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

LEARNING LATENT STRUCTURE OVER DEEP FUSION MODEL OF MILD COGNITIVE IMPAIRMENT

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

Many computational models have been developed to understand Alzheimer's disease (AD) and its precursor - mild cognitive impairment (MCI) using non-invasive neural imaging techniques, i.e. magnetic resonance imaging (MRI) based imaging modalities. Most existing methods focused on identification of imaging biomarkers, classification/ prediction of different clinical stages, regression of cognitive scores, or their combination as multi-task learning. Given the widely existed individual variability, however, it is still challenging to consider different learning tasks simultaneously even they share a similar goal: exploring the intrinsic alteration patterns in AD/MCI patients. Moreover, AD is a progressive neurodegenerative disorder with a long preclinical period. Besides conducting simple classification, brain changes should be considered within the entire AD/MCI progression process. Here, we introduced a novel deep fusion model for MCI using functional MRI data. We integrated autoencoder, multi-class classification and structure learning into a single deep model. During the modeling, different clinical groups including normal controls, early MCI and late MCI are considered simultaneously. With the learned discriminative representations, we not only can achieve a satisfied classification performance, but also construct a tree structure of MCI progressions.

Index Terms—autoencoder, multi-class classification, deep fusion, fMRI

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the greatest known risk factor is increasing age[1]. As the sixth-leading cause of death in the United States, AD is the only disorder that cannot be prevented, cured or even slowed[2]. Almost all AD patients experience the stage of mild cognitive impairment (MCI) including early MCI and late MCI, and more than half of MCI patients will convert to AD eventually. Therefore, using non-invasive imaging techniques, such as magnetic resonance imaging (MRI) based imaging modalities, to understand how MCI patients progress is of great interest. In the last decades, many computational methods have been developed to understand AD/MCI, such as voxel-based morphometry[3] (VBM), tensor-based morphometry[4] (TBM), tract-based spatial statistics[5] (TBSS), connectome analysis[6,7] and shallow/deep model-based machine learning algorithms[8, 9]. They have been applied to different tasks: classification of different clinical groups (normal aging versus AD/MCI), prediction of AD conversion, regression of cognitive scores or identification of objective imaging biomarkers. Though these tasks have different learning targets, they are highly related since they share a similar goal: exploring the intrinsic alteration patterns associated with the development of the disorder and clinical diagnosis. Unfortunately, current methods have difficulties in considering these related tasks simultaneously, which is however essential to understand the fundamental of how MCI develops and progresses.

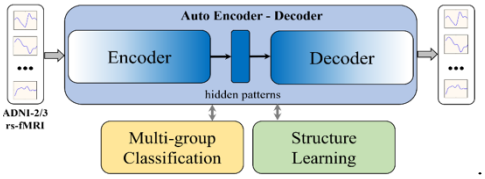

Fig. 1. Overview of the proposed framework.

Here, we proposed a novel deep fusion model to study multiple clinical groups including normal controls (NC), early MCI (EMCI) and late MCI (LMCI) at the same time, using functional MRI data from Alzheimer's Disease Neuroimaging Initiative[10] (ADNI). Fig.1 illustrates the overview of our framework. We integrate three different learning tasks into a single deep model: 1) Autoencoder – identify hidden patterns that are essential to represent the original data. 2) Classification – perform multi-class classification of NC/EMCI/LMCI. 3) Structure learning – learn a graph structure to model the relationship among different clinical groups. By intervening the identification of basis representation (encoder/decoder) under the guidance from classification and structure learning, we can effectively combine these three tasks/models together to explore the intrinsic patterns of MCI patients. Our results indicate that with the learned deep representation, we not only can

achieve 73% accuracy for multi-class classification, but also derive a tree structure for MCI progression.

2. METHOD

2.1. Data

2.1.1. Acquisition and preprocessing

The rs-fMRI data used in this work was obtained from Alzheimer's Disease Neuroimaging Initiative[10] (ADNI 2/3). In total there are 386 subjects (253 CN, 45 EMCI and 88 LMCI) involved in this study. The range of image resolution in X and Y dimensions was from 2.29mm to 3.31mm. The range of slice thickness was from 3.3mm to 3.4mm, TE=30ms, and there are two kinds of TR (3s and 607ms). The number of volumes (time points) is 197. We applied standard pre-processing procedures including skull removal, spatial smoothing, slice time correction, temporal pre-whitening, global drift removal and band pass filtering (0.01-0.1 Hz). We adopted Destrieux atlas[11] for ROI labeling and made the average signal within each ROI as the representative.

