Letter

Tensor Distribution Regression Based on the 3D Conventional Neural Networks

Lin Chen, *Member, IEEE*, and Xin Luo, *Senior Member, IEEE*

Dear Editor,

This letter presents a novel tensor-distribution-regression model based on 3D conventional neural networks (3D-TDR) with an application to clinical score prediction in aging-related diagnosis. The estimation of clinical scores of subjects using brain magnetic resonance imaging (MRI) helps understand the pathological stage of dementia. However, clinical scores prediction is still unsolved due to the reasons of: 1) Analyzing the whole-brain MRI is extremely difficult as the high-dimensional MRI data contains millions of voxels; 2) The clinical scores prediction is formulated as a one-dimensional regression issue in the current deep-learning-based algorithms, which ignores the implicit label information between subjects with different score levels. Motivated by the above discoveries, the proposed 3D-TDR model innovatively establishes the following three-fold ideas: a) incorporating a tensor regression layer (TRL) into a 3D conventional neural network (3D-CNN) to enable its extraction of more discriminative structural changes from the high-dimensional whole-brain magnetic resonance (MR) data; b) adopting the label distribution learning (LDL) to fully utilize the label correlation among the MR images, thus emphasizing the diversity of subjects' scores; and c) combining the TRL and LDL for an end-to-end deep learning framework, thereby achieving jointly low-rank feature extraction and clinical scores prediction. Experimental results on two real-world MRI datasets of two typical clinical prediction tasks indicate that the 3D-TDR outperforms the benchmark and state-of-the-art models. The proposed 3D-TDR model can achieve significant accuracy gain in dementia score and brain age prediction.

MRI is a non-invasive imaging technique with specific effects in clinical practice for assessing and diagnosing neurologic diseases of the brain [1]. Aging-related clinical score prediction based on the MRI data, e.g., clinical dementia rating sum of boxes (CDR-SB) [2] or chronological brain age, is vital for understating the stage of dementia pathology [3]. However, accurately estimating the clinical scores is a highly challenging task because high-dimensional brain MRI is a 3D neuroimaging data containing millions of voxels, making some brain regions irrelevant to dementia. Numerous machine-learning approaches [4], [5], especially deep-learning-based models [6], have been recently studied to improve aging-related score prediction relying on conventional regression [4] and CNN-based models [2], [7]. Existing models mostly adopt hand-craft morphology features [5], [8] or preselected brain regions [9].

Related work: Earlier attempts at dementia prediction are generally based on hand-craft MRI features at the voxel level for dementia diagnosis and regression [4]. Some studies also focus on feature extraction from the regions of interest (ROIs) in the brain, which is known to be informative. For instance, Lei *et al*. [8] implement the dementia scores regression based on a sparse learning technique using the selected brain gray matter density. However, these hand-craft-based features extracted following predefined patterns are independent of prediction models, leading to suboptimal results.

Corresponding author: Xin Luo.

Citation: L. Chen and X. Luo, "Tensor distribution regression based on the 3D conventional neural networks," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 7, pp. 1628–1630, Jul. 2023.

L. Chen is with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China (e-mail: chenlin@cigit.ac.cn).

X. Luo is with the College of Computer and Information Science, Southwest University, Chongqing 400715, China (e-mail: luoxin21@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2023.123591

With the rapid progress of deep learning, CNNs have been widely adopted in dementia prognosis. Liu *et al*. [2] propose a joint dementia classification and score regression framework with discriminative anatomical landmarks. Liu *et al*. [6] implement a weakly supervised densely neural network (wiseDNN) to predict the future clinical metrics using local MRI patches. Zhu *et al*. [10] capture regional MRI biomarkers with a dual attention mechanism. Peng *et al*. [11] propose a simple 3D-CNN architecture to predict brain ages using segmented gray and white matters. Cheng *et al*. [3] introduce a two-stage cascade densely network to estimate the brain age gaps for identifying the AD progressive stages.

However, most existing methods rely on predefined landmarks or local brain region extraction to enhance the CNN's partial feature expression, while how to fully utilize the whole-brain MR data for accurate clinical score prediction remains a thorny issue.

Problem statement: Supposing that a training set $\{\mathbf{X}_i, y_i\}_{i=1}^{S}$ with S subjects as $\mathbf{X}_i$ be the whole-brain MR scan for the i-th subject and y_i be the corresponding ground-truth clinical score. We are interested in precisely estimating y_i based on $\mathbf{X}_i$ via a machine learning program, which usually contains three main steps, i.e., 1) MR image preprocessing, 2) brain structural features extraction, and 3) clinical score regression model construction.

