

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Neuroimaging-based Artificial Neural Network Predicts Conversion of Cognitive Impairment Spectrum in Alzheimer's Disease

Permalink

<https://escholarship.org/uc/item/7w70x77v>

Author

Gao, Xiao

Publication Date

2018

Peer reviewed|Thesis/dissertation

Neuroimaging-based Artificial Neural Network Predicts Conversion
of Cognitive Impairment Spectrum in Alzheimer's Disease

by

Xiao Gao

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Imaging

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Neuroimaging-based Artificial Neural Network Predicts Conversion of Cognitive Impairment Spectrum in Alzheimer's Disease

Xiao Gao

Abstract

Alzheimer's Disease (AD) represents the most frequent (60-80%) subtype of dementia and is one kind of progressive spectrum disorder without effective treatment so far. In the last decades, great efforts from all over the community have been made on the early diagnosis of AD at its preclinical stage, Mild Cognitive Impairment (MCI). Recently, a series of machine learning studies have successfully constructed several computational models in predicting conversion of cognitive impairment but seldom foresee beyond 4 years. Thanks to Alzheimer's Disease Neuroimaging Initiative (ADNI) database, in this study we extracted cognition feature from several clinical outcomes. We then took advantage of structural MRI data and one Network Diffusion Model (NDM) raised by our group for subject-specific prediction of future cognition features. One supervised classification neural network was trained with ground-truth baseline and time-of-interest data but applied with predicted future cognition features. This established machine learning framework has demonstrated descent sensitivity and specificity in prediction of MCI-to-AD conversion (0.890 ± 0.083 and 0.923 ± 0.045) and healthy control (HC)-to-AD conversion (0.900 ± 0.074 and 0.744 ± 0.154) 5 years post baseline. To the best of our knowledge, we are the very first groups working on long-term prediction of AD spectrum conversion from both HC and MCI.

Table of Contents

1. Introduction	1
2. Methods	8
<i>Data collection and grouping</i>	8
<i>Cognition feature-extraction</i>	11
<i>Construction of atrophy-factor vector and diffusion-rate vector</i>	11
<i>Feature-extraction of atrophy-factor and diffusion-rate</i>	13
<i>Regression of future cognition-feature</i>	14
<i>Regression of future atrophy-factor and atrophy-feature</i>	15
<i>Prediction of future diagnosis</i>	16
3. Results	18
<i>Demographic characteristics of different training datasets</i>	18
<i>Cognition feature-extraction</i>	19
<i>Feature-extraction of atrophy-factor and diffusion-rate</i>	20
<i>Regression of future log-odds</i>	20
<i>Regression of future atrophy-features</i>	21
<i>Prediction of future AD spectrum diagnosis</i>	22
4. Discussion	28
References	30
Appendix	33

1. Introduction

Recently, a connectome-based Network Diffusion Model (NDM) proposed by our group was able to successfully estimate the progression of regional atrophy over time for patients with Alzheimer's Disease (AD) and other neurodegenerative disorders (Raj et al. 2015). The goal of this study is to go beyond the estimation of atrophy patterns and leverage the NDM's prognostic capabilities to predict the subjects' future clinical cognitive states via *deep learning* (i.e., using several artificial neural network architectures for feature-extraction, feature-regression, and classification of future mental states). One of our primary objectives is to predict if/when susceptible individuals with mild cognitive impairment or even normal cognition will ultimately be diagnosed with AD.

This study poses an innovative machine learning framework that may aid in (i) clinical decisions, (ii) in pharmacological and non-pharmacological interventions during pre-dementia stage, (iii) in the interpretation of the underlying mechanisms of AD progression, and (iv) in helping dementia patients and their families to make timely plans for future.

Background and Significance

Dementia is an umbrella term for a broad range of progressive neurological disorders where multiple high-level brain functions are impaired, such as memory, orientation, comprehension, calculation, learning, language capability and judgement. Deficits can also take place in other areas, such as emotional control, social behavior and motivation. In 2015, over 47 million people around the world were estimated to be living with dementia, and this number might exceed 130 million by 2050 (Prince et al. 2015). Dementia incorporates a series of neurodegenerative disorders that

are characterized by progressive loss of cognitive function, among which Alzheimer's Disease (AD) represents the most frequent (60-80%) subtype.

