The Journal of Supercomputing
https://doi.org/10.1007/s11227-022-04668-0

MRI-based model for MCI conversion using deep zero-shot transfer learning

Fujia Ren[1,2] · **Chenhui Yang**[1] · **Y. A. Nanehkaran**[3]

Accepted: 18 June 2022

Abstract

This study describes a deep zero-shot transfer learning model (DZTLM) for predicting mild cognitive impairment (MCI) in patients with Alzheimer's disease (AD). The proposed DZTLM combines ResNet and deep subdomain adaptation network (DsAN) blocks with a simple data augmentation and transfer technique, Elastic-Mixup. We test the DZTLM using 3D gray matter images segregated from structural MRI as input. Ablation experiments are conducted to evaluate the proposed model and compare it with existing approaches. Experiments demonstrate that the DsAN network coordinating Elastic-Mixup enhances the accuracy of MCI-AD prediction by more than 18% compared with a standard 3D ResNet50 classifier. The Elastic-Mixup technique contributes more than 16% to this increase in prediction accuracy. Elastic-Mixup also enhances the sensitivity of recognition for stable MCI. When labeled samples are scarce, the unsupervised DZTLM outperforms a semi-supervised transfer learning model. The DZTLM achieves comparable outcomes to existing models despite the absence of tagged MRI data.

Keywords MCI conversion prediction · Deep zero-shot transfer learning · Augmentation · Domain adaptation

✉ Chenhui Yang
chyang@xmu.edu.cn

Fujia Ren
fujia_ren@qq.com

Y. A. Nanehkaran
artavil20@gmail.com

[1] School of Informatics, Xiamen University, 422 Siming Road South, Xiamen 361005, Fujian, China

[2] School of Big Data and Computer Science, Guizhou Normal University, 270 Waihuandonglu Guiyang, Guiyang 50001, Guizhou, China

[3] College of Information Engineering, Yancheng Teachers University, 50 Kaifang Avenue, Yancheng 224002, Jiangsu, China

Published online: 22 July 2022

1 Introduction

Computer-aided diagnosis techniques (CAD) for predicting moderate cognitive impairment (MCI) advancing to Alzheimer's disease (AD) have emerged as an important topic in intelligent healthcare [1]. AD is a degenerative brain ailment with complex pathophysiology that is currently preventive but incurable [2]. People with late-stage AD exhibit clinical symptoms such as progressive memory loss and subsequent dementia, as well as neuropathological deposition of senile plaques and neurofibrillary tangles [3]. As a result, CAD for late-stage AD using behavioral scales and related biomarkers is simple and accurate, reaching above 99 percent diagnosis accuracy [4, 5], whereas CAD for MCI is challenging. More importantly, by the time MCI reaches the late stage, most neurons are irreversibly dead. Thus, the most efficient method for delaying AD is to enhance early diagnosis performance. Currently, diagnostic models for predicting MCI progression are largely based on developing machine learning [6–8] and deep neural networks. Deep diagnostic models, in particular, have received considerable interest because of their self-training ability and strong performance.

Numerous medical image modalities are currently being employed in deep diagnostic models of MCI [9]. Among them, magnetic resonance imaging (MRI) and positron emission tomography (PET) are widely used for brain imaging, providing vital information on the brain's structure and functions [10]. Because MRI scans are noninvasive, safe, and affordable, they are often chosen over PET scans. More crucially, several studies [11, 12] have established that MRI cues are critical in categorizing MCI as progressive MCI (pMCI) or stable MCI (sMCI). However, most MRI-based models focus on supervised learning using labeled training data. There are always challenges to achieving good results for MCI prediction tasks. MRI enables the assessment and monitoring of changes in the size of certain brain areas. In the early stage of AD, prior to obvious clinical signs, MRI scans aid in detecting the structural atrophy caused by cellular damage, axonal degeneration, or synaptic damage. However, the MRI features that distinguish MCI subtypes have become obscured. As an illustration, Fig. 1 compares the region of interest (ROI) distributions in AD and normal control (NC) cases vs. sMCI and pMCI using t-SNE. The distinction between sMCI and pMCI is more uncertain than the distinction between AD and NC. Additionally, because compiling a trustworthy dataset of well-labeled MCI images is more challenging, the MCI data scale is smaller than that of AD and NC data [13]. Thus, advanced models and procedures should focus on resolving the dataset issues. There are always challenges to improving MCI diagnosis performance.

To address these issues and the lack of research on zero-shot learning, this study develops a deep zero-shot transfer learning model (DZTLM) for MCI conversion prediction. We show that this model works well, even without credible MCI samples. The unique contributions of this study are listed below. (1) The proposed DZTLM is the first to combine a deep subdomain adaptation network (DsAN) [14] and 3D ResNet to diagnose MCI conversion. The network essentially self-learns the relationship between the original domain and the target

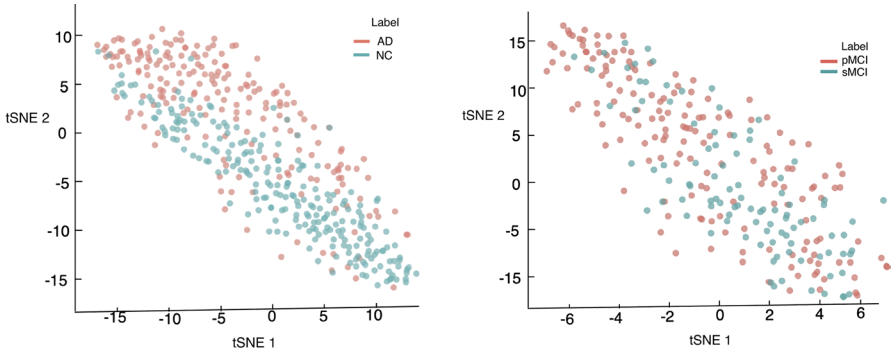

Fig. 1 t-SNE visualizations of the ROIs in AD and NC datasets vs. MCI datasets. ROIs—Regions of interest; AD–Alzheimer's disease; NC—normal control; sMCI—stable mild cognitive impairment; pMCI—progressive mild cognitive impairment

subdomains via a simple and effective mechanism for aligning the multiple network levels. This produces more consistent migration between subdomains on the feature layer. (2) A simple and improved version of Mixup, Elastic-Mixup, is designed to augment and transfer samples between subdomains simultaneously. (3) Rather than relying entirely on either the sample migration strategy or feature migration model, we combine them. (4) Comparing the effects of the model structure and data strategy on MCI diagnosis performance clearly shows that the data strategy makes a greater contribution. The effectiveness of our strategy without MCI labeling is comparable to that of previous approaches in identifying sMCI (p>5%) as a precursor to MCI detection.