Considering the first step, we start with performing a standard preprocessing pipeline to transform the raw MR data into a standard format via the FMRIB's software library (FSL) package [12], including inhomogeneity correction, brain skull stripping, intensity normalization, and nonlinear registration to the MNI-152 template. Thus, the processed MRIs are resampled to the same size of 91×109×91 (spatial resolution of 2 mm^3) to form the input of the prediction model.

Afterward, the second and third steps are embedded into an end-to-end framework for a joint feature and regression learning based on the following learning objective:

$$\arg\min_{\mathbf{W}} \frac{1}{S}\sum_{i=1}^{S}(y_i - f(\mathbf{X}_i;\mathbf{W}))^2 \tag{1}$$

where $f(\mathbf{X}_i;\mathbf{W})$ is the predicted score with learnt $\mathbf{W}$. However, existing models suffer from the following two issues:

1) Conventional CNN-based methods usually adopt the local regions or patches of brain MRIs for model training based on predefined landmarks [6], [7]. However, empirical predefined brain regions cannot coordinate with a deep-learning framework well, leading to performance degradation.

2) Meanwhile, direct regression of the continuous variables based on (1) cannot fully utilize the correlation between subjects with different levels of scores [13].

This work aims to improve the performance of aging-related score prediction based on whole-brain 3D-MR images by carefully addressing the above two issues.

Proposed 3D-TDR: A novel 3D-CNN-based prediction model, i.e., 3D-TDR, is presented in this letter for accurately predicting the clinical scores, whose structure is illustrated in Fig. 1. It mainly consists of four components: 1) the 3D-CNN backbone, 2) the tensor regression layer, 3) the label distribution learning, and 4) the total training loss. We will go over the specifics of each component.

1) 3D-CNN backbone: The brain MRI can be visualized as a 3D matrix whose dimension is $H\times W\times D$ as H and W be the height and width of MRI while D be the imaging sequence. The 3D-CNN can be utilized to explore the overall structure alterations of the brain. As depicted in Fig. 1, we present a shallow 3D-CNN backbone, which is stacked with four blocks, and each block consists of a 3D-CNN layer with 3×3×3 kernels, a BN layer, a ReLu activation, and a 3D-max-pooling layer. Finally, we can obtain the 4D feature maps $\mathcal{F}$ with the size of $c\times h\times w\times d$ by our 3D-CNN backbone for each input brain MR scan, where $c = 512$ is the number of channels, and $h = H/32$, $w = W/32$, $d = D/32$ are the height, width, and the slices of the 3D-CNN feature maps, respectively.

2) Tensor regression layer (TRL): Tensor decomposition is an efficient way to low-rank representation for high-dimensional data while maintaining its multidimensional structures [14], [15]. This paper leverages a trainable weight tensor to extract a low-rank representation for 3D-CNN-based features via a TRL [16]. Let $F \in \mathbb{R}^{S\times c\times h\times w\times d}$ be the set of the activation 3D-CNN features and $\mathbf{Y} \in \mathbb{R}^{S\times 1}$ be the cor-

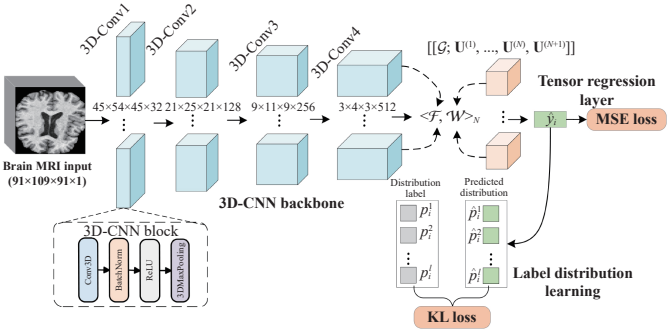

Fig. 1. The structure of the proposed 3D-TDR model.

