Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Bagging Ensembles for the Diagnosis and Prognostication of Alzheimer's Disease

Peng Dai, Femida Gwadry-Sridhar, Michael Bauer
Department of Computer Science
University of Western Ontario
and Robarts Research
London, ON, Canada
{pdai5, fgwadrys, bauer}@uwo.ca

Michael Borrie, for the ADNI*
Division of Geriatric Medicine
University of Western Ontario
London, ON, Canada
michael.borrie@sjhc.london.on.ca

Abstract

Alzheimer's disease (AD) is a chronic neurodegenerative disease, which involves the degeneration of various brain functions, resulting in memory loss, cognitive disorder and death. Large amounts of multivariate heterogeneous medical test data are available for the analysis of brain deterioration. How to measure the deterioration remains a challenging problem. In this study, we first investigate how different regions of the human brain change as the patient develops AD. Correlation analysis and feature ranking are performed based on the feature vectors from different stages of the pathologic process in Alzheimer disease. Then, an automatic diagnosis system is presented, which is based on a hybrid manifold learning for feature embedding and the bootstrap aggregating (Bagging) algorithm for classification. We investigate two different tasks, i.e. diagnosis and progression prediction. Extensive comparison is made against Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT) and Random Subspace (RS) methods. Experimental results show that our proposed algorithm yields superior results when compared to the other methods, suggesting promising robustness for possible clinical applications.

Introduction

Alzheimers disease (AD) is a genetically complex, progressive and fatal neurodegenerative disease. According to the 2014 World Alzheimer report, there are an estimated 44 million people worldwide living with dementia with about 60% of the cases having Alzheimers disease at a total cost of over US$605 billion (WHO 2012). The disease presents tremendous challenges to public health, health care delivery, social services, and to the family. It is well known that the initial AD pathology develops in situ while the patient is cognitively normal. At some point in time, sufficient brain damage accumulates to result in cognitive symptoms and impairment. Clinical diagnosis of AD often includes establishing the presence of dementia, amnesia and a deficit in one or more cognitive functions. AD is typically present with subtle recent memory loss and subsequent progressive memory impairment, followed by global cognitive decline and death. Increasingly, more evidence points to the fact that AD can be active several years prior to appearance of any signs of cognitive defects (WHO 2012). Consequently, there is a pressing need to identify AD at an earlier stage in order to ensure that treatment can be initiated early, where it may have a greater chance of success in delaying disease progression.

The diagnosis of AD is complex which involves a variety of medical tests including physical exam, neurological exam, mental status tests and brain imaging. Clinicians usually assess these factors based on personal experience and subjective assessment. The development of machine learning algorithms offers the potential for diagnostically relevant analysis techniques, which can assist the clinicians to better understand the information underlying the medical variables. Automatic diagnosis of AD can be formulated as a multiclass classification problem. It is particularly challenging due to the inherent difficulty in distinguishing between normal aging, mild cognitive impairment (MCI), and early signs of AD. Besides, with the advancement of medical technology, a large amount of medical test data are available for the analysis of AD, including Magnetic resonance imaging (MRI), Positron emission tomography (PET) scan, gene sequence, etc. The medical test data are multivariate heterogeneous, which makes comparison and analysis highly challenging.

In this paper, we propose a general framework for the diagnosis and (progression) prediction of Alzheimer's disease based on Bagging trees ensemble. A hybrid manifold learning algorithm is proposed for feature embedding, which is based on an optimal neighborhood similarity graph generated by the Probabilistic Global Distance Metric Learning (PGDM) (Xing et al. 2003). Thus, the multivariate arbitrary length data sequences are embedded into a manifold that has all the local properties of Euclidean space. Then, the bagging trees algorithm is used for classification, i.e. automatic diagnosis and AD progression prediction. Another important contribution of our work is the analysis of how various regions of the brain changes while the patient develops AD. Alzheimer's disease is closely related to the atrophy of various brain regions. Our analysis is based on both intra- and

*Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

inter-patient brain volume changes. The intra-patient analysis shows how the human brain changes as the same patient develops AD, while the inter-patient analysis helps to show the similarity between different patients.