2.1.2. Feature extraction: local signal patterns

Given a preprocessed fMRI BOLD signal, we ran a slide window of size d to extract all segments. Previous study[12] suggested that using 14 time points (TR=2s) the dynamic patterns could be effectively captured. Therefore, we set a slightly larger d=20 in this study to ensure that no useful patterns will be missed. Then a greedy clustering approach is conducted to partition similar segments into clusters over all segments. In total, we generated 18,556 representative patterns (clusters). Further, we applied a discriminative graph embedding method[13] to learn a latent space, which was treated as the input of our deep fusion model. The reason of this step is to remove the noise of the input signals and make the input signal patterns more discriminative.

2.2. Deep representation guided by classification and structure learning.

Our proposed deep fusion model aims to explore the intrinsic patterns of MCI patients. It contains three learning tasks which are corresponding to three different models: autoencoder, multi-class classification and structure learning. Section 2.2.1-2.2.3 introduce three models separately and their joint modeling is discussed in 2.2.4.

2.2.1. Autoencoder for hidden representation learning

Given a set of data $\{x_1, \dots, x_n\}$, let f be the nonlinear mapping that transforms x to a latent representation. To obtain a "good" latent representation, we also expect that the basis representation can somehow reconstruct the original signals by removing certain noise. Autoencoder[14] is a proper model to learn a nonlinear representation of data by introducing the other function g as the decoder, which maps f(x) back to its original space. In general, the structures of layers for encoder and decoder are set symmetrically. The objective function of the autoencoder is formulated by minimizing the reconstruction error using the mean square error with respect to {f, g} given by:

$$L_a(f,g) = \frac{1}{n}\sum_i \left\| x_i - g\big(f(x_i)\big)\right\|^2. \quad (1)$$

2.2.2. Multi-class classification.

We can leverage the nonlinear representation $f(x)$ in (1) for better classification problem. Let $y_i \in \{1, \cdots, C\}$ be the class labels of the input data with C classes. For the use of deep learning, each label is transformed into a row vector of the one-hot representation with the c^{th} entry for the appearance of the c^{th} class in y_i, that is $y_{i,c} = 1$ if $y_i = c$ otherwise zeros. To connect the data representation with class labels, a classification layer is required to build on hidden representation $f(x)$ as the input. Accordingly, the model parameters $\{W_c\}$ and $\{b_c\}$ are designed for the classification output for each predicted class. Moreover, the function $\{h_c\}$ is set to be the softmax function with respect to the c^{th} class in order to form a probabilistic output of a given input. Finally, we minimize the categorical crossentropy loss for multi-class classification based on the class labels and the probabilistic outputs, and use a deep fully connected network to model f. Given the input data and its labels, the minimization problem with respect to $\{f, W, b\}$ is formulated as:

$$L_c(f,W,b) = -\frac{1}{n}\sum_i \sum_c y_{i,c}\, log\big(h_c(W_c f(x_i) + b_c)\big). \quad (2)$$

2.2.3. Latent representation denoising and graph structure learning.

As one of three major objectives in the proposed deep fusion model, structure learning aims to find a latent graph structure that can represent the intrinsic patterns of MCI patients. This latent representation also serves for both autoencoder and classification, so it is possible that the representation was contaminated by some noise introduced by the other two objectives. This might prevent the current model from learning a desired graph structure. Here, we introduce a density-based model for further denoising the latent representation. First, the fuzzy c-means model[15] is presented with the number of centers equal to the number of data points. Let z_k be the k^{th} center of the latent representation of data, and $p_{i,k}$ is the probability of assigning $f(x_i)$ to the k^{th} cluster. The clustering model with the above special setting leads to the effect of moving points to the high-density region of data, so it is able to effectively remove data noise that causes the points away from the high density regions. In other words, z_i is the destination of moving $f(x_i)$ by the clustering algorithm. Parameter σ is the bandwidth of the estimated density function. For graph