responding clinical scores for training. We try to estimate the regression weight tensor $\mathcal{W} \in \mathbb{R}^{c\times h\times w\times d\times 1}$ with the fixed low rank (R_1,…, R_N, R_{N+1}) and a bias $\mathbf{b} \in \mathbb{R}^{S\times 1}$, where $N = 4$ denotes the way count of $\mathcal{F}$. We have

$$\mathbf{Y} = \langle \mathcal{F}, \mathcal{W} \rangle_N + \mathbf{b},\ \mathcal{W} = \left[\mathcal{G}; \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}, \mathbf{U}^{(N+1)} \right] \tag{2}$$

where $\mathcal{G} \in \mathbb{R}^{R_1\times \cdots R_N \times R_{N+1}}$ denotes the core tensor $\mathcal{W}$ achieved with the Tucker decomposition, and $\mathbf{U}^{(k)}$ is the projection vector as $\left\{ \mathbf{U}^{(k)} \in \mathbb{R}^{I_k \times R_k} | k \in [1, N] \right\}$ and $\mathbf{U}^{(N+1)} \in \mathbb{R}^{1\times R_{N+1}}$. Note that $\langle \mathcal{F}, \mathcal{W} \rangle_N \in \mathbb{R}^{S\times 1}$ denotes the generalized inner-product along the first (respectively last) modes of $\mathcal{F}$ (respectively $\mathcal{W}$) as

$$\langle \mathcal{F}, \mathcal{W} \rangle_N = \mathcal{F}_{[1]} \times \mathcal{W}_{[N+1]}. \tag{3}$$

Notably, the weight tensor $\mathcal{W}$ can be regarded as a differentiable neural network layer that is updated via gradient backpropagation. The low-rank values of (R_1, …, R_N, R_{N+1}) are fixed as (r, 3, 4, 3, 1), where r is a hyper-parameter to be discussed in the experiments. The mean square error (MSE) loss is adopted to measure the errors between the predicted $\hat{y}_i$ and the ground-truth y_i as

$$\ell_{mse} = \frac{1}{s} \sum_{i=1}^{S} (y_i - \hat{y}_i)^2 \tag{4}$$

where $\hat{y}_i \in \hat{\mathbf{Y}}$ denotes the predicted score of $\mathbf{X}_i$ following (2).

3) Label distribution learning (LDL): LDL is further utilized to explore the correlation among the MR images with similar outputs, thereby enhancing the prediction performance. For instance, the correlation between AD subjects with CDR-SB scores 6 and 8 should be stronger than that between normal control subjects with CDR-SB scores 0 and 1, in terms of structural changes in brain tissue that reflect dementia (e.g., the atrophy of the frontoparietal cortex and hippocampus). Thus, LDL tries to convert a one-dimensional regression problem to an LDL problem [17] by emphasizing the diversity between different levels of scores. Given an instance $\mathbf{X}_i$ with label y_i, we adopt $\mathbf{T} = \{t_i^0, t_i^1, \ldots, t_i^l\}$ to represent the ordered label vector obtained by splitting the ground truth of scores into several equally-spaced intervals, where l is the max-value of the clinical scores. Support a distribution vector label $\mathbf{P}_i = \{p_i^0, p_i^1, \ldots, p_i^l\}$ as $p_i^k \in \mathbf{P}_i$ be the probability of score $y_i = t_i^k$ (i.e., $p_i^k = Pr(y_i = t_i^k)$), for $0 \le k \le l$. In general, the ground-truth distribution $\mathbf{P}_i$ can be generated with a normal distribution as

$$p_i^k = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t_i^k - y_i)^2}{2\sigma} \right) \tag{5}$$

where σ is a hyper-parameter will be investigated in the experiments. By making $\mathbf{z}_i$ be the output after the TRL for the i-th sample, we adopt softmax to transfer predicted $\mathbf{z}_i$ to its label distribution $\hat{\mathbf{P}}_i$ as

$$\hat{p}_i^k = \frac{\exp(z_i^k)}{\sum_k \exp(z_i^k)} \tag{6}$$

where $\hat{p}_i^k \in \hat{\mathbf{P}}_i$, and $z_i^k \in \mathbf{z}_i$ as $0 \le k \le l$. The expected clinical score value $\hat{y_i}$ produced by TRL can be denoted as $\hat{y}_i = \sum_k \hat{p}_i^k t_i^k$. By using the KL divergence, we can estimate the dissimilarity between ground-truth $\mathbf{P}_i$ and predicted $\hat{\mathbf{P}}_i$ as

$$\ell_{kl} = \frac{1}{S} \sum_{i=1}^{S} \sum_k p_i^k \ln \frac{p_i^k}{\hat{p}_i^k}. \tag{7}$$

