Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Healthy Cognitive Aging: A Hybrid Random Vector Functional-Link Model for the Analysis of Alzheimers Disease

Peng Dai,[1,2,3] **Femida Gwadry-Sridhar,**[1,3] **Michael Bauer,**[1] **Michael Borrie,**[4] **Xue Teng**[3]
for the Alzheimers Disease Neuroimaging Initiative (ADNI)*
[1]Department of Computer Science, University of Western Ontario, London, ON, Canada
[2]Robarts Research, London, ON, Canada
[3]Pulse Infoframe Inc., London, ON, Canada
[4]Division of Geriatric Medicine, University of Western Ontario, London, ON, Canada
peng.dai.ca@ieee.org, {fgwadrys, bauer}@uwo.ca, michael.borrie@sjhc.london.on.ca, xt.biomedical@gmail.com

Abstract

Alzheimer's disease (AD) is a genetically complex neurodegenerative disease, which leads to irreversible brain damage, severe cognitive problems and ultimately death. A number of clinical trials and study initiatives have been set up to investigate AD pathology, leading to large amounts of high dimensional heterogeneous data (biomarkers) for analysis. This paper focuses on combining clinical features from different modalities, including medical imaging, cerebrospinal fluid (CSF), etc., to diagnose AD and predict potential progression. Due to privacy and legal issues involved with clinical research, the study cohort (number of patients) is relatively small, compared to thousands of available biomarkers (predictors). We propose a hybrid pathological analysis model, which integrates manifold learning and Random Vector functional-link network (RVFL) so as to achieve better ability to extract discriminant information with limited training materials. Furthermore, we model (current and future) cognitive healthiness as a regression problem about age. By comparing the difference between predicted age and actual age, we manage to show statistical differences between different pathological stages. Verification tests are conducted based on the Alzheimers Disease Neuroimaging Initiative (ADNI) database. Extensive comparison is made against different machine learning algorithms, i.e. Support Vector Machine (SVM), Random Forest (RF), Decision Tree and Multilayer Perceptron (MLP). Experimental results show that our proposed algorithm achieves better results than the comparison targets, which indicates promising robustness for practical clinical implementation.

Introduction

According to the 2015 World Alzheimer report, there are an estimated 46 million people worldwide living with dementia at a total cost of over $818 billion, which is estimated to increase to a trillion dollar by 2018 (Alzheimers Disease International, 2015). Alzheimers disease (AD) is one of the most common causes for dementia, accounting for about 60% of the total. The disease presents a tremendous burden and challenge to public health, health care delivery, social services and the family (Alzheimers Disease International, 2015). AD usually develops in situ while the patient is cognitively normal. At some point in time, sufficient brain damage accumulates to result in cognitive symptoms, which may further deteriorate to disability and ultimately death. There is currently no effective cure to reverse the damages caused by Alzheimer's. Treatments are mainly to ease cognitive symptoms, delay progression and improve quality of life via assistive technologies. Therefore, it is crucial to diagnose or predict AD as early as possible so as to allow treatments start early, which helps patients to maintain cognitive functionality.

Clinical diagnosis of AD often includes establishing the presence of dementia, amnesia and a deficit in one or more cognitive functions, such as skilled movements (limb apraxia), language (aphasia) or executive function (e.g., planning, attention and abstract reasoning)(Scott and Barrett 2007; American Psychiatric Association, 2013). The diagnosis process is complex, which involves a number of assessments, e.g. medical history review, physical examination, neurological examination, cognitive testing, laboratory testing and brain imaging. Physicians usually evaluate the above mentioned tests based on experience with quantitative guidelines. It is very challenging especially for early AD patients without clear cognitive symptoms.

