2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

JOINTLY ANALYZING ALZHEIMER'S DISEASE RELATED STRUCTURE-FUNCTION USING DEEP CROSS-MODEL ATTENTION NETWORK

Lu Zhang[1], Li Wang[1,2], Dajiang Zhu[1]

[1]Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA
[2]Mathematics, University of Texas at Arlington, Arlington, TX, USA

ABSTRACT

Reversing the pathology of Alzheimer's disease (AD) has so far not been possible, a more tractable way may be having the intervention in its earlier stage, such as mild cognitive impairment (MCI) which is considered as the precursor of AD. Recent advances in deep learning have triggered a new era in AD/MCI classification and a variety of deep models and algorithms have been developed to classify multiple clinical groups (e.g. aged normal control - CN vs. MCI) and AD conversion. Unfortunately, it is still largely unknown what is the relationship between the altered functional connectivity and structural connectome at individual level. In this work, we introduced a deep cross-model attention network (DCMAT) to jointly model brain structure and function. Specifically, DCMAT is composed of one RNN (Recurrent Neural Network) layer and multiple graph attention (GAT) blocks, which can effectively represent disease-specific functional dynamics on individual structural network. The designed attention layer (in GAT block) aims to learn deep relations among different brain regions when differentiating MCI from CN. The proposed DCMAT shows promising classification performance compared to recent studies. More importantly, our results suggest that the MCI related functional interactions might go beyond the directly connected brain regions.

Index Terms— Attention, Alzheimer’s disease, GAT

1. INTRODUCTION

As an irreversible, progressive brain disorder, Alzheimer’s disease (AD) is currently ranked as the sixth leading cause of death in the United States[1]. Although the pathology cannot be reversed after established, diagnosing at its earlier stage - mild cognitive impairment (MCI - precursor of AD and converts to AD at approximately 10% to 15% per year), is a more feasibly way for AD patients and their family to plan for the future. Therefore, establishing a set of objective and non-invasive imaging biomarkers is extremely critical.

Previous studies reported a variety of approaches for early detection of AD/MCI, such as machine learning-based algorithms[2,3], tract-based spatial statistics[4,5] and voxel-based analysis[6,7]. Recently, benefiting from the development of deep learning methods, deep models have been applied on AD/MCI prediction and classification. For example, Convolutional Neural Network (CNN)[8] was used to conduct AD classification based on structural MRI and PET image. Graph Convolutional Network (GCN)[9] was adopted for modeling the MCI conversion. These deep learning models have shown potential in large-scale brain image analysis for disorders comparing to previous linear/shallow methods. On the other hand, brain structure and functional are closely related[10] and AD-induced brain deterioration is a systemic process. It is suggested that AD progression tends to involve both brain structural and functional alterations[11]. Thus, only focusing on structural or functional abnormalities may be suboptimal. Unfortunately, due to the lack of deep models that can effectively integrate both brain structural and functional knowledge, it is still largely unknown what is the relationship between the altered functional connectivity and structural connectome at individual level.

In this work, using multimodal brain imaging data in ADNI 3 as a test-bed, we introduced a deep cross-model attention network (DCMAT) to jointly model brain structure and function. Our contributions include: 1) DCMAT is composed of one RNN (Recurrent Neural Network) layer and multiple graph attention (GAT) blocks containing a GCN layer and an attention layer, which can effectively represent MCI-specific functional dynamics on individual structural network; 2) we create structural connectome for different individuals to initialize the adjacency matrix of GCN layer. Thus, the functional interactions can be examined within the topology of brain structural connectome; 3) the designed attention layer aims to learn “deep relations” among different brain regions when differentiating MCI from CN. The proposed DCMAT shows promising classification performance (98.6% for overall accuracy) compared to recent studies. More importantly, our results suggest that the MCI related functional interactions might go beyond the directly connected brain regions.