4) Total training loss: As shown in Fig. 1, the total loss consists of the MSE loss depending on the TRL and the KL loss of LDL, who are combined for jointly training the proposed 3D-TDR in an end-to-end manner. More specifically, for the input training samples with size S, 3D-TDR's total loss function is given as

$$\ell_{\text{total}} = \ell_{\text{mse}} + \ell_{kl} = \frac{1}{S} \sum_{i=1}^{S} \left((y_i - \sum_k \hat{p}_i^k t_i^k)^2 + \sum_k p_i^k \ln \frac{p_i^k}{\hat{p}_i^k} \right). \tag{8}$$

Experiments: We evaluate our method on two public databases, including ADNI (https://adni.loni.usc.edu/) and OASIS (http://www.oasis-brains.org/). For cognitive score estimation experiments, we collected a total of 1474 subjects from ADNI-1 (810 subjects) and ADNI-2 (664 subjects), along with the baseline MR scans and CDR-SB scores at four different time-points (i.e., baseline, month-06, month-12, and month-24, denoted as BL, M06, M12, and M24). We perform the twofold cross-validation on ADNI-1 and ADNI-2 (i.e., the models are first trained on ADNI-1 and then tested on ADNI-2. Afterwards, training and testing sets are reserved), respectively, to evaluate the generalization ability of the proposed model. For the brain age estimation test, we collect 1379 subjects with 2100 MR scans, and the dataset is divided into 8:1:1 for training, validation, and testing, respectively. We implement the tested models with Python 3.6 and Pytorch 1.10 on a single NVIDIA GeForce 1080Ti-11G GPU. The root mean square error (RMSE) and correlation coefficient (CC) are selected as the accuracy metrics.

The proposed model is compared with the following benchmarks and state-of-the-art models: 1) the landmark-based morphological features (LMF) based on SVR [9], 2) the baseline 3D-CNN, 3) weakly supervised densely connected neural network (wiseDNN) [6], 4) the 3D-CNN with ranking learning (rankCNN) [18], 5) the simple fully convolutional network (SFCN) [11], 6) the two-stage-age-network (TSAN) with the ranking loss [3], 7) the 3D-CNN with LDL (3D-LDL) [17], and h) the 3D-CNN with TRL (3D-TRL) [16]. Note that the proposed 3D-TDR is vastly different from the model in [16] in the following aspects: 1) We reform the single value regression in 3D-TRL as the multiple discrete labels distribution predictions; 2) We embed the KL loss of LDL into our 3D-TDR for exploring the diversity among subjects with different clinical values.

We report the RMSE and CC of CDR-SB obtained by the tested models on ADNI and OASIS databases in Table 1, Figs. 2 and 3, respectively. From them, we find that:

1) The deep-learning-based methods gain better CDR-SB prediction results than the LMF method on the ADNI dataset, which adopts the hand-crafted features extracted from MRI. The main reason is that the process of expert-defined features from MRI in LMF, is independent of the training for the disease prognosis predictor.

2) The proposed 3D-TDR outperforms its peers in both CC and RMSE on two databases. 3D-TDR has the lowest mean RMSE of 1.959 across four time-points on the ADNI database, indicating its significant performance gain over its peers. Although the wiseDNN can remove the redundancy voxels in brain MRI via the landmark-based-technology, 3D-TDR's RMSE (2.976) is still much lower than that of wiseDNN (3.412) in predicting CDR-SB at M24 on ADNI-1. Similar results are observed for the brain age estimation task. All these demonstrate that our proposed method is suitable for whole-brain MRI-based clinical score prediction.

3) In comparison with the baseline 3D-CNN, 3D-TRL achieves lower RMSE, and 3D-LDL achieves higher CC on most cases. 3D-TRL and 3D-LDL take the second place in terms of RMSE and CC, respectively, on the OASIS dataset. These results indicate that 3D-

Table 1. Comparison Results of RMSE Achieved by the Tested Methods on ADNI Datasets. The Best Values are Highlighted in Boldface. * Means That Our 3D-TDR Performs Significantly Better Than the Tested Method Using Paired t-test ($p < 0.05$)