structure learning, the reversed graph embedding is introduced, where S is the matrix representation of a graph and G is the set of graphs of certain type (i.e. tree). By combining the two objectives, the graph structure of data is learned over the denoised latent representation $Z = [\, z_k \,]$ by minimizing the following objective function:

$$L_s(f,P,Z,S) = \sum_{i,k} p_{i,k}\left(\|f(x_i) - z_k\|^2 + \sigma\, log\left(p_{i,k}\right)\right) + \frac{\gamma}{\beta}\sum_{k,k'} s_{k,k'}\|z_k - z_{k'}\|^2 \quad (3)$$

where assignment matrix $P = \left[\, p_{i,k} \,\right]$ satisfies constraints: $\sum_k p_{i,k} = 1, p_{i,k} > 0, \forall i,k.$ Parameters β and γ are introduced to tradeoff the objectives of clustering and structure learning. Moreover, we consider the minimum-cost spanning trees $S = [s_{k,k'}] \in G$ for modeling the progression of MCI patients.

2.2.4. Joint model for classification, nonlinear representation and graph structure.

By combining the three objective functions (1)-(3), we propose the following joint optimization problem to achieve the aforementioned purposes in a unified framework:

$$\min_{f,g,Z,S,W,b} L_c(f,W,b) + \alpha L_a(f,g) + \beta L_s(f,P,Z,S) \quad (4)$$
$$s.t. \quad \sum_k p_{i,k} = 1, p_{i,k} > 0, S \in G,$$

where hyper-parameters α, β are the regularization parameters to trade-off the contributions of the three objectives.

The alternating method is employed to solve problem **(4)**. We partition model parameters to three groups $\{f,g,W,b\}$, Z, and $\{P,S\}$ The problem **(4)** with respect to $\{f,g,W,b\}$ by fixing others is reformulated as:

$$\min_{f,g,W,b} -\frac{1}{n}\sum_i\sum_c y_{i,c}\, log\left(h_c(W_c f(x_i) + b_c)\right) + \alpha\frac{1}{n}\sum_i \left\|x_i - g\left(f(x_i)\right)\right\|^2 + \beta\sum_i\left(\|f(x_i)\|^2 - 2\, u_i^T f(x_i)\right) \quad (5)$$

where $u_i = \sum_k p_{i,k} z_k,\ \forall i$. The stochastic gradient descend method can be used to solve the above problem and the gradient with respect to all network parameters can be obtained by the backpropagation approach through the network. The problem **(4)** with respect to Z is a quadratic programming, which can be rewritten as:

$$\min_Z \beta\, trace(Z\Lambda Z^T - 2ZA^T) - 2\Upsilon trace(ZLZ^T) \quad (6)$$

Where $A = [a_1, \dots, a_k] \in R^{n\times k}$ with $a_k = \sum_i p_{i,k} f(x_i),\ \forall k$, Λ is a diagonal matrix with the k^{th} entry $\sum_i p_{i,k}$, and graph Laplacian matrix $L = diag(S1) - S$ on graph S. As a result, we have the closed form solution as

$$Z = U\left(V + \frac{2\gamma}{\beta}L\right)^{-1} \quad (7)$$

The problem **(4)** with respect to $\{P,S\}$ can be solved independently due to the decoupling of two variables. By applying the Lagrange duality, we have the closed form solution, $\forall i,k$ given by:

$$p_{i,k} = \frac{exp\left(-\frac{\|f(x_i) - z_k\|^2}{\sigma}\right)}{\sum_k exp\left(-\frac{\|f(x_i) - z_k\|^2}{\sigma}\right)} \quad (8)$$

In addition, S is obtained by the Kruskal's algorithm. Since problem **(4)** is a nonlinear optimization problem, the initialization for variables $\{f,g,W,b\}$ becomes very important. Following the work in[16], we conduct pre-training and fine-tuning steps. In the pre-training, we first train the classification model for $\{f,W,b\}$ and then train the decoder g by fixing the encoder and the classification layer by setting $\beta = \gamma = 0$. Once $\{f,g,W,b\}$ are all initialized, we do fine-tuning for all variables together. Finally, by initializing $z_i = f(x_i)$, we alternate **(6)**, **(7)**, **(8)** and the Kruskal's algorithm for S until convergence or some predefine number of iterations.