2 Related work

Using medical imaging data, breakthroughs in machine learning [6, 8, 15], deep learning models [16–18], and computational frameworks [19–21] have resulted in significant advancements in MCI conversion prediction research. Random subset feature selection, minimum redundancy maximum correlation, and sparse line regression feature selection based on stationary selection have been used [22] to select discriminative features in an iterative combination of MRI and network measurements in a support vector machine (SVM) classifier. The authors of [23] used a novel dynamic morphological feature and improved parameters to complete feature selection, and then classified the disease using SVM. Reference [24] describes the use of MRI to determine the thickness and volume of 66 unique brain areas and the application of particle swarm optimization to choose and construct new feature vectors, eventually producing a fusion classifier combining Bayesian, SVM, and K-nearest neighbor algorithms. Reference [25] introduces the hippocampal texture and value as input characteristics for an SVM classifier, while [26] aggregates the structural characteristics of the most discriminative ROIs to create a thorough assessment of brain deterioration using an ensemble learning methodology. In [26],

a novel Bayesian latent space model is established to exploit the spatial information contained in the cortical surface thickness.

Based on the original MRI, 2D deep models can be constructed using prior knowledge of AD [23, 27], and relatively efficient deep 3D convolutional neural network (CNN) diagnostic models can also be developed [14, 28]. Along with the aforementioned classic machine learning models, deep models for predicting MCI changes have become widespread over the last decade. These models employ iteratively trained deep neural networks to extract meaningful information from high-dimensional medical images. For example, deep models have been established using CNNs to extract image features and integrated with SVM classifiers [29, 30]. The multi-view separable pyramid network [31] learns representations from PET scans' axial, coronal, and sagittal perspectives to provide complementary information that is subsequently merged to make joint judgments. Reference [32] reports a Wasserstein–generative adversarial network (GAN) model that takes a single coronal slice from a patient's baseline T1 scan and generates synthetically aged brain pictures. Reference [33] describes how highly informative slices and gray matter can be extracted using entropy and EICA, leading to the creation of a CNN model based on inception blocks. In [34], the authors develop a modified version of the 3D ResNet framework with an age-corrected architecture.

Transfer learning techniques are effective in overcoming the overfitting problem associated with relatively complex models when the number of available direct training samples is limited, and they are gradually being applied to solve medical image-related tasks [35]. However, these novel strategies are mainly concerned with sample or feature changes. Therefore, the authors of [36, 37] placed a premium on sample space adaptation; for example, [37] used an unsupervised optimization method to weight the sample space instances according to their age and gender. Various studies [38–41] have used adaptive feature learning to predict early AD progression with NC auxiliary data. For example, [38] used the GM tissue volume from 93 ROIs in the auxiliary target domains (i.e., AD and NC), normalized by the total intracranial volume (estimated as the sum of the GM, WM, and CSF volumes in all ROIs), to migrate a feature for a given subject to the MCI target domain. References [42–44] discuss how several layers of the deep network can be fine-tuned to transfer the feature space between domains. The research outlined in [45] uses fMRI to develop a deep adaptive framework that is learned from the source to eliminate noise in the target labels. In addition, [44, 46] report on frozen features collected from ImageNet to create a fine-tuned convolutional network architecture. The VGG architecture was found to outperform other state-of-the-art architectures. Mehmood et al. [44] trained a 2D VGG network on ImageNet using augmentation techniques. ResNet29, created by Jinhyeong et al. [47], works by shrinking and shortening neurons trained on AD and NC data. Anees et al. [48] developed a modified version of ResNet trained on AD and NC, whereas [49] used convolutional autoencoders (CAEs) for classification problems involving AD vs. NC, with transfer learning applied to solve the MCI-AD prediction problem. A 2018 paper [50] presented a domain transfer learning strategy for structural MRI and diffusion tensor imaging modalities. The method begins with a model that has been pretrained on structural MRI data, followed by training on average diffusivity data. Unfortunately, few researchers [43, 51, 52] have discussed

zero-shot transfer learning approaches for MCI conversion prediction, which is a form of unsupervised learning without any labeled target samples. In a recent paper [43], a sequential CNN was constructed for MCI classification to extract the total weights from NC-AD data.

Table 1 lists various existing models, datasets, training techniques, and classifiers. Generally, the MCI-AD prediction problem requires advanced and novel techniques. Supervised learning models are highly dependent on the quality and quantity of the training data, and this is particularly true when dealing with MCI. Semi-supervised learning models based on transfer learning are gaining traction as a means of circumventing the data size constraint. However, most contemporary deep transfer learning techniques require fine-tuning to transfer features, but this is still constrained by the characteristics of the labeled MCI data. To the best of our knowledge, there have been

Table 1 Comparison with studies for MCI conversion prediction

References	Learning approach	Modality	Input	Classifer
Guo et al. [53]	SL	MRI	cortical	SVM
Sørensen et al. [24]	SL	MRI	hippocampal	SVM
Dai et al. [26]	SL	MRI	cortical	Bayes
Varatharajah et al. [54]	SL	FDG-PET Genetic CSF MRI	ROI	SVM, MKL
Popuri et al. [25]	SL	sMRI	ROI	multiple classifiers
Zhang et al. [22]	SL	MRI	cortical	SVM
Shen et al. [29]	SL	sMRI	GM	SVM+CNN
Basheera et al. [33]	SL	sMRI	GM slices	CNN with inception blocks
Pan et al.[31]	SL	PET	Slices	Pyramid Network
Wegmayr et al. [32]	SL	MRI	One coronal slice	Wasserstein-GAN
Pan et al. [55]	SL	sMRI	slices	CNN
Aderghal et al. [50]	SSL	DTI sMRI	a Hippocampal slice	2DCNN
Naz et al. [46]	SSL	MRI	Whole slices	VGG
prakash et al. [41]	SSL	MRI	cortical	3D Inception
Atif et al. [44]	SSL	MRI	the gray matter	2D VGG
Jinhyeong et al. [47]	SSL	MRI	Grey matter	Assembled 2D CNN
Anees et al. [48]	SSL	MRI	Whole brain	3D ResNet29
Ohet al. [49]	SSL	MRI	Grey matter MRI3D modified ResNet	
Loris et al. [56]	SSL	MRI	Whole brain MR	3D ICAE(Tl)I
Cheng et al. [38]	SSL	Pet MRI CSF	SVM	
Ethan et al. [43]	USL	sMRI	3D MRIs	3D CNN
Our work	USL	sMRI	3D GM	3D Resnet-DsAN block

[a] Note: SL, Supervised learning; SSL, Semisupervised learning; USL, Unsupervised learning

no previous studies on MCI transformation prediction using subdomain relations. Therefore, the proposed DZTLM has enormous potential for development when subdomain relationships are considered. In addition to the studies mentioned above, the rapid growth of computing power, especially cloud computing and the Internet of Things (IoT)[57–59], is expanding and influencing the healthcare industry. Several prospective studies[60–63] predict Alzheimer's disease using more and earlier data and multimodal data. These studies investigate the possibility of artificial intelligence in improving Alzheimer's disease diagnosis.

