certainty maximization enhance cross-domain decision reliability and generalization, achieving unsupervised domain adaptation for AD auxiliary diagnosis.

The main contribution of this paper are as follows:

1. We designed a Multi-Scale Feature Aggregation (MSFA) module that integrates feature information from different hierarchical levels. This module enhances the model's perception capability for diverse lesion scales and structural variations in brain regions, thereby improving its adaptability to multi-scale targets.
2. We propose a metric learning module that integrates both domain and category feature differences. This approach generates dynamic prototype features based on category

clustering and computes fused domain feature prototypes and category feature prototypes from the source and target domains. The proposed metric learning method achieves cross-domain feature alignment, enabling the model to better distinguish between different categories while aligning data distributions across domains, thereby significantly enhancing classification performance in cross-domain tasks.

3. We introduce a classification certainty maximization strategy to construct a joint adversarial mechanism comprising a domain discriminator and a classification discrepancy discriminator. This leads to the proposal of a dual adversarial learning domain adaptation module that combines both domain and category information, achieving synergistic optimization of domain adaptation through coordinated domain adversarial and category adversarial learning.
4. Extensive experiments were conducted on four benchmark datasets: ADNI-1, ADNI-2, ADNI-3, and AIBL, for unsupervised target domain four-classification tasks. These tasks include Alzheimer's Disease (AD) versus Normal Control (NC), Mild Cognitive Impairment (MCI) versus NC, AD versus MCI, and AD versus MCI versus NC. The proposed JDC-DA method achieves the best overall performance compared with several state-of-the-art domain adaptation methods.

2. Related Works

In this section, we review studies related to JDC-DA.

2.1. Magnetic Resonance Imaging-Based Diagnosis of Alzheimer's Disease

Accurate Alzheimer's disease diagnosis enables effective intervention and treatment. Structural MRI has become essential in clinical practice due to its non-invasiveness, rapid acquisition, and cost-effectiveness. MRI-based AD diagnosis classifies subjects into cognitive stages—Normal Control, Mild Cognitive Impairment, and Alzheimer's Disease—using structural sequences, constituting a classification problem addressed through machine learning and deep learning approaches.

Traditional methods rely on handcrafted feature extraction followed by classification using Bayesian classifiers, SVMs, logistic regression, and clustering. Deep learning integrates feature extraction and classification through hierarchical neural networks, constructing higher-level features from basic components.

Deep learning has superseded most machine learning approaches in image classification due to automatic feature extraction, high accuracy, and strong generalization. Various AD classification methods employ attention mechanisms, feature fusion, and weak supervision in CNN and Transformer architectures. Representative works include Wu et al.'s weakly supervised attention approach [8], Wang et al.'s Asymmetry-Enhanced Attention Network [9], Li et al.'s 3D CNN with contrastive learning [10], Liang et al.'s multi-scale model with self-distillation [11], Hu et al.'s Conv-Swinformer [12], and Gan et al.'s lightweight architecture [13].

Deep learning requires substantial annotated data and assumes identical training-test distributions. However, domain shifts across datasets and clinical settings necessitate adaptation techniques to enhance model performance and generalizability.

2.2. Unsupervised Domain Adaptation for Medical Image Analysis

Domain adaptation techniques offer promising solutions to domain shift in MRI analysis, with growing applications in medical imaging tasks including detection, classification, and segmentation. Unsupervised domain adaptation has gained significant attention in Alzheimer's disease diagnosis [14–16] by enabling knowledge transfer from labeled source domains to unlabeled target domains, facilitating diagnosis without target domain labels.

The core objective involves mitigating distribution discrepancies through domain-invariant feature learning. Mainstream approaches employ two primary strategies: metric learning that minimizes distribution divergence metrics to align features, and adversarial training that uses domain discriminators to generate domain-indistinguishable features. Both methodologies have proven effective in domain adaptation tasks.

For feature alignment, Cai et al. [17] proposed a prototype-guided multi-scale domain adaptation framework (PMDA) for AD diagnosis, comprising: multi-scale feature extraction with 3D convolution and self-attention; prototype learning to reduce outliers and align domains; and adversarial adaptation for feature alignment. Zhang et al. [18] developed a source-free framework using multi-center prototypes and pseudo-labels, integrating CNN-Transformer architectures with balanced sampling strategies to generate accurate pseudo-labels while preserving source domain privacy.

In adversarial training, Generative Adversarial Networks [19] have been widely adopted. Jung et al. [20] introduced a cGAN model with attention-based 2D generators and 3D discriminators for synthesizing continuous 2D slices, demonstrating superior image quality. Zhang et al. [21] proposed an unsupervised conditional adversarial consistency network (UCAN) with cyclic feature adaptation for domain adaptation.

Beyond conventional alignment methods, Han et al. [22] employed broad learning systems with multi-scale feature extraction for AD diagnosis, while Nguyen et al. [23] developed an ensemble framework combining 3D-ResNet with XGBoost, integrating imaging features with demographic and cognitive data for final diagnosis.

3. Methods

Domain discriminator-based adaptive algorithms, while successfully applied in numerous medical diagnostic tasks, exhibit significant limitations in current unsupervised domain adaptation implementations. Most existing methods concentrate exclusively on global domain alignment between source and target domains while neglecting essential category-level alignment. These approaches typically employ two adversarial components: a domain discriminator and a feature generator. Both source and target samples pass through a shared feature generator, while the discriminator distinguishes between domains, and the generator attempts to deceive the discriminator by minimizing inter-domain distribution differences.

This alignment process fundamentally fails to adequately utilize the relationship between samples and specific decision boundaries, consequently hindering the acquisition of discriminative features. In Alzheimer's disease auxiliary diagnosis, for instance, the generator may produce ambiguous features near decision boundaries, as it primarily optimizes for domain distribution similarity without considering whether these features effectively discriminate between Normal Control (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) categories.

Using binary classification for illustration, feature generation relying solely on domain alignment increases diagnostic uncertainty and compromises accuracy, as shown in Figure 1a. Achieving optimal unsupervised classification performance requires simultaneous global domain alignment and category-level alignment to enable accurate cross-domain classification, demonstrated in Figure 1b. Therefore, in Alzheimer's disease diagnosis, global distribution alignment alone proves insufficient for extracting discriminative features; integration with task-specific decision boundaries becomes imperative for optimizing feature generation.

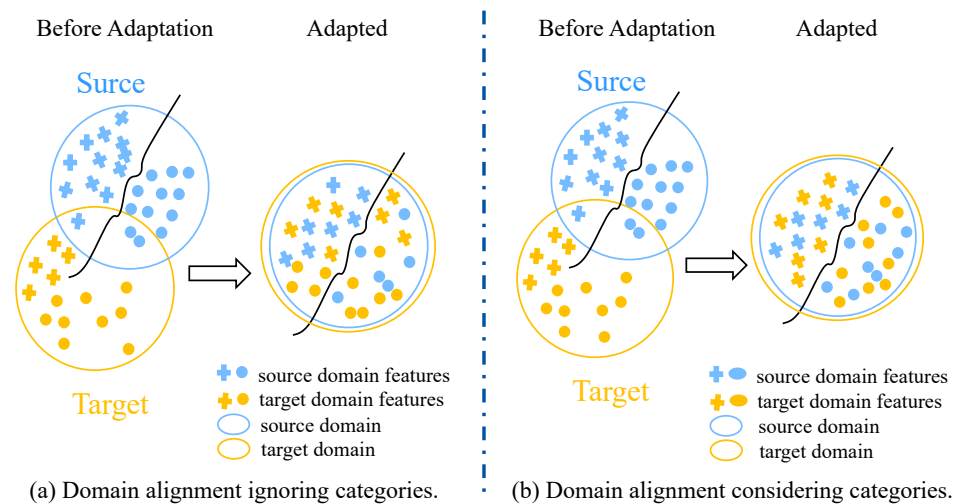


Figure 1. Diagram of two domain alignment methods.

To address the performance degradation in Alzheimer's disease diagnosis caused by domain shifts in MRI data from different sources and acquisition domains, we investigate unsupervised domain adaptation strategies for target domains. Building upon multi-scale feature extraction from original images, we integrate the advantages of two mainstream unsupervised domain adaptation approaches: metric learning and adversarial learning. We propose a metric learning method for feature alignment that incorporates both domain-level and category-level feature differences. Furthermore, we introduce a classification certainty maximization strategy and develop a dual adversarial learning framework based on joint domain and category alignment. Finally, we construct an unsupervised auxiliary diagnosis algorithm for Alzheimer's disease that achieves joint domain and category adaptation through our proposed dual-domain adaptation approach.

We denote the source domain data as $D_s = \{x_{si}\}_{i=1}^{n_s}$, which represents MRI image data for training, where x_{si} is a source domain sample and y_{si} is its corresponding label. The target domain data is denoted as $D_t = \{x_{tj}\}_{j=1}^{n_t}$, which comprises MRI images from different domains for training and testing, where x_{tj} is a target domain sample and remains unlabeled. Let n_s and n_t represent the sample sizes of the source and target domains, respectively, with both domains sharing identical feature and label spaces. We design an unsupervised domain adaptation (UDA) algorithm framework that jointly utilizes labeled source domain data and unlabeled target domain data during training, thereby significantly improving the model's generalization capability and robustness in the target domain.

3.1. Overall Model Architecture

We integrate the advantages of two mainstream UDA approaches, metric learning and adversarial learning, to develop a comprehensive framework for Alzheimer's disease diagnosis. We employ multi-scale feature aggregation to enhance the model's perception capability for diverse lesion scales and structural variations across brain regions. This approach effectively mitigates the impact of resolution and scale discrepancies across different domains, such as MRI data collected from various hospitals, while simultaneously improving the model's adaptability to these variations. We generate dynamic prototype features based on category clustering using the multi-scale aggregated features. By computing and fusing both domain feature prototypes and category feature prototypes from source and target domains, we propose a novel metric learning method that integrates both domain-level and category-level feature differences. This enables effective cross-domain feature alignment at both global and category levels. Furthermore, we introduce a classification certainty maximization strategy that leverages both joint certainty among classifiers and

individual classifier certainty to guide model optimization. This strategy facilitates the construction of a joint adversarial mechanism comprising a domain discriminator and a classification discrepancy discriminator, leading to the development of a dual adversarial learning method based on joint domain and category alignment.

The proposed unsupervised joint adversarial domain adaptation algorithm for Alzheimer's disease auxiliary diagnosis consists of three core components: a multi-scale feature extraction and aggregation module, a metric learning module that integrates domain and category feature differences, and a joint domain and category dual adversarial learning module, as illustrated in Figure 2.

