Published in final edited form as:

Mach Learn Med Imaging. 2017 September ; 10541: 132–140. doi:10.1007/978-3-319-67389-9_16.

Feature Learning and Fusion of Multimodality Neuroimaging and Genetic Data for Multi-status Dementia Diagnosis

Tao Zhou, Kim-Han Thung, Xiaofeng Zhu, and Dinggang Shen

Department of Radiology, BRIC, University of North Carolina, Chapel Hill, USA

Abstract

In this paper, we aim to maximally utilize multimodality neuroimaging and genetic data to predict Alzheimer's disease (AD) and its prodromal status, i.e., a multi-status dementia diagnosis problem. Multimodality neuroimaging data such as MRI and PET provide valuable insights to abnormalities, and genetic data such as Single Nucleotide Polymorphism (SNP) provide information about a patient's AD risk factors. When used in conjunction, AD diagnosis may be improved. However, these data are heterogeneous (e.g., having different data distributions), and have different number of samples (e.g., PET data is having far less number of samples than the numbers of MRI or SNPs). Thus, learning an effective model using these data is challenging. To this end, we present a novel **three-stage deep feature learning and fusion framework**, where the deep neural network is trained stage-wise. Each stage of the network learns feature representations for different combination of modalities, via effective training using **maximum number of available samples**. Specifically, in the first stage, we learn latent representations (i.e., high-level features) for each modality independently, so that the heterogeneity between modalities can be better addressed and then combined in the next stage. In the second stage, we learn the joint latent features for each pair of modality combination by using the high-level features learned from the first stage. In the third stage, we learn the diagnostic labels by fusing the learned joint latent features from the second stage. We have tested our framework on Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset for multi-status AD diagnosis, and the experimental results show that the proposed framework outperforms other methods.

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia that often affect individuals over 65 years old. According to a recent report from Alzheimer's Association, the total estimated prevalence of AD is expected to be 60 million worldwide over the next 50 years. As there is no cure for AD, the accurate diagnosis of AD and especially its prodromal stage, i.e., mild cognitive impairment (MCI), which can be further categorized into progressive MCI (p-MCI) and stable MCI (s-MCI), is highly desirable in clinical application.

In search of an accurate biomarker for AD, data from different types of modalities have been collected and investigated [6,14,17]. Among these modalities, neuroimaging techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Topography (PET), are able

to provide anatomical and functional information about the brain, respectively. Many studies in recent years have shown that fusing the information from multiple neuroimaging modalities will enhance the performance of AD diagnosis [3]. On the other hand, genomewide association studies have identified a series of genetic variations (e.g., Single Nucleotide Polymorphism (SNP)) that are associated with AD. In addition, the association between the SNP genotype data and the neuroimaging phenotype features has also been studied [1,9]. Through the genetic information from SNPs, we can assess the risk factor of an individual likely to get AD, while through the brain structural and functional information gathered from neuroimaging data, we can assess the current brain abnormality of an individual. Both types of information are complementary to each other, and, when used together, may be able to improve the AD diagnosis.

Machine learning approaches have been widely used in neuroimaging-based AD diagnosis. Their performances generally depend on how well they can transform the neuroimages into representations that can be effectively used for AD identification. This transformation is commonly called feature learning or representation, which involves techniques like feature extraction and selection. For example, works in [8,15] use sparse learning to select discriminative features for AD diagnosis. On the other hand, deep learning, which has been proven effective in many classification applications, has the ability to learn high-level abstraction (or latent patterns) from the data. Because of this, deep learning has been widely applied in biomedical research, achieving promising results. For example, Xiao et al. [16] used stacked deep polynomial networks to learn multimodal neuroimaging features for AD diagnosis. Fakoor et al. [4] used unsupervised deep learning methods such as Stack Auto-Encoder (SAE) to enhance cancer diagnosis using gene expression data, and Suk et al. [13] and Liu et al. [11] adopted SAE to discover the latent feature representation from region-of-interest (ROI) based features for AD diagnosis. In addition, some related works [11–13] focused on learning high-level features from multimodality data. However, all these methods used only samples with complete multimodality data to train the model, which may seriously jeopardize the effectiveness of deep learning, a process which generally relies on large amount of data for training a good model.

In this study, we will use multimodality neuroimaging data (i.e., MRI and PET) and genetic data (i.e., SNP). There are two challenges on how to maximally utilize and fuse the information from these multimodality data for AD diagnosis. The first challenge is related to the data heterogeneity, as the neuroimaging and genetic data often have different data distributions, different numbers of features, and different levels of discriminative ability to AD diagnosis (e.g., SNP data in its raw form are less effective in AD diagnosis). Thus, using these data directly by simple concatenation will result in an inaccurate prediction model. The second challenge is the small sample size issue caused by incomplete data, i.e., not all samples have all the three modalities. As deep learning performs better when it is trained using a large number of samples, the small sample size issue may severely affect its performance.