Method	Tested on ADNI-2				Tested on ADNI-1				Mean
	BL	M06	M12	M24	BL	M06	M12	M24	
LMF	1.772	2.099	2.268	2.413	1.922	2.394	2.694	4.009	2.446*
3D-CNN	1.585	1.726	2.173	2.241	1.640	2.393	2.246	3.265	2.159*
wiseDNN	1.585	1.688	2.024	2.334	1.619	2.016	2.442	3.412	2.146*
rankCNN	1.535	1.910	2.042	2.190	1.689	1.934	2.250	3.277	2.098*
TSAN	1.552	1.685	2.046	2.193	1.562	1.957	2.124	3.078	2.025*
SFCN	1.533	1.673	2.033	2.133	1.577	1.928	2.173	3.040	2.011*
3D-LDL	1.534	1.637	2.101	2.179	1.533	1.978	2.147	3.083	2.024*
3D-TRL	1.521	1.660	2.013	2.202	1.530	1.929	2.108	3.066	2.004*
3D-TDR	**1.509**	**1.617**	**1.953**	**2.129**	**1.520**	**1.882**	**2.088**	**2.976**	**1.959**

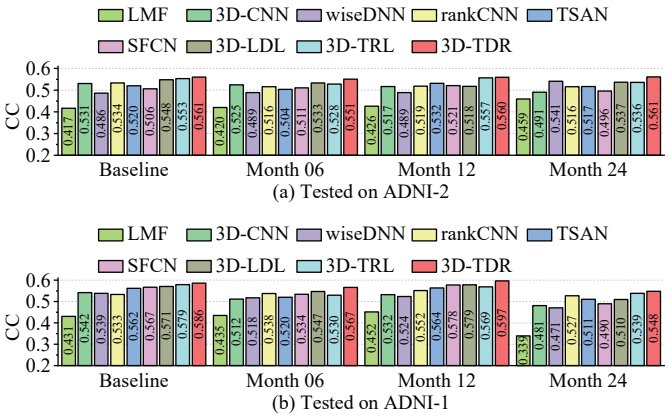

Fig. 2. Compression results of CCs on the ADNI dataset.

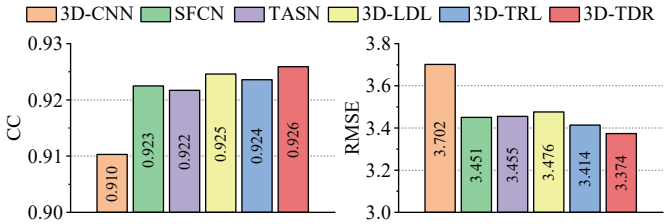

Fig. 3. Brain age prediction results in RMSE and CC on the OASIS dataset.

TDR with low-rank tensor decomposition can extract more useful structure information from the high-dimensional CNN features, while 3D-LDL can boost the generality of 3D-CNN by exploiting the inherent relationship among different clinical values. Our 3D-TDR method combined these two components achieves the best CC and RMSE values.

Moreover, we report the 3D-TDR's performance on ADNI-2 as r increases in the scale of [32, 512] and σ in the scale of [0.5, 2.5], as illustrated by Fig. 4. We see that the lowest RMSE can be generally achieved as $r = 128$ and $\sigma = 1$, which are consistent with our settings.

In addition, we replace our shallow 3D-CNN with the deeper 3D-ResNet. The results are shown in Table 2. From it, we see that the 3D-ResNet34 outperforms the other ResNet models, but our shallow backbone (with only 2 M parameters) presents much lower RMSE than the best ResNet model does, which is similar with that reported in [11]. The possible reason is that the 3D-Resnet has no pre-trained weights like in the 2D-Resnet (i.e., pre-trained on the ImageNet).

Conclusions: This paper presents a new tensor-distribution-regression model based on 3D-CNN to fully exploit the 3D-structure brain features and establish the relationship among different clinical values. Experiments on two public databases demonstrate that the proposed 3D-TDR method can predict clinical scores more accurately than the tested state-of-the-art methods. However, lots of computational resources are still consumed with the 3D-CNN model. In the future, we plan to explore more advanced model compression technology to improve the efficiency of the proposed model.

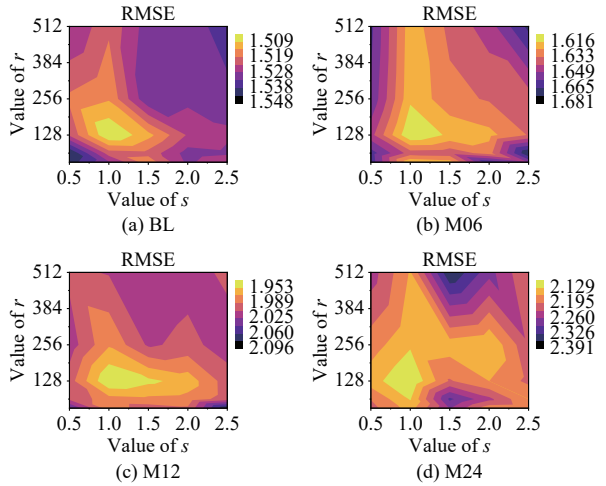

Fig. 4. The effort of hyper-parameters of r and σ. All the results are obtained by training and evaluating the models on ADNI-1 and ADNI-2, respectively.