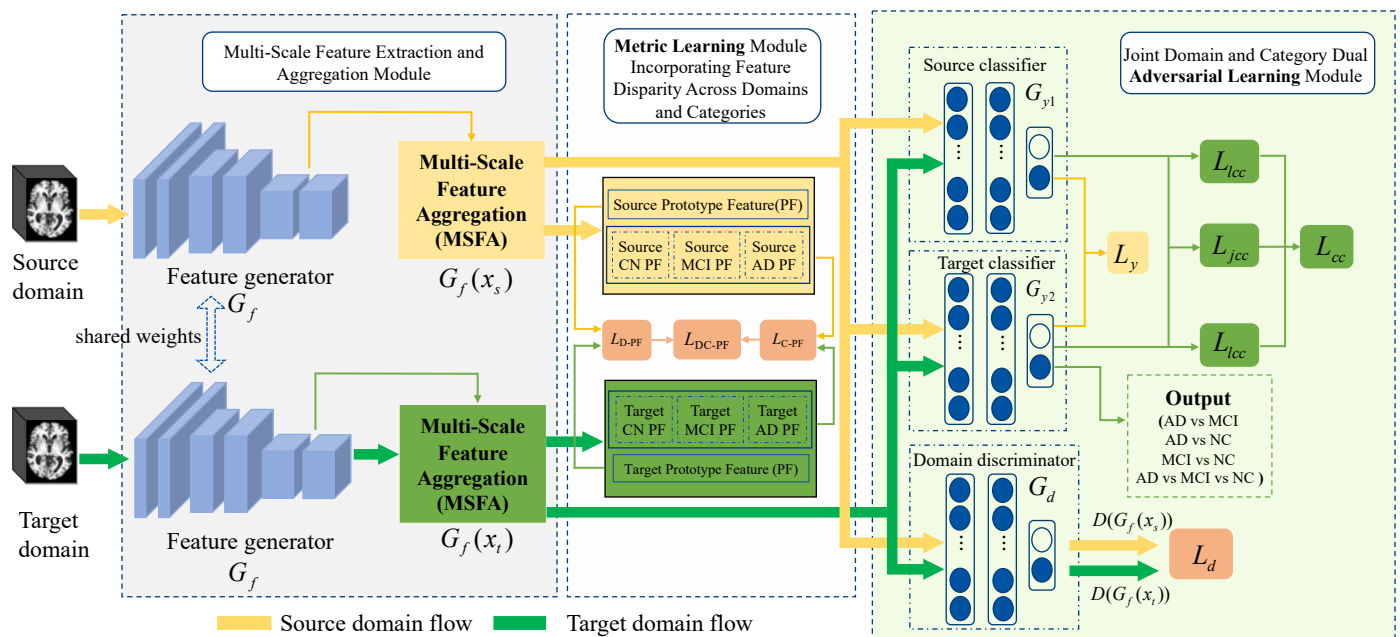


Figure 2. Overall framework of the proposed algorithm.

Our proposed framework comprises three integrated components:

1. **Multi-Scale Feature Extraction and Aggregation Module (MSFA):** We extract and aggregate multi-scale features from both source and target domains to enhance the representational capacity and discriminative power of the fused features.
2. **Incorporating Feature Disparity Across Domains and Categories Metric Learning Module (ADCML):** We construct dynamic prototypes for both domains and categories to compute feature disparities, enabling cross-domain feature alignment at both global and category levels.
3. **Joint Domain and Category Dual Adversarial Learning Module (DCDAL):** We implement a classification certainty maximization strategy and establish a joint adversarial mechanism comprising a domain discriminator and a classification discrepancy discriminator.

By integrating the advantages of two mainstream UDA paradigms, metric learning and adversarial learning, we achieve joint domain and category adaptation through coordinated feature metric learning and dual adversarial learning. This integrated approach enables effective unsupervised cross-domain auxiliary diagnosis for Alzheimer's disease, simultaneously addressing both domain-level and category-level distribution shifts.

3.2. MSFA: Multi-Scale Feature Extraction and Aggregation Module

Data distributions across different domains, such as MRI data from various hospitals, may exhibit significant variations due to differences in resolution or scale. We employ multi-scale feature extraction techniques to enhance model adaptability to these variations,

thereby mitigating the negative impact of inter-domain differences on model performance. In Alzheimer's disease diagnosis, pathological brain changes typically manifest at multiple scales: macroscopic alterations may include global structural atrophy such as hippocampal volume reduction, while microscopic changes may involve subtle local texture variations in gray and white matter signals. Consequently, single-scale feature extraction methods often fail to simultaneously capture information at these different granularities. Multi-scale feature extraction enables the concurrent capture of both large-scale structural changes and local detailed characteristics, thereby constructing more discriminative feature representations. This approach not only improves classification performance but also enhances diagnostic comprehensiveness and accuracy, providing substantial support for early detection and precise diagnosis of Alzheimer's disease.

However, during multi-scale feature extraction, as network depth increases, the model's capacity for extracting high-level semantic features improves at the expense of capturing low-level spatial information. Furthermore, the extended path between low-level features and high-level outputs impedes adequate optimization of low-level features. This high-level feature alignment approach often results in suboptimal performance when adapting to small-scale targets requiring detailed spatial information.

To address these limitations and enable joint adversarial learning across different feature hierarchies, we design a Multi-Scale Feature Aggregation (MSFA) module. This module integrates feature maps generated from both conv5 and conv6 layers, combining hierarchical feature information to preserve high-level semantics while fully utilizing low-level spatial details, thereby significantly enhancing domain adaptation performance. Through this design, the model achieves better balance between high-level semantic features and low-level spatial information, improves adaptability to multi-scale targets, and further optimizes cross-domain task performance.

The schematic diagram of this module is shown in Figure 3. The workflow proceeds as follows: first, we resize the low-level feature maps F_l extracted from the multi-scale feature extractor to match the dimensions of the high-level feature maps F_h through resampling. Subsequently, we concatenate the feature maps from different hierarchies. The concatenated feature maps then undergo transformation through a function composed of 1×1 convolution and batch normalization, enabling multi-scale information interaction and correlation to complete "feature fusion". Finally, we employ a Channel Attention Module (CAM) for "feature enhancement" to further optimize the feature representation. The channel attention module significantly enhances the effectiveness of fused features by modeling inter-dependencies between different channels. Specifically, we first perform global average pooling on the fused feature maps along spatial dimensions to extract global descriptor information for each channel. Through processing with nonlinear activation functions and sigmoid function, we generate channel attention weights. These weights reflect the importance of each channel, with higher values indicating more critical features. Finally, we perform element-wise multiplication between the generated channel weights and the original fused features, thereby suppressing irrelevant or redundant channel information while strengthening focus on key features. This channel attention mechanism enhances information beneficial for the classifier, consequently improving the model's representational capacity and discriminative power for multi-scale fused features.

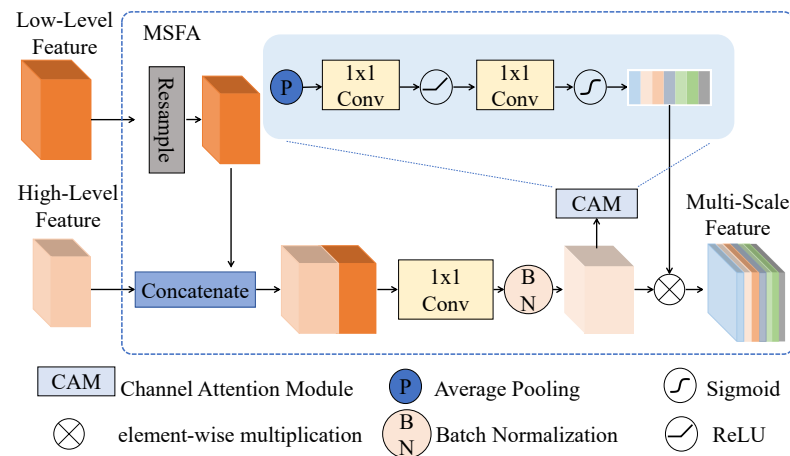


Figure 3. Schematic diagram of multi-scale feature aggregation module (MSFA).

We denote the features obtained from source and target domain samples after processing through the MSFA module as $G_f(x_s)$ and $G_f(x_t)$, respectively. By utilizing the Multi-Scale Feature Aggregation module, we effectively aggregate and enhance multi-scale information, providing robust support for domain adaptation learning across multiple scales.

3.3. ADCML: Cross-Domain and Cross-Category Metric Learning Module

As introduced in Section 1, metric learning-based domain adaptation has emerged as an active research focus in recent years. The fundamental objective of metric learning is to learn an appropriate distance metric that directly captures similarity measurements between samples, thereby providing novel perspectives for domain adaptation tasks. The core concept involves learning a shared embedding space between source and target domains to enhance cross-domain sample consistency. Specifically, metric learning methods define distance functions or similarity metrics to learn a unified feature space, bringing the data distributions of source and target domains closer together and consequently significantly improving model generalization in the target domain.

To excavate more discriminative information from cross-domain relationships and ensure that samples from the same category are closer while those from different categories are farther apart in the feature embedding space, we propose a metric learning module that integrates both domain feature disparities and category feature discrepancies. This module combines the advantages of prototype learning and metric learning, enabling the model to better distinguish between different categories while aligning data distributions across domains, thereby substantially enhancing classification performance in cross-domain tasks. Furthermore, this module can partially alleviate the instability and gradient explosion issues commonly encountered in subsequent adversarial learning domain adaptation, consequently strengthening model stability and robustness.

The module constructs average feature prototypes for source and target domain samples to compute domain feature differences, while establishing average feature prototypes for positive and negative category samples across domains to calculate category feature discrepancies.

We compute the average feature prototypes for the source and target domains as follows:

$$F_s = \frac{1}{n_s} \sum_{i=1}^{n_s} G_f(x, i)$$

$$F_t = \frac{1}{n_t} \sum_{j=1}^{n_t} G_f(x, j)$$
(1)

where F_s and F_t represent the average feature prototypes of the source and target domains, respectively, and $G_f(x, i)$ and $G_f(x, j)$ denote the feature vectors of the i -th source domain sample and the j -th target domain sample obtained through the multi-scale feature aggregation module.

The domain feature disparity is computed as:

$$d_D = \|F_s - F_t\|^2 \quad (2)$$

We calculate the domain prototype feature disparity metric loss L_{D-PF} from the domain feature disparity d_D :

$$L_{D-PF}(\theta_f) = \sqrt{d_D} \quad (3)$$

where θ_f denotes the network parameters of the multi-scale feature extraction and aggregation module, i.e., the feature generation network.

Through this formulation, we obtain the domain feature metric loss. Subsequently, we compute the category feature disparity metric loss by constructing sample feature prototypes for the three categories (NC, MCI, and AD) in both the source and target domains, and then calculating the category feature prototype disparity metrics between the source and target domains.

We compute the category feature prototypes for the source domain as follows:

$$\begin{aligned} F_{s_CN} &= \frac{1}{n_{s_P}} \sum_{i=1}^{n_{s_CN}} G_f(x, i) \\ F_{s_MCI} &= \frac{1}{n_{s_MCI}} \sum_{j=1}^{n_{s_MCI}} G_f(x, j) \\ F_{s_AD} &= \frac{1}{n_{s_AD}} \sum_{k=1}^{n_{s_AD}} G_f(x, k) \end{aligned} \quad (4)$$

where F_{s_CN} , F_{s_MCI} , and F_{s_AD} represent the category average feature prototypes for NC, MCI, and AD samples in the source domain during training, while n_{s_CN} , n_{s_MCI} , and n_{s_AD} denote the sample counts for the three categories in the source domain.

For target domain samples lacking labels, we employ k-means clustering to obtain category average feature prototypes. The clustering procedure is summarized as follows: (1) Initialization: initialize cluster centers using the three source domain feature prototypes F_{s_CN} , F_{s_MCI} , and F_{s_AD} ; (2) Assignment: assign each feature vector from the feature extraction and aggregation module to the nearest cluster center; (3) Update: recalculate cluster centers as the mean of all points in each cluster; (4) Iteration: repeat assignment and update steps until all target domain samples are processed. This yields the three target domain feature prototypes F_{t_CN} , F_{t_MCI} , and F_{t_AD} .