With recent advances in artificial intelligence, evidence has shown that effective application of machine learning algorithms can greatly improve the efficiency of many tasks. Machine learning offers valuable tools for advanced diagnostic techniques, which can assist the clinicians to better understand the information underlying the high dimensional heterogeneous medical variables. Diagnosis of AD can be formulated as a classification problem. The problem is particularly challenging due to the inherent difficulty in distinguishing between normal aging, mild cognitive impairment (MCI), and early signs of AD. For example, patients with dementia may not complain of cognitive difficulty owing to

*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. A complete list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

loss of self-awareness, while patients with depression often complain of memory difficulties and seek medical attention of their own initiative (Scott and Barrett 2007).

Because of the fast development of medical technology, a number of medical tests have been developed for AD analysis, which yields a large amount of high dimensional heterogeneous data. However, due to privacy and legal issues of clinical research, the study cohort (i.e. the number of available patients) is relatively small. If well trained, state-of-the-art machine learning algorithms, e.g. deep learning, can usually achieve very good performance (Hinton et al. 2012), which on the other hand require a large amount of training materials, i.e. large cohort. Therefore, we have to balance between algorithm complexity and required data. In this paper, we propose a hybrid system which combines manifold learning and random vector functional link network (RVFL) to achieve better ability to capture high dimensional nonlinear information from clinical data. Different from traditional Artificial Neural Network (ANN), RVFL sets most of the parameters completely at random, which do not need to be tuned during training. Besides, manifold learning is integrated as part of the system, which helps to construct a better representation of the high dimensional heterogeneous data. Combined with manifold learning, RVFL is able to obtain a satisfying approximation of the original problem. It is particularly suitable for problems where only limited data are available.

Related Work

Alzheimer's disease causes progressive damages to the human brain, causing massive brain cell death and thus atrophy in various brain regions. Magnetic resonance imaging (MRI) techniques utilize strong magnetic fields to form anatomical images of the body, which provides a valuable tool to directly observe brain changes such as cerebral atrophy or ventricular expansion. Therefore, MRI has become one of the most widely used means to assist AD diagnosis. A large amount of work has been done about applying image processing techniques to MRIs. For example, Keraudren et. al. proposed to use Scale-Invariant Feature Transform (SIFT) to analyze brain atrophy (Keraudren et al. 2013). Another important approach is to establish 3D brain model and extract volumetric information of various brain regions. Freesurfer is one of the most commonly used package for the analysis and visualization of structural and functional neuroimaging data (Dale, Fischl, and Sereno 1999). In our current implementation, brain volume together with genome and demographics (age, gender, education) forms the feature vector.

The diagnosis of Alzheimer's disease can be formulated as a classification problem, where the clinical diagnosis can serve as labels and the high dimensional medical variables can serve as features. Therefore, a number of related work have been reported during the past decades. Lebedev et. al. utilized the random forest algorithm for AD diagnosis (Lebedev et al. 2014). Lopez et. al. utilized support vector machine (SVM) to detect early signs of AD (López et al. 2011). Feature selection algorithms, e.g. statistical significance, are widely used for dimension reduction in clinical studies. Recently, manifold learning algorithms are introduced to relevant studies. 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 (Lin and Zha 2008). Instead of removing redundant feature dimensions, manifold learning algorithms construct a low dimensional representation based on the original data. Lopez et al. implements PCA as part of their system(López et al. 2011). Dai et. al. proposed an improved isometric mapping algorithm for feature embedding and utilized ensemble learning algorithms for similar tasks (Dai et al. 2015; 2016a).

Recently, deep learning has become one of the most powerful machine learning techniques, which has shown superior performance in various practical applications, e.g. natural language processing and image recognition (Hinton et al. 2012; LeCun, Bengio, and Hinton 2015). Inspired by its promising performance, researchers have been trying to implement deep learning in dementia research. Li et al. proposed a hybrid system which combined principal component analysis (PCA) and deep learning autoencoder to extract discriminative features for AD diagnosis (Li et al. 2015). Payan et al. proposed a deep learning algorithm based on 3D convolution of MRI images (Payan and Montana 2015). Dai et. al. utilized multilayer perceptron (MLP) for AD diagnosis and prognosis (Dai et al. 2016b). It has to be noted that all the above mentioned work mainly implement a relative 'easy' or 'shallow' version of the deep learning algorithms. There are only a small number of hidden layers and barely show any complex structures, e.g. convolution layer. This is mainly due to the fact that the study cohort is relatively small compared with the imaging database used in deep learning studies. Therefore, we have to balance between algorithm complexity and the issues caused by limited training data. In this paper, we study Random vector functional link network (RVFL), in which only the output weights are chosen as adaptable parameters, while the remaining parameters are constrained to random values independently selected in advance (Husmeier 1999). RVFL simplifies the artificial neural network as a linear regression problem on top of a series of randomly assigned transition functions (hidden layers), which is an efficient approximation of the original nonlinear optimization problem.