2. METHOD

2.1. Dataset

In this work, all the data used is obtained from Alzheimer’s Disease Neuroimaging Initiative (ADNI 3)[12]. We began

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on June 20,2022 at 01:36:15 UTC from IEEE Xplore. Restrictions apply.

with 661 subjects, which have both DTI and rs-fMRI data. The DTI data is 2.0mm isotropic, TE=56ms, TR=7.2s, and the gradient directions is 54. For rs-fMRI, the range of image resolution in X and Y dimensions is from 2.29mm to 3.31mm. The range of slice thickness is from 3.3mm to 3.4mm, TE=30ms, TR=3s, and there are 197 volumes (time points) for each subject. We applied the same standard pre-processing procedures in [13]. After the pre-processing and image quality check, we have 209 subjects (116 CN/93 MCI) for experiments.

2.2. Data organization

We adopted Destrieux[14] atlas (as nodes) to create brain structural connectivity matrices, $\boldsymbol{A} \epsilon \boldsymbol{R}^{N \times N}$, weighted by number of fibers connecting two regions. N=148 is the number of brain regions. We used the averaged fMRI BOLD signals, $\boldsymbol{F} \epsilon \boldsymbol{R}^{N \times T}$ as temporal feature. To acquire more training data, we divided the original signal into multiple segments and each segment contains 45 time points. Previous study[15] indicates that 20 time points (40s) incline to be sufficient to capture major functional dynamics in resting state fMRI.

2.3. Deep cross-model attention network (DCMAT)

DCMAT is designed for classifying aged normal controls (CN) from MCI patients and the classification performance is the driving force for model training. It is composed of one RNN layer and multiple graph attention (GAT) blocks (**Fig. 1**). Each GAT block contains an attention layer and a graph convolution layer. The details of RNN layer, attention layer and graph convolution layer will be discussed in section 2.3.1, 2.3.2 and 2.3.3 respectively.

2.3.1. Long Short Term Memory Networks layer (LSTM)

RNN[16,17] is a class of deep neural networks which can model dynamic temporal characteristics of sequences by maintaining internal hidden states. LSTM[18, 19] extends the basic RNN by adding three gates: forget gate f, input gate i and output gate o, which enable LSTM to capture long-term dependency in a sequence, and make it easier to optimize[20]. LSTM formulas are listed below: σ is the activation function (e.g. ReLU in this work), matrices $\{W_f, W_i, W_c, W_o\}$ and vectors $\{b_f, b_i, b_c, b_o\}$ are parameters which need to be learned.

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot [h_{t-1}, \boldsymbol{F_t}] + b_f\right) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, \boldsymbol{F_t}] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, \boldsymbol{F_t}] + b_c) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, \boldsymbol{F_t}] + b_o) \\ h_t &= o_t * tanh(C_t) \end{aligned} \tag{1}$$

In our experiment design, we used LSTM to explore the dynamic temporal features of fMRI signals. The weights of LSTM are updated according to the classification results, so the learned dynamic temporal features are likely to play important roles in differentiating CN and MCI. That is, by passing through the RNN layer, the outputs (hidden states) represent the most useful temporal features for each brain region when classifying MCI patients.

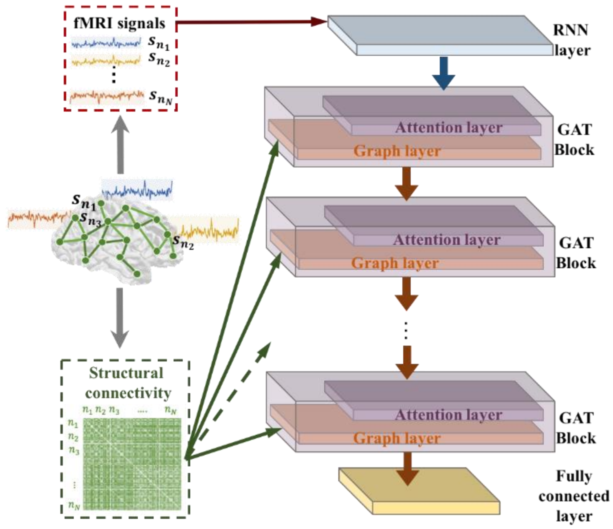

Fig. 1. Architecture of the proposed deep cross-model attention network (DCMAT). The fMRI signals extracted from different brain regions are used as inputs for the RNN layer to capture the dynamic temporal features; then, the output of RNN layer together with structural connectivity are fed into multiple GAT blocks to 1) calculate disease-based functional correlations and reorganize dynamic temporal features at each attention layer; 2) model the brain structural and functional information simultaneously using structural connectivity matrix and reorganized dynamic temporal features at GCN layer (graph layer).