In practical implementation, due to the large size of 3D MRI data and hardware limitations, we set a relatively small batch size (8 in this work). Directly computing feature means for all samples in a category would require excessive computational resources. To address this challenge and enable end-to-end training, we adopt a dynamic computation approach for feature prototypes. Specifically, we first calculate the average feature prototypes for each category within the current batch, then compute a weighted average with the prototypes from previous iterations to obtain updated category prototypes. This moving average representation of category prototypes reduces noise impact and enhances model generalization.

After obtaining category feature prototypes for both source and target domains, we compute the category feature disparities as:

$$\begin{aligned}d_{CN} &= \|F_{s_{CN}} - F_{t_{CN}}\|^2 \\d_{MCI} &= \|F_{s_{MCI}} - F_{t_{MCI}}\|^2 \\d_{AD} &= \|F_{s_{AD}} - F_{t_{AD}}\|^2\end{aligned}\quad (5)$$

The category prototype feature disparity metric loss L_{C-PF} is calculated as:

$$L_{C-PF}(\theta_f) = \frac{1}{3} \left(\sqrt{d_{CN}} + \sqrt{d_{MCI}} + \sqrt{d_{AD}} \right) \quad (6)$$

We formulate the integrated domain and category feature disparity metric loss as:

$$L_{DC-PF}(\theta_f) = L_{D-PF} + \mu L_{C-PF} \quad (7)$$

where L_{DC-PF} represents the integrated feature disparity metric loss, and μ denotes a weighting coefficient. To gradually enhance the reliability of target domain sample categories during training, we design μ as an adaptive weighting coefficient that varies between 0 and 1 with increasing training iterations:

$$\begin{aligned}\mu &= \frac{2}{1 + \exp(-\beta \times p)} - 1 \\p &= \frac{E}{E_{\max}}\end{aligned}\quad (8)$$

where β is a constant, E indicates the current training epoch, and $E_{\max}$ represents the total number of training epochs.

For the Alzheimer's disease auxiliary diagnosis task, we propose a metric learning method that integrates both domain and category feature disparities. This approach effectively excavates complex structural information from imaging data, captures global and category-level features, and enhances distribution alignment capability between source and target domains, thereby improving diagnostic accuracy and model generalization under cross-domain conditions.

3.4. DCDAL: Joint Domain and Category Dual Adversarial Learning

Considering the critical role of domain-level knowledge and category information in successful domain adaptation, we propose a joint domain and category dual adversarial learning method. This approach achieves synergistic optimization through domain adversarial learning via a domain discriminator and category adversarial learning via a dual-classifier mechanism. Our architecture comprises three core components: a feature generator (the multi-scale feature extraction and aggregation module), a domain discriminator, and a discrepancy discriminator composed of dual task-specific label predictors. These components collectively form a minimax game for both domain and category adversarial learning.

The adversarial learning between the feature generator and domain discriminator ensures effective domain alignment, while the adversarial learning between the feature generator and discrepancy discriminator enables the generation of target features that approximate the source class support, thereby achieving category alignment.

In our designed joint domain and category dual adversarial learning module, we incorporate a feature generator G_f , a domain discriminator G_d , and two label classifiers G_{y1} and G_{y2} for the source and target domains. The generator G_f primarily perturbs the data source to make the source and target feature domains mutually similar. The domain

discriminator G_d mainly identifies and distinguishes between the source and target domains, implementing domain adversarial learning from a global perspective. The two label classifiers perform classification on both source and target domains, and we implement a classification certainty maximization strategy to achieve category adversarial learning. Building upon the domain discriminator and dual-classifier discriminator, we establish a minimax game between these two modules, collectively realizing global domain adversarial learning and category information-based adversarial learning.

3.4.1. Domain-Adversarial Domain Adaptation

The domain discriminator G_d is designed to distinguish the origin of samples (source or target domain). The adversarial learning between the feature generator and domain discriminator produces the domain classification loss. The domain discriminator functions as a binary classifier, where $d_i = 1$ when $x_i \in D_s$ and $d_i = 0$ when $x_i \in D_t$. This adversarial mechanism effectively explores and utilizes inter-domain discriminatory information, enabling the feature generator to learn domain-invariant features and reduce domain discrepancy. The domain adversarial learning loss is formulated as:

$$L_d(\theta_f, \theta_d) = \frac{1}{n_s + n_t} \sum_{i=1}^{n_s+n_t} CE(G_d(G_f(x_i)), d_i) \quad (9)$$

where L_d represents the domain classification loss, θ_f and θ_d denote the parameters of the feature generator G_f and domain discriminator G_d respectively, and $CE(\cdot, \cdot)$ indicates the cross-entropy loss operation.

3.4.2. Source Domain Classification Loss

We employ two label classifiers to classify the source domain data and compute the source domain label classification loss. In this work, the two classifiers utilize different initialization methods, enabling them to mitigate misclassification issues caused by insufficient feature extraction during domain alignment, despite potentially producing different predictions. The source domain label classification loss is calculated as follows:

$$L_y(\theta_f, \theta_{y1}, \theta_{y2}) = \frac{1}{2n_s} \sum_{j=1}^2 \sum_{i=1}^{n_s} CE(G_{yj}(G_f(x_{s,i})), y_{si}) \quad (10)$$

where L_y denotes the label classification loss, θ_{y1} and θ_{y2} represent the parameters of the two label classifiers G_{y1} and G_{y2} respectively, and $CE(\cdot, \cdot)$ indicates the cross-entropy loss operation. It should be noted that since only the source domain samples possess labels, L_y is computed exclusively using source domain samples.

Our discrepancy discriminator performs dual tasks: ensuring correct predictions for source samples while synthesizing discriminative target features near the source class support by maximizing the classification certainty information from both label predictors. This design effectively addresses the feature ambiguity and category uncertainty issues that may arise when using a single classifier during domain alignment. Specifically, while global alignment with a single classifier may lead to insufficient feature extraction and consequent misclassification, the dual-classifier mechanism provides additional discriminative information to mitigate these limitations.

3.4.3. Classification Certainty Maximization Strategy

In joint adversarial domain adaptation, we utilize two label classifiers to guide the model in learning domain-invariant features. Conventional methods primarily focus on the consistency between classifier outputs while overlooking the classification certainty

of individual classifiers. Consequently, existing approaches may potentially misguide the classifiers toward ambiguous outputs that compromise discriminability. To address this limitation, we propose a classification certainty maximization strategy that considers both the joint certainty between classifiers and the individual certainty of each classifier.

The dual task-specific label predictors in joint adversarial domain adaptation operate under the premise that a model can correctly classify samples when both classifiers produce consistent outputs. However, most methods merely minimize the output discrepancy between classifiers while ignoring their output certainty, which may yield ambiguous predictions where each class receives equal probability.

For instance, when constraining the two classifiers' outputs using L1-distance, specifically $\|p_1 - p_2\|_1$ where p_1 and p_2 represent the probability outputs from both classifiers, both classifiers producing predictions of [0.5, 0.5] for Alzheimer's disease would satisfy the L_1 -distance constraint despite failing to provide meaningful classification.

We therefore extend this hypothesis by incorporating classification certainty, proposing that the model can correctly classify samples when both classifiers G_{y1} and G_{y2} produce consistent and certain outputs for the same sample. We maintain that certain outputs can effectively guide the model to extract discriminative features from samples, and accordingly propose a joint classification certainty metric formulated as:

$$L_{jcc} = \sqrt{\sum_{k=1}^K \delta_k(C_1(G(X^t))) \cdot \delta_k(C_2(G(X^t)))} \quad (11)$$

where L_{jcc} denotes the joint categorical certainty loss function, $C_1(\cdot)$ and $C_2(\cdot)$ represent the prediction outputs from classifiers G_{y1} and G_{y2} for target domain samples, and $\delta_k(\cdot)$ indicates the softmax probability for the k -th class. Notably, L_{jcc} is upper-bounded by 1 since it represents the sum of dot products between softmax-normalized classifier outputs. Higher values indicate greater joint classification certainty between the classifiers.

The aforementioned equation serves to quantify the consistent certainty between both classifiers' outputs. In terms of consistency, when significant discrepancies exist between the classifiers, their predictions for the same category inevitably exhibit substantial variance, resulting in diminished joint classification certainty. Regarding certainty, even with minimal divergence between classifiers, ambiguous classification can still lead to reduced joint classification certainty. Therefore, the joint classification certainty metric simultaneously considers both the consistency and certainty of the outputs from both classifiers.

Furthermore, focusing exclusively on the outputs between classifiers may still yield ambiguous predictions when the two classifiers substantially diverge, particularly in their highest-confidence prediction categories. For instance, with prediction values of [0.8, 0.2] and [0.1, 0.9], both classifiers' outputs might be optimized toward [0.5, 0.5] to increase joint classification certainty, thereby compromising each classifier's individual discriminative capability. Consequently, we maintain that the independent certainty of each classifier's output should also be preserved, which we define as each classifier's joint classification certainty with itself:

$$L_{lcc} = \sqrt{\sum_{k=1}^K \delta_k(C_1(G(X^t)))^2 + \sum_{k=1}^K \delta_k(C_2(G(X^t)))^2} \quad (12)$$

where L_{lcc} represents the classifier independence loss. As evident from Equation (12), both L_{jcc} and L_{lcc} are necessary since L_{jcc} ensures the discriminative capability of the entire model on the target domain, while L_{lcc} preserves the certainty of each classifier's output to avoid ambiguous predictions. However, determining the appropriate weight distribution

between L_{jcc} and L_{lcc} requires careful consideration. Excessive weighting of L_{jcc} would undermine the purpose of introducing L_{lcc} , while disproportionate weighting of L_{lcc} could impair the overall discriminative capability of both classifiers. Through gradient derivation, we establish that $\nabla L_{lcc} \leq \sqrt{2} \nabla L_{jcc}$.

Intuitively, we focus on individual classifier certainty while ensuring the gradient influence of L_{lcc} does not exceed that of the joint classification certainty L_{jcc} . We therefore compute the classification certainty maximization loss as follows:

$$L_{cc} = \sqrt{2}L_{jcc} + L_{lcc} \quad (13)$$

where L_{cc} represents the classification certainty maximization loss, L_{jcc} denotes the joint classification certainty loss, and L_{lcc} indicates the individual classifier independence loss. Thus, our proposed classification certainty maximization strategy incorporates both the joint certainty among classifiers and the individual certainty of each classifier.

3.5. Overall Objective Function and Parameter Optimization

We construct an unsupervised Alzheimer's disease auxiliary diagnosis model based on joint adversarial domain adaptation. By incorporating multi-scale feature aggregation to enhance the model's perception of diverse lesion scales and structural variations in brain regions, we integrate the advantages of metric learning and adversarial learning. We propose a metric learning method that combines domain feature disparities and category feature discrepancies, introduce a classification certainty maximization strategy, and develop a domain adaptation approach with joint domain and category dual adversarial learning. This framework improves the reliability and generalization capability of cross-domain decision-making. The complete objective function is formulated as:

$$L(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d) = L_y(\theta_f, \theta_{y1}, \theta_{y2}) + \gamma L_{DC-PF} + \lambda_1 L_d(\theta_f, \theta_d) + \lambda_2 L_{cc}(\theta_f, \theta_{y1}, \theta_{y2}) \quad (14)$$

where L_y represents the label classification loss, L_{DC-PF} denotes the feature metric loss, L_d indicates the domain classification loss, and L_{cc} represents the classification certainty loss. The coefficients γ and λ are trade-off parameters with $\lambda = \lambda_1 + \lambda_2 \leq 2$, $\lambda_1 \leq 1$, and $\lambda_2 \leq 1$.