The remainder of this paper is structured as follows. Firstly, we present a brief review of current development of automatic diagnosis system. Special focus is put on the application of manifold learning and ensemble algorithms in Alzheimer's Disease (AD). Then, we introduce the research problems followed by detailed description of our proposed algorithm. Finally we present detailed experiment setup as well as extensive comparison and draw conclusions.

Related Work

Automatic diagnosis of Alzheimer's disease can be formulated as a classification problem. For a straight forward approach, patients are represented as high dimensional feature points (vector) consisting of various medical test variables. Different machine learning algorithms can be used to classify the feature points into the corresponding categories, e.g. Normal, mild cognitive impairment (MCI), and Alzheimer's Disease (AD). AD causes progressive structural damage to the human brain. Magnetic resonance imaging (MRI) provides a chance to directly observe brain changes such as cerebral atrophy or ventricular expansion. Therefore, MRI images become one of the most widely used source of data. Some researchers try to extract brain atrophy information directly from the MRI images. Keraudren et al. proposed to localize the fetal brain in MRI using Scale-Invariant Feature Transform (SIFT) features (Keraudren et al. 2013). Another important approach is to extract brain volume based on a unified 3D brain model. An important examples in this category is the CIVET project from the McConnell Brain Imaging Centre (BIC) (Ad-Dabbagh et al. 2006; Sherif et al. 2014). In our current implementation, brain volune information together with genome and demographics (age, gender, education) forms the feature vector.

In order to effectively extract information from high dimensional data, manifold learning algorithms are usually utilized for latent space embedding. Conventional manifold learning refers to nonlinear dimensionality reduction methods based on the assumption that [high-dimensional] input data are sampled from a smooth manifold so that one can embed these data into the [low dimensional] manifold while preserving some structural (or geometric) properties that exist in the original input space (Ho, Dai, and Rudzicz 2015; Lin and Zha 2008). An appropriate manifold can help to reduce redundant features and prepare the data for further processing.

Generally speaking, manifold learning algorithms can be divided into two broad cagetories, i.e. unsupervised and supervised. Principal component analysis (PCA), for example, is one of the most widely used unsupervised manifold learning approach. PCA embeds the original data into a subspace, whose bases span along directions with largest variance (Lin and Zha 2008). Independent Component Analysis (ICA) and Multidimensional Scaling (MDS) are also examples of this category, which have been applied in Alzheimer's disease diagnosis by Yang et al. (Yang et al. 2011) and Park et al. (Park and Seo 2011), respectively. For unsupervised approaches, they are usually formulated using prior knowledge about the training data. For example, Xing et al. proposed to learn a robust distance metric by introducing the similar/dissimilar constraint (Xing et al. 2003). Other examples of this category include Wolz et al.'s neighborhood embedding approach (Wolz et al. 2010) and Sparse Bayesian Learning (Shen et al. 2010). After manifold learning, the high dimensional feature vectors (corresponding to various patients at different time steps) are embedded into a latent space, leading to a low dimensional representation. A classification algorithm is then applied for classification. A variety of pattern recognition algorithms have been applied in the biomedical area, such as Random Forest (Gray et al. 2013; Set 2013), SVM (López et al. 2011; Yang et al. 2011), and Artificial Neural Network (Keraudren et al. 2013).

Methods

Problem Formulation

As patients develop AD, the human brain shows gradual atrophy in various regions, which can be measured by the brain volume changes. Given the brain volume information of N_p participants, for each participant u, we measure the volumetric information of various brain regions, e.g. 3rd ventricle, 4th ventricle, right/left brain fornix, right/left globus pallidus, etc., denoted as $\mathbf{P}_u = \{f_1, f_2, \cdots\}$. This process is repeated every 6 months, which will form a $R \times F$ feature matrix, $\mathbf{B}_{R\times F}$, where F is the number of features obtained in each test and R is the number of repetition.