3 Research model and framework

This section describes the improved Elastic-Mixup, which is capable of performing both data augmentation and migration, as well as our proposed DZTLM.

3.1 Elastic-mixup

Data augmentation is frequently used to generalize the performance of deep models. Based on Mixup [64], the Elastic-Mixup module performs sample transformation and sample expansion. As indicated in Eq. (1), the Mixup technique generates new samples by linear numerical insertion.

$$\begin{aligned} \bar{x} &= \lambda x_i + (1-\lambda)x_j \\ \bar{y} &= \lambda y_i + (1-\lambda)y_j \end{aligned} \tag{1}$$

where (x_i, y_i) and (x_j, y_j) are two randomly selected samples from the source dataset; $\lambda \in [0, 1]$ is the ratio of generated data to selected data; and y is a single heat. This technique enlarges the sample space by weighting the sample data and their associated class labels simultaneously. Notably, Mixup no longer generates data that are labeled with a single class. The loss functions for both classes are recalculated throughout the training phase, and the scale factors of the data mixture are used as weights, which are then weighted as the final loss.

Instead of ignoring the prior knowledge and expanding the dataset through a uniform interpretation of the training data, we modify the Mixup method to give Elastic-Mixup. As shown in Eq. (2), the generated data $\bar{x}$ are the same as in Eq. (1), but a hyperparameter p is added to adjust the displacement of the label $\bar{y}$.

$$\begin{aligned} \bar{y} = &\frac{1 - sign[\lambda y_i + (1-\lambda)y_j - p]}{2} \\ &+ \frac{1 + sign[\lambda y_i + (1-\lambda)y_j - p]}{2}[\lambda y_i + (1-\lambda)y_j] \end{aligned} \tag{2}$$

The reason for this modification is that AD and NC data have different distributions from pMCI and sMCI data. The ROI distributions for these four categories of data are shown in Fig. 2. The AD and pMCI data distributions are more similar than the NC and sMCI data distributions. This is because $p(\text{AD} \mid p\text{MCI}) > p(\text{NC} \mid s\text{MCI})$

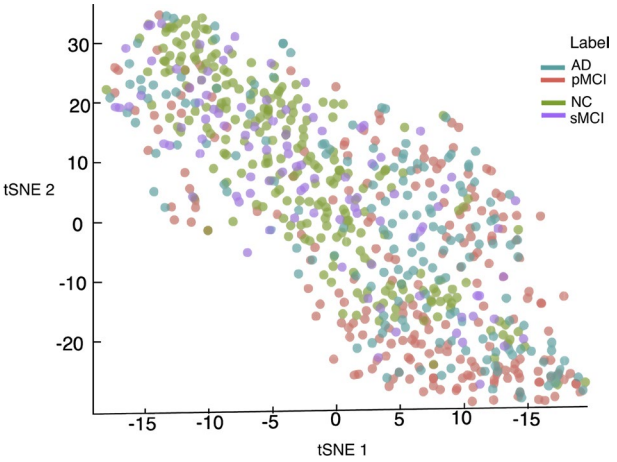

Fig. 2 t-SNE visualizations of the ROIs of pMCI, AD, NC, and sMCI data. ROIs—Regions of interest; AD–Alzheimer's disease; NC—normal control; sMCI—stable mild cognitive impairment; pMCI— progressive mild cognitive impairment

indicates that $p_T(x \mid y) \neq p_S(x \mid y)$. When the parameter p in Eq. (2) is set to 0.8, a segmental function of this value generates a data marker. Thus, if the value of this parameter is greater than 0.8, it is replaced with 1; otherwise, it remains unchanged. The data distributions generated by empirical risk minimization (ERM) [65], Mixup, and Elastic-Mixup are shown in Fig. 3. It can be seen that the proposed Elastic-Mixup method generates the shifted data.

3.2 Model and loss function

To improve the accuracy of MCI conversion classification, the DsAN block is used. This is an unsupervised adaptive network that extracts the target and source domain data in the feature space to reduce the feature differences. Figure 4 illustrates a 50-layer 3D ResNet [16] with attention blocks and an adaptation block. The

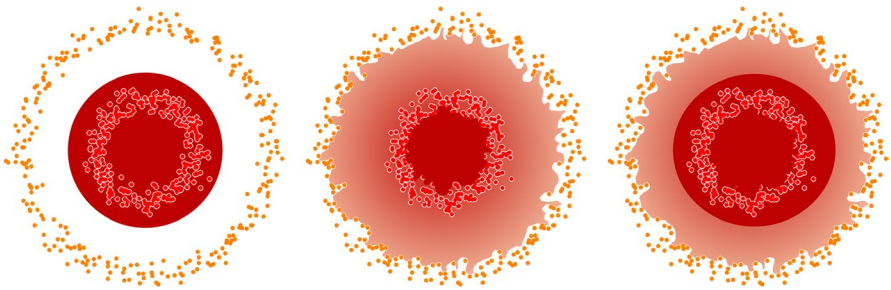

Fig. 3 Distributions of data generated by ERM, Mixup, and Elastic-Mixup on AD and NC datasets. Orange represents the distribution of the AD data; red represents the distribution of the NC data; shaded section represents the possibility of AD or NC

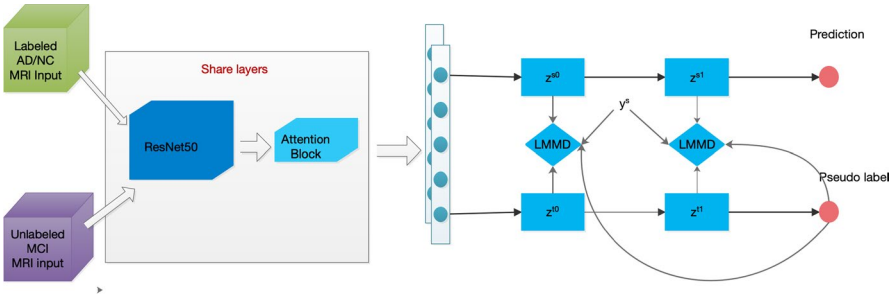

Fig. 4 Architecture of the proposed 3D deep model

convolutional block attention module [66] is used as a lightweight general attention structure to decompose the learning into channel attention learning and spatial attention learning, which ensures low cost and better performance. Adaptation layers are built on DAN [67] and DsAN [14], which apply the local maximum mean difference (LMMD) through a feedforward network. Equation (3) expresses the specific calculation.