Based on the comprehensive analysis above, we formulate the optimization objectives for different loss components. To enhance target domain classification accuracy, we minimize the label classification loss L_y and metric loss L_{DC-PF} as shown in Equations (15) and (16). For the domain classification loss L_d , we maximize it to enable the feature generator to extract domain-invariant features while improving the discriminative capability of the domain discriminator, as expressed in Equation (17). Regarding the classification certainty loss L_{cc} , we maximize it to ensure both classifiers achieve maximum classification certainty to some extent, as specified in Equation (18). This formulation establishes a dual adversarial domain adaptation framework with two minimax games.

We optimize the network parameters according to the following formulations:

$$\hat{\theta}_f = \arg \min_{\theta_f} L(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_s, x_t) \quad (15)$$

$$\hat{\theta}_f, \hat{\theta}_{y1}, \hat{\theta}_{y2} = \arg \min_{\theta_f, \theta_{y1}} L(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_s) \quad (16)$$

$$\hat{\theta}_d = \arg \max_{\theta_d} L(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_s, x_t) \quad (17)$$

$$\hat{\theta}_{y1}, \hat{\theta}_{y2} = \arg \max_{\theta_{y1}, \theta_{y2}} L(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_t) \quad (18)$$

where $\hat{\theta}_f$ represents the optimal parameters for training the feature generator G_f . Equation (15) trains the feature generator G_f by minimizing the feature disparity metric loss between source and target domains. $\hat{\theta}_{y_1}$ and $\hat{\theta}_{y_2}$ denote the optimal parameters for training the two classifiers G_{y_1} and G_{y_2} using source domain data. Equation (16) trains both the feature generator G_f and the two label classifiers G_{y_1} and G_{y_2} by minimizing the source domain label classification loss based on features generated by the feature generator. $\hat{\theta}_d$ indicates the optimal parameters for training the domain discriminator G_d . Equation (17) trains the domain discriminator G_d by maximizing the domain discriminator loss to distinguish between source and target data. Equation (18) trains the two label classifiers G_{y_1} and G_{y_2} by maximizing the classification certainty loss for target domain data.

Through this dual minimax joint adversarial game, the model achieves effective training. The feature extractor acquires domain-invariant features that preserve discriminative information, while the domain discriminator and two label classifiers accomplish domain adaptation through joint domain and category dual adversarial learning. This approach enables cross-domain alignment from global to category levels, ultimately realizing unsupervised Alzheimer's disease classification diagnosis in cross-modal target domains.

4. Result and Analysis

In this section, we conduct an experimental evaluation of our method and compare it with other approaches.

4.1. Datasets and Data Processing

We evaluate the proposed method using four benchmark datasets with baseline structural MRI: (1) Alzheimer's Disease Neuroimaging Initiative (ADNI-1) [24], (2) ADNI-2, (3) ADNI-3, and (4) Australian Imaging Biomarkers and Lifestyle Study of Aging database (AIBL) [25]. To ensure independent evaluation, we remove subjects that appear simultaneously in ADNI-1, ADNI-2, and ADNI-3 from the latter two datasets. To prevent classification bias caused by class imbalance, we select approximately balanced numbers of samples from all three categories in both source and target domains. The demographic characteristics of the studied subjects are presented in Table 1. In each of our cross-domain experiments, the source domain samples are labeled, while the target domain samples are unlabeled and are used exclusively for model validation. The number of source domain samples is larger than that of the target domain. A portion of the source domain data is utilized for model pre-training, and the remaining part is employed for training the domain adaptation network. For the unlabeled target domain samples, 80% are used for training the domain adaptation network, and the remaining 20% are reserved for model validation. Importantly, the data used for training are strictly excluded from validation.

For consistency across the four datasets, we merge both stable Mild Cognitive Impairment (sMCI) and progressive Mild Cognitive Impairment (pMCI) categories into a single MCI category. To prevent data leakage, we utilize only the initial examination data for each subject, specifically the first available 1.5T/3T T1-weighted structural MRI scans. Data from different time points of the same subject never appear simultaneously in both training and testing sets.

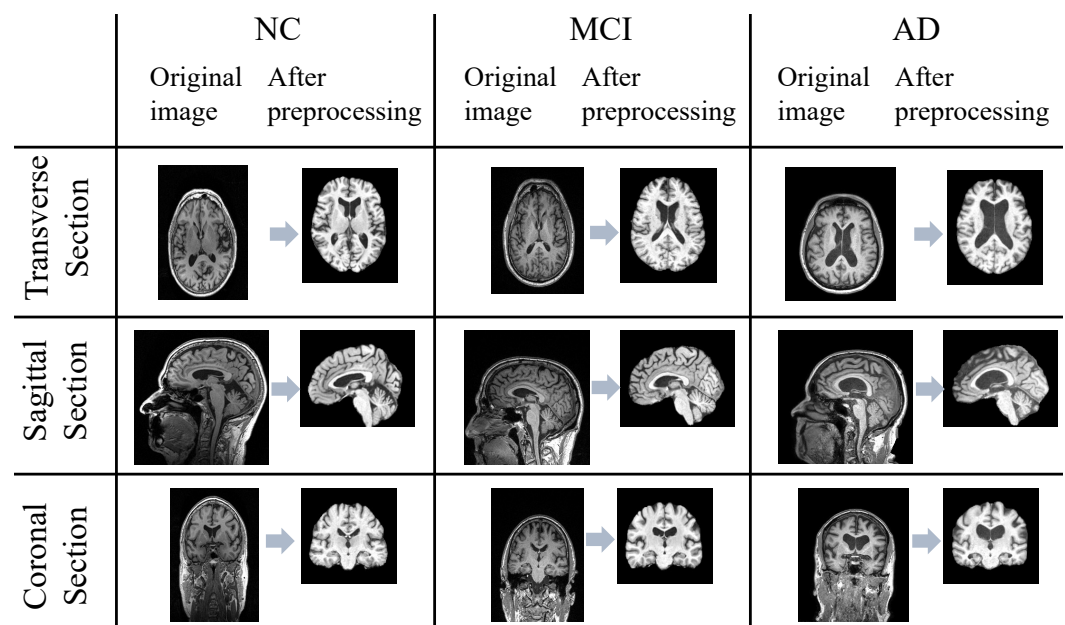
The imaging data employed in this study comprise ADNI-1 with 1.5T scanners and the other three datasets with 3T T1-weighted structural MRI. Original MRI scans from ADNI and AIBL datasets exhibit varying spatial resolutions, with the most common native resolutions of T1-weighted MRI being $1 \times 1 \times 1.2$ mm or $1.2 \times 1.2 \times 1.2$ mm, and partially $1 \times 1 \times 1$ mm, resulting in different voxel counts across scans. Additional variations arise from differences in positioning, brain size and location within images, and intensity variations across multi-center scanner models.

Table 1. Demographic information of the studied subjects included in the datasets.

Datasets	Category	Number of Subjects	Gender (Male/Female)	Age (Mean $\pm$ Std)
ADNI-1	AD	148	81/67	75.7 $\pm$ 7.5
	MCI	151	90/61	74.7 $\pm$ 7.1
	NC	154	82/72	75.9 $\pm$ 6.6
ADNI-2	AD	128	66/62	74.2 $\pm$ 7.9
	MCI	126	67/59	72.6 $\pm$ 7.4
	NC	131	69/62	73.3 $\pm$ 6.5
ADNI-3	AD	57	32/25	74.2 $\pm$ 11.3
	MCI	62	35/27	72.3 $\pm$ 7.6
	NC	65	35/30	71.1 $\pm$ 7.2
AIBL	AD	68	30/38	73.3 $\pm$ 7.7
	MCI	67	34/33	74.7 $\pm$ 7.1
	NC	73	34/39	73.1 $\pm$ 6.1

To mitigate the impact of these external factors on Alzheimer’s disease recognition performance, we implement the following preprocessing pipeline for all original images: (1) resampling all images to a uniform spatial resolution (voxel size of $1 \times 1 \times 1 \text{ mm}^3$) [26]; (2) skull stripping [27]; (3) intensity correction using the N4 algorithm [28]; (4) linear registration of MRI images to the Montreal Neurological Institute (MNI) standard brain template AAL (MNI152, resolution of $1 \times 1 \times 1 \text{ mm}^3$) [29]; (5) center cropping to remove background regions (non-brain areas), reducing computational complexity while ensuring the registered brain remains centered; and (6) intensity normalization to the range $[0, 1]$. The final processed MRI scans all achieve a spatial resolution of $182 \times 218 \times 182$.

Figure 4 demonstrates representative examples of original and preprocessed sections from identical spatial locations for three subjects with varying native resolutions: Normal Control ($256 \times 256 \times 170$), Mild Cognitive Impairment ($240 \times 256 \times 176$), and Alzheimer’s Disease ($256 \times 256 \times 211$). For each subject, we utilize three-dimensional structural MRI scans for both training and testing procedures.

**Figure 4.** Representative original and preprocessed MRI scans from the NC, MCI, and AD subject groups.

4.2. Experimental Setup

In order to verify the effectiveness of the proposed method, we conducted experiments on four classification tasks, including AD vs. NC, MCI vs. NC, AD vs. MCI, and AD vs. MCI vs. NC. We adopted four metrics of accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under receiver operating characteristic (ROC) curve (AUC) for evaluation. For each metric, a higher value indicates better classification performance. The metrics are defined as follows:

$$\begin{aligned} ACC &= \frac{TP + TN}{TP + TN + FP + FN} \\ SEN &= \frac{TP}{TP + FN} \\ SPE &= \frac{TN}{TN + FP} \end{aligned} \quad (19)$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative values respectively.

Among the four datasets, ADNI-3 and AIBL contain relatively fewer samples and therefore serve exclusively as target domains in our domain adaptation experiments. We first pre-train the model using source domain samples before initiating adversarial training, which necessitates larger sample sizes in source domains compared to target domains.

We evaluate four domain adaptation configurations for each classification task: (1) ADNI-1 → ADNI-2: ADNI-1 as source domain and ADNI-2 as target domain (2) ADNI-2 → ADNI-1: ADNI-2 as source domain and ADNI-1 as target domain (3) ADNI-1 + ADNI-2 → ADNI-3: combined ADNI-1 and ADNI-2 as source domain and ADNI-3 as target domain (4) ADNI-1 + ADNI-2 → AIBL: combined ADNI-1 and ADNI-2 as source domain and AIBL as target domain.

To ensure fair comparison with state-of-the-art unsupervised domain adaptation methods, we implement all algorithms using their publicly available source code. All experiments are conducted on 4 × NVIDIA RTX3090Ti GPUs with 24 GB memory. We employ the Adam optimizer with an initial learning rate of 0.001, which is reduced by a factor of 0.1 every 10 epochs. The training process spans 100 epochs with a batch size of 8. We set all loss weights γ , λ_1 , and λ_2 to 1 to maintain balanced optimization. To enhance source domain classification accuracy and accelerate subsequent domain adaptation, we pre-train both the MRI multi-scale feature extraction and aggregation module and the two label classifiers G_{y1} and G_{y2} for 30 epochs prior to formal domain adaptation training.