There are generally two problems in Alzheimer's disease research, i.e. diagnosis and prediction. Diagnosis helps to identify if the patient is cognitively normal, MCI or AD (see Problem 1). Prediction is to tell whether the patient will show cognitive decline or stay at the current mental status (see Problem 2). For example, if the patient is MCI, the prediction task is to determine if the patient will stay at MCI or further deteriorate to AD. In our current work, the prediction task includes three possible outcomes, i.e. (1) deteriorate later (in more than 6 months); (2) deteriorate immediately (in 6 months); (3) stay at current stage.

Problem 1 *(Diagnosis) Given the feature points,* $\mathbf{B}$*, how to decide the patient's mental status, i.e. Normal, Mild Cognitive Impairment (MCI), or Alzheimer's disease (AD)?*

Problem 2 *(Prediction) Given the feature points,* $\mathbf{B}$*, and diagnosis label,* D*, for a specific patient* u*, how to predict if the patient will stay at the same stage or deteriorate?*

Figure 1 shows the diagram of our proposed algorithm. Our proposed automatic diagnosis system mainly consists of three parts, i.e. hybrid manifold learning, bootstrap aggregating (Bagging), and decision trees (for each subset).

Hybrid Manifold Learning

Manifold learning algorithms are widely used in many areas for dimension reduction and effective feature embedding. An embedding is a representation of a (topological) object (such as a manifold or a graph) in a certain space, such that

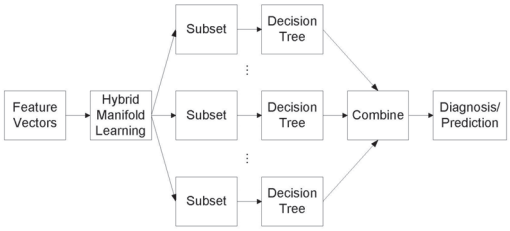


Figure 1: Schematic overview of the proposed methodology.

its (topological or structural) properties are preserved (Ho, Dai, and Rudzicz 2015; Lee and Verleysen 2007). Manifold learning algorithms have been applied in many different areas to find appropriate coordinate embeddings (Dai, Ho, and Rudzicz 2015; Cox and Cox 2001). It can be designed based on different criteria. For example, principal component analysis (PCA) is designed to find a manifold whose basis spans along the directions with the largest variance (Jolliffe 1997).

In our present application, the Neighborhood Preserving Embedding (NPE) algorithm is adopted for manifold learning. NPE is a linear approximation to Locally Linear Embedding (LLE) (Roweis and Saul 2000). It is designed to preserve the local neighborhood structure (Xiaofei He et al. 2005). K nearest neighbors (KNN) is utilized to construct a neighborhood graph. Let W denote the weight matrix with W_{ij} having the weight of the edge from node i to node j, and 0 if there is no such edge. The weights on the edges can be computed by minimizing the following objective function,

$$min \sum_i ||\mathbf{x}_i - \sum_j W_{ij}\mathbf{x}_j||^2 \tag{1}$$

Then, the NPE embedding can be calculated by solving the generalized eigenvector problem (Roweis and Saul 2000)

$$XMX^T\mathbf{a} = \lambda XX^T\mathbf{a} \tag{2}$$

where

$$X = (\mathbf{x}_1, \cdots, \mathbf{x}_m)$$
$$M = (I - W)^T(I - W)$$
$$I = diag(1, \cdots, 1)$$

In our proposed algorithm, the distance between different data points is calculated by the optimal Mahalanobis distance learned from Xing et al.'s metric learning framework. According to Xing (Xing et al. 2003), a proper metric can be calculated by:

$$\begin{aligned} \min_{\mathbf{A}} \quad & \textstyle\sum_{(x_i,x_j)\in\mathbf{S}} ||x_i - x_j||_A^2 \\ \text{s.t.} \quad & \textstyle\sum_{(x_i,x_j)\in\mathbf{D}} ||x_i - x_j||_{\mathbf{A}} \geq 1 \end{aligned} \tag{3}$$

where $||\cdot||_{\mathbf{A}}$ is the Mahalanobis distance metric; $\mathbf{A}$ is a positive semi-definite matrix; $\mathbf{S}$ and $\mathbf{D}$ are the similar and dissimilar training sets, respectively.