DsAN makes a significant contribution by implementing relevant subdomain adaptation rather than focusing exclusively on global adaptation. The following loss function is used in DsAN:

$$\min_{f} \frac{1}{n_s} \sum_{i=1}^{n_s} J(f(x_i^s), y_i^s) + \lambda E_c[\hat{d}(p^{(c)}, q^{(c)}] \tag{3}$$

where $D(s)$ and $D(c)$ denote the source and target domains, respectively; $p^{(c)}$ and $q^{(c)}$ denote the distributions of the source and target domains, respectively. The final loss function combines the classification loss function J and the domain adaptation loss function D, and the parameter λ represents the ratio of domain adaptation to classification loss. DsAN converts the maximum mean discrepancy (MMD) domain adaptation loss to the LMMD domain adaptation loss. The LMMD [14] is calculated as

$$d_H(p, q) = \frac{1}{c} \sum_{c=1}^{C} \left\| \sum_{x_i^s \in D_s} w_i^{sc} \phi(x_i^s) - \sum_{x_j^t \in D_t} w_i^{tc} \phi(x_j^t) \right\| \tag{4}$$

where x_i^s and x_j^t denote instances of the source and target domains, respectively; c denotes the number of categories; wi^{sc} and w_j^{tc} denote the c-th category of the source and target domains, respectively; $\phi()$ is used to map the original sample to the regenerative kernel Hilbert space; and $d_H(p, q) = 0$ if $p = q$.

This study uses AD data, NC data, and their true labels. Moreover, the unlabeled MCI data are pre-classified into two subdomains and then pseudo-labeled. To determine the difference between the source and target domains, the LMMD is calculated. Samples from the same class will be more closely related after the feedforward network has been trained for several iterations through backpropagation. The

multi-kernel technique can be used to decompose the distance calculation into inner products. Additionally, the MMD calculation has a computational complexity of $O(n^2)$, which is prohibitively expensive in deep learning; thus, this study uses MK-MMD unbiased estimation, which reduces the computational complexity to $O(n)$.

4 Experimental design and performance evaluation

This section provides an overview of the data collection and experimental design of this study. We then conduct an ablation test to validate the proposed solutions. Finally, we compare our method against more sophisticated techniques.

4.1 Data preparation

Experiments were conducted using the Alzheimer's Disease Neuroimaging Initiative (ADNI) MRI dataset, obtained from AD, NC, and MCI groups. MCI is classified into two types: sMCI (stable mild cognitive impairment without conversion over 36 months) and pMCI (progressive mild cognitive impairment without conversion over 36 months). The ADNI dataset was collected and distributed by a nonprofit organization of the National Institute on Aging (NIA). It contains data related to brain disorders, including serial MRI and PET images, clinical and neuropsychological assessments, and other biomarkers. The MRI images were preprocessed, including skull stripping and alignment, as well as "spatial normalization, masking, and N3 correction." Additionally, because GM atrophy is a critical biomarker of early AD [68], this study used the CAT12 [69] toolbox to automatically segment the skull to obtain GM images with a size of $121 \times 145 \times 121$ voxels. Table 2 summarizes the dataset's characteristics.

4.2 Experimental settings

Our model was written in Python and executed on a computer equipped with a Nvidia GTX20080Ti GPU. The batch size of the experiment was set to eight during the training phase, and the ADAM optimization algorithm was used. After 100 epochs of training, the model reached a stable convergence point after 45 epochs. To evaluate the model's performance, we used the accuracy (Acc), specificity (Spe), and the area under the receiver operating characteristic (ROC) curve (AUC). The

Table 2 The characteristics of the NIA dataset

Subject	AD	NC	pMCI	sMCI
Number of subjects	200	229	164	100
Gender (Male/Female)	103/97	118/111	97/67	66/34
Age range	55-91	59-90	55-89	57-90

NC—normal cohort; AD—people with Alzheimer's disease; pMCI—progressive mild cognitive impairment; sMCI—stable mild cognitive impairment

ROC curve was constructed using the false positive rate as the abscissa and the true positive rate as the ordinate values. The variables TP, TN, FP, and FN in Eq. (5) denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. The sensitivity (Sen) and Spe correspond to the prediction accuracy of sMCI and pMCI, respectively, because sMCI is represented by positive values and pMCI is represented by negative values in our experiments.

$$\begin{aligned} \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{Spe} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \end{aligned} \tag{5}$$

4.3 Results

We conducted ablation experiments to determine the effect of the data strategy, backbone selection, and DsAN modules in the model on MCI diagnostic performance. In addition, our approach was compared against existing state-of-the-art methods. The model was fed a preprocessed 3D GM image, and the experimental results were averaged over ten rounds of fivefold cross-validation.

4.3.1 Ablation experiment results

We first examine the example strategy. The basic 3D-ResNet classification model was trained using AD/NC data to predict the MCI transformation. This experiment compared four techniques: direct usage of AD and NC, regular rotation and scaling, Mixup, and Elastic-Mixup. Table 3 summarizes the results of the four training procedures. As can be observed, despite the fact that we set very modest values for the distortion amplitude at the conclusion of the process, rotation and scaling of the 3D images fail owing to the limited data, which introduces additional noise. The Elastic-Mixup algorithm somewhat enhances the model's prediction accuracy and generalizability. Additionally, it gives a Sen score that is 7% higher than that of Mixup, demonstrating a considerable increase in the rate of pMCI detection. The hyperparameter p in Eq. (2) affects the final outcomes, with experimental evidence indicating that setting $p = 0.8$ yields the best results.

