MTFIL-Net: automated Alzheimer's disease detection and MMSE score prediction based on feature interactive learning

Jin Liu, Xu Tian, Jianxin Wang, Rui Guo, Hulin Kuang*
Hunan Province Key Lab on Bioinformatics,
School of Computer Science and Engineering, Central South University, Changsha, China
**Corresponding author: Hulin Kuang, Email: hulinkuang@csu.edu.cn*

2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669563

***Abstract*—Automatic detection of Alzheimer's disease (AD) is conducive to intervention in the disease progression. MMSE score prediction can reveal the development of AD. In recent years, some studies have designed multi-task methods for AD detection and MMSE score prediction to take advantage of the correlation between them. However, how to use the correlation between the two task features is still a problem. To address this challenge, we propose a multi-task feature interactive leanrning network (MTFIL-Net) to perform AD detection and MMSE score prediction. First, we interact the features acquired by CNNs corresponding to the two tasks to take advantage of the feature correlation between the two tasks. The interaction module extracts the shared features of the two tasks and concatenate them with the features of the two task. Then, we design a joint loss based on cross entropy and smooth L1 function. We use the distribution of MMSE scores to dynamically adjust the relationship between the two tasks. We validate our method with subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We use the ADNI1 dataset for training and testing, and used the ADNI2 dataset as an external validation set. Our proposed MTFIL-Net reached an ACC of 0.86 for AD detection and a correlation coefficient of 0.67 for MMSE score prediction on the ADNI1 dataset, and reached an ACC of 0.85 for AD detection and a correlation coefficient of 0.66 for MMSE score prediction on the ADNI2 dataset. Experiment results show that MTFIL-Net effectively utilizes the correlation between AD and MMSE score.**

***Index Terms*—Alzheimer's disease detection, MMSE score, Multi-task learning, Feature interaction, Adaptive joint loss**

I. Introduction

Alzheimer's disease (AD) [1] is a common neurodegenerative disease. It is the most predominant type of dementia in the elderly population [2]. In order to intervene in the disease as early as possible, it is necessary to develop computer-assisted automatic methods for accurate AD detection [3]. As the disease progresses, AD patients often suffer from memory, cognitive and motor impairments [4]. The mini-mental state examination (MMSE) is conducted through question and answer methods, which are mainly used to judge the state of thinking, cognition and behavior of patients [5]. This scale contains 30 questions, each with 1 point.

In recent years, many studies have explored various brain diseases including AD [6], autism spectrum disorders [7], and glioma [8] through medical imaging such as MRI. CNN-based AD diagnosis methods through MRI include 2D slice level, 3D patch level, region of interest (ROI) level and 3D subject level methods. The advantage of the 2D CNN method is that there are many successful methods in the field of image classification, such as ResNet [9] and VGGNet [10]. To use the spatial information between slices, some researchers segmented the 3D MRI into smaller blocks as input for the 3D CNN model. Zhang et al. [11] proposed a landmark-based feature extraction method for AD detection. This method identifies brain regions with significant differences between AD and CN, and uses features extracted from these regions for AD detection. As same as some methods of obtaining ROIs based on prior knowledge, this kind of methods is easy to lose information. 3D subject level methods use the entire MRI image of subject as input to the CNN. Qiu et al. [12] proposed an AD detection method that includes a fully convolutional network (FCN) and a separate classification model. FCN is used to determine which voxels are more effective for AD detection. Then, they use these voxels to complete AD detection. The MMSE score can reflect progression of dementia, many studies have tried to predict the MMSE score. Huang et al. [13] proposed a regression forest based on a soft split strategy. This soft splitting strategy makes the prediction result of the model a probabilistic fusion of the results of multiple branches of the decision tree, which allows the model to make more accurate predictions. Tabarestani et al. [14] tried to constrain the parameters of the tasks through different regular items to adjust the relationship between the prediction of multiple clinical scores including MMSE. Liu et al. [15] tried to predict clinical information including MMSE scores through weakly supervised deep learning.