If $\mathbf{A}$ is diagonal, Equation (3) can be solved using the equivalent form,

$$\begin{aligned} g(\mathbf{A}) \quad = \quad & \sum_{(x_i,x_j)\in\mathbf{S}} ||x_i - x_j||_{\mathbf{A}}^2 \\ & - \log\left(\sum_{(x_i,x_j)\in\mathbf{D}} ||x_i - x_j||_{\mathbf{A}}\right), \end{aligned} \tag{4}$$

which can be solved using the Newton-Raphson method (Xing et al. 2003).

Bagging Ensembles

Bagging or Bootstrap aggregating is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms, which reduces variance and helps to avoid overfitting (Breiman 1996). It is usually applied to decision tree methods.

Given a training set $\mathbf{T}$ of size n, bagging generates m new training sets T_i, each of size n, by sampling from $\mathbf{T}$ uniformly. By sampling with replacement, some observations may be repeated in each T_i (Breiman 1996). This kind of sample is known as a bootstrap sample. The m models are fitted using the above m bootstrap samples and combined by averaging the output (for regression) or voting (for classification) (Breiman 1996). Bagging helps to reduce variance and thus improve performance. In our present implementation, we utilize bagging with decision trees for classification.

Results and Discussion

In this section, detailed descriptions about the database and experimental settings are presented. Extensive analysis is made to show how the brain degrades as the patient develops Alzheimer's disease. Two different sets of experiments are given to show the performance of our proposed algorithm. Comparison is made against state-of-the-art methods.

Data acquisition and pre-processing

Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials (Weiner et al. 2012).

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many

co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org [1].

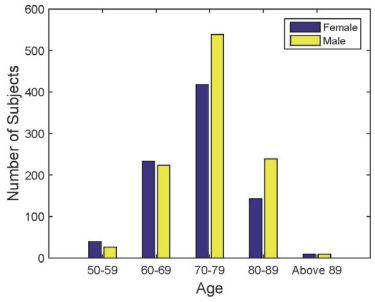


Figure 2: Participant distribution.

Figure 2 gives the participant distribution of ADNI. The Mini Mental State Examination (MMSE) score is used as an indicator for mental status. The Mini Mental State Examination (MMSE) is the most commonly used test for complaints of problems with memory or other mental abilities, which consists of a series of questions and tests, each of which scores points if answered correctly. The MMSE tests a number of different mental abilities, including a person's memory, attention and language (Weiner et al. 2012).

The multivariate heterogeneous clinical data available for analysis is high dimensional. For example, there are about 200 brain volumes extracted from each set of MRI images repeated every 6 months leading to 6∼8 repetition for each valid participant. This means that without further processing the MRI images results in more than 1000 features for a single participant. In our present implementation, the variables recorded at a specific time are formulated as a feature point.

Experimental setup

As described in previous sections, there are multiple feature points for the same patient corresponding to the patients' different visits to an ADNI site. After removing invalid entries, there are 2158 data points, with 586 normal records, 1006 MCI records, and 416 AD records. We randomly select 50 normal, 50 CI, and 50 AD data points as the testing set. The rest (i.e. 536 normal, 956 MCI, 366 AD) are left as the training set (2008 training points and 150 testing points). Comparison is made against Support Vector Machines (SVM) and three typical ensemble learning algorithms, i.e. Random Subspace (RS) with KNN (Ho 1998), Random Forest (Breiman 2001) and Decision Trees.

Relative improvement is defined as

$$R_{im} = \frac{r_p - r_t}{r_t} \times 100\% \tag{5}$$

where R_{im} is the relative improvement; r_p is the recognition rate of our proposed algorithm; r_t is the recognition rate of the comparison target.

For single patient AD progression analysis, we choose 10 patients from the entire dataset. The selection criteria are

1. There are at least 5 consecutive MRI scans for the patient;
2. The patient shows mental status conversion (e.g. Normal to MCI);
3. The mental status conversion should not happen at the last MRI scan.

The above mentioned requirements are to guarantee that there are enough meaningful data for analysis.