Table 3 Comparison results of augmentation methods

Augmentation method	ACC%	Sen%	Spe%	Acu
No data augmentation	82.45	54.72	92.27	0.74
Scaling and rotation	62	50.28	65.88	0.65
Mixup	83.25	61.39	92.51	0.78
Elastic-mixup	85.77	71.78	91.43	0.82

Acc, Accuracy; Sen, Sensitivity; Spe, Specificity

We also examined the effects of using Elastic-Mixup in the data and feature spaces. Figure 5 illustrates the results when Elastic-Mixup was used in several places, including the input layer, hidden layer, and fully connected layer. The results show that the optimal performance was achieved when Elastic-Mixup was used in the original input layer. As a result, Elastic-Mixup was exclusively used in the input layer in subsequent experiments.

We compared three 3D model structures: ResNet, ResNet-DAN, and ResNet-DsAN. All models in this experiment used the Elastic-Mixup technique. According to Table 4, DsAN does not improve the diagnostic performance significantly. The effects of changing the p value of Elastic-mixup on the 3D ResNet-DsAN and ResNet models are depicted in Fig. 6. The red line represents the 3D ResNet results, while the blue line represents 3D ResNet-DsAN results. The red line is

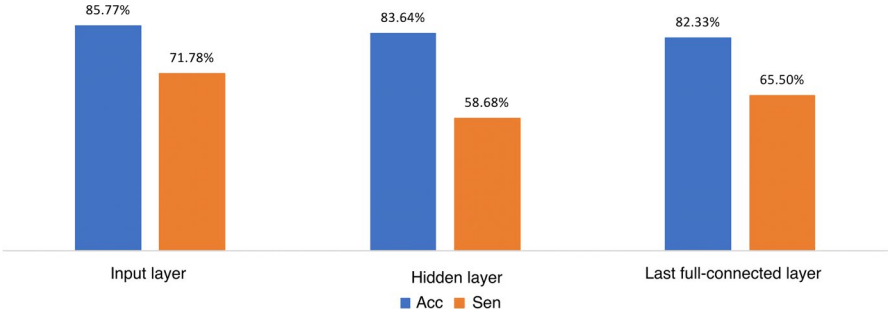

Fig. 5 Sen and Spe scores when using Elastic-mixup in different positions of the deep model

Table 4 Comparison results of the three advanced 3D networks

Network	Acc%	Sen%	Spe%	ACU
Resnet-50	82.45	54.72	92.27	0.74
Resnet-DAN	81.24	58.11	90.13	0.72
Resnet-DsAN	85.16	64.13	93.13	0.78

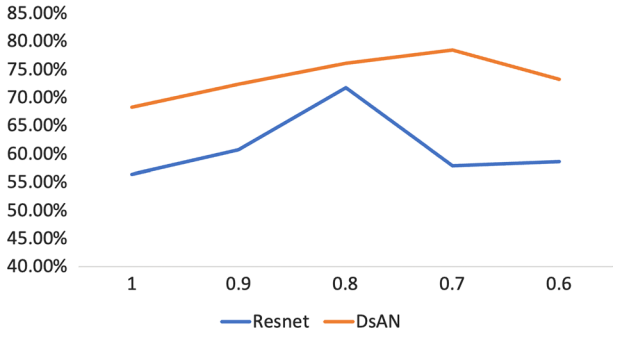

Fig. 6 Trend of accuracy with respect to the Elastic-mixup parameter p for 3D ResNet (red) and ResNet-DsAN (blue). Note that this is only the result of one fold of the data

steeper than the blue line, implying that combining the two methods results in more consistent network diagnostic performance.

Next, ablation experiments were conducted to investigate the effects of the training data, augmentation strategy, and network structure on the prediction results. From the results presented in Table 5, the following observations can be made: (1) MCI prediction trained on the original AD and NC data obtains the lowest prediction accuracy. (2) The application of the fine-tuning method to the labeled MCI data improves the accuracy slightly, which is consistent with previous results [43]. The most obvious difference is that Sen outperforms Spe. This may be related to the fact that the distributions of NC and sMCI data are different from those of AD and pMCI data. Additionally, superior performance is achieved by applying fine-tuning to the last layer rather than the hidden layer, so only the fine-tuning results for the last layer are presented here. (3) Training can be performed directly using labeled MCI data, but the results depend on the quality and quantity of the labels. A lack of training data often results in large standard deviations and may cause network overfitting. (4) The proposed Elastic-Mixup method leads to significantly better prediction results, particularly in terms of Sen. This may be because the method changes the distribution of the difference. Consequently, the effectiveness of the proposed method and the improved network when Elastic-Mixup and DsAN are combined has been validated.

4.3.2 Comparison with several state-of-the-art methods

Finally, the proposed approach was compared with several methods that also use MRI, as listed below.

(1) A modified version of the 3D ResNet framework [34]. (2) A 3D CNN with age correction [28]. (3) A 3D-CNN-SVM classifier trained on MCI [30]. (4) A 2D VGG model pertained on ImageNet and using standard augmentation techniques

Table 5 Ablation experiment results

No	Learning appoach	Training data	Model	Augmentation	Acc%	Sen%	Spe%	Acu
①	SL	Labled sMC pMCI	3D-Resnet	No	68.68	54.99	82.12	0.66
②	SL	llabled AD,NC labled sMC pMCI	3D-Resnet	No	70.22%	54.28	83.2	0.68
③	SSL(Finetune)	labled AD,NC labled sMC pMCI	3D-Resnet	No	82.45	54.72	92.27	0.74
④	USL(ZL)	labeled AD NC	3D-Resnet	Elastic-MIxup	85.77	71.78	91.43	0.82
⑤	USL(ZTL)	(labeled AD, NC	3D-Resnet-DsAN	Elastic-Mixup	**87.16**	**78.11**	**92.40**	**0.85**

[a]SL, Supervised learning; SSL, Semisupervised learning; USL, Unsupervised learning

[b]ZL, Zero-shot learning; ZTL, Zero-shot transfer learning

[44]. (5) A CAE with parameters learned from AD and NC data using unsupervised learning techniques, then fine-tuned on pMCI and sMCI data [49]. (6) ResNet29, which was established by shrinking and shortening the model built in advance on AD and NC data [47]. (7) A modified version of 3D ResNet that uses neighborhood relations trained on AD and NC data [48]. (8) A collection of 2D CNNs pretrained on generic images such as AlexNet, GoogleNet, ResNet, and Inception-v3 [56]. (9) A zero-shot learning model based on a sequential CNN that only uses the total weights from the NC-AD classification [43].