In the last decades, the overall biomedical community became greatly interested on the early diagnosis of dementia. The concept of mild cognitive impairment (MCI) was first introduced 30 years ago by the Mayo Clinic during one longitudinal study of AD (Reisberg et al. 1988), to identify the intermediate stage of cognitive impairment that is often, but not always, a transitional phase from cognitive changes of normal ageing to those typically found in dementia. Since then, many thousands of studies have been reported, in a huge effort to identify biomarkers that could reliably predict the conversion from MCI to AD, including a variety of randomized controlled trials of medications (Petersen et al. 2005; Feldman et al. 2007; Thal et al. 2005; Winblad et al. 2008). One successful identification of such a biomarker would allow future clinical research to refine patient sub-grouping and to assess whether potential clinical interventions may change the natural history of the disease at its very preclinical stage.

AD Spectrum Biomarkers and Neuropsychology Examination

Fox et al. (Fox 1999) were the first group to use sMRI to report cerebral atrophy in a longitudinal study of asymptomatic individuals at high risk of the familial AD¹. Other biomarkers for dementia were also proposed, such as functional MRI (fMRI), Diffusion Tensor Imaging (DTI), Arterial Spin Labeling (ASL), and F-FDG PET (Mintun et al. 2006; Mosconi et al. 2006; Johnson et al. 2012; Ewers et al. 2013; Dubois et al. 2016). Although these more advanced neuroimaging methods may provide a more detailed description of metabolic/functional changes on the demented

¹ which is one autosomal dominant dementia characterized by early-onset before 65 years of age

brain, sMRI-based atrophy analysis remains as the major routine method for most AD clinical studies due to its accessibility and low cost.

In addition to neuroimaging biomarkers, a comprehensive neuropsychological examination of the patient also plays an important role in the clinical diagnosis and staging of dementia. The grading for one kind of clinical test can reflect the cognitive performance of susceptible subjects within certain domains, such as memory, executive functions, language, attention and visuospatial skills. So far, however, there has been no specific neuropsychological biomarker capable of independently providing a comprehensive outline of mental state and accurate long-term prognosis. One sturdy strategy of combining the diagnostic and prognostic power of variable neuropsychological indices is needed for better delineation of each subject's cognition state.

Machine Learning prediction of AD spectrum conversions

Machine Learning is "the scientific discipline that focuses on how computers learn from data. It arises at the intersection of statistics, taking its advantage on efficient computing algorithms, and seeks to learn relationships between phenomenon and hypothesis" (Rahul et al. 2015). If we learn an underlying rule from past truth, it becomes possible for us to predict future events with certain confidence. Due to decades of continuous studies on AD, great hope has been placed on machine learning algorithms to transfer a large array of multi-disciplinary data into future clinical practice and to predict AD spectrum conversions. The main machine learning attempts until 2015 were summarized by (Moradi et al. 2015):

"(...) Recently, several computational studies have been reported on predicting the conversion to AD in MCI patients by utilizing various types of ADNI data such as MRI (e.g. Ye et al., 2011; Filipovych and Davatzikos, 2011; Batmanghelich et al., 2011), positron emission tomography (PET) (Cheng et al., 2012; Shaffer et al., 2013),

cerebrospinal fluid (CSF) biomarkers (Cheng et al., 2012; Davatzikos et al., 2011; Shaffer et al., 2013), and demographic and cognitive information (...) ".

Table 1: Relevant results from preceding studies

Earlier studies	Performance in MCI-AD prediction		Conversion time
	Sensitivity	Specificity	
Salvatori et al. (2018)	83%	87%	0-24 months
Moradi et al. (2015)	87%	74%	0-36 months
Misra et al. (2013)	67%	69%	0-48 months

Most of the studies listed in **Table 1** extract features from baseline neuroimaging and neuropsychological test data to detect MCI-to-AD conversion early on. **Fig. 1** provides a schematic for their general pattern regarding the progressive degeneration of cognitive state in AD. The area covering the cognition feature plane encompasses all extracted features from neuroimaging, neuropsychological or demographic data. Such a respective from preceding studies is somewhat limited, since a non-longitudinal dataset could hardly inform the progression-rate of one subject's cognitive impairment. Thus, their results may be regarded as one kind of stereotyped average-estimation of future cognitive state based on baseline input.