For the automatic diagnosis problem, the neuroimaging and biological data from 822 ADNI participants (229 normal patient, 405 MCI patients, and 188 AD patients) are chosen for verification tests. All MRIs were sagittal T1-weighted scans. The scans were collected using a 1.5 T GE Signa scanner with an MR-RAGE acquisition sequence. We excluded all invalid records (with missing/void feature entries), resulting in totally 2158 high dimensional data points.

The MRI images are processed by the CIVET (Ad-Dabbagh et al. 2006) through the CBRAIN platform (Sherif et al. 2014). CIVET is a human brain image-processing pipeline for fully-automated corticometric, morphometric and volumetric analyses of magnetic resonance (MR) images (Ad-Dabbagh et al. 2006). The MRI images are firstly transformed to stereotaxic space, followed by tissue classification, which are then registered to a unified brain model. Subsequently a series of brain volume information is extracted from the 3D brain model, which are used as discriminative features (Ad-Dabbagh et al. 2006). Based on different models used in brain model construction, various brain volumes can be calculated, such as the volume for the 3rd ventricle, 4th ventricle, right/left brain fornix, right/left frontal, right/left globus pallidus, right/left occipital, etc(Collins et al. 1994). For a complete list of brain volumes available please refer to (Zijdenbos, Forghani, and Evans 2002).

Other information provided in the ADNI dataset includes medical history, genome, and biospecimen. In our study, the Apolipoprotein E (ApoE) and medical history are used for further analysis. Apolipoprotein E (ApoE) is a class of apolipoprotein that is essential for the normal catabolism of triglyceride-rich lipoprotein constituents (Sadigh-Eteghad, Talebi, and Farhoudi 2012). The E4 variant of ApoE is the largest known genetic risk factor for late-onset sporadic Alzheimer's disease (AD) in a variety of ethnic groups (Sadigh-Eteghad, Talebi, and Farhoudi 2012).

Experimental results

Feature Analysis There are a large number of features corresponding to different time steps of the patients. Since

[1] More details please refer to Acknowledgement section

genome and demographics (age, gender, education) do not change over time, we mainly focus on the brain volumes extracted from MRI images. In our present implementation, four different sets of brain volumes are extracted based on the CIVET software through the CBrain portal, i.e. ANIMAL segmentation (Collins et al. 1995), DKT surface parcellation (Klein and Tourville 2012), and AAL surface parcellation (Tzourio-Mazoyer et al. 2002).

Table 1 shows the paired t-test results of the feature vectors derived from the five different visits of a sample patient (#0195). As the the patient develops AD there are significant changes in various regions of brain. Moreover, the feature vectors corresponding to the first 3 visits show relatively less difference, while the scans from last two visits are significantly different from the rest. This indicates that the patient's brain deteriorates at an increasing speed after the diagnosis of AD, i.e. accelerated deterioration.

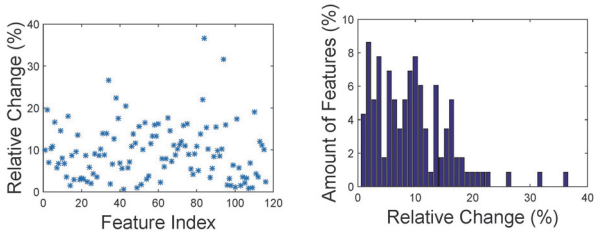
Table 1: Significance test (p-value) based on feature vectors from different visits.

	1st Vis.	2nd Vis.	3rd Vis.	4th Vis.	5th Vis.
1^{st} V.	1	-	-	-	-
2^{nd} V.	0.3	1	-	-	-
3^{rd} V.	0.6	0.1	1	-	-
4^{th} V.	10^{-6}	10^{-8}	10^{-6}	1	-
5^{th} V.	10^{-9}	10^{-5}	10^{-11}	10^{-23}	1
Diag.	Normal	Normal	AD	AD	AD

The results in Table 1 describe how the feature vectors as a whole change as the patient develops Alzheimer's disease. However, how individual feature change is also of vital importance. Figure 3(a) gives the relative change results (defined in Equation (6)). For visualization purposes, only the relative changes in terms of ANIMAL and AAL segmentations (totally 116 points) are given.

$$C_r = \left| \frac{V_n - V_o}{V_o} \right| \times 100\% \tag{6}$$

where $|\cdot|$ is the absolute value operator, V_n is the new volume, and V_o is the original volume.