In Table 6, the worst performance is given by method (3). This is because the whole brain images contain too much noise and the direct training dataset is relatively small. Similar results are obtained for methods (4) and (7), because both use GM as features and apply migration techniques. The accuracy of the proposed DZTLM reaches 87%, which is an improvement of nearly 5% over the next-best method. The DZTLM also achieves the best Sen and Spe scores. An important reason for this is the use of the sample transfer technique instead of fine-tuning.

4.4 Discussion

Our study began by comparing the predictive ability of several classical deep backbones for MCI conversion using MRI data. The baseline dataset included individuals with AD, NC, pMCI, and sMCI. The validation trials demonstrated that several deep neural architectures perform similarly. Although DenseNet is better than ResNet in terms of performance, we chose ResNet as the backbone due to its straightforward

Table 6 Comparison with the state-of-art methods for MCI classification

Model	Training data	Input	Acc%	Sen%	Spe%
(1) Improved 3D ResNet	Labeled MCI	Hippocampal MRI	75	77.8	81.3
(2) 3D CNN	MCI	Whole brain MRI	79.50	68.80	86.10
(3) 3D CNN-SVM	MCI	Whole brain MRI	62.00	64.40	68
(4) 2D VGG	Imagnet	The gray matter (GM) tissue	83.75	82.0	85.13
(5) Assembled 2D CNN	General image MCI	Grey matter MRI	70.1	65.35	72.22
(6) 3D ResNet29 t	Labeled ADvsNC Labeled MCI	Whole brain MRI	82.4	72	83.56
(7) 3D modified ResNet	ADvsNC LabeledMCI	Gray matter MRI	83	82.0	85.13
(8) 3D ICAE(Tl)	Unlabled ADvsNC Labeled MCI	Whole brain MRI	73.95	70.71	77.46
9) Sequential 3D CNN	AD vsNC	The entire MRI volumes	74.2	68.2	72.3
Proposed method	Labeled ADvsNC Unlabeled MCI	Grey matter MRI	87.16	78.11	92.40

(1) (2) (3) are supervised learning. (4) (5) (6) (7) are semi-supervised learning methods that use the fine-tuning technique, with (4) (5) using general images as the source domain and (6) (7) using AD and CN. (8) is unsupervised with a fine-tuning technique, using AD and NC as source domains. (9) is unsupervised learning

structure and low memory consumption. Additionally, an ablation experiment demonstrated the critical role of the data approach in increasing the prediction accuracy.

This proposed method uses the Elastic-Mixup data transfer strategy for unsupervised zero-shot scenery learning. As most existing research is based on semi-supervised or supervised learning, traditional data augmentation techniques such as rotation, scaling, or the original Mixup are prevalent. However, the rotation angle and scale ratio of 3D images significantly affect their performance. Additionally, data distortion may impair the detection of clues in diseased brain regions. Although Mixup enhances the performance of supervised and unsupervised learning, our experiments demonstrated that Elastic-Mixup predicts MCI more accurately when only using AD and NC data.

Numerous existing deep models are pretrained on ImageNet or AD and NC data, with the final layers tuned using a small amount of labeled MCI data. By utilizing Elastic-Mixup, the proposed DZTLM outperformed most previous semi-supervised approaches. This is consistent with previous results [49]. Possible explanations include disguised AD and NC characteristics containing more clues and more stability than MCI characteristics. The fine-tuning technique transfers features by sharing part layers and adjusting the final layer of each part. Inspired by the concept of feature transfer, we placed Elastic-Mixup in several different feature layers. The results indicated that Elastic-Mixup performed better on data than on feature layers. As a result, we hypothesize that the performance of feature learning with sparse data is unstable. This is the first time, to our knowledge, that DsAN has been used to analyze MRI data for the MCI task. The DsAN module determines the difference between the source and the target using the LMMD. As with the fine-tuning technique, relying solely on DsAN results in less marginal improvement than using data transfer. However, research indicates that DsAN minimizes the accuracy bias across many parameter values. As stated previously, the hyperparameter reflects a hypothesis about the relationship between the source and target data. Due to its low generalization error, DsAN learns effectively on the final few layers. Our findings demonstrate that, even in the absence of labeled MCI data, the proposed model consistently performs well when representing target and source data.

This work has inherent limitations that could be investigated further in the future. First, we performed the MCI transformation prediction using only the MRI data from the ADNI collection. The size of this dataset is restricted, and the single-image patterns and the test set are insufficiently diverse. Despite using normalization and reducing the model's complexity, there is still a risk of overfitting. Naturally, fusing additional modalities with larger datasets will improve the MCI diagnostic performance. Second, the data transfer suggested here makes use of the Elastic-Mixup hyperparameter. This hyperparameter is determined by the data transfer threshold between AD and NC. The grid search method could be used to determine the best value. However, there is no guarantee of obtaining the optimal parameter combination, and the iteration process requires time. Hence, future research will attempt to automate the parameter learning. Finally, the visualization model communicates the model's findings and assists in identifying locations with disease lesions. Because our data contain information about the entire brain without significant structural deformation, visualizing the model

may allow critical clues to be identified from MRI. Additionally, the visualization of data and models serves as a guide for our future work.

5 Conclusion

In this study, we used unsupervised learning to create a unique 3D-ResNet-DsAN-based DZTLM, while improving traditional data augmentation and incorporating sample migration. The study's findings are therapeutically noteworthy because, among other things, MCI training data are severely inadequate due to patient privacy problems, racial differences, illness features, and annotation reliability. Moreover, our method is novel in that it integrates data and model transfer techniques. Additionally, the suggested model and approach investigate the same subdomain distribution.

The experimental results presented in this paper show that the DsAN module and the Elastic-Mixup data scaling method improve the MCI conversion prediction performance. Elastic-Mixup makes a greater contribution to this improvement than DsAN, significantly boosting the ability to detect sMCI. The combination of the two leads to improved diagnostic accuracy and robustness. Interestingly, comparison testing showed that utilizing a 3D-ResNet-based supervised learning model enhances accuracy by about 18% and that the lack of labeled data does not produce inferior performance to existing approaches. However, there is a risk of overfitting due to data shortages. In future work, we hope to increase the generality of our model and apply it to various illnesses.

Acknowledgements Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12- f2–0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. HoffmannLa Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson&Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. The ADNI data are disseminated by the Laboratory for Neuro-Imaging at the University of Southern California. This work was supported by the National Key Research Development Program of China (Grant No. 2020YFB1313703), the National Natural Science Foundation of China (Grant No.6002304), and the Natural Science Foundation of Fujian Province of China (Grant No. 2020J05002).