Methods

Problem Formulation

As the patient develops AD, there are pathological changes in various regions of the human brain, which can be measured by volumetric changes. Besides, medical history, laboratory testing, physical examination and cognitive testing are all closely related to the final diagnosis (American Psychiatric Association, 2013). All those medical variables forms the original feature vector, $\mathbf{f}$. There are totally N_p participants in the study. For each participant u, the diagnostic analysis is repeated (follow-up medical tests) every 6

months, which will form a $r \times n$ feature vector, $\mathbf{f}_{1\times(r\cdot n)}$, where n is the number of features obtained in each test and r is the number of follow-up tests.

There are generally two problems in Alzheimer's disease research, i.e. diagnosis and prognosis. Diagnosis intends to identify if the patient is cognitively normal or AD (see Problem 1). Prognosis is to tell how the patient will evolve in AD pathology (see Problem 2). For example, if the patient is currently healthy, the prognosis task is to determine if the patient will stay healthy or likely develop AD.

Problem 1 *(Diagnosis) Given different patients, described as feature* $\mathbf{B}$*, how to decide the patient's pathological status, e.g. Healthy, Mild Cognitive Impairment (MCI), or Alzheimer's disease (AD)?*

Problem 2 *(Prognosis) Given a patient, described as* $\mathbf{B}$*, and his/her historical mental status label,* D*, how to predict if the patient will stay at the same stage or progress in the pathological path?*

The aging process can be understood as an interactive process between the human body and the environment, described as a sequence of medical variables. Diagnosis is to reveal the current cognitive status and thus a classification problem. For prognosis, a straight forward approach is to construct a time series model, e.g. Hidden Markov Model, to capture the temporal evolution trajectory. Nevertheless, due to the lack of valid data, it's very difficult to fully train advanced time series models. We formulate the prognosis problem as a classification problem based on the sequence of clinical diagnosis from ADNI. The prognosis is generated as 'progression' and 'no progression'. In the prognosis task, we group the current and preceding observation, $\{\mathbf{B}_{t,f}, \mathbf{B}_{t-1,f}\}$, to form the new feature vector so as to account for the temporal cognitive changes.

Problem 3 *(Healthy Aging) Given different patients, described as feature* $\mathbf{B}$*, what causes some people to develop AD while other people remain healthy?*

We investigate a third problem about healthy aging within our proposed framework. While aging, various functions of the human body gradually degrades. The healthy aging problem intends to explain the difference that features various aging pathways, i.e. healthy or dementia. Figure 1 shows the diagram of our proposed system. Our proposed system mainly consists of two parts, i.e. manifold learning and Random vector functional link network (RVFL), which will be discussed in detail in the following sections.

Manifold Learning

The available clinical data are high dimensional heterogeneous, which are obtained from different sources, e.g. medical imaging and blood tests. It is of vital importance to preprocess the data so as to remove noise, normalize scaling factors, etc. Another important step is to reduce feature dimension. Because of the curse of dimensionality, the data required to fully represent the hidden mechanism increase exponentially as the feature dimension increases. Therefore, dimension reduction is one of the most simple and effective way to boost system performance.