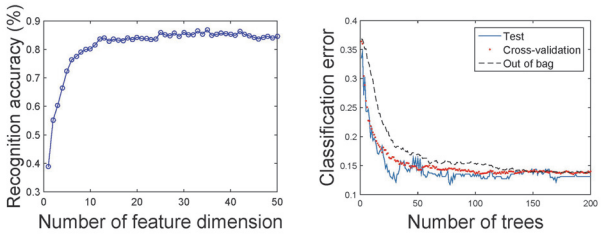


(a) Relative change (b) Histogram

Figure 3: The brain volumes (i.e. features) change.

It can be seen that different features show different patterns as the patient develops AD. The 84th and 94th feature show large relative change ($> 30\%$). They correspond to the opercular part of right inferior frontal gyrus and right median cingulate together with paracingulate gyri. On the other hand, most of the features show small relative change. For exmaple, the relative changes for 56.9% of the features are less than 10%. In addition, about 3.44% of the features nearly stayed unchanged ($C_r < 1\%$). The 42nd feature shows the least change (0.51%) corresponding to left middle frontal gyrus orbital part.

Automatic Diagnosis Our proposed algorithm mainly consists of two parts, manifold learning and bagging ensemble classification. Since the medical variables are high dimensional with strong correlation, we perform dimension reduction during the manifold learning step, i.e. taking only the leading dimensions for classification.



(a) Dimension optimization (b) Performance

Figure 4: Classification results for different conditions.

Figure 4(a) shows the exprimental results as we change the dimension. It can be seen that at very small feature dimension the classification accuracy (i.e. diagnosis accuracy) is very low. As the number of selected dimension increases, the classification accuracy significantly improves, from below 40% to over 80% at 10 selected dimensions. The classification accuracy further improves as the dimension increases. The optimal result is achieved at dimension 35. Table 2 gives the confusion matrix for our proposed algorithm. It can be seen that our proposed algorithm yields an average accuracy of 86.0%. In particular, for Normal and AD, our proposed algorithm yields very promising results, 91.4% and 93.1%. The false positive rate for AD is 35.1%. However, AD is mostly misclassified as MCI. Our proposed approach is mainly based on medical imaging. In terms of medical imaging the transition from MCI to AD is a gradual process. There is no clear border line between late MCI and early AD. The relative high false positive rate is caused by the subjective diagnosis label provided by clinicians. Moreover, with additional medical variables, e.g. Mini Mental State Examination (MMSE), the false positive rate between MCI and AD can be improved. For MCI, the classification accuracy is 81.7%. However, the misclassification rate is 5.7%. This means that our proposed algorithm can satisfyingly identify most of MCI patients, 94.3%. Figure 4(b) shows the performance of our proposed algorithm based on independent test set, cross validation and out-of-bag data. It can be seen that with more than 140 trees the experimental results converge. Our proposed algorithm shows comparable results in cross validation, test set and out-of-bag data.

Table 2: Confusion Matrix.

		Target Class			Accuracy
		Normal	MCI	AD	
Pred. Class	Normal	**508**	38	10	91.4%
	MCI	77	**949**	136	81.7%
	AD	1	19	**270**	93.1%
sensitivity		86.7%	94.3%	64.9%	86.0%

Table 3: Recognition results for comparison targets (%).

	Proposed	SVM	RS	RF	DT
Accuracy	86.0	83.83	85.93	84.21	67.37
Rel. Imp.	-	2.59	0.08	2.13	27.65

Table 4: Confusion Matrix.

		Target Class		Rate (Accuracy)
		Decline	Stay	
Predicted Class	Decline	**49**	1	98.0%
	Stay	9	**366**	97.6%
Rate (sensitivity)		84.5%	99.7%	97.7%

Extensive comparison is made against Support Vector Machine (SVM), Random Subspace (RS), Random Forest (RF) and Decision Tree (DT) methods. The raw data vectors are processed by mean and variance normalization before being processed by each of the above mentioned algorithms. Table 3 gives the experimental results for all the comparison targets.