Table 2. The Effort of 3D-CNN Backbones in Terms of RMSE. All the Results are Obtained by Training Models on ADNI-1 and Tested on ADNI-2

Method	RMSE				Mean	Parameters
	BL	M06	M12	M24		
3D-CNN (Ours)	**1.509**	**1.617**	**1.953**	**2.129**	**1.802**	**2.05M**
3D-ResNet10	1.581	1.753	2.141	2.298	1.943	14.56M
3D-ResNet18	1.582	1.709	2.088	2.287	1.917	33.37M
3D-ResNet34	1.573	1.699	2.065	2.218	1.889	63.67M

Acknowledgments: This work was supported by the National Nature Science Foundation of China (61902370, 62272078, 82030066), and the CAAI-Huawei MindSpore Open Fund (CAAIXSJLJJ-2021-035A).

References

[1] B. Zhang and Z. Zhu, "Linearized proximal alternating direction method of multipliers for parallel magnetic resonance imaging," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 4, pp. 763–769, 2017.

[2] M. Liu, J. Zhang, E. Adeli, *et al*., "Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis," *IEEE Trans. Biomedical Engineering*, vol. 66, no. 5, pp. 1195–1206, May 2019.

[3] J. Cheng, Z. Liu, H. Guan, *et al*., "Brain age estimation from MRI using cascade networks with ranking loss," *IEEE Trans. Medical Imaging*, vol. 40, no. 12, pp. 3400–3412, Dec. 2021.

[4] J. H. Morra, Z. Tu, L. G. Apostolova, *et al*., "Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls," *Neuroimage*, vol. 45, no. 1, pp. S3–S15, 2009.

[5] B. Lei, M. Yang, P. Yang, *et al*., "Deep and joint learning of longitudinal data for Alzheimer's disease prediction," *Pattern Recognition*, vol. 102, p. 107247, 2020.

[6] M. Liu, J. Zhang, *et al*., "Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores," *IEEE Trans. Cybernetics*, vol. 50, no. 7, pp. 3381–3392, Jul. 2020.

[7] C. Lian, M. Liu, J. Zhang, *et al*., "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[8] B. Lei, P. Yang, T. Wang, *et al*., "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybernetics*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.

[9] J. Zhang, Y. Gao, Y. Gao, *et al*., "Detecting anatomical landmarks for fast Alzheimer's disease diagnosis," *IEEE Trans. Medical Imaging*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.

[10] W. Zhu, L. Sun, *et al*., "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Medical Imaging*, vol. 40, no. 9, pp. 2354–2366, Sept. 2021.

[11] H. Peng, W. Gong, C. F. Beckmann, *et al*., "Accurate brain age prediction with lightweight deep neural networks," *Medical Image Analysis*, vol. 68, p. 101871, 2021.

[12] M. Jenkinson, C. F. Beckmann, T. E. Behrens, *et al*., "FSl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[13] Y. Lei, H. Zhu, J. Zhang, *et al*., "Meta ordinal regression forest for medical image classification with ordinal labels," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1–15, Jul. 2022.

[14] X. Luo, H. Wu, Z. Wang, *et al*., "A novel approach to large-scale dynamically weighted directed network representation," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 44, no. 12, pp. 9756–9773, Dec. 2022.

[15] C. Leng, H. Zhang, *et al*., "Total variation constrained non-negative matrix factorization for medical image registration," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 5, pp. 1025–1037, May 2021.

[16] J. Kossaifi, Z. C. Lipton, A. Kolbeinsson, *et al*., "Tensor regression networks," *J. Machine Learning Research*, vol. 21, no. 1, pp. 4862–4882, 2020.

[17] B.-B. Gao, C. Xing, C.-W. Xie, *et al*., "Deep label distribution learning with label ambiguity," *IEEE Trans. Image Processing*, vol. 26, no. 6, pp. 2825–2838, Jun. 2017.

[18] H. Qiao, L. Chen, and F. Zhu, "Ranking convolutional neural network for Alzheimer's disease mini-mental state examination prediction at multiple time-points," *Computer Methods Programs Biomedicine*, vol. 213, p. 106503, 2022.