4.3. Comparison with State-of-Art Domain Adaptation Methods

To validate the effectiveness of the proposed JDC-DA method, we compare our approach with 3DResNet50 [30] and four state-of-the-art domain adaptation methods: AD²A [31], ATM [32], DCAN [33], and RCE [34]. These competing methods are briefly introduced as follows: (1) 3DResNet50 [30]: We employ ResNet50 as the baseline model for supervised learning on the source domain, which is directly applied to prediction tasks on the target domain without performing feature alignment. Given the three-dimensional nature of MRI data, we replace all two-dimensional convolution blocks in ResNet50 with 3D convolution blocks. (2) AD²A [31]: This attention-guided deep domain adaptation framework for multi-site MRI data employs adversarial training using a domain discriminator and feature extractor, requiring no category label information from the target data. (3) ATM [32]: This method leverages both adversarial training and metric learning advantages. Through the Maximum Density Divergence loss function, it simultaneously minimizes inter-domain divergence and maximizes intra-class density, effectively achieving cross-domain feature alignment. (4) DCAN [33]: This approach aligns conditional distributions by minimizing Conditional Maximum Mean Discrepancy while extracting dis-

criminative information from the target domain through mutual information maximization between samples and their predicted labels. (5) RCE [34]: This method employs both clean and noisy classifiers to estimate the noise transition matrix. The clean classifier assigns pseudo-labels for target data, while the noisy classifier trains on noisy target samples and derives optimal parameters through a closed-form solution, enhanced by a pre-trained domain predictor. Table 2 presents a comparative summary of four state-of-the-art domain adaptation methods, highlighting the core ideas, contributions, and limitations associated with each method.

Table 2. Comparison of representative UDA methods for multi-site MRI-based Alzheimer’s disease diagnosis.

Method	Core Idea	Strengths	Limitations	Tasks
AD ² A [31]	Attention-guided adversarial adaptation for multi-site MRI to mitigate inter-site domain discrepancy and improve cross-site diagnosis.	Clinically targeted for harmonization; leverages attention to emphasize domain-invariant representations.	Task-specific to medical imaging; relies on labeled source data and limited attention interpretability.	Binary AD vs. NC; MCI vs. NC; AD vs. MCI
ATM [32]	Introduces Maximum Density Divergence with adversarial tight matching to reduce inter-domain discrepancy and enhance intra-class compactness.	Novel discrepancy metric; strong benchmark performance; improved adversarial training balance.	Higher design/tuning complexity; comparatively higher implementation barrier.	Multi-class AD vs. NC; MCI vs. NC; AD vs. MCI; AD vs. MCI vs. NC
DCAN [33]	Aligns conditional distributions (CMMD) and maximizes target mutual information, rather than only matching marginal distributions.	Mitigates negative transfer; stable non-GAN optimization; naturally extensible to partial UDA.	Sensitive to pseudo-label quality; theoretical novelty less centered on new discrepancy definitions.	Multi-class AD vs. NC; MCI vs. NC; AD vs. MCI; AD vs. MCI vs. NC
RCE [34]	Risk-consistent self-training with clean/noisy classifiers, noise-transition estimation, and uncertainty-guided soft pseudo-labels.	Directly addresses pseudo-label noise; principled risk-consistency formulation.	Multi-module training pipeline is complex; source-sample selection strategy can affect robustness.	Multi-class AD vs. NC; MCI vs. NC; AD vs. MCI; AD vs. MCI vs. NC
JDC-DA (Ours)	Joint metric-adversarial adaptation that aligns both domain-level and category-level discrepancies via dual discriminators.	Integrates metric and adversarial paradigms; improves robustness and generalization in cross-domain AD diagnosis.	Training complexity increases due to coupled modules; broader disease/domain transfer requires further validation.	Multi-class AD vs. NC; MCI vs. NC; AD vs. MCI; AD vs. MCI vs. NC

We evaluate all methods on cross-domain problems where source domain samples contain labels while target domain samples remain unlabeled. For experimental setup, we use a portion of source domain samples for model pre-training and the remainder for domain adaptation network training. For target domain unlabeled samples, we allocate 80% for domain adaptation network training and 20% for model validation. We conduct comparative experiments on four classification tasks: AD versus NC, MCI versus NC, AD versus MCI, and AD versus MCI versus NC.

To ensure fair comparison, we employ the identical multi-scale feature extraction module of JDC-DA across all methods while maintaining consistent training strategies, including learning rate and number of training epochs.

4.3.1. AD vs. NC Classification

Table 3 presents the performance results of different methods in the task of AD vs. NC classification. As shown in Table 3, the proposed JDC-DA model achieves the best performance across all four Domain Adaptation scenarios. Furthermore, our JDC-DA demonstrates superior overall performance in the “ADNI-1 + ADNI-2 → ADNI-3/AIBL” scenario compared to single-source domain scenarios. Figure 5 shows the ROC curves of the compared algorithms for AD vs. NC classification under the four domain adaptation settings.

Table 3. Results of six methods in AD vs. NC classification under four Domain Adaptation settings.

Source Domain → Target Domain	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
ADNI-1 → ADNI-2	3DResNet50 [30]	81.03	79.63	83.47	81.07
	AD ² A [31]	88.65	85.16	90.02	89.26
	ATM [32]	90.43	88.15	89.07	89.66
	DCAN [33]	89.88	89.27	87.29	90.50
	RCE [34]	89.17	88.87	90.18	91.25
	JDC-DA (Ours)	91.03	90.42	91.02	93.43
ADNI-2 → ADNI-1	3DResNet50 [30]	80.15	78.85	82.55	80.37
	AD ² A [31]	87.26	84.94	88.75	88.37
	ATM [32]	89.55	87.94	89.21	89.21
	DCAN [33]	88.24	88.25	86.58	88.52
	RCE [34]	88.56	87.95	90.11	89.76
	JDC-DA (Ours)	89.17	90.05	90.21	91.95
ADNI-1 + ADNI-2 → ADNI-3	3DResNet50 [30]	82.08	80.43	85.39	82.13
	AD ² A [31]	91.41	78.16	92.58	91.69
	ATM [32]	91.55	88.64	89.93	92.02
	DCAN [33]	90.67	89.72	88.66	91.69
	RCE [34]	90.18	90.45	91.35	92.47
	JDC-DA (Ours)	92.16	92.05	92.67	94.03
ADNI-1 + ADNI-2 → AIBL	3DResNet50 [30]	81.55	80.01	84.07	81.58
	AD ² A [31]	90.05	86.74	90.13	91.45
	ATM [32]	91.04	88.35	89.48	91.78
	DCAN [33]	90.15	89.36	88.14	91.31
	RCE [34]	89.69	90.01	90.79	92.02
	JDC-DA (Ours)	91.77	91.16	91.72	93.68

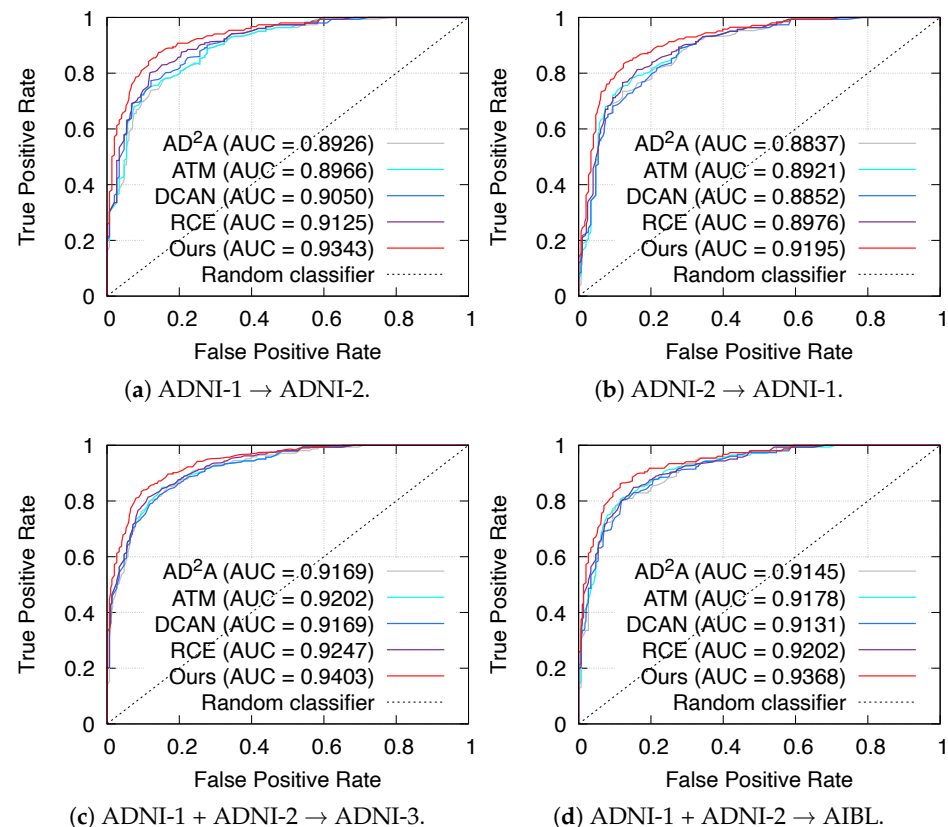


Figure 5. ROC curves of the compared algorithms for AD vs. NC classification under the four domain adaptation settings.

4.3.2. AD vs. MCI Classification

Table 4 presents the performance results of different methods in the task of AD vs. MCI classification. From Table 2, we can see that among the compared methods, the proposed JDC-DA model achieves the best performance of ACC, SEN, SPE and AUC across all four Domain Adaptation scenarios. Compare with AD vs. NC classification, AD vs. MCI classification is more challenging. The difficulty may be attributable to the subtle and often indistinguishable neuroimaging features of MCI on structural MRI, owing to its status as an early, prodromal stage of Alzheimer's disease.

Table 4. Results of six methods in AD vs. MCI classification under four Domain Adaptation settings.

Source Domain → Target Domain	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
ADNI-1 → ADNI-2	3DResNet50 [30]	68.23	71.55	67.38	69.52
	AD²A [31]	76.48	74.46	79.65	77.93
	ATM [32]	79.32	78.56	79.33	80.22
	DCAN [33]	78.61	75.96	79.65	79.58
	RCE [34]	78.95	79.22	76.45	79.65
	JDC-DA (Ours)	80.77	79.17	80.23	82.28
ADNI-2 → ADNI-1	3DResNet50 [30]	68.12	70.25	63.88	69.34
	AD²A [31]	73.47	71.55	78.36	72.41
	ATM [32]	77.63	75.87	78.25	77.98
	DCAN [33]	76.94	78.83	73.52	77.26
	RCE [34]	78.15	79.13	76.33	78.51
	JDC-DA (Ours)	79.32	79.86	79.57	81.75

Table 4. Cont.