3. RESULTS

3.1. Experimental setting.

The classification model is a fully connected multilayer perceptron with dimensions d-50-10-C, where d is the dimensionality of the input data and C is the number of classes. Except the classification layer, the rest of layers are the encoder. The decoder is a mirrored version of the encoder. The first layer and classification layer are applied the linear activation function and the second layer is applied a *ReLU* activation function. During the training, we apply *L1* regularization for the first layer with regularization parameter 5 and *L2* regularization for the second layer with parameter 0.5. The parameters in each layer are initialized following the *Xavier scheme*. The Adam optimizer is used to train the deep learning model **(6)** with standard learning rate 0.0001 and momentum rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For small data size, the batch size is set to be the number of data in the training set.

3.2. Discriminative signal patterns.

We have found a series of discriminative signal patterns that represent the most intrinsic differences among three clinical

groups. We name them as hidden signal patterns. **Fig. 2(a)** shows the top 25 hidden signal patterns that are most discriminative during both multi-class classification and latent structure learning. Because we partitioned the brain cortex into 150 regions and the representative signal of each region can be recovered using these hidden signal patterns. We can compute how many discriminative hidden patterns "contained" in each brain region by simply counting. **Fig. 2(b)** shows one randomly selected subject and all brain regions are encoded by the normalized result of how many discriminative signal patterns involved. The top five regions containing the most discriminative patterns are right G_precentral, left G&S_transv-frontopol, left G&S_frontomargin, left G&S_paracentral and left G_temporal_middle. All these regions have been reported for abnormal functional activities in AD/MCI[17,18]. The overall classification and reconstruction results are summarized in **Table 1**. In general, 73% is so far one of the best performance for multi-class classification (NC vs. EMCI vs. LMCI) using fMRI only. A recent study achieved similar accuracy for NC vs. MCI vs. AD with multi-modality data[8].

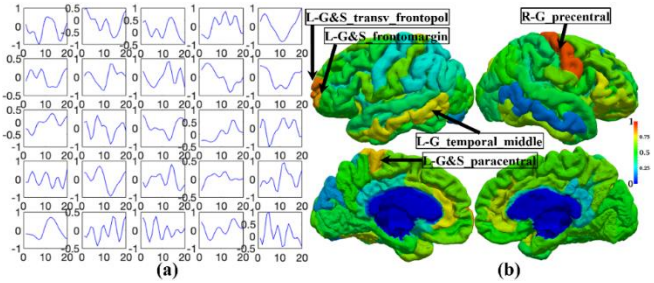

Fig. 1. (a) Top 25 discriminative hidden signal patterns found by the proposed method. **(b)** Normalized result of how many discriminative patterns contained in each brain region. The brain regions are from Destrieux atlas[11].

Table 1. Performance of multi-class classification and autoencoder.

Accuracy (NC vs. EMCI vs. LMCI)		Reconstruction error	
Training	**Testing**	**training**	**testing**
68%	73%	0.0015	0.0034

3.3. Latent structure - MCI progression tree.

Besides autoencoder and multi-class classification, we aim to learn a latent structure that represents the intrinsic relationship between normal people and MCI patients. Fig. 3 illustrates three views of the learned latent structure. Different colors represent different clinical groups including NC (*green*), EMCI (*yellow*) and LMCI (*red*). Individual subject is encoded as colored bubble. In general, the learned structure displays an interesting pattern of MCI progression: in the center, NC and EMCI are mixed together (black circle). For most branches (black arrows), they start from mixing of NC and EMCI (*green and yellow*), goes through EMCI (*yellow*) and eventually ends with LMCI (*red*). Therefore, we call it a "progression tree" of MCI. It should be noted that instead of using multiple clinical stages of the same group of subjects (i.e. longitudinal data), the achieved "progression tree" is learned by cross-sectional data from different clinical diagnostic groups.