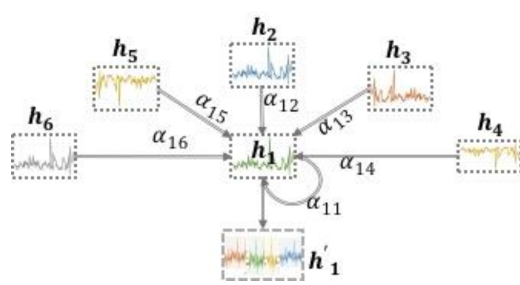

Fig. 2. An example of the mechanism of attention layer. Feature $\boldsymbol{h}_1$ is reorganized by linear combination of features $\boldsymbol{h}_j$ (the coefficient α_{1j} is the learned attention coefficients).

2.3.2. Attention layer

In this paper we attempted to explore the influence of MCI on not only temporal characteristic of single brain region but also functional correlations between different brain regions. To achieve this goal, we designed an attention mechanism: the input of attention layer is the dynamic temporal functional features captured by the previous RNN layer, $\mathbf{h} =$

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on June 20,2022 at 01:36:15 UTC from IEEE Xplore. Restrictions apply.

$\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_N\}$, $\boldsymbol{h}_i \epsilon \boldsymbol{R}^F$, where N is the number of brain regions, and F is the number of functional features. In order to obtain high-level features, a shared linear transformation with transformation matrix $\mathbf{W} \epsilon \boldsymbol{R}^{F' \times F}$, was applied, where F' is the number of output features. We performed a shared attention mechanism $\boldsymbol{att}$: $\boldsymbol{R}^{2F'} \times \boldsymbol{R}^{2F'} \rightarrow \boldsymbol{R}$ to calculate attention coefficients:

$$\alpha_{ij} = \boldsymbol{att}(\boldsymbol{W}h_i, \boldsymbol{W}h_j) = ReLU\left(\boldsymbol{a}^T \left[\boldsymbol{W}h_i || \boldsymbol{W}h_j\right]\right) \quad (2)$$

Where $\boldsymbol{a}^T \epsilon \boldsymbol{R}^{2F'}$ is a weight vector, $||$ is concatenation operation and α_{ij} indicates the importance of feature j to feature i. Since the weight vector is updated based on classification results, the learned attention coefficients reflect hidden relations among brain regions. The proposed attention mechanism is implemented by a single-layer feedforward neural network and parameterized by weight vector $\boldsymbol{a}^T$. Finally, after computing the attention coefficients, we reorganized each feature (**Fig. 2**):

$$\boldsymbol{h}_i' = ReLU\left(\textstyle\sum_j \alpha_{ij} \boldsymbol{W}\boldsymbol{h}_j\right) \quad (3)$$

2.3.3. Graph convolution layer (GCN layer)

The attention layer only considers the data from brain functional domain. In this section we plan to further model the relations between the learned functional features and individual structural connectome. We introduced a GCN[21, 22] layer, with individual structural connectivity as the graph topology (adjacency matrix) and the output from attention layer as features (**Fig. 3**).

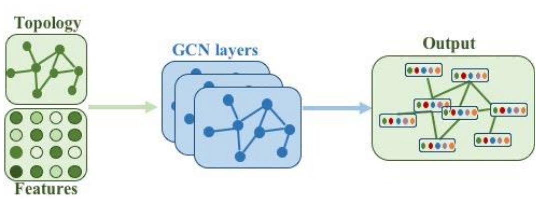

Fig. 3. Graph convolution layers (GCN layers). Each GCN layer can combine features from first-order neighbor (directly connected nodes), and by stacking multiple layers, features from distant neighbors (higher-order neighbor) can be integrated.