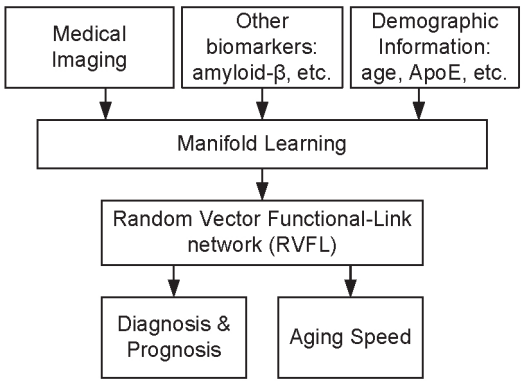


Figure 1: Schematic overview of the proposed methodology.

Manifold learning algorithms are designed to construct a low dimensional representation, which preserves the topological or structural properties of the original data (Lee and Verleysen 2007; Ho, Dai, and Rudzicz 2016). In this paper, we compare the performance of different manifold learning algorithms in our RVFL based framework, including Principal Component Analysis (PCA), Neighborhood Preserving Embedding (NPE) (Xiaofei He et al. 2005), Locality Preserving Projections (LPP) (Xiaofei He 2003) and stochastic proximity embedding (SPE). NPE and SPE are based on neighborhood graph. Our experimental results show that LPP shows better performance in our current framework.

Locality Preserving Projection (LPP) is a linear approximation of the nonlinear Laplacian Eigenmap (Belkin and Niyogi 2001). A neighborhood graph is firstly constructed with weights defined as

$$W_{i,j} = e^{-\frac{\|f_i - f_j\|}{t}} \tag{1}$$

where f_i is the feature vector for different patients

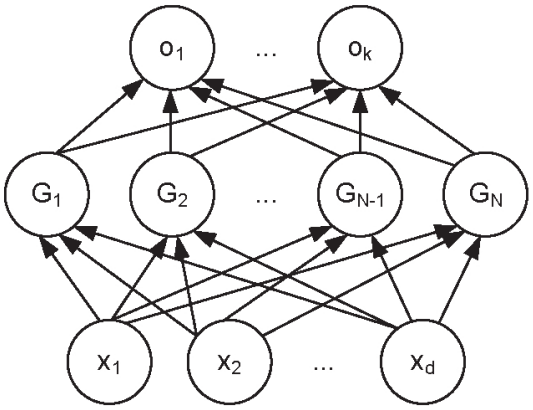
Then, LPP embedding can be calculated as the generalized eigenvector problem

$$\mathbf{FLF}^T\mathbf{a} = \lambda\mathbf{FDF}^T\mathbf{a} \tag{2}$$

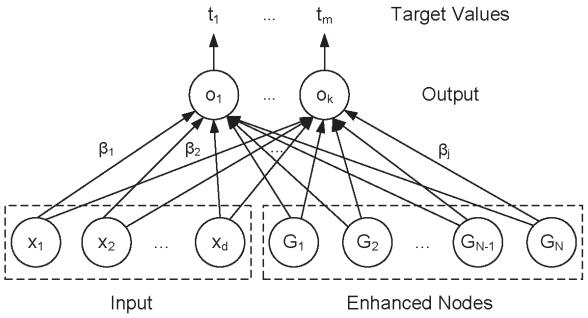
where $\mathbf{D}$ is a diagonal matrix whose entries are column sums of $\mathbf{W}$, $D_{ii} = \Sigma_j W_{ji}$. $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix.

Random vector functional-link (RVFL) network

The idea of functional link network was suggested by Pao and co-workers in 1988 (Klassen, Pao, and Chen 1988). A typical artificial neural network consists of a linear link of inputs together with nonlinear activation functions, while Pao and co-workers suggested that a link can also be nonlinear. In a semi-linear net, the pattern vector at any layer is multiplied linearly by a matrix of link weights to yield the vector input to the next layer. Pao suggested that a nonlinear functional transform be carried out along a nonlinear functional link to yield a new pattern vector in a larger space (Klassen, Pao, and Chen 1988). Since functional link network incorporates nonlinearity by variations of additional input nodes, generally 'flat' nets with no hidden layers are



(a) Hidden Layer Neural Network: d Input Nodes, N hidden nodes, k output Nodes.