Since the correlation between the MMSE score and AD, Duc et al. [16] tried to perform AD detection and MMSE score prediction at the same time. They developed a 3D CNN architecture for AD detection, and used multiple regression methods for MMSE score prediction. They used the consistency of the data distribution. Liu et al. [17] used a patch-based method for joint learning of AD detection and MMSE score prediction. They first identified the effective anatomical landmarks from the MRI image in a data-driven manner, and then extracted multiple image patches around these landmarks for joint AD detection and MMSE score prediction. EI-Sappagh et al. [18] tried to use multi-modal data including

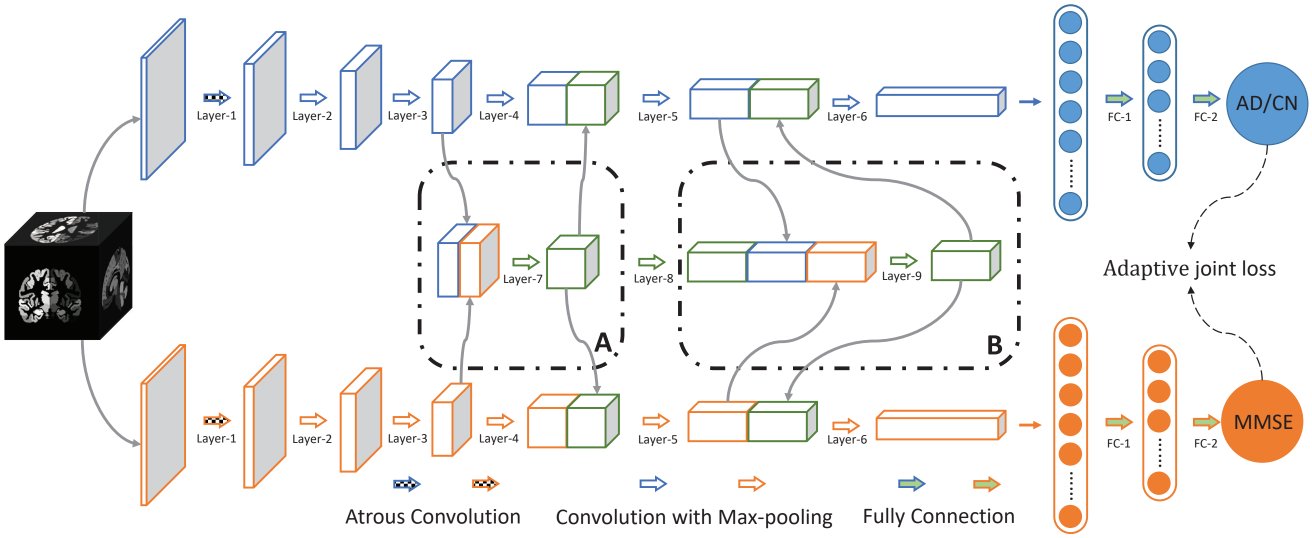

Fig. 1. Schematic diagram of MTFIL-Net structure. MTFIL-Net mainly includes three parts: the networks that perform AD detection and MMSE score prediction respectively and interactive modules A and B. The Layer-i or FC-i in the two task networks respectively represent the i-th layer or the i-th fully connected layer in the network used to implement these two tasks.

MRI neuroimaging data, genetic data and other statistical data to complete multiple tasks including AD detection and MMSE score prediction. The above methods generally used only one network backbone for joint learning. This made it impossible for the model to obtain features that are only effective for a single task. On the other hand, these methods did not adjust the relationship between the two tasks.

In this study, we propose multi-task feature interactive learning network (MTFIL-Net) with adaptive joint loss for AD detection and MMSE score prediction. The network structure is shown in Figure 1. First, we use the same 3D CNN for the two tasks. Then, we add interaction modules between the backbone networks of the two tasks, which extract the shared features of the two tasks. Finally, we design a adaptive joint loss for AD detection and MMSE score prediction. We use the ADNI (http://adni.loni.usc.edu/) dataset to validate the method. The ADNI1 dataset is used for training and testing, and the ADNI2 dataset is used for external verification.