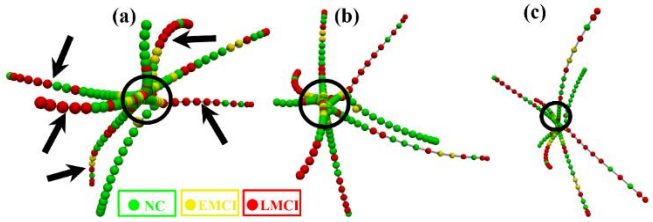

Fig. 2. The latent structure learned forms a MCI progression tree.

3.4. Convergence analysis.

The major optimization procedure in **(4)** has three parts: autoencoder, multi-class classification and structure learning. To examine the convergence of the proposed model, we plot the changes of the objective values in **Fig. 4**. We observed that the classification model converges well in pretraining stage during the first 30 iterations and the alternating method for the entire model also converges tens of iterations.

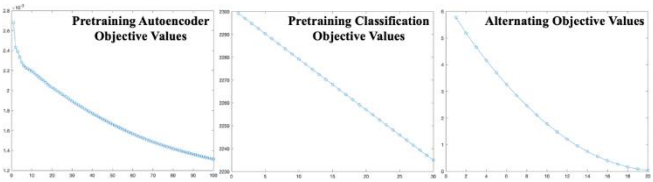

Fig. 3. The convergence results for pre-training classifier/decoder and the alternating method.

4. CONCLUSION

In this paper, we proposed a novel framework to integrate autoencoder, multi-class classification and structure learning as a deep fusion model. We aim to identify a group of basis representations of the original fMRI signals through autoencoder. At the same time, these representations are not only discriminative for different clinical groups including NC, EMCI and LMCI, but also can be used to learn a latent structure based on the relationship between different groups. Our results suggest that with hidden signal patterns we can achieve 73% accuracy in multi-class classification and the latent structure can form a progression tree for MCI. Both the discriminative hidden patterns and the latent structure are derived from our proposed deep fusion model and they have the potential to provide a new insight to understand the mechanism of AD/MCI development.

5. REFERENCES

[1] Hebert, L.E., Weuve, J., Scherr, P.A., Evans, D.A.: Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. Neurology. 80, 1778–1783 (2013).

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

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

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

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

[6] Yu, R., et al.: Weighted graph regularized sparse brain network construction for MCI identification. Pattern Recognit. 90, 220–231 (2019).

[7] Zhu, D., et al.: Connectome-scale assessments of structural and functional connectivity in MCI. Hum. Brain Mapp. 35, 2911–2923 (2014).

[8] Tong, T., et al.: Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. Pattern Recognit. 63, 171–181 (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 (MICCAI). pp. 177–185. Springer, Cham (2017).

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

[11] Destrieux, C., et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage. 53, 1–15 (2010).

[12] Zhang, X., et al.: Characterization of task-free and task-performance brain states via functional connectome patterns. Med. Image Anal. 17, 1106–1122 (2013).

[13] Wang, Li, Paul M. Thompson, and Dajiang Zhu.: Analyzing Mild Cognitive Impairment Progression via Multi-view Structural Learning. International Conference on Information Processing in Medical Imaging. Springer, Cham, 2019.

[14] Hinton, G.E., Zemel, R.S.: Autoencoders, Minimum Description Length and Helmholtz Free Energy. In: Cowan, J.D., Tesauro, G., and Alspector, J. (eds.) Advances in Neural Information Processing Systems 6. pp. 3–10. Morgan-Kaufmann (1994).

[15] Bezdek, J.C., Ehrlich, R., Full, W.: FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. 10, 191–203 (1984). https://doi.org/10.1016/0098-3004(84)90020-7.

[16] Erhan, D., Bengio, et al.: Why Does Unsupervised Pre-training Help Deep Learning? J. Mach. Learn. Res. 11, 625–660 (2010).

[17] Risacher, S.L., Saykin, A.J.: Neuroimaging and Other Biomarkers for Alzheimer's Disease: The Changing Landscape of Early Detection. Annu. Rev. Clin. Psychol. 9, 621–6ss48 (2013).

[18] Weiner, M.W., et al.: The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers. Dement. 8, S1-68 (2012).