In this paper, we propose a novel three-stage deep feature learning and fusion framework to address the above challenges, as shown in Fig. 1. Specifically, we train our networks stage-wise, where, at each stage, we learn the latent data representations (high-level features) for

different combination of modalities by using the maximum number of available samples. Specifically, in the first stage, we learn high-level features for each modality independently, so that the heterogeneity between modalities can be better addressed and different modality data can be more comparable in the latent representation space. For instance, for SNPs with high dimensionality, we use more hidden layers (compared to MRI and PET data), but with gradually decreased number of neurons for subsequent layers, to reduce its dimensionality while capturing its latent features, so that it is more comparable with the latent features learned from the neuroimaging data. In other words, we address the data heterogeneity issue by learning latent features through progressive mapping via multiple hidden layers. In the second stage, we learn joint features for each modality combination by using the high-level latent features learned from the first stage. In the third stage, we learn the diagnostic labels by fusing the learned joint features from the second stage. It is worth noting that to effectively train each stage of the network, one must use the maximum number of available samples. For example, in the first stage, to learn the high-level latent features from MRI data, we use all the available MRI data; in the second stage, to learn the joint high-level features from MRI and PET data, we use all samples with both MRI and PET data. By this way, the small-sample-size and incomplete dataset issues are partially addressed.

2 Methodology

Our goal is to learn a multi-status AD diagnostic model that can identify different statuses of AD, i.e., Normal Control (NC – no dementia), MCI (prodromal status of AD, which can be further categorized to p-MCI and s-MCI), and AD. We achieve this by training a three-stage Deep Neural Network (DNN) as shown in Fig. 1, using MRI, PET and SNP data. We train our networks stage-wise, since we want to learn the latent features from different combinations of modalities by best using all the available samples for each combination. The details of this framework are discussed below.

Stage 1

Independent single-modality deep feature learning. As described earlier, three modalities have different number of samples, e.g., the PET data has far less number of samples than numbers of MRI and SNPs. Our aim in this stage is to learn discriminative high-level features from each modality by using all its available training samples. We first employ DNN architecture on each independent modality, with the respective network architecture depicted as **Stage 1** in Fig. 1. In each DNN architecture, there are multiple fully-connected layers and one output layer (e.g., soft-max classifier). The output layer consists of three neurons for the case of three-class classification (i.e., AD/MCI/NC classification task), or four neurons for the case of four-class classification (i.e., AD/p-MCI/s-MCI/AD classification task). The label information from the training data at the output layer is also used to guide the learning of the network weights. After training, the outputs of hidden layers can be regarded as the high-level features for each of different modalities. Note that we use the maximum number of available samples for each modality in this stage, for effective training. For example, assume that we have N subjects, where only N_1 subjects contain MRI data, N_2 subjects contain PET data, and N_3 subjects contain SNP data. The conventional multimodality model uses only the subjects with all three modality data, which

is much less than $\min(N_1, N_2, N_3)$. On the other hand, by using our proposed framework, we can use all the N_1 , N_2 and N_3 samples to train three separate deep learning models for three modalities, respectively. It is expected that, by using more samples in training, our model can have better ability for feature representation.

Stage 2

Joint two-modality deep feature learning. As mentioned, different combinations of two modalities (i.e., MRI and PET, MRI and SNP, PET and SNP) are used to learn their joint or shared representations. The aim of this stage is to fuse complementary information from different modalities to further improve the performance of the model. The architecture used in this study is depicted in Fig. 2. There are a total of three DNN architectures, one for each pair of modalities. Specifically, the outputs from hidden layers in **Stage 1** are regarded as intermediate inputs in **Stage 2**, and the weights from **Stage 1** can be regarded as initial weights to initialize the DNN architecture for **Stage 2**. In addition, use three outputs to train each DNN architecture. Two of the outputs are used to guide the learning of high-level features from two different modalities, while the other is used to guide the learning of joint high-level features for the two modalities. Using three outputs simultaneously during the network training can balancing the high-level features learning at different level of modality combinations.

Stage 3

Joint three-modality deep feature learning and fusion. After training the networks in **Stage 2**, we can obtain joint representations between any pair of modalities, which can then be used to get a final prediction. The architecture used in this stage is depicted as **Stage 3** in Fig. 1. In **Stage 3**, we use the learned joint high-level features from **Stage 2** as input and the target labels as output. Obviously, in **Stage 3**, we can use only the samples with complete MRI, PET and SNP data to train the remaining network, and fine-tune the whole network. Afterward, we obtain the diagnostic label for each new sample, based on the output of the last classification layer (i.e., soft-max classifier in **Stage 3**).