The contributions of this study is summarized as follows:

- We propose modules A and B to implement the feature interaction between the two network backbones for AD detection and MMSE scoring.
- We propose the joint loss of the two tasks and dynamically adjust it through the distribution of MMSE score.

II. Method

A. Overview of our method

The MTFIL-Net we proposed is mainly composed of three parts, which are the two CNNs that implement AD detection and MMSE score prediction and the interaction modules between them. The interactive modules perform feature interaction between this two tasks in the feature extraction stage.

B. CNN structure for single task

We use the same shallow network for AD detection and MMSE score prediction. We choose to build a shallower network based on AlexNet [19] as the backbone network for each task. Our network contains six layers composed of convolution and max-pooling and two fully connected layers. A larger convolution kernel can extract more information. However, directly increasing the size of the convolution kernel puts the network at a higher risk of overfitting. Therefore, we use atrous convolution [20] with an dilation rate of 2 in the layer-1 of the network backbone of the two tasks. Atrous convolution expands the receptive field of the filter without increasing the parameters of the network, which further increases the information obtained by the filter.

C. Interactive feature learning between tasks

We propose two interaction modules named A and B for multi-task learning. The features extracted by the first two layers of the network are usually closely related to the high information density areas of the input image such as edges [21]. Therefore, our model starts performing feature interactions between from the output of layer-3 of the CNNs for this two tasks.

Module A takes the features extracted by the layer-3 of the two task networks as input, it contains a convolution and a max-pooling. Module A will return the acquired features to the network of the two tasks, which makes it possible to get the features that are effective for both tasks in a learning-based manner. The module A concatenates the shared features with the features acquired at the layer-4 of the network of the two tasks and inputs them to the layer-5.

Module B obtains the output of the layer-5 of the two networks as input, it contains a convolution only. Module B also obtain the shared features of the two networks. The

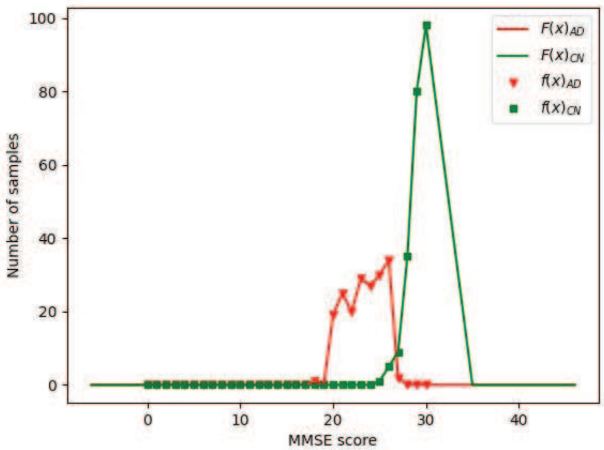

Fig. 2. MMSE score distribution with different groups in ADNI1 dataset

shared features acquired by the previous interaction module may be able to guide the the shared feature extraction of the subsequent interaction modules. Therefore, the shared features acquired by module A is also the input of module B. In order to unify the dimensions, the shared features of module A need to undergo necessary convolution and pooling before feeding into B. Module B concatenates the acquired shared features with the features output by the layer-5 of the two task networks. The layer-6 performs feature fusion and feature selection on the shared features and the features acquired by the task network itself.

D. Joint loss for multi-task learning

A loss function related to these AD detection and MMSE score prediction is required.

1) The loss of AD detection task: In the AD detection task, we use the cross-entropy loss:

$$loss_i^{AD}(y_i, p_i) = -[y_i log(p_i) + (1 - y_i)log(1 - p_i)] \quad (1)$$

where y_i denotes the label of sample i, and p_i denotes the probability that the predicted sample i is a positive sample.