A GCN layer creates a hidden representation for each graph vertex by combining features from its neighbors. By stacking more layers, features from distant neighbors can be integrated. Therefore, the output of GCN layers considers both brain structural connectome and functional features at the same time. The two-layer GCN is given by:

$$G(\boldsymbol{A}, \boldsymbol{F}) = ReLU(\boldsymbol{A}\, ReLU(\boldsymbol{AFW_0})\boldsymbol{W}_1) \quad (4)$$

$\boldsymbol{W}$ is weight matrix of each layer, $\boldsymbol{A}$ is the adjacency matrix and $\boldsymbol{F}$ is feature matrix. In our model, one GAT block is composed of an attention layer and a GCN layer. In general, the attention layer prepares the features reflecting brain functional dynamics, and the GCN layer tries to capture the latent relations among the functional features on individual brain structural network. Through stacking multiple GAT blocks, we can have a more comprehensive representation of both brain structure and function simultaneously. At last, the outputs of GAT are fed into a fully-connected layer for classification, and the classification results are fed into the cross entropy loss function to optimize the entire model.

3. RESULTS

Given the rs-fMRI signal and structural connectivity (section 2.2), we conducted MCI related brain structure-function analysis proposed in section 2.3. In order to explore their deep relations, we stacked three GAT blocks (if enough data are available, more GAT blocks can be stacked).

3.1. Top 10 and top 100 MCI-related correlations.

In this work, we show top 10 and top 100 MCI-related functional connectome at each GAT layer and the related brain regions in **Fig. 4** and **Fig. 5**. According to section 2.3.2, since the attention coefficients are learned from MCI classification results, the larger coefficient implies stronger relation to MCI. Different GAT layer focuses on different depths of brain structural network.

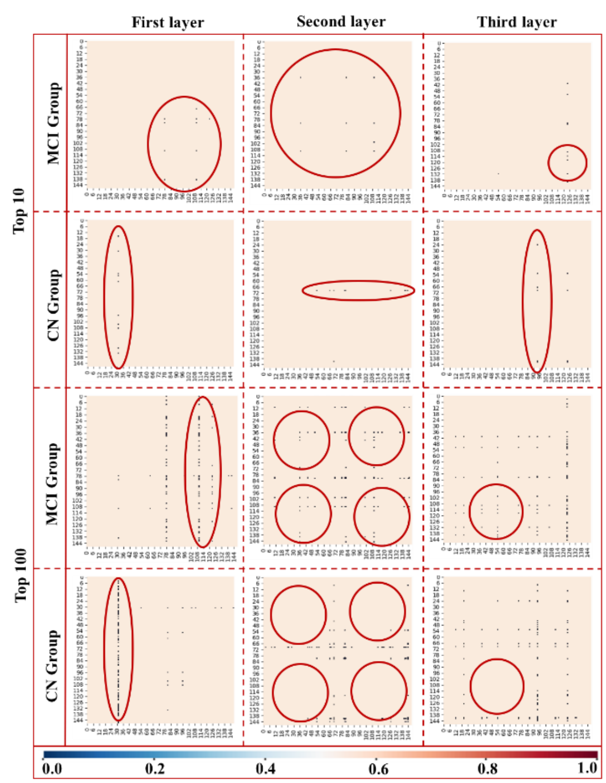

Fig. 4. Top 10 and top 100 MCI-related functional correlations of each GAT layer of CN and MCI groups.

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on June 20,2022 at 01:36:15 UTC from IEEE Xplore. Restrictions apply.

From **Fig. 4** we can see that these top functional connectomes show different spatial patterns on two groups (see highlighted circles). The deeper layer tends to capture more disease specific relations. **Fig. 5** shows the distribution and connectivity of these brain regions on brain map, the radius of the bubbles is proportional to the coefficients.