3 Experimental Results and Analysis

3.1 Dataset

We use MRI, PET and SNP data from the ADNI dataset¹ for multi-status AD diagnosis. We use the baseline data from a total of 805 subjects, including 190 AD, 389 MCI and 226 normal controls (NC) subjects. All the subjects have MRI data, while only 736 subjects have SNP data, and 360 subjects have PET data. The MCI subjects can be further categorized into two subclasses, i.e., p-MCI and s-MCI, which are retrospectively determined by monitoring the disease status of MCI subjects after certain period of time (e.g., 24 months). However, since some MCI subjects dropped out of the study after the baseline scans, their sublabels (p-MCI or s-MCI) can not be determined. Hence, the total number of p-MCI (i.e., 205) and s-MCI (i.e., 157) subjects does not match with the total number of baseline MCI subjects. The detailed demographic information of the baseline subjects are summarized in Table. 1.

¹<http://www.loni.usc.edu/ADNI>.

After preprocessing steps (including skull stripping, cerebellum removal, intensity correction, tissue segmentation, etc.), the MRI and PET images were divided into 93 ROIs, and the gray matter volumes and the average PET intensity values of these ROIs were calculated as MRI and PET features, respectively. In addition, for SNP data, according to the AlzGene database², we only select SNPs that belong to the top AD gene candidates, and those selected SNP data have 3023 features.

3.2 Experimental Setup

In this section, we evaluate the effectiveness of the proposed deep feature learning and fusion framework by considering two classification tasks: (1) AD vs. MCI vs. NC (i.e., three-class classification task) and (2) AD vs. p-MCI vs. s-MCI vs. NC (i.e., four-class classification task). For each classification task, a twentyfold cross-validation technique was used for evaluation. Specifically, 5% of the subjects were randomly selected as testing samples, and the remaining subjects were used as training samples. 10% of the samples from the training set were used as validation set to select parameters of the networks. In order to reduce randomness, the process was performed 10 times. The final classification results were computed by averaging the results from all the experiments. In the first stage, we built a two-hidden-layer (i.e., 32-16) DNN for MRI and PET data, respectively, and a three-hidden-layer (i.e., 128-64-16) DNN for SNPs data. In the following stages, the outputs of the last hidden layer of the previous network were used as inputs, and a two-hidden-layer (i.e., 32-16) DNN was used for the all the modality combinations in the second and third stages. Furthermore, l_1 and l_2 regularizations were imposed on the weight matrices of the networks.

We compared the proposed framework with three popular dimension reduction methods, i.e., Principal Component Analysis (PCA) [7], Canonical Correlation Analysis (CCA) [5], and Lasso [10]. To fuse the three modalities, we concatenated the feature vectors of the multimodality data into a single long vector for the above three comparison methods. In addition, we also included SAE [13] as another deep feature learning method for comparison. For this method, we obtained SAE-learned features from each modality independently, and then concatenated all the learned features into a single long vector. As a baseline method, we further included the result for the experiment using just the original features without any feature selection (denoted as “Original”). We finally used SVM classifier from LIBSVM toolbox [2] to perform classification for all the above comparison methods. For each classification task, we use grid search to determine the best parameters for the feature selection algorithms and the classifier, based on their performances on the validation set. For instance, the best soft margin parameter C of SVM classifier was determined by grid searching from $\{10^{-4}, \dots, 10^4\}$.

Figure 3 shows the classification performance of all the competing methods for two multi-class classification tasks. It can be clearly seen that our proposed method outperforms all other methods, showing the advantage of deep feature learning and high-level feature fusion of our method. Comparing with another deep feature learning strategy, i.e., SAE, its learned high-level features did not perform as well as ours, probably due to the fact the SAE is an

²www.alzgene.org.

unsupervised feature learning method that did not consider label information. In addition, the good performance of our proposed framework could also be due to stage-wise feature learning strategy, which uses the maximum number of available samples for training. For fair comparison, we also include results for our proposed framework that only use samples with complete multimodality data, which are denoted as “Ours-c” in Fig. 3. From Fig. 3, we can see that our proposed framework that uses all available samples (“Ours”) still performs better than its degraded version “Ours-c”, which uses only complete multimodality data.