2) The loss of MMSE score prediction task: For the MMSE score prediction task, we use a smooth L1 loss function, which combines the mean square error and the mean absolute error by using a piecewise function. Compared with the L1 function and the L2 function, it makes the training process more stable.

$$loss_i^{MMSE}(m_i, M_i) = \begin{cases} 0.5(m_i - M_i)^2 & |m_i - M_i| < 1 \\ |m_i - M_i| - 0.5 & \text{otherwise} \end{cases} \quad (2)$$

where m_i and M_i denotes the predicted value and the true value of MMSE score of sample i.

3) Adaptive parameter: As is shown in the Figure 2, the MMSE score has a strong correlation with AD. For the convenience of observation, we first fit the original discrete distribution function $f(x)$ with the continuous piecewise function $F(x)$.

First, we get a new intermediate function $f'(x)$. $f'(x)$ contains all points in $f(x)$. If $f(0)$ or $f(30)$ is not 0, then add $f'(-d)$ or $f(30+d)$ equal to 0 into $f'(x)$. d is the range between the maximum value and the minimum value of x when $f(x)$ is not 0:

$$d = max(\{x|f(x) > 0,\ 0 \le x \le 30\}) \\ - min(\{x|f(x) > 0,\ 0 \le x \le 30\}) \quad (3)$$

where $max(U)$ and $min(U)$ denote functions to obtain the maximum and minimum values of the set U. Then, we construct a linear function between the discrete points in the function $f'(x)$ and make it $F(x)$. Finally, it needs to extend the domain to the entire real number interval. At this time, its function value is always 0. The formula of $F(x)$ is as follows:

$$F(x) = \begin{cases} 0 & x \le s \ or \ x \ge e \\ f'(\lfloor x \rfloor) + \frac{f'(\lceil x \rceil) - f'(\lfloor x \rfloor)}{x - \lfloor x \rfloor} & s < x < e \end{cases} \\ s = min(A), \quad e = max(A) \quad (4)$$

where A denotes the domain of definition of the function $f'(x)$, s and e are the minimum and maximum values in A. After the above formula 4, we can get the distribution function $F_{AD}(x)$ and $F_{CN}(x)$.

We need a weight to adjust the MMSE score to predict the loss. When the prediction result is significantly different from the true value, the model needs to pay more attention to the MMSE score prediction task. Therefore, we designed γ_i to measure the difference between m_i and M_i:

$$\gamma_i = \frac{|\int_{M_i}^{m_i} F_c(x)dx|}{AU(F_c)}, c \in \{AD, CN\} \quad (5)$$

where m_i and M_i denote the predicted and true values of the MMSE score respectively, $AU(F_c)$ is the area under the function F_c, and c denotes the category of sample i. When m_i is close to M_i, the integration interval formed by m_i and M_i is narrower, and γ_i will be smaller. The γ_i value ranges from 0 to 1, which will cause us to underestimate the loss of the MMSE score prediction task. Therefore, we used a logarithmic function to calculate ω_i as the final weight of the MMSE score prediction:

$$\omega_i = -log(1 - \gamma_i + \epsilon) \quad (6)$$

where ϵ is a bias to ensure that the log function does not calculate 0.

4) The joint loss of MTFIL-Net: The model implements multi-task learning by minimizing the weighted combination of cross-entropy loss and smoothing L1 loss:

$$Loss_{JL} = \frac{1}{N}\sum_{i=1}^{N}(loss_i^{AD}(y_i, p_i) + \omega_i \, loss_i^{MMSE}(m_i, M_i)) \quad (7)$$

III. Experiments and Results

A. Data

In this study, we select subjects from the ADNI dataset to validate our proposed MTFIL-Net. The samples we used

TABLE I
DEMOGRAPHIC INFORMATION OF THE SUBJECTS INVOLVED IN THIS STUDY (MEAN)

Dataset	Group	Gender(male/female)	Age(year)	MMSE
ADNI1	AD	101/86	75.43	23.35
	CN	119/109	76.00	29.11
ADNI2	AD	55/29	75.74	21.90
	CN	73/64	74.88	28.93

include 187 AD patients and 228 CN subjects from the ADNI1 dataset and 84 AD patients and 137 CN subjects from the ADNI2 dataset. Table I summarizes the detailed information of these subjects.