(b) Random vector functional link network (RVFL): d Input Nodes, N Random Neurons (equivalent to hidden nodes), k output Nodes.

Figure 2: Diagram of (a) hidden-layer net and (b) functional-link network architectures.

sufficient for most of practical tasks (Klassen, Pao, and Chen 1988).

Random vector functional link network (RVFL) is one of the practical implementations of functional link network. It's a multilayer perceptron (MLP) in which only the output weights are chosen as adaptable parameters, while the remaining parameters are constrained to random values independently selected in advance (Husmeier 1999). Standard single layer neural network can be modeled as

$$\mathbf{o}_i = \sum \beta_j G(\mathbf{A}_j \mathbf{x} + b_j) \tag{3}$$

The random-vector version of the functional-link net generates $\mathbf{A}$ and $\mathbf{b}$, randomly, and learn only β.

Figure 2(b) shows the net architecture of a single hidden layer RVFL net. Although random vector (or feature) have been generally believed to be less powerful than learned features, it has shown reasonably success in many practical applications (Saxe et al. 2011; Rahimi and Recht 2008). Recently, Huang et. al. further improved RVFL to extreme learning machine (ELM), which achieves very promising results with simple network structure (Huang, Zhu, and Siew 2006). Random vector can significantly simplify the algorithm complexity, since a large amount of the parameters are randomly selected and do not need to be tuned. The RVFL can also be considered to consist of input and output layers, but no hidden layers. An input layer has enhanced input values which are created by various functional links with original input values.

The key advantage of RVFL related approaches is the ability to obtain promising results with limited data, where state-of-the-art algorithms, e.g. deep learning, probably cannot properly trained. In our present implementation, we use the Extreme Learning Machine (ELM) version of RVFL [1].

Healthy Aging

Although people are in different aging path (healthy or dementia), there will always be degradation in various functions of daily living. Patients with dementia 'moves faster' in the aging process than the healthy aging counterparts. Therefore, a straight forward approach to evaluate the cognitive health is to study the 'pathological' aging status.

A regression model is constructed based on RVFL to estimate the patient's age.

$$A = f(\mathbf{B}) \tag{4}$$

where $\mathbf{B}$ is the feature matrix consisting relevant medical variables. The model is trained using the healthy participants. We assume the predictive power of our proposed algorithm is reasonably well. Therefore, the predicted age can reflect the actual status of the patient's brain. When applied to an AD patient, it reflects how old the patient should be if he/she is healthy.

Then, we study the difference between the predicted age and the actual age.

$$A_{dif} = A_p - A_{real} \tag{5}$$

where A_{real} is the actual age; A_p is the predicted age. Based on our study, A_{dif} follows Normal distribution. It reflects the aging speed of the target patient and shows different statistical properties for different AD stages. More details will be given in the results section.

Results and Discussion

In this section, detailed descriptions about the database and experiment settings are presented. Extensive comparison is made to show how the performance of our proposed algorithm.

Data acquisition and pre-processing

Verification tests are performed based on 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

[1]The implementation codes are provided by the authors at https://github.com/dclambert/Python-ELM (Huang, Zhu, and Siew 2006) .

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). (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. For up-to-date information, see www.adni-info.org .

Experiment setup

As described in previous sections, there are multiple feature points for the same patient corresponding the patients' different visits to ADNI site. After removing invalid entries, there are totally 2158 data points, with 636 Healthy Control (HC) records, 1056 MCI records, and 466 AD records. Ten fold cross validation is used in our experiments. Comparison is made against Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (Breiman 2001) and Decision Tree. The optimal result of MLP is achieved with 2 hidden layers. SVM is implemented with Radial basis function (RBF) kernel.