To further analyze the benefit of neuroimaging and genetic data fusion, Fig. 4 illustrates the performance of our proposed framework for different combinations of modalities. From Fig. 4, it can be seen that our proposed framework performs better when using MRI+PET data if compared with the other two data combinations (i.e., MRI+SNP, PET+SNP). However, when using all the three modalities, genetic data can further improve the classification performance.

4 Conclusion

In this paper, we present a novel three-stage deep learning framework to integrate multimodality neuroimaging and genetic data for multi-status AD diagnosis. We address the data heterogeneity issue in our framework by learning the latent feature representations for each modality before fusing them for AD diagnosis. We also use three-stage learning strategy to effectively learn the latent features for different modality combinations, using maximum number of available samples. Furthermore, small sample size issue due to incomplete dataset is addressed by using all available samples at the respective stages of the network learning. Experimental results using ADNI dataset have clearly demonstrated the effectiveness of our proposed framework in comparison with other state-of-the-art methods.

Acknowledgments

This research was supported in part by NIH grants EB006733, EB008374, EB009634, MH100217, AG041721 and AG042599.

References

1. Biffi A, Anderson CD, Desikan RS, et al. Genetic variation and neuroimaging measures in Alzheimer disease. *Arch. Neurol.* 2010; 67(6):677–685. [PubMed: 20558387]
2. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Tech. (TIST)*. 2011; 2(3):27.
3. Dai Z, et al. Discriminative analysis of early Alzheimer’s disease using multimodal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage*. 2012; 59(3):2187–2195. [PubMed: 22008370]
4. Fakoor, R., Ladhak, F., Nazi, A., Huber, M. ICML. 2013. Using deep learning to enhance cancer diagnosis and classification.
5. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* 2004; 16(12):2639–2664. [PubMed: 15516276]
6. Hinrichs C, et al. Predictive markers for ad in a multi-modality framework: an analysis of MCI progression in the adni population. *Neuroimage*. 2011; 55(2):574–589. [PubMed: 21146621]
7. Jolliffe, I. Principal Component Analysis. Wiley Online Library; 2002.

8. Lei B, Yang P, Wang T, Chen S, Ni D. Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. *IEEE Trans. Cybern.* 2017; 47:1102–1113. [PubMed: 28092591]
9. Lin D, Cao H, Calhoun V, Wang Y. Sparse models for correlative and integrative analysis of imaging and genetic data. *J. Neurosci. Methods.* 2014; 237:69–78. [PubMed: 25218561]
10. Liu, J., Chen, J., Ye, J. Large-scale sparse logistic regression. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM; 2009. p. 547-556.
11. Liu S, Liu S, Cai W, Che H, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* 2015; 62(4):1132–1140. [PubMed: 25423647]
12. Rastegar, S., Soleymani, M., Rabiee, H., Mohsen Shojaei, S. *CVPR*. 2016. MDL-CW: a multimodal deep learning framework with cross weights; p. 2601-2609.
13. Suk H, Lee S, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Struct. Funct.* 2015; 220(2):841–859. [PubMed: 24363140]
14. Thung KH, et al. Neurodegenerative disease diagnosis using incomplete multimodality data via matrix shrinkage and completion. *NeuroImage.* 2014; 91:386–400. [PubMed: 24480301]
15. Ye J, Farnum M, Yang E, et al. Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurol.* 2012; 12(1):46. [PubMed: 22731740]
16. Zheng, X., Shi, J., Li, Y., Liu, X., Zhang, Q. Multi-modality stacked deep polynomial network based feature learning for Alzheimer's disease diagnosis. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*; IEEE; 2016. p. 851-854.
17. Zhu X, et al. Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Trans. Biomed. Eng.* 2016; 63(3):607–618. [PubMed: 26276982]

Fig. 1.
Proposed diagnostic framework for AD diagnosis by deep feature learning.

Fig. 2.

Stage 2 includes training three DNN architectures for joint latent representation learning of any pair of modalities (i.e., MRI and PET, MRI and SNP, PET and SNP). Note that for each DNN architecture, there are three (soft-max) outputs which are used simultaneously to train the network.

Fig. 3. Comparison of classification accuracy for the three-class classification task (left) and the four-class classification task (right).

Fig. 4.

Comparison of classification accuracy of the proposed framework for AD/MCI/NC and AD/p-MCI/s-MCI/NC by using with different modality combinations.

Table 1

Demographic information of the baseline subjects in the study.

	Female/male	Education	Age	MMSE
NC	108/118	16.0±2.9	75.8±5.0	29.1±1.0
MCI	138/251	15.6±3.0	74.9±7.3	27.0±1.8
AD	101/89	14.7±3.1	75.2±7.5	23.3±2.0
Total	347/458	15.5±3.0	75.2±6.8	26.7±2.7