It can be seen that our proposed algorithm yields much better results than all the comparison targets. The relative improvements are 2.59% over SVM, 0.08% over RS, 2.13% over RF, and 27.65% over DT. Our proposed algorithm shows consistently better results than all the comparison targets.

AD progression prediction

Another important problem in Alzheimer's disease research is the prediction of AD progression (or cognitive decline), i.e, to predict whether a patient is to stay at the current mental status or deteriorate to the next stage (Problem 2). It has to be noted that AD is the final stage of our study. Therefore, AD patients cannot convert to other stages. In our cognitive decline prediction experiments, AD patients are excluded.

Because the conversion happens mostly in patients at late stages of MCI, prediction of cognitive decline or conversion becomes a binary prediction between normal and late MCI (or early AD). When applied to cognitive decline classification, our proposed algorithm gets very promising results. Table 4 gives the confusion matrix. It can be seen that our proposed algorithm achieves a nearly perfect result, 97.6%. This is consistent with the results in Table 2. The binary classification task between Normal and AD also yields an average accuracy of over 95%.

Table 5: Confusion Matrix.

		Target Class			Rate (Accuracy)
		$>$ 6m	6m	Stay	
Predicted Class	$>$ 6m	**14**	18	1	42.4%
	6m	15	**2**	0	11.8%
	Stay	4	5	**366**	97.6%
Rate (sensitivity)		42.4%	8.0%	99.7%	86.1%

Because in ADNI patients are examined every 6 months, it is possible to predict whether a patient will move to the next stage of AD in a specific timeline. For a more practical approach, the possible prediction is set to (1) conversion in more than 6 months; (2) conversion in 6 month; (3) stay in current stage. Table 5 shows the confusion matrix. The accuracy for 'Stay' is 97.6%. However, the accuracies for predicting the timeline is very low, 42.4% for long term decline and 11.8% for immediate decline. It has to be noted that in most of the cases our proposed algorithm can successfully identify if a patient will deteriorate to the next stage but cannot predict when conversion will happen. This is mainly due to the lack of training materials. Theoretically, one patient can at most have 2 transition, i.e. Normal/MCI and MCI/AD. In practice, the study subjects usually have only 1 transition during the study. This is because the patient is already at the MCI stage when entering the study. It can be seen from Table 4 that there are only 58 conversion data (33 immediate conversion and 25 delayed conversion) compared with 367 stay data.

Conclusion

This study suggests volumetric changes show different patterns for different sides of the brain. Bagging decision trees can successfully classify the patients into the corresponding diagnosis group. When it comes to cognitive decline prediction, our proposed algorithm can satisfyingly identify the risk of deterioration (97.6%) but cannot predict the exact conversion time (11.8%). Future work will be focused on the prediction of conversion.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). Please refer to http://adni.loni.usc.edu/ for more details.

References

Ad-Dabbagh, Y.; Einarson, D.; Lyttelton, O.; Muehlboeck, J.-S.; Mok, K.; Ivanov, O.; Vincent, R.; Lepage, C.; Lerch, J.; Fombonne, E.; and Evans, A. 2006. The CIVET Image-Processing Environment: A Fully Automated Comprehensive Pipeline for Anatomical Neuroimaging Research. In *12th Annual Meeting of the Organization for Human Brain Mapping*.

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24(2):123–140.

Breiman, L. 2001. Random Forests. *Machine Learning* 45(1):5–32.

Collins, L.; Neelin, P.; Peters, T.; and Evans, A. 1994. Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space. *Journal of Computer Assisted Tomography* 18:192 – 205.

Collins, D. L.; Holmes, C. J.; Peters, T. M.; and Evans, A. C. 1995. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping* 3(3):190–208.

Cox, T. F., and Cox, M. A. 2001. *Multidimensional Scaling*. Chapman and Hall/CRC, 2nd edition.

Dai, P.; Ho, S.-S.; and Rudzicz, F. 2015. Sequential behavior prediction based on hybrid similarity and cross-user activity transfer. *Knowledge-Based Systems* 77:29–39.