Experimental results

Aging Speed

Alzheimer's disease is a geriatric disease, and thus age is a strong risk factor of the disease. Since aging is (to the best of our knowledge) is inevitable, what makes the difference is aging speed. Aging speed is defined as the difference between structural age (predicted age) and the demographic age (actual age). Higher aging speed indicates more likelihood (or vulnerability) to dementia. We fit the proposed algorithm into a regression task to estimate the patient age based on the healthy control set. Then, we calculate the estimation difference, A_{dif} in Equation (5). Figure 3 shows the results. It can be seen that at younger age (e.g. < 70) the predicted age tends to be smaller than the actual age, while at older age (e.g. > 70) the predicted age tends to be larger than the actual age. This is mainly due to the fact that age is a strong indicator of AD. Therefore, there are more occurrences in the senior population. In our current experiment settings, $mean(A_{dif}) = -1.33$ and $std(A_{dif}) = 8.72$. Based on normality tests, i.e. Kolmogorov-Smirnov test and Anderson-Darling test, the difference between predicted age and real age can be modeled by a normal distribution.

Mean estimation difference, $mean(A_{dif})$, for HC, MCI and AD are -1.02, 1.65 and 2.44, respectively. It can be seen that HC participants tend to be older than the predicted age, while MCI and AD patients tend to be younger than the predicted age. The physical meaning of the results is that healthy people possess a younger brain (or nearly at the same age). Nevertheless, the brain of an AD patient seems

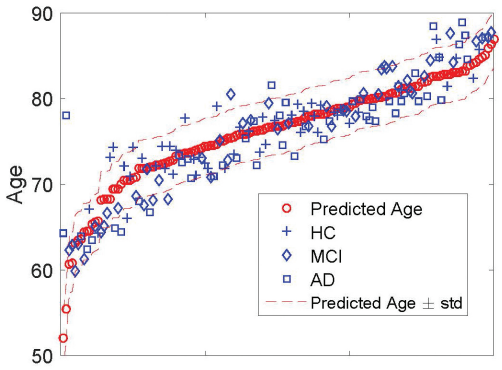


Figure 3: Predicted age Vs. actual age.

older than the actual age. Besides, MCI patients are more concentrated to the predicted age, while AD patients are more scattered. Different pathological phases show different statistical properties. Although HC, MCI and AD participants show different mean estimation difference, the corresponding standard deviations are relatively large, 4.35, 6.31 and 6.71. There are large overlap between neighboring categories. This is mainly due to the complex pathology of AD. Although severe brain damage can lead to cognitive disorder, there is no direct causal relationship between structural anomaly and dementia symptoms in geriatric cohorts. The underlying anatomical mechanism of dementia (cognitive disorder) is still unclear. The majority of the input features are brain volumes extracted from MRIs. Therefore, our proposed framework is more suitable to identify structural anomaly. When it comes to cognitive disorder, cognitive assessment scores, e.g. Mini Mental State Examination (MMSE), may be more suitable, since they directly answer all the criteria (in the form of interactive questionnaires) of clinical dementia diagnosis. However, the objective of our research project is to help identify risk factors associated with brain structural changes and other related biomarkers, while cognitive assessments treat internal pathological changes as a black box. This work offers a valuable tool to model the aging process as different pathways featured by various aging speed, which can be calculated based on anatomical properties of the brain. It explains how brain pathological aging affects the aging pathways of different patients (Problem 3).

Automatic Diagnosis One of the most important problems in AD study is diagnosis. The key contribution of this paper is an automatic diagnosis system based on RVFL network. Manifold learning is incorporated as part of the system to remove noise and construct low dimensional representation of the original high dimensional data. Figure 4 shows the comparison results. In the experiments all the results are based on RVFL. We compare the results from principal component analysis (PCA), Neighborhood Preserving Embedding (NPE), Locality Preserving Projections (LPP) and stochastic proximity embedding (SPE). It can be seen that generally manifold learning algorithms improves the performance of the classification algorithm. In particular,

NPE and LPP improve the system performance by about 2%. The optimal results are achieved at about 40 selected dimensions.