Source Domain → Target Domain	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
ADNI-1 + ADNI-2 → ADNI-3	3DResNet50 [30]	74.19	72.73	70.55	73.85
	AD ² A [31]	82.15	75.96	81.88	78.95
	ATM [32]	81.34	81.12	77.54	79.28
	DCAN [33]	82.46	80.45	79.67	80.12
	RCE [34]	82.01	79.96	80.22	81.54
	JDC-DA (Ours)	83.56	81.04	82.04	84.72
ADNI-1 + ADNI-2 → AIBL	3DResNet50 [30]	74.33	73.87	69.58	71.36
	AD ² A [31]	79.22	76.58	80.29	78.67
	ATM [32]	79.53	78.52	77.44	77.92
	DCAN [33]	78.68	79.33	76.21	78.05
	RCE [34]	79.68	80.01	75.48	80.13
	JDC-DA (Ours)	81.14	80.55	81.66	82.58

4.3.3. MCI vs. NC Classification

Table 5 presents the performance results of different methods in the task of MCI vs. NC classification. From Table 5, we can see that the proposed JDC-DA model still achieves the best performance across all four Domain Adaptation scenarios.

Table 5. Results of six methods in MCI vs. NC classification under four Domain Adaptation settings.

Source Domain → Target Domain	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
ADNI-1 → ADNI-2	3DResNet50 [30]	66.52	70.69	69.88	68.33
	AD ² A [31]	77.65	75.82	78.99	78.03
	ATM [32]	79.25	80.14	77.45	79.18
	DCAN [33]	78.91	77.85	78.98	79.77
	RCE [34]	78.55	80.28	77.23	79.89
	JDC-DA (Ours)	80.88	81.57	80.75	81.06
ADNI-2 → ADNI-1	3DResNet50 [30]	67.33	66.88	65.27	68.02
	AD ² A [31]	72.16	71.43	77.41	72.58
	ATM [32]	74.55	77.93	76.95	75.22
	DCAN [33]	76.11	77.59	74.55	76.87
	RCE [34]	78.06	79.25	77.11	78.69
	JDC-DA (Ours)	80.06	80.75	79.98	80.45
ADNI-1 + ADNI-2 → ADNI-3	3DResNet50 [30]	71.35	69.94	70.03	71.89
	AD ² A [31]	78.33	73.97	81.92	78.78
	ATM [32]	79.96	81.64	78.69	80.03
	DCAN [33]	80.82	81.03	79.97	80.96
	RCE [34]	81.13	80.24	80.85	80.94
	JDC-DA (Ours)	81.96	82.68	81.87	82.06
ADNI-1 + ADNI-2 → AIBL	3DResNet50 [30]	70.93	71.67	69.86	71.01
	AD ² A [31]	78.21	75.55	80.93	78.74
	ATM [32]	79.61	77.33	78.54	78.67
	DCAN [33]	78.98	79.04	75.39	78.33
	RCE [34]	80.13	81.22	75.49	80.58
	JDC-DA (Ours)	81.34	82.51	81.37	81.95

4.3.4. AD vs. MCI vs. NC Classification

In the AD versus MCI versus NC three-class classification task, Tables 6 and 7 present comparative ACC metrics for AD and MCI classifications, respectively, across four domain adaptation scenarios. We note that AD²A [31] employs a binary classification framework for AD versus NC and consequently does not appear in these tables. The three-class classification problem presents greater complexity than binary classification, which explains the relatively lower ACC metrics observed across all four domain adaptation scenarios. From Tables 6 and 7, we observe that among the compared methods, the proposed JDC-DA model achieves superior performance across all four domain adaptation configurations.

Table 6. Comparison of ACC (%) metrics for AD classification in AD vs. MCI vs. NC tasks under four domain adaptation scenarios.

Method	ADNI-1 → ADNI-2	ADNI-2 → ADNI-1	ADNI-1 + ADNI-2 → ADNI-3	ADNI-1 + ADNI-2 → AIBL
3DResNet50 [30]	68.15	68.07	71.28	70.85
ATM [32]	75.81	75.56	76.34	76.13
DCAN [33]	75.97	75.78	76.66	76.27
RCE [34]	76.45	75.93	77.04	76.93
JDC-DA (Ours)	78.39	77.11	79.12	78.78

Table 7. Comparison of ACC (%) metrics for MCI classification in AD vs. MCI vs. NC tasks under four domain adaptation scenarios.

Method	ADNI-1 → ADNI-2	ADNI-2 → ADNI-1	ADNI-1 + ADNI-2 → ADNI-3	ADNI-1 + ADNI-2 → AIBL
3DResNet50 [30]	68.32	67.69	70.55	69.98
ATM [32]	75.48	74.79	76.22	75.45
DCAN [33]	75.65	74.92	76.13	75.54
RCE [34]	75.81	75.06	76.82	76.27
JDC-DA (Ours)	77.74	76.83	78.35	78.06

Figure 6 presents the confusion matrix comparison of four algorithms for AD versus MCI versus NC classification under the ADNI-1 → ADNI-2 domain adaptation configuration. We observe that the proposed JDC-DA model maintains superior performance among the compared methods. Particularly, misclassifying AD or MCI cases as NC represents missed diagnoses in clinical practice, and our method achieves the lowest misdiagnosis rate in this critical aspect.

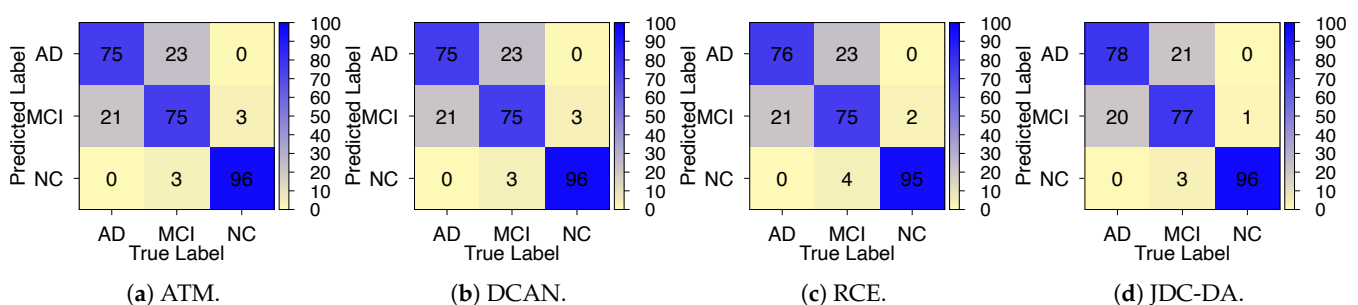


Figure 6. Comparative analysis of confusion matrices for four algorithms in AD vs. MCI vs. NC classification under ADNI-1 → ADNI-2 domain adaptation setting.

Based on the comprehensive experimental results, we demonstrate that the proposed method achieves excellent performance across all four classification tasks: AD versus NC, AD versus MCI, MCI versus NC, and AD versus MCI versus NC. Specifically, our approach attains 92.16% accuracy in the AD versus NC classification task, representing a 10.08% improvement over the supervised learning baseline using 3DResNet50. For the AD versus MCI classification task, our method achieves 83.56% accuracy. The proposed method also demonstrates outstanding performance in the MCI versus NC classification task with 81.96% accuracy. In the challenging three-class classification task of AD versus MCI versus NC, our method obtains the highest classification accuracy for both AD and MCI categories while producing the fewest missed diagnoses.

Compared with existing domain adaptation methods, the proposed algorithm exhibits significant advantages across all key evaluation metrics, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC) for the classification tasks.

5. Discussion

Here, we provide discussions on JDC-DA.

5.1. Ablation Study

We propose a joint adversarial domain adaptation framework for auxiliary diagnosis of Alzheimer's disease. Our approach incorporates multi-scale feature aggregation to enhance the model's perception of diverse lesion scales and structural variations across brain regions. By integrating the advantages of two mainstream unsupervised domain adaptation methodologies—metric learning and adversarial learning—we develop a feature alignment method that combines both domain and category feature disparities. Furthermore, we introduce a classification certainty maximization strategy and propose a dual adversarial learning approach based on joint domain and category alignment, thereby improving the reliability and generalization capability of cross-domain decision-making.

To validate the effectiveness of individual components in JDC-DA, we employ an adversarial domain adaptation framework as our baseline model for comparative analysis. This baseline comprises three core components: (1) a feature extractor that extracts discriminative features from input images, shared between source and target domains; (2) two classifiers that perform classification predictions on the extracted features, maximizing prediction discrepancies on target domain samples to guide the feature extractor in learning domain-invariant discriminative features; and (3) a domain discriminator that distinguishes between source and target domain samples, employing an adversarial learning mechanism to optimize the feature extractor for generating domain-invariant features. In our experiments, this baseline model utilizes six 3D convolutional layers for feature extraction, with both classifiers and the domain discriminator implemented as three-layer fully connected networks.

To verify the performance improvements achieved by our proposed modules—the Multi-Scale Feature Aggregation module (MSFA), the Across Domains and Categories Metric Learning module (ADCML), and the classification certainty maximization module (Lcc) from the joint domain and category dual adversarial learning (note that domain adversarial learning is already included in the baseline)—we design a comprehensive ablation study. Starting from the baseline model, we progressively incorporate different modules for performance comparison: (a) baseline model; (b) baseline with MSFA; (c) baseline with ADCML; (d) baseline with Lcc; and (e) baseline with all three modules (MSFA, ADCML, and Lcc). We conduct comparative experiments on three binary classification tasks (AD versus NC, AD versus MCI, MCI versus NC) and one three-class classification

task (AD versus MCI versus NC) under the ADNI-1 → ADNI-2 domain adaptation setting. The comparative results are presented in Table 8.

Table 8. Results of module effectiveness verification under ADNI-1 → ADNI-2 Domain Adaptation setting.

Classification Task	Model	ACC (%)	SEN (%)	SPE (%)	AUC (%)
AD vs. NC	Baseline	81.21	81.55	82.01	83.56
	Baseline + MSFA	84.78	83.72	83.35	85.28
	Baseline + ADCML	87.03	85.57	85.43	87.82
	Baseline + L_{cc}	89.11	87.93	88.15	89.97
	Baseline + MSFA + ADCML + L_{cc}	91.03	90.42	91.02	93.43
AD vs. MCI	Baseline	70.23	70.07	71.11	72.04
	Baseline + MSFA	73.17	72.85	73.39	74.43
	Baseline + ADCML	75.33	74.89	75.21	77.96
	Baseline + L_{cc}	77.84	76.33	77.55	79.33
	Baseline + MSFA + ADCML + L_{cc}	80.77	79.17	80.23	82.28
MCI vs. NC	Baseline	71.19	72.70	71.27	73.28
	Baseline + MSFA	73.21	74.94	73.18	75.15
	Baseline + ADCML	75.13	76.36	75.56	77.33
	Baseline + L_{cc}	77.86	78.41	77.61	78.84
	Baseline + MSFA + ADCML + L_{cc}	80.88	81.57	80.75	81.06
AD vs. MCI vs. NC	Baseline	69.03	68.83	69.58	70.25
	Baseline + MSFA	71.12	70.88	71.23	72.17
	Baseline + ADCML	72.59	72.34	73.45	74.62
	Baseline + L_{cc}	73.92	73.98	76.02	77.04
	Baseline + MSFA + ADCML + L_{cc}	76.59	77.55	79.26	80.68

Through comparative analysis of experimental results, we observe that the baseline model (a) demonstrates suboptimal performance across all evaluation metrics for the three classification tasks. Upon integrating the multi-scale feature aggregation module (b), we achieve diagnostic accuracy improvements of 3.57%, 2.94%, 2.02%, and 2.09% for the four classification tasks respectively compared to the baseline network, with simultaneous AUC value increases of approximately 2 percentage points. These findings indicate that the multi-scale feature aggregation module effectively enhances the model's capacity for modeling image structures at different scales, improves feature representation, and consequently enhances diagnostic performance.