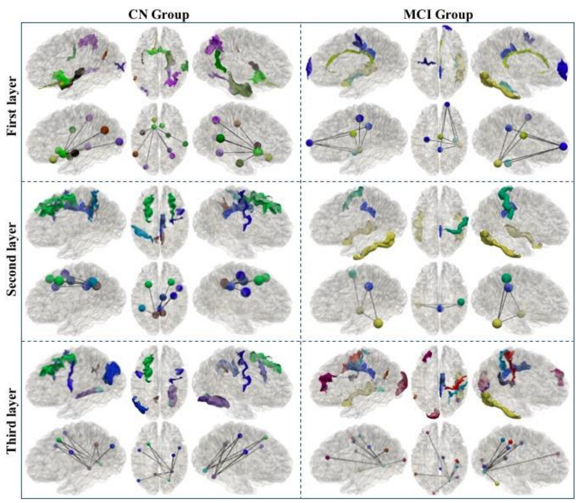

Fig. 5. Brain map of top 10 MCI-related functional correlations of each GAT layer of CN and MCI group.

3.2. Coefficients of different GAT Blocks.

In our experiment, we used three GAT blocks to explore the MCI-related functional correlations. Different layers consider different depths of brain connections. **Fig. 6 left column** shows the whole attention coefficient matrices for each of three GAT layers of both CN and MCI groups. For better visualization, we also show coefficients between median and maximum of each matrix at **right column**.

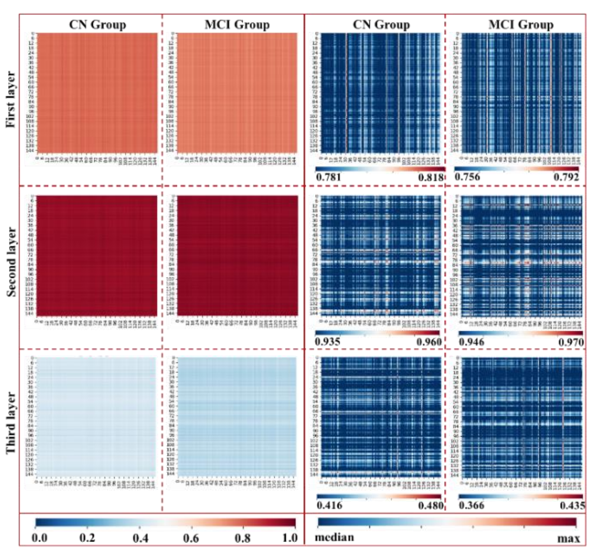

Fig. 6. Attention coefficient matrices of three GAT layers of CN and MCI groups.

From **Fig. 6** we can see that the attention coefficients of the second layer have the largest values and the attention coefficients of the third layer have the smallest values. Note that the attention coefficient represents the influence between brain regions. Thus, this result suggests that the disease specific functional interactions may go beyond the first-order neighbors (brain regions with direct structural connections) and incline to propagate along with structural connectivity to influence the second-order neighbors (two regions are connected via another region). However, this affects may limit to the third-order.

3.3. Classification performance.

The classification performance of our model and other widely used methods in the literatures are listed in **Table 1**. Given the similar sample size, our results outperform most recent studies.

Table 1. Classification performance.

Study	Sample size	Method	Acc
Payan and Montana, 2015 [23]	755AD, 755MCI, 755CN	Sparse Auto-edcoder with CNN	MCI vs CN 92.1%
Sarraf et al, 2016[24]	91AD, 211CN	CNN	AD vs CN 98.84%
Hosseini-Asl et al, 2016 [25]	70AD, 70MCI, 70CN	3D convolutional auto-encoder	MCI vs CN 94.2%
Suk et al, 2017 [26]	186AD, 226CN, 167 converters MCI, 226 stable MCI	Multiple sparse regression with CNN	MCI vs CN 66.78%-73.02%
Lu Zhang et al, 2019 [13]	116CN, 93MCI	GCRNN	MCI vs CN 97.7%
DCMAT with **two-layer** GAT (**ours**)	116CN, 93MCI	DCMAT	MCI vs CN 98.6%
DCMAT with **three-layer** GAT (**ours**)	116CN, 93MCI	DCMAT	MCI vs CN 98.3%

4. CONCLUSION

In this work, we introduced a deep cross-model attention network (DCMAT) to jointly analyze MCI-related functional interactions and individual structural connectome. We used RNN, GCN and attention layer to model brain temporal dynamics, structural connectome and latent structure-function associations, respectively. Besides the promising classification performance, DCMAT suggests that the MCI related functional interactions may go beyond the first-order neighbors (brain regions) which have direct connections.