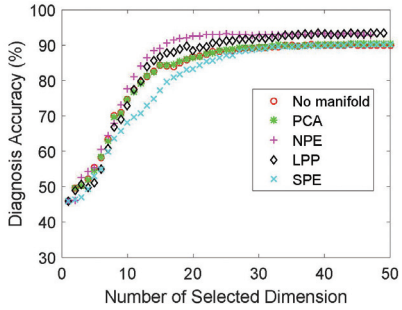


Figure 4: Comparison of different manifold learning algorithms.

Table 1 gives the experimental results. It can be seen that our proposed algorithm achieves very promising results, overall 93.28% accuracy. The precisions for HC, MCI and AD are 92.91%, 92.21% and 96.64%, respectively. The sensitivities for HC, MCI and AD are 94.81%, 95.36% and 86.48%, respectively. Although the sensitivity for AD is relatively low, 86.48%, AD is mainly misclassified as MCI. It is not very harmful. It has to be noted that the progression of AD is a gradual process, which may take decades. The benchmarks between different pathology phases are relatively vague. Therefore, classification of the transition stages is very challenging.

Table 1: Confusion Matrix, ten-fold cross validation.

		Predicted Class			Rate
		HC	MCI	AD	
True Class	HC	**603**	31	2	94.81%
	MCI	37	**1007**	12	95.36%
	AD	9	54	**403**	86.48%
Rate		92.91%	92.21%	96.64%	93.28%

Extensive comparison is made against Support Vector Machine (SVM), Random Subspace (RS), Multilayer Perceptron (MLP), Random Forest and Decision Tree. Table 2 shows the results for comparison targets. It can be seen that our proposed algorithm achieves superior results. The improvements are 9.45%, 3.95%, 9.07%, and 25.91%, respectively. Besides, we also show that with only medical imaging data ('Imaging' in Table 2), the proposed system achieves 84.13% accuracy. The integration of multiple sources of medical data show clear synergy and added value to the overall performance.

Prognosis

Another important problem in Alzheimer's disease research is prognosis, i.e. the prediction of AD progression (Problem 2). In this task, only patients with clear pathological progression are included in the experiments. Besides, since we utilize coupled observation as input, only those patients with at least 2 consecutive observation at the same stage are chosen. There are totally 425 valid records in the prognosis task.

Table 2: Recognition results for comparison targets (%).

	SVM	MLP	RF	DT	Imaging
Accuracy	83.83	89.33	84.21	67.37	84.13
Rel. Imp.	9.45	3.95	9.07	25.91	9.15

Table 3: Confusion Matrix for prognosis, ten-fold cross validation.

		Predicted Class		Rate
		Progression	No Prog.	
True Class	Progression	**32**	26	55.17%
	No Prog.	1	**366**	99.73%
Rate		96.97%	93.37%	93.65%

Table 3 shows the confusion matrix for the prognosis task. It can be seen that our proposed algorithm achieves a nearly perfect result to predict no progression, 99.73%. However, when it comes to progression, the results are two-fold. On the one hand, the proposed algorithm achieves very high precision, 96.97%. On the other hand, the sensitivity is relatively low, 55%. This is mainly due to the fact that clinical diagnosis of various stages of Alzheimer's consists of many subjective decisions. Moreover, the anatomical structures of the brain is closely correlated to dementia symptoms. However, there is no clear qualitative causal relationship between structure change and the corresponding symptoms. Besides, as shown in the table, the data in the prognosis problem are extremely unbalanced. The data for 'progression' is less than 0.1 of the 'No Progression' category, which makes different classifiers tend to overfit the 'No Progression' class.

Conclusion

In this study, we show that Random Vector Functional-link (RVFL) network is very suitable for Alzheimer's disease analysis, due to its ability to incorporate nonlinear relationship with a single layer structure. The proposed algorithm achieves very promising results in the diagnosis task. On the other hand, our proposed algorithm achieves very high prognosis precision with relatively low sensitivity. This may due to the complex nature of Alzheimer's disease pathology. Moreover, we present a novel analysis framework to study the aging speed of the participants, which clearly follows the biological aging process. Preliminary results are presented to show the potential of aging speed as a strong indicator for diagnosis and prognosis applications as well as a tool to assess future machine learning algorithms. Future work will be focused on the investigation of a complete aging model to describe different aging styles.