B. Experimental settings

The FSL [22] software package is used to preprocess the MRI images. The data preprocessing process includes registration of the original image to the MNI152 standard template, skull removal and voxel intensity normalization. We perform 10-fold cross-validation with random division on the ADNI1 dataset. In each experiment, we only use the training set to construct $F_{CN}(x)$ and $F_{AD}(x)$. The ϵ is 10^{-4}. We choose the model that performs best on the validation set for testing. The performance of the AD detection task is first concern when choosing the optimal model. The parameter optimization algorithm we used is the Adam method [23]. The number of epochs is 300. The learning rate is 10^{-4}. The batch size is 2. All experiments is performed on NVIDIA GTX2080.

C. Evaluation metrics

For AD detection task, we use accuracy(ACC), sensitivity (SEN), specificity(SPE) and area under ROC curve(AUC) value to validate our model. For the MMSE score prediction task, we use root mean square error($RMSE$) and the Pearson correlation coefficient(r) as the metrics.

D. Comparison of MTFIL-Net with the state-of-the-art methods

We compare MTFIL-Net with some methods for the AD detection task. We choose Random Forest (RF), AlexNet [19] and VGGNet [10] for comparison. In the random forest method, we use voxels of MRI image as the features of the samples for AD detection. For AlexNet and VGGNet, we input the entire MRI image of the patient to detect AD. We also select some state-of-the-art methods for comparison. Zhang et al. [11] used a data-driven method to identify brain regions with differences between AD and CN, and used the features extracted from these regions to perform AD detection. Qiu et al. [12] proposed an AD detection method based on Fully Convolutional Network (FCN). They selected a part of voxels to participate in classification through FCN. Duc et al. [16] used 3D CNN to perform the AD classification. As shown in Table II, our proposed MTFIL-Net achieves a ACC of 0.85, a SEN of 0.81, a SPE of 0.80 and a AUC of 0.95. Compared with other methods, we are in a leading position in most metrics.

TABLE II
COMPARISON OF AD DETECTION PERFORMANCE OF MTFIL-NET RANDOM FOREST (RF), ALEXNET, VGGNET AND THE STATE-OF-THE-ART METHODS USING ADNI1 DATASET

Methods	ACC	SEN	SPE	AUC
RF	0.73	0.68	0.77	0.78
AlexNet [19]	0.71	0.63	0.78	0.78
VGG [10]	0.77	0.65	0.87	0.86
Zhang et al. [11]	0.83	0.81	0.87	-
Qiu et al. [12]	0.83	0.76	0.88	0.92
Duc et al. [16]	0.85	**0.98**	0.67	-
MTFIL-Net	**0.86**	0.81	**0.90**	**0.95**

TABLE III
COMPARISON OF MMSE SCORE PREDICTION PERFORMANCE OF MTFIL-NET, ALEXNET AND THE THE STATE-OF-THE-ART METHODS USING ADNI1 DATASET

Methods	RMSE	r
AlexNet [19]	3.71	0.58
Zhu et al. [24]	4.58	0.66
Duc et al. [16]	2.83	0.57
Liu et al. [17]	**2.37**	0.56
MTFIL-Net	2.65	**0.67**

We also compare MTFIL-Net with some methods for the MMSE score prediction task. We choose AlexNet [19] for comparison. Same as AD detection task, the completed MRI image is input into AlextNet to predict MMSE score. Then, we compare our proposed MTFIL-Net with some state-of-the-art methods. Zhu et al. [24] designed a new loss function based on matrix similarity, which is related to the high-level information inherent in the target corresponding matrix. For MMSE score prediction, Duc et al. [16] used SVM-based recursive feature elimination and tree regression with group independent component analysis. Liu et al. [17] used the multi-channel learning method to predict the MMSE score. As shown in Table III, our proposed MTFIL-Net achieves a $RMSE$ of 2.65 and a r value of 0.67. Compared with these methods, our method has the highest r, and $RMSE$ is also lower than most methods.