Further incorporation of the classification certainty loss function (c) yields additional improvements in both SPE and SEN values across the three tasks. This demonstrates that the classification certainty loss function reduces prediction uncertainty by increasing confidence in target domain sample predictions, thereby enhancing classifier robustness and reliability.

Finally, the combined integration of multi-scale feature aggregation, the metric learning module incorporating both domain and category feature disparities (ADCML), and the classification certainty loss function (e) with the baseline model yields optimal cross-domain diagnostic performance. Specifically, we achieve 91.03% ACC in AD versus NC classification, representing a 9.82% improvement over the baseline. Additional improvements include 10.57% for MCI versus NC, 9.69% for AD versus MCI, and 7.56% for AD versus MCI versus NC classification. The other three metrics—SEN, SPE, and AUC—show improvements of approximately 9–10 percentage points.

These results collectively demonstrate that the multi-scale feature aggregation module enhances lesion recognition capability through integration of multi-scale brain imaging

features; the ADCML module improves cross-domain sample consistency and category discriminability; and the classification certainty loss function reduces classification uncertainty through increased prediction confidence. The synergistic interaction among these three components significantly enhances performance in AD, MCI, and NC classification tasks.

5.2. Grad-CAM Visualization

In Alzheimer's disease auxiliary diagnosis research, enhancing the interpretability of diagnostic models is crucial for ensuring their reliability in clinical practice. To evaluate the practical effectiveness of our proposed method, we employ the Grad-CAM visualization technique [35] and randomly select three cases from the target domain: NC (RID: 018_S_0425), MCI (RID: 021_S_0178), and AD (RID: 027_S_1254). The visualization results across three orthogonal planes for one representative layer are presented in Figure 7.

The color bar adjacent to the visualization maps indicates the model's attention intensity for each region, where dark blue represents minimal attention (activation values approaching 0.0) and red indicates regions of highest attention (activation values reaching 0.3, 0.4, or 0.75). The varying numerical ranges in the color bars for AD, MCI, and NC images reflect the model's differential attention patterns across disease stages: NC images (0.0 to 0.3) demonstrate weak model attention; MCI images (0.0 to 0.4) show emerging attention to potential pathological regions; AD images (0.0 to 0.75) reveal heightened sensitivity to characteristic lesion areas.

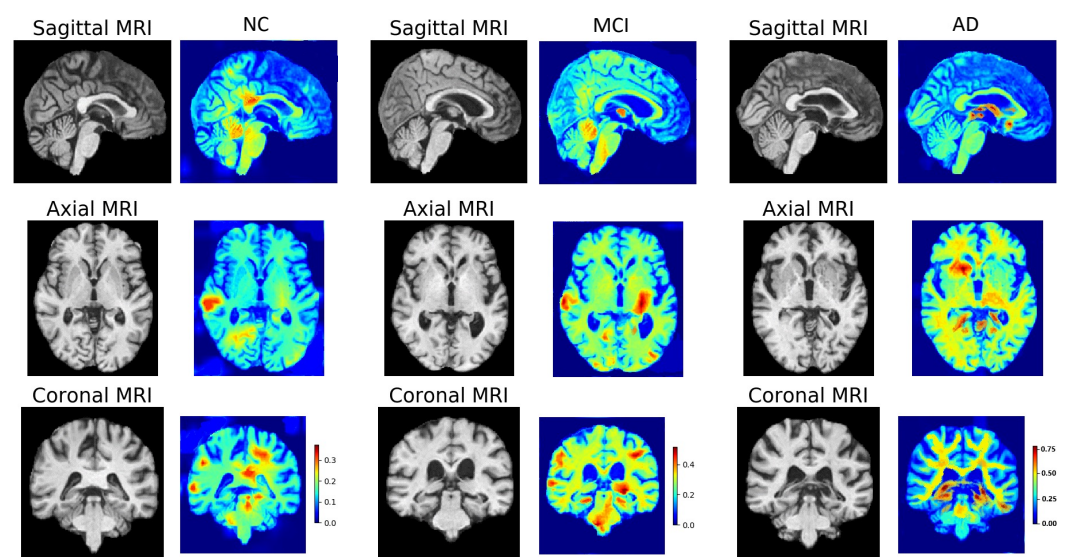


Figure 7. Grad-CAM visualization of target domain image.

Figure 7 demonstrates that brain MRI reveals progressive alterations as cognitive status evolves from NC to MCI and finally to AD. Normal brains maintain structural integrity without significant atrophy, while MCI subjects begin to exhibit localized mild atrophy or density changes. AD patients display marked structural degeneration, particularly in memory-related regions such as the hippocampal area.

Concurrently, our Grad-CAM visualizations show the model's attention regions progressively shifting from homogeneous distributions to specific high-contribution areas. The salient regions in class activation maps expand progressively with disease advancement. These visualization results not only intuitively present pathological characteristics of neurodegenerative changes but also reveal the decision-making rationale of the deep learning model during disease identification, thereby providing interpretable imaging evidence for auxiliary diagnosis.

5.3. Feature Distribution Analysis via T-SNE Visualization

To validate the effectiveness of feature alignment in our proposed method, we conduct visual analysis of features from both source and target domains. We select image data from three categories—NC, MCI, and ADisease—across source and target domains. These data are processed through both the pre-trained 3DResNet50 model and our proposed JDC-DA algorithm for feature extraction and alignment. We then employ t-SNE technique [36] to reduce the dimensionality of the extracted features for visualization. The comparative results are presented in Figure 8.

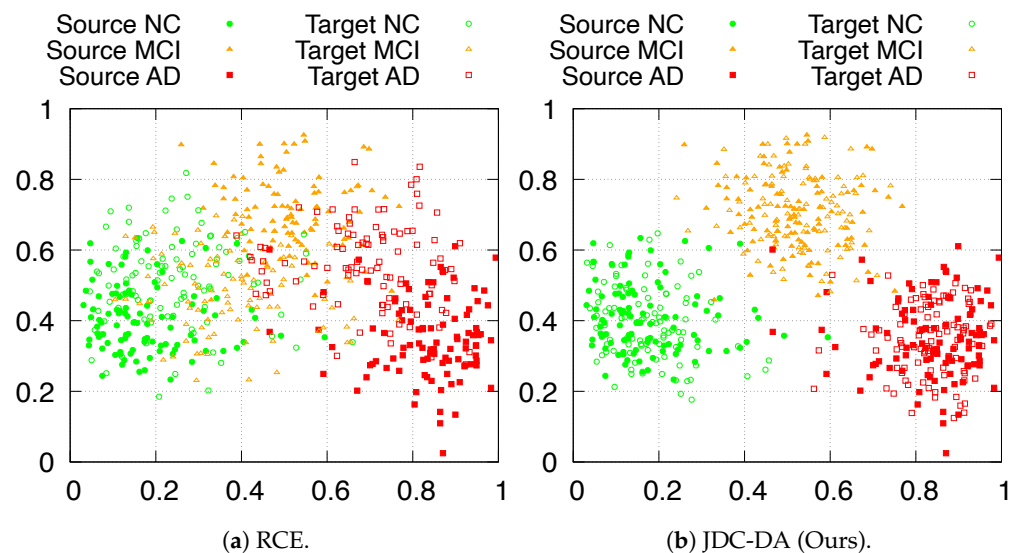


Figure 8. Feature distribution visualization map for source domain and target domain.

Figure 8 demonstrates that for unlabeled target domain data, the traditional 3DResNet model exhibits significant category confusion among the feature distributions of NC, MCI, and AD classes, resulting in suboptimal classification performance. In contrast, our proposed algorithm achieves substantial feature alignment between source and target domains in the feature space. The feature distributions of all three categories (NC, MCI, and AD) show effective category alignment across domains, resulting in well-defined inter-class boundaries in the target domain that facilitate accurate classification. These findings provide compelling evidence for the superiority of our method in feature alignment, effectively reducing both domain-shift and category distribution discrepancies, thereby significantly enhancing the model's generalization capability in cross-domain tasks.

5.4. Analysis of Model Training and Inference Speed

The loss curves of our model for the four classification tasks—AD vs. NC, MCI vs. NC, AD vs. MCI, and AD vs. MCI vs. NC—are shown in Figure 9 (under the ADNI-1 → ADNI-2 cross-domain setting).

As observed from the training loss curves, our model achieves rapid convergence during training.

In our source-domain → target-domain cross-domain experiments, different datasets are used, and the dataset sizes vary across settings. Therefore, the total training time is not a fixed quantity and is not suitable for direct comparison. Instead, we report the per-case training time per epoch and the per-case inference time for volumes of size $182 \times 218 \times 182$.

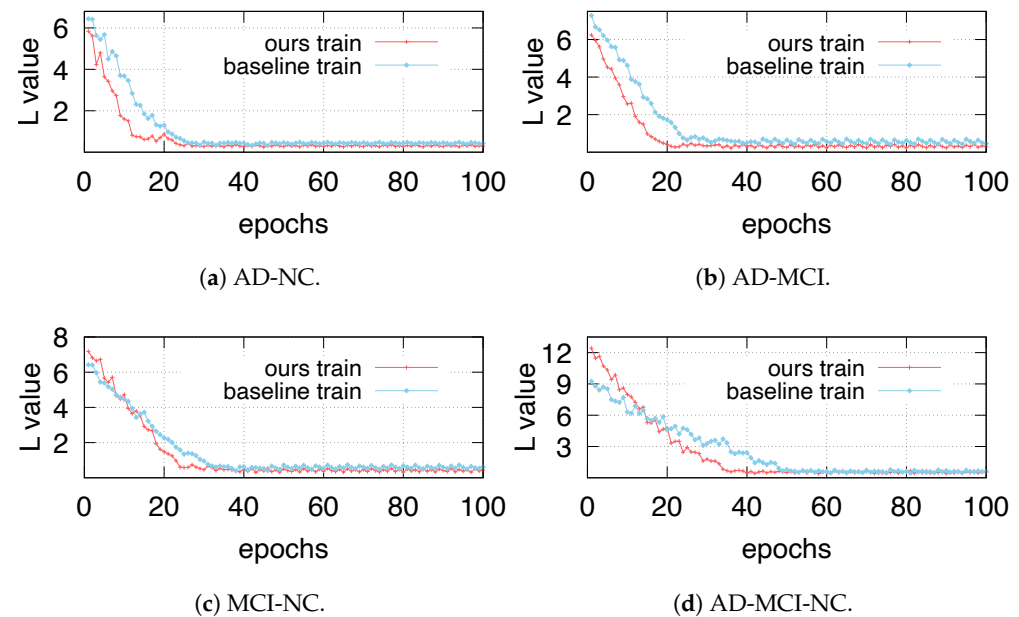


Figure 9. Train loss of 4 diagnosis tasks on ADNI-1→ADNI-2 datasets.

A comparison of computational cost between our method and existing methods is presented in Table 9. Here, *TTime* denotes the per-case training time within one epoch (minutes), and *ITime* denotes the per-case inference time (seconds). All experiments were conducted on an Ubuntu workstation equipped with 4× NVIDIA RTX 3090 Ti GPUs (24 GB memory per GPU).