Acknowledgments

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

Alzheimers Disease International,. 2015. *World Alzheimer Report 2015: The Global Impact of Dementia*.

American Psychiatric Association,. 2013. *Diagnostic and statistical manual of mental disorders*. 5th edition edition.

Belkin, M., and Niyogi, P. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*, 585–591. MIT Press.

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

Dai, P.; Gwadry-Sridhar, F.; Bauer, M.; and Borrie, M. 2015. A hybrid manifold learning algorithm for the diagnosis and prognostication of Alzheimer's disease. In *AMIA 2015 Annual Symposium, San Francisco, CA, USA, 14-18 Nov*.

Dai, P.; Gwadry-Sridhar, F.; Bauer, M.; and Borrie, M. 2016a. Bagging Ensembles for the Diagnosis and Prognostication of Alzheimer's Disease. In *the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Phoenix, Arizona, USA, 12-17 Feb*.

Dai, P.; Gwadry-Sridhar, F.; Bauer, M.; and Borrie, M. 2016b. Longitudinal Brain Structure Changes in Health/MCI Patients: A Deep Learning Approach for the Diagnosis and Prognosis of Alzheimer's Disease. In *Alzheimer's Association International Conference (AAIC),Toronto, ON, Canada, 24-27 July*.

Dale, A.; Fischl, B.; and Sereno, M. I. 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage* 9(2):179 – 194.

Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.-r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; and Kingsbury, B. 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine* 29(6):82–97.

Ho, S.-S.; Dai, P.; and Rudzicz, F. 2016. Manifold learning for multivariate variable-length sequences with an application to similarity search. *IEEE Transactions on Neural Networks and Learning Systems* 27(6):1333–1344.

Huang, G.-B.; Zhu, Q.-Y.; and Siew, C.-K. 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70(1):489–501.

Husmeier, D. 1999. *Neural Networks for Conditional Probability Estimation*. Springer-Verlag London.

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.

Klassen; Pao; and Chen. 1988. Characteristics of the functional link net: a higher order delta rule net. In *IEEE International Conference on Neural Networks*, 507–513 vol.1. IEEE.

Lebedev, A. V.; Westman, E.; Van Westen, G. J. P.; Kramberger, M. G.; Lundervold, A.; Aarsland, D.; Soininen, H.; Koszewska, I.; Mecocci, P.; Tsolaki, M.; Vellas, B.; Lovestone, S.; and Simmons, A. 2014. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage. Clinical* 6:115–25.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *Nature* 521(7553):436–444.

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

Li, F.; Tran, L.; Thung, K.-H.; Ji, S.; Shen, D.; and Li, J. 2015. A Robust Deep Model for Improved Classification of AD/MCI Patients. *IEEE journal of biomedical and health informatics* 19(5):1610–6.

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.

Payan, A., and Montana, G. 2015. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *CoRR* abs/1502.02506.

Rahimi, A., and Recht, B. 2008. Random features for large-scale kernel machines. In Platt, J. C.; Koller, D.; Singer, Y.; and Roweis, S. T., eds., *Advances in Neural Information Processing Systems 20*. Curran Associates, Inc. 1177–1184.

Saxe, A.; Koh, P. W.; Chen, Z.; Bhand, M.; Suresh, B.; and Ng, A. Y. 2011. On random weights and unsupervised feature learning. In Getoor, L., and Scheffer, T., eds., *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 1089–1096. New York, NY, USA: ACM.

Scott, K. R., and Barrett, A. M. 2007. Dementia syndromes: evaluation and treatment. *Expert review of neurotherapeutics* 7(4):407–22.

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.

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.

Xiaofei He, P. N. 2003. Locality Preserving Projections.