E. Ablation experiment

We further explore the effectiveness of multi-task feature interactive learning and adaptive joint loss. In subsequent experiments, we add the ADNI2 dataset as an external test set. The ADNI2 dataset will be directly test on the optimal model trained on the ADNI1 dataset.

1) Multi-task feature interactive learning: We remove the interactive modules in MTFIL-Net and name the CNNs that implement the AD detection and the MMSE scoring ST-Net$_{AD}$ and ST-Net$_{MMSE}$. We also compare with common multi-task learning method named MT-Net. In this model, two tasks share a network backbone and branch at the end of the network. The results are shown in Table IV. Our proposed MTFIL-Net are better than other methods in the most metrics.

2) Adaptive joint loss: We explore the impact of adaptive joint loss on model performance. In this experiment, we fix ω to 1 of MTFIL-Net and name it MTFIL-Net$_{\omega=1}$. The result

TABLE IV
COMPARISON OF AD DETECTION AND MMSE SCORE PREDICTION PERFORMANCE OF MTFIL-NET, ST-NET$_{AD}$, ST-NET$_{MMSE}$ AND MT-NET

Dataset	Methods	ACC	SEN	SPE	AUC	RMSE	r
ADNI1	ST-Net$_{AD}$	0.77	0.71	0.84	0.88	-	-
	ST-Net$_{MMSE}$	-	-	-	-	3.57	0.42
	MT-Net	0.80	0.79	0.82	0.89	3.34	0.37
	MTFIL-Net	**0.86**	**0.84**	**0.90**	**0.95**	**2.65**	**0.67**
ADNI2	ST-Net$_{AD}$	0.73	0.60	**0.92**	0.87	-	-
	ST-Net$_{MMSE}$	-	-	-	-	3.93	0.30
	MT-Net	0.80	**0.87**	0.89	0.92	4.03	0.35
	MTFIL-Net	**0.85**	0.85	0.86	**0.93**	**3.08**	**0.66**

TABLE V
COMPARISON OF AD DETECTION AND MMSE SCORE PREDICTION PERFORMANCE OF MTFIL-NET AND MTFIL-NET$_{\omega=1}$

Datasets	Methods	ACC	SEN	SPE	AUC	RMSE	r
ADNI1	MTFIL-Net$_{\omega=1}$	0.80	0.76	0.85	0.92	3.31	0.46
	MTFIL-Net	**0.86**	**0.80**	**0.91**	**0.94**	**2.65**	**0.67**
ADNI2	MTFIL-Net$_{\omega=1}$	0.84	0.84	0.85	**0.93**	3.82	0.45
	MTFIL-Net	**0.85**	**0.85**	**0.86**	**0.93**	**3.08**	**0.66**

is shown in Table V. On the ADNI1 dataset and ADNI2 dataset, the $RMSE$ of the model decreased by 0.66 and 0.74, respectively, and the r value increased by 0.21 both. The adaptive parameters greatly reduce the prediction error of the MMSE score.

IV. CONCLUSION

In summary, we propose MTFIL-Net to implement AD detection and MMSE score prediction. The experiment results not only show that our method has better performance than several state-of-the-art methods, but also show the effectiveness of multi-task feature interactive learning and adaptive joint loss based on MMSE score distribution. MTFIL-Net might provides a general framework for disease detection and prediction of disease-related clinical scores.

ACKNOWLEDGMENTS

This work is funded partially by the National Natural Science Foundation of China under Grant No.62172444, No.62102454, No.61877059, the 111 Project(No.B18059), the Hunan Provincial Science and Technology Program(No.2018WK4001), and the Hunan Provincial Science and Technology Innovation Leading Plan(No.2020GK2019).