Gray, K. R.; Aljabar, P.; Heckemann, R. A.; Hammers, A.; and Rueckert, D. 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* 65:167–75.

Ho, S.-S.; Dai, P.; and Rudzicz, F. 2015. Manifold Learning for Multivariate Variable-Length Sequences With an Application to Similarity Search. *IEEE transactions on neural networks and learning systems* PP(99):1.

Ho, T. K. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8):832–844.

Jolliffe, I. T. 1997. *Principal Component Analysis*. Springer-Verlag, New York, Inc.

Keraudren, K.; Kyriakopoulou, V.; Rutherford, M.; Hajnal, J. V.; and Rueckert, D. 2013. Localisation of the brain in fetal MRI using bundled SIFT features. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 16(Pt 1):582–9.

Klein, A., and Tourville, J. 2012. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience* 6:171.

Lee, J. J. A., and Verleysen, M. 2007. *Nonlinear dimensionality reduction*. Springer.

Lin, T., and Zha, H. 2008. Riemannian manifold learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30(5):796–809.

López, M.; Ramírez, J.; Górriz, J.; Álvarez, I.; Salas-Gonzalez, D.; Segovia, F.; Chaves, R.; Padilla, P.; and Gómez-Río, M. 2011. Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. *Neurocomputing* 74(8):1260–1271.

Park, H., and Seo, J. 2011. Application of multidimensional scaling to quantify shape in Alzheimer's disease and its correlation with Mini Mental State Examination: a feasibility study. *Journal of neuroscience methods* 194(2):380–5.

Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–6.

Sadigh-Eteghad, S.; Talebi, M.; and Farhoudi, M. 2012. Association of apolipoprotein E epsilon 4 allele with sporadic late onset Alzheimer's disease. A meta-analysis. *Neurosciences (Riyadh, Saudi Arabia)* 17(4):321–6.

2013. Random Forest and Gene Networks for Association of SNPs to Alzheimers Disease. *Advances in Bioinformatics and Computational Biology* 8213.

Shen, L.; Qi, Y.; Kim, S.; Nho, K.; Wan, J.; Risacher, S. L.; and Saykin, A. J. 2010. Sparse bayesian learning for identifying imaging biomarkers in AD prediction. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 13(Pt 3):611–8.

Sherif, T.; Rioux, P.; Rousseau, M.-E.; Kassis, N.; Beck, N.; Adalat, R.; Das, S.; Glatard, T.; and Evans, A. C. 2014. CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Frontiers in neuroinformatics* 8:54.

Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; and Joliot, M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15(1):273–89.

Weiner, M.; Veitch, D.; Aisen, P.; Beckett, L.; Cairns, N.; Green, R.; Harvey, D.; Jack, C.; Jagust, W.; Liu, E.; Morris, J.; Petersen, R.; Saykin, A.; Schmidt, M.; Shaw, L.; Siuciak, J. A.; Soares, H.; Toga, A.; and Trojanowski, J. 2012. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 8(1 Suppl):S1–68.

WHO. 2012. Dementia: a public health priority. Technical report.

Wolz, R.; Aljabar, P.; Hajnal, J. V.; Hammers, A.; and Rueckert, D. 2010. LEAP: learning embeddings for atlas propagation. *NeuroImage* 49(2):1316–25.

Xiaofei He; Deng Cai; Shuicheng Yan; and Hong-Jiang Zhang. 2005. Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, 1208–1213 Vol. 2. IEEE.

Xing, E.; Ng, A.; Jordan, M.; and Russell, S. 2003. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems*, volume 15.

Yang, W.; Lui, R. L. M.; Gao, J.-H.; Chan, T. F.; Yau, S.-T.; Sperling, R. A.; and Huang, X. 2011. Independent component analysis-based classification of Alzheimer's disease MRI data. *Journal of Alzheimer's disease : JAD* 24(4):775–83.

Zijdenbos, A. P.; Forghani, R.; and Evans, A. C. 2002. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE transactions on medical imaging* 21(10):1280–91.