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on June 20,2022 at 01:36:15 UTC from IEEE Xplore. Restrictions apply.

5. REFERENCES

[1] Alzheimer's Association 2019 Alzheimer's Disease Facts and Figures Report, https://www.alz.org/alzheimers-dementia/facts-figures.

[2] Jiang, X., et al.: Intrinsic Functional Component Analysis via Sparse Representation on Alz-heimer's Disease Neuroimaging Initiative Database. Brain Connect. 4, 575–586 (2014).

[3] Tong, Tong, et al.: Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. Pattern recognition 63, 171-181 (2017).

[4] Vemuri, P. et al.: Accelerated vs. unaccelerated serial MRI based TBM-SyN measurements for clinical trials in Alzheimer's disease. Neuroimage 113, 61–69 (2015).

[5] Smith, S. M. et al.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505 (2006).

[6] Ashburner, J. & Friston, K. J.: Voxel-Based Morphometry-The Methods. Neuroimage 11, 805–821 (2000).

[7] Thompson, P. M., Apostolova, L. G.: Computational anatomical methods as ap-plied to ageing and dementia. Br. J. Radiol. 80, S78–S91 (2007).

[8] Liu, Mingxia, et al.: Deep multi-task multi-channel learning for joint classification and regression of brain status. In International conference on medical image computing and computer-assisted intervention, Springer, Cham, pp. (3-11). (2017).

[9] Parisot, S., et al.: Spectral graph convolutions for population-based disease prediction. In International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, Cham. pp. (177-185), (2017).

[10] Honey CJ, Sporns O, Cammoun L, Gigandet X, Thiran JP, Meuli R, Hagmann P (2009) Predicting human resting-state functional connectivity from structural connectivity. Proceedings of the National Academy of Sciences, 106(6):2035–2040.

[11] Franzmeier N, Dyrba M (2017) Functional brain network architecture may route progression of Alzheimer's disease pathology. Brain, 140(12):3077–3080.

[12] ADNI | Alzheimer's Disease Neuroimaging Initiative, http://adni.loni.usc.edu/.

[13] Lu Zhang, et al.: A Cascaded Multi-Modality Analysis in Mild Cognitive Impairment. 10th International Workshop on Machine learning in Medical Imaging (MLMI 2019), held in Conjunction with MICCAI 2019, in pressing.

[14] Destrieux, Christophe, et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Nuroimage 53(1), 1-15 (2010).

[15] Zhang, Xin, et al.: Characterization of task-free and task-performance brain states via functional connectome patterns. Medical image analysis 17(8) 1106-1122. (2013).

[16] Cho, K., et al.: Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proc. Empirical Methods in Natural Language Processing, pp. 1724– 1735, 2014.

[17] Sutskever, I., et al.: Sequence to Sequence Learning with Neural Networks. In Proc. Advances in Neural Information Processing Systems 27, pp. 3104–3112, 2014.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] Sutskever, Ilya, Vinyals, Oriol, and Le, Quoc V. V. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems 27, pp. 3104–3112. 2014.

[20] Wang, Jiang, et al. "Cnn-rnn: A unified framework for multi-label image classification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[21] Wu, Zonghan, et al. "A comprehensive survey on graph neural networks." arXiv preprint arXiv:1901.00596 (2019).

[22] Zhang, Ziwei, Peng Cui, and Wenwu Zhu. "Deep learning on graphs: A survey." arXiv preprint arXiv:1812.04202 (2018).

[23] Payan A, Montana G.: Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks[J]. arXiv preprint arXiv:1502.02506, 2015.

[24] Sarraf S, Tofighi G.: DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI[J]. BioRxiv, 2016: 070441.

[25] Hosseini-Asl E, Gimel'farb G, El-Baz A.: Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network[J]. arXiv preprint arXiv:1607.00556, 2016.

[26] Suk H I, Lee S W, Shen D, et al.: Deep ensemble learning of sparse regression models for brain disease diagnosis[J]. Medical image analysis, 2017, 37: 101-113.

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on June 20,2022 at 01:36:15 UTC from IEEE Xplore. Restrictions apply.