REFERENCES

[1] M. A. DeTure and D. W. Dickson, “The neuropathological diagnosis of alzheimer’s disease,” *Molecular neurodegeneration*, vol. 14, no. 1, pp. 1–18, 2019.

[2] P. Yi, L. Jin, T. Xu, L. Wei, and G. Rui, “Hippocampal segmentation in brain mri images using machine learning methods: A survey,” *Chinese Journal of Electronics*, vol. 30, no. 5, pp. 793–814, 2021.

[3] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. Frackowiak, “Automatic classification of mr scans in alzheimer’s disease,” *Brain*, vol. 131, no. 3, pp. 681–689, 2008.

[4] H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang *et al.*, “Single-cell transcriptomic analysis of alzheimer’s disease,” *Nature*, vol. 570, no. 7761, pp. 332–337, 2019.

[5] J. Liu, Y. Pan, F.-X. Wu, and J. Wang, “Enhancing the feature representation of multi-modal mri data by combining multi-view information for mci classification,” *Neurocomputing*, vol. 400, pp. 322–332, 2020.

[6] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, “Classification of alzheimer’s disease using whole brain hierarchical network,” *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 2, pp. 624–632, 2016.

[7] Y. Wang, J. Liu, Y. Xiang, J. Wang, Q. Chen, and J. Chong, “Mage: Automatic diagnosis of autism spectrum disorders using multi-atlas graph convolutional networks and ensemble learning,” *Neurocomputing*, 2021, doi: 10.1016/j.neucom.2020.06.152.

[8] J. Cheng, J. Liu, H. Yue, H. Bai, Y. Pan, and J. Wang, “Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric mri images,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020, doi: 10.1109/TCBB.2020.3033538.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[11] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, “Detecting anatomical landmarks for fast alzheimer’s disease diagnosis,” *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2524–2533, 2016.

[12] S. Qiu, P. S. Joshi, M. I. Miller, C. Xue, X. Zhou, C. Karjadi, G. H. Chang, A. S. Joshi, B. Dwyer, S. Zhu *et al.*, “Development and validation of an interpretable deep learning framework for alzheimer’s disease classification,” *Brain*, vol. 143, no. 6, pp. 1920–1933, 2020.

[13] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, D. Shen, A. D. N. Initiative *et al.*, “Longitudinal clinical score prediction in alzheimer’s disease with soft-split sparse regression based random forest,” *Neurobiology of aging*, vol. 46, pp. 180–191, 2016.

[14] S. Tabarestani, M. Aghili, M. Eslami, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, “A distributed multitask multimodal approach for the prediction of alzheimer’s disease in a longitudinal study,” *NeuroImage*, vol. 206, p. 116317, 2020.

[15] M. Liu, J. Zhang, C. Lian, and D. Shen, “Weakly supervised deep learning for brain disease prognosis using mri and incomplete clinical scores,” *IEEE transactions on cybernetics*, vol. 50, no. 7, pp. 3381–3392, 2019.

[16] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, “3d-deep learning based automatic diagnosis of alzheimer’s disease with joint mmse prediction using resting-state fmri,” *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, 2020.

[17] M. Liu, J. Zhang, E. Adeli, and D. Shen, “Joint classification and regression via deep multi-task multi-channel learning for alzheimer’s disease diagnosis,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1195–1206, 2018.

[18] S. El-Sappagh, T. Abuhmed, S. R. Islam, and K. S. Kwak, "Multimodal multitask deep learning model for alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, 2020.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[21] X. Zhang, L. Han, W. Zhu, L. Sun, and D. Zhang, "An explainable 3d residual self-attention deep neural network for joint atrophy localization and alzheimer's disease diagnosis using structural mri," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[22] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] X. Zhu, H.-I. Suk, and D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in ad diagnosis," *NeuroImage*, vol. 100, pp. 91–105, 2014.