Data availability This article is supported by data that can be found in ADNI, which can be accessed from http://adni.loni.usc.edu/.

Declarations

Conflict of interest The authors stated that they have no conflicts of interest regarding this work. We certify that we have no commercial or other affiliations with the submitted work that would create a conflict of interest.

References

1. Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rossor M, Thal L, Winblad B (2001) Current concepts in mild cognitive impairment. Arch Neurol 58(12):1985–1992
2. de la Torre JC (2010) Alzheimer's disease is incurable but preventable. J Alzheimer's dis 20(3):861–870
3. Tramutola A, Triplett JC, Di Domenico F, Niedowicz DM, Murphy MP, Coccia R, Perluigi M, Butterfield DA (2015) Alteration of mtor signaling occurs early in the progression of Alzheimer disease (ad): analysis of brain from subjects with pre-clinical ad, amnestic mild cognitive impairment and late-stage ad. J neurochem 133(5):739–749
4. Pellegrini E, Ballerini L, Hernandez MdCV, Chappell FM, González-Castro V, Anblagan D, Danso S, Muñoz-Maniega S, Job D, Pernet C et al (2018) Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimer's Dement Diagn Assess Dis Monit 10:519–535
5. Jo T, Nho K, Saykin AJ (2019) Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci 11:220
6. Ye Q, Huang P, Zhang Z, Zheng Y, Fu L, Yang W (2021) Multiview learning with robust double-sided twin svm. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3088519
7. Fu L, Li Z, Ye Q, Yin H, Liu Q, Chen X, Fan X, Yang W, Yang G (2022) Learning robust discriminant subspace based on joint l2, p-and l2, s-norm distance metrics. IEEE Trans Neural Netw and Learn Syst 33(1):130–144
8. Ye Q, Li Z, Fu L, Zhang Z, Yang W, Yang G (2019) Nonpeaked discriminant analysis for data representation. IEEE Trans Neural Netw Learn Syst 30(12):3818–3832
9. Feng Q, Ding Z (2020) MRI radiomics classification and prediction in Alzheimer's disease and mild cognitive impairment: a review. Curr Alzheimer Res 17(3):297–309
10. Marcus C, Mena E, Subramaniam RM (2014) Brain pet in the diagnosis of Alzheimer's disease. Clin Nuclear Med 39(10):413
11. Petersen RC (2010) Alzheimer's disease: progress in prediction. The Lancet Neurol 9(1):4–5
12. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R et al (2014) Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurol 13(6):614–629
13. Risacher SL, Saykin AJ, Wes JD, Shen L, Firpi HA, McDonald BC (2009) Baseline MRI predictors of conversion from MCI to probable ad in the ADNI cohort. Curr Alzheimer Res 6(4):347–361
14. Zhu Y, Zhuang F, Wang J, Ke G, Chen J, Bian J, Xiong H, He Q (2020) Deep subdomain adaptation network for image classification. IEEE Trans Neural Netw Learn Syst 32(4):1713–1722
15. Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2(3):1–21
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778. IEEE
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708. IEEE
18. Dong S, Wang P, Abbas K (2021) A survey on deep learning and its applications. Comput Sci Rev 40:110379
19. Chen X, Yang L, Chen Z, Min G, Zheng X, Rong C (2022) Resource allocation with workload-time windows for cloud-based software services: a deep reinforcement learning approach. IEEE Trans Cloud Comput. https://doi.org/10.1109/TCC.2022.3169157
20. Huang G, Chen X, Zhang Y, Zhang X (2012) Towards architecture-based management of platforms in the cloud. Front Comput Sci 6(4):388–397

21. Chen X, Zhang J, Lin B, Chen Z, Wolter K, Min G (2021) Energy-efficient offloading for DNN-based smart ioT systems in cloud-edge environments. IEEE Trans Parallel Distrib Syst 33(3):683–697
22. Zhang T, Liao Q, Zhang D, Zhang C, Yan J, Ngetich R, Zhang J, Jin Z, Li L (2021) Predicting MCI to ad conversation using integrated SMRI and RS-FMRI: machine learning and graph theory approach. Front Aging Neurosci 13:429
23. Kruthika K, Maheshappa H, Initiative ADN et al (2019) Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. Inform Med Unlocked 14:34–42
24. SøRensen L, Igel C, Liv Hansen N, Osler M, Lauritzen M, Rostrup E, Nielsen M (2016) Initiative ADN, the Australian imaging biomarkers, of ageing, LFS: early detection of Alzheimer's disease using MRI hippocampal texture. Hum Brain Mapp 37(3):1148–1161
25. Popuri K, Ma D, Wang L, Beg MF (2020) Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBl, OASIS, and MIRIAD databases. Hum Brain Mapp 41(14):4127–4147
26. Dai N, Kang H, Jones GL, Fiecas MB, Initiative ADN (2021) A bayesian latent spatial model for mapping the cortical signature of progression to Alzheimer's disease. Can J Stat 49(1):46–62
27. Farooq A, Anwar S, Awais M, Rehman S (2017) A deep CNN based multi-class classification of alzheimer's disease using MRI. In: 2017 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–6. IEEE
28. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, Guo G, Xiao M, Du M, Qu X et al (2018) Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front Neurosci 12:777
29. Shen T, Jiang J, Li Y, Wu P, Zuo C, Yan Z (2018) Decision supporting model for one-year conversion probability from mci to ad using CNN and SVM. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 738–741. IEEE
30. Feng W, Halm-Lutterodt NV, Tang H, Mecum A, Mesregah MK, Ma Y, Li H, Zhang F, Wu Z, Yao E et al (2020) Automated MRI-based deep learning model for detection of Alzheimer's disease process. Int J Neural Syst 30(06):2050032
31. Pan X, Phan T-L, Adel M, Fossati C, Gaidon T, Wojak J, Guedj E (2020) Multi-view separable pyramid network for ad prediction at MCI stage by 18 f-FDG brain pet imaging. IEEE Trans Med Imaging 40(1):81–92
32. Wegmayr V, Hörold M, Buhmann JM (2019) Generative aging of brain MRI for early prediction of MCI-AD conversion. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). 1042–1046. IEE
33. Basheera S, Ram MSS (2021) Deep learning based Alzheimer's disease early diagnosis using T2w segmented gray matter MRI. Intl J Imaging Syst Technol 31(3):1692–1710
34. Li H, Habes M, Wolk DA, Fan Y, Initiative ADN et al (2019) A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimer's Dement 15(8):1059–1070
35. Valverde JM, Imani V, Abdollahzadeh A, De Feo R, Prakash M, Ciszek R, Tohka J (2021) Transfer learning in magnetic resonance brain imaging: a systematic review. J Imaging 7(4):66
36. Cheng B, Liu M, Suk H-I, Shen D, Zhang D (2015) Multimodal manifold-regularized transfer learning for mci conversion prediction. Brain Imaging Behav 9(4):913–926
37. Wachinger C, Reuter M, Initiative ADN et al (2016) Domain adaptation for Alzheimer's disease diagnostics. Neuroimage 139:470–479
38. Cheng B, Liu M, Zhang D, Munsell BC, Shen D (2015) Domain transfer learning for MCI conversion prediction. IEEE Trans Biomed Eng 62(7):1805–1817
39. Cheng B, Zhang D, Shen D (2012) Domain transfer learning for mci conversion prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention pp. 82–90. Springer
40. Cheng B, Liu M, Shen D, Li Z, Zhang D (2017) Multi-domain transfer learning for early diagnosis of Alzheimer's disease. Neuroinformatics 15(2):115–132
41. Prakash D, Madusanka N, Bhattacharjee S, Kim C-H, Park H-G, Choi H-K (2021) Diagnosing alzheimer's disease based on multiclass mri scans using transfer learning techniques. Current medical imaging 17(12):1460–1472
42. Aderghal K, Afdel K, Benois-Pineau J, Catheline G, Initiative ADN et al (2020) Improving alzheimer's stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities. Heliyon 6(12):05652