Table 9. Comparison of computational cost between our method and existing methods.

	3DResNet50 [30]	AD ² A [31]	ATM [32]	DCAN [33]	RCE [34]	Ours
TTime (min)	0.81	1.52	1.63	1.57	1.83	1.71
ITime (s)	1.00	0.97	0.93	0.93	0.97	0.87

Our method has a relatively complex training pipeline, including a multi-scale feature aggregation module, a metric learning module, and an adversarial learning module; consequently, its training time is comparatively longer. However, during inference, features extracted by the multi-scale feature aggregation module are directly fed into the target-domain classifier, resulting in a streamlined inference pipeline. Among all compared methods, our model achieves the fastest inference speed, which is advantageous for rapid Alzheimer’s disease diagnosis in practical applications.

6. Conclusions

To address the performance degradation in Alzheimer’s disease diagnosis caused by domain shifts in multi-source MRI data, we integrate the advantages of two mainstream unsupervised domain adaptation approaches, metric learning and adversarial learning, and propose a Joint Adversarial Domain Adaptation framework for Alzheimer’s Disease auxiliary diagnosis (JDC-DA). The proposed framework comprises three key components: (a) a Multi-Scale Feature Aggregation module that integrates hierarchical feature information to enhance the model’s perception of diverse lesion scales and structural variations across brain regions, thereby improving adaptability to multi-scale targets; (b) an Across Domains and Categories Metric Learning module that generates dynamic prototype features based on category clustering, computes fused domain and category feature prototypes

from source and target domains, and achieves cross-domain feature alignment through integrated domain and category feature disparities, enabling the model to better distinguish between categories while aligning data distributions across domains; and (c) a Joint Domain and Category Dual Adversarial Learning module that incorporates a classification certainty maximization strategy to establish a joint adversarial mechanism with domain and classification discrepancy discriminators, achieving synergistic optimization of domain adaptation through coordinated domain adversarial and category adversarial learning.

We evaluate our method on 1230 structural MRI scans from four benchmark datasets (ADNI-1, ADNI-2, ADNI-3, and AIBL) for unsupervised target domain classification tasks, including AD versus NC, MCI versus NC, AD versus MCI, and AD versus MCI versus NC. Experimental results demonstrate that the proposed JDC-DA method achieves superior overall performance compared to several state-of-the-art domain adaptation approaches. Our findings confirm the effectiveness of JDC-DA for unsupervised target domain classification of AD, MCI, and CN, suggesting its potential value in clinical applications for Alzheimer's disease diagnosis.

Author Contributions: Conceptualization, Y.S. and Y.Z.; methodology, Y.S., Y.Z., Y.W. and G.Y.; software, Y.S., Y.Z. and G.Y.; validation, Y.S., Y.Z., Y.W. and G.Y.; formal analysis, Y.S. and Y.Z.; investigation, Y.S., Y.W. and Y.Z.; resources, Y.S., G.Y. and Y.Z.; data curation, Y.S. and Y.Z.; writing—original draft preparation, Y.S.; writing—review and editing, Y.S., Y.Z. and Y.W.; visualization, Y.S. and Y.Z.; supervision, Y.W. and G.Y.; project administration, Y.W. and G.Y.; funding acquisition, Y.W. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (Grant Nos. 61871106, 62441231), Key R&D Projects of Liaoning Province, China (Grant No. 2024JH2/102500015), and the Fundamental Research Funds for the Central Universities of Ministry of Education (Grant No. N25BSS034).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Sorour, S.E.; Abd El-Mageed, A.A.; Albarrak, K.M.; Alnaim, A.K.; Wafa, A.A.; El-Shafeiy, E. Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101940. [\[CrossRef\]](#)
2. Zong, Y.; Zuo, Q.; Ng, M.K.P.; Lei, B.; Wang, S. A new brain network construction paradigm for brain disorder via diffusion-based graph contrastive learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10389–10403. [\[CrossRef\]](#)
3. Zuo, Q.; Wu, H.; Chen, C.P.; Lei, B.; Wang, S. Prior-guided adversarial learning with hypergraph for predicting abnormal connections in Alzheimer's disease. *IEEE Trans. Cybern.* **2024**, *54*, 3652–3665. [\[CrossRef\]](#)
4. El-Assy, A.; Amer, H.M.; Ibrahim, H.; Mohamed, M. A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Sci. Rep.* **2024**, *14*, 3463. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Menagadevi, M.; Devaraj, S.; Madian, N.; Thiagarajan, D. Machine and deep learning approaches for alzheimer disease detection using magnetic resonance images: An updated review. *Measurement* **2024**, *226*, 114100. [\[CrossRef\]](#)
6. Folego, G.; Weiler, M.; Casseb, R.F.; Pires, R.; Rocha, A. Alzheimer's disease detection through whole-brain 3D-CNN MRI. *Front. Bioeng. Biotechnol.* **2020**, *8*, 534592. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Fei, X.; Wang, J.; Ying, S.; Hu, Z.; Shi, J. Projective parameter transfer based sparse multiple empirical kernel learning machine for diagnosis of brain disease. *Neurocomputing* **2020**, *413*, 271–283. [\[CrossRef\]](#)
8. Wu, X.; Gao, S.; Sun, J.; Zhang, Y.; Wang, S. Classification of Alzheimer's disease based on weakly supervised learning and attention mechanism. *Brain Sci.* **2022**, *12*, 1601. [\[CrossRef\]](#)

9. Wang, C.; Wei, Y.; Li, J.; Li, X.; Liu, Y.; Hu, Q.; Wang, Y.; Alzheimer's Disease Neuroimaging Initiative. Asymmetry-enhanced attention network for Alzheimer's diagnosis with structural magnetic resonance imaging. *Comput. Biol. Med.* **2022**, *151*, 106282. [\[CrossRef\]](#)
10. Li, J.; Wei, Y.; Wang, C.; Hu, Q.; Liu, Y.; Xu, L. 3-D CNN-based multichannel contrastive learning for Alzheimer's disease automatic diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5008411. [\[CrossRef\]](#)
11. Liang, X.; Wang, Z.; Chen, Z.; Song, X. Alzheimer's disease classification using distilled multi-residual network. *Appl. Intell.* **2023**, *53*, 11934–11950. [\[CrossRef\]](#)
12. Hu, Z.; Li, Y.; Wang, Z.; Zhang, S.; Hou, W.; Alzheimer's Disease Neuroimaging Initiative. Conv-Swinformer: Integration of CNN and shift window attention for Alzheimer's disease classification. *Comput. Biol. Med.* **2023**, *164*, 107304. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Gan, Y.; Lan, Q.; Huang, C.; Su, W.; Huang, Z. Dense convolution-based attention network for Alzheimer's disease classification. *Sci. Rep.* **2025**, *15*, 5693. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Yu, L.; Liu, J.; Wu, Q.; Wang, J.; Qu, A. A siamese-transport domain adaptation framework for 3D MRI classification of gliomas and Alzheimer's diseases. *IEEE J. Biomed. Health Inform.* **2023**, *28*, 391–402. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Fiasam, L.D.; Rao, Y.; Sey, C.; Aggrey, S.E.; Kodjiku, S.L.; Agyekum, K.O.B.O.; Razafindratovolohy, A.; Adjei-Mensah, I.; Ukwuoma, C.C.; Sam, F. DAW-FA: Domain-aware adaptive weighting with fine-grain attention for unsupervised MRI harmonization. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 102157. [\[CrossRef\]](#)
16. Wilson, G.; Cook, D.J. A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 51. [\[CrossRef\]](#)
17. Cai, H.; Zhang, Q.; Long, Y. Prototype-guided multi-scale domain adaptation for Alzheimer's disease detection. *Comput. Biol. Med.* **2023**, *154*, 106570. [\[CrossRef\]](#)
18. Zhang, Q.; Cai, H.; Long, Y.; Yu, S. Multicentric prototype and pseudo-labeling based source-free domain adaptation for Alzheimer's disease classification. *Biomed. Signal Process. Control* **2025**, *103*, 107483. [\[CrossRef\]](#)
19. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; NeurIPS: San Diego, CA, USA, 2014.
20. Jung, E.; Luna, M.; Park, S.H. Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression. *Pattern Recognit.* **2023**, *133*, 109061. [\[CrossRef\]](#)
21. Zhang, J.; Liu, M.; Pan, Y.; Shen, D. Unsupervised conditional consensus adversarial network for brain disease identification with structural MRI. In *Machine Learning in Medical Imaging*; International Workshop on Machine Learning in Medical Imaging; Springer: Cham, Switzerland, 2019; pp. 391–399.
22. Han, R.; Liu, Z.; Chen, C.P. Multi-scale 3D convolution feature-based broad learning system for Alzheimer's disease diagnosis via MRI images. *Appl. Soft Comput.* **2022**, *120*, 108660. [\[CrossRef\]](#)
23. Nguyen, D.; Nguyen, H.; Ong, H.; Le, H.; Ha, H.; Duc, N.T.; Ngo, H.T. Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. *IBRO Neurosci. Rep.* **2022**, *13*, 255–263. [\[CrossRef\]](#)
24. Jack, C.R., Jr.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; L. Whitwell, J.; Ward, C.; et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2008**, *27*, 685–691. [\[CrossRef\]](#)
25. Ellis, K.A.; Bush, A.I.; Darby, D.; De Fazio, D.; Foster, J.; Hudson, P.; Lautenschlager, N.T.; Lenzo, N.; Martins, R.N.; Maruff, P.; et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int. Psychogeriatr.* **2009**, *21*, 672–687. [\[CrossRef\]](#)
26. Sled, J.G.; Zijdenbos, A.P.; Evans, A.C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **2002**, *17*, 87–97. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Wang, Y.; Nie, J.; Yap, P.T.; Shi, F.; Guo, L.; Shen, D. Robust deformable-surface-based skull-stripping for large-scale studies. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*; International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer: Berlin/Heidelberg, Germany, 2011; pp. 635–642.
28. Shamoun, D.P.; Bron, E.E.; Lelieveldt, B.P.; Smits, M.; Klein, S.; Staring, M.; Initiative, A.D.N. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front. Neuroinform.* **2014**, *7*, 50. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Mazziotta, J.; Toga, A.; Evans, A.; Fox, P.; Lancaster, J.; Zilles, K.; Woods, R.; Paus, T.; Simpson, G.; Pike, B.; et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2001**, *356*, 1293–1322. [\[CrossRef\]](#)
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
31. Guan, H.; Liu, Y.; Yang, E.; Yap, P.T.; Shen, D.; Liu, M. Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification. *Med. Image Anal.* **2021**, *71*, 102076. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Li, J.; Chen, E.; Ding, Z.; Zhu, L.; Lu, K.; Shen, H.T. Maximum density divergence for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3918–3930. [\[CrossRef\]](#)

33. Ge, P.; Ren, C.X.; Xu, X.L.; Yan, H. Unsupervised domain adaptation via deep conditional adaptation network. *Pattern Recognit.* **2023**, *134*, 109088. [[CrossRef](#)]
34. Ding, F.; Li, J.; Tian, W.; Zhang, S.; Yuan, W. Unsupervised domain adaptation via risk-consistent estimators. *IEEE Trans. Multimed.* **2023**, *26*, 1179–1187. [[CrossRef](#)]
35. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
36. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

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