43. Ocasio E, Duong TQ (2021) Deep learning prediction of mild cognitive impairment conversion to alzheimer's disease at 3 years after diagnosis using longitudinal and whole-brain 3d mri. PeerJ Computer Science 7:560
44. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, Maqsood M, Yaqub M (2021) A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. Neuroscience 460:43–52
45. Zhou S, Cox CR, Lu H (2019) Improving whole-brain neural decoding of FMRI with domain adaptation. In: International workshop on machine learning in medical imaging, pp. 265–273. Springer
46. Naz S, Ashraf A, Zaib A (2022) Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. Multimed Syst 28(1):85–94
47. Bae J, Stocks J, Heywood A, Jung Y, Jenkins L, Hill V, Katsaggelos A, Popuri K, Rosen H, Beg MF et al (2021) Transfer learning for predicting conversion from mild cognitive impairment to dementia of Alzheimer's type based on a three-dimensional convolutional neural network. Neurobiol Aging 99:53–64
48. Abrol A, Bhattarai M, Fedorov A, Du Y, Plis S, Calhoun V, Initiative ADN et al (2020) Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. J Neurosci Methods 339:108701
49. Oh K, Chung Y-C, Kim KW, Kim W-S, Oh I-S (2019) Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci Rep 9(1):1–16
50. Aderghal K, Khvostikov A, Krylov A, Benois-Pineau J, Afdel K, Catheline G (2018) Classification of Alzheimer disease on imaging modalities with deep CNNS using cross-modal transfer learning. In: 2018 IEEE 31st international symposium on computer-based medical systems (CBMS), pp. 345–350. IEEE
51. Gamberger D, Lavrač N, Srivatsa S, Tanzi RE, Doraiswamy PM (2017) Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. Sci Rep 7(1):1–12
52. Nezhadmoghadam F, Martinez-Torteya A, Treviño V, Martínez E, Santos A, Tamez-Peña J, Initiative ADN et al (2021) Robust discovery of mild cognitive impairment subtypes and their risk of Alzheimer's disease conversion using unsupervised machine learning and gaussian mixture modeling. Curr Alzheimer Res 18(7):595–606
53. Guo M, Li Y, Zheng W, Huang K, Zhou L, Hu X, Yao Z, Hu B (2020) A novel conversion prediction method of MCI to ad based on longitudinal dynamic morphological features using ADNI structural MRIS. J Neurol 267(10):2983–2997
54. Varatharajah Y, Ramanan VK, Iyer R, Vemuri P (2019) Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. Sci Rep 9(1):1–15
55. Pan D, Zeng A, Jia L, Huang Y, Frizzell T, Song X (2020) Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning. Front Neurosci 14:259
56. Nanni L, Interlenghi M, Brahnam S, Salvatore C, Papa S, Nemni R, Castiglioni I, Initiative ADN, et al. (2020) Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer's disease. Front Neurol 345
57. Chen X, Li M, Zhong H, Ma Y, Hsu C-H (2021) Dnnoff: offloading DNN-based intelligent ioT applications in mobile edge computing. IEEE transactions on industrial informatics.18(4):2820-9
58. Chen X, Hu J, Chen Z, Lin B, Xiong N, Min G (2021) A reinforcement learning-empowered feedback control system for industrial internet of things. IEEE Trans Ind Inform 18(4):2724–2733
59. Huang G, Luo C, Wu K, Ma Y, Zhang Y, Liu X (2019) Software-defined infrastructure for decentralized data lifecycle governance: principled design and open challenges. In: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 1674–1683. IEEE
60. Kaur PD, Sharma P (2020) Ic-smart: iotcloud enabled seamless monitoring for Alzheimer diagnosis and rehabilitation system. J Ambient Intell Humaniz Comput 11(8):3387–3403
61. Varatharajan R, Manogaran G, Priyan MK, Sundarasekar R (2018) Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. Clust Comput 21(1):681–690
62. Yadav P, Kumar P, Kishan P, Raj P, et al. (2021) Development of pervasive iot based healthcare monitoring system for alzheimer patients. In: Journal of physics: Conference Series, vol. 2007, p. 012035. IOP Publishing
63. Sharma S, Dudeja RK, Aujla GS, Bali RS, Kumar N (2020) Detras: deep learning-based healthcare framework for ioT-based assistance of Alzheimer patients. Neural Comput Appl 17:1–3

64. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2017) Mixup: beyond empirical risk minimization. arXiv preprint arXiv:1710.09412
65. Vapnik V (1992) Principles of risk minimization for learning theory. In: Advances in neural information processing systems, pp. 831–838. Springer
66. Woo S, Park J, Lee J-Y, So Kweon I (2018) CBAM: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19. Springer
67. Long M, Cao Y, Wang J, Jordan M (2015) Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning, pp. 97–105. PMLR
68. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to ad conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 32(12):2322.e19
69. Gaser C, Kurth F (2017) Manual computational anatomy toolbox-cat12. University of Jena, Structural brain mapping Group at the Departments of Psychiatry and Neurology