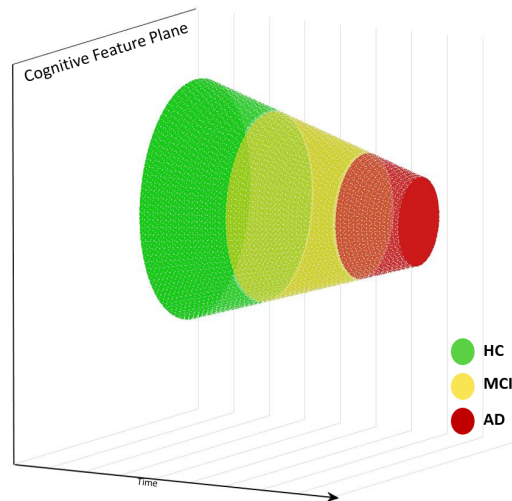


Figure 1: General disease progression pattern.

However, to our current knowledge, none of these studies have provided an accurate long-term prognosis beyond 48 months after baseline. Moreover, they focus solely on the prediction of MCI-to-AD conversion and disregard Healthy Control (HC)-to-MCI and HC-to-AD conversions, which account for a significant portion of the overall conversions.

One of main challenges in studying AD-spectrum conversions is that each subject's long-term cognitive state change may vary, and they typically do not follow a stereotypical progression. See **Fig. 2** for a schematic of this concept.

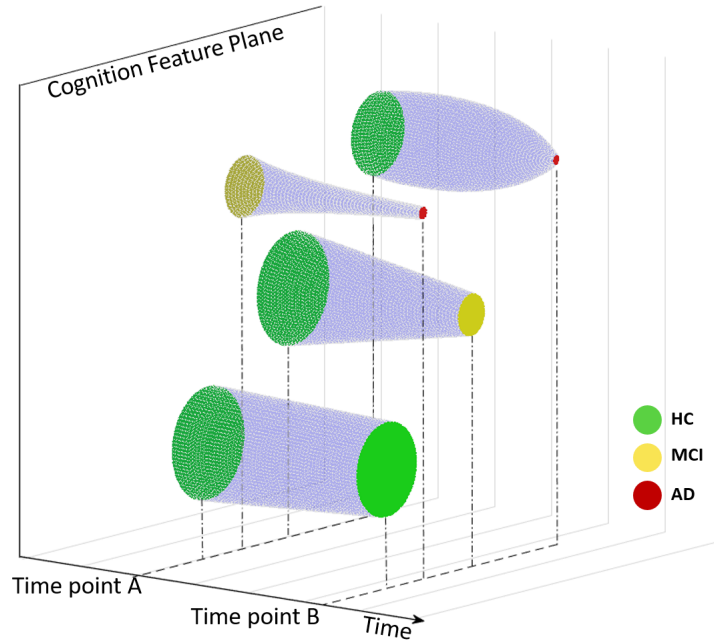


Figure 2: Subject-specific AD progression patterns hardly follow a general/stereotypical pattern.

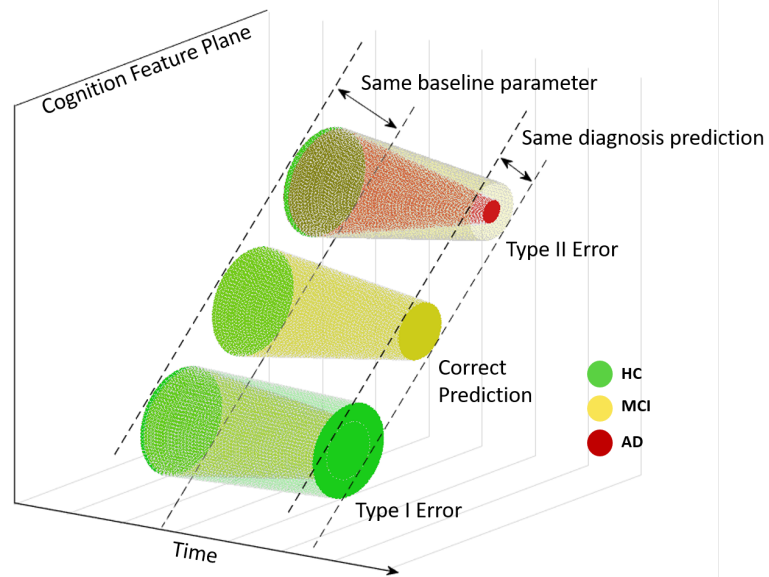


Figure 3: Three outcomes of diagnosis prediction using general progression model.

While applying a stereotypical progression model to each patient's baseline data may be acceptable for short-term predictions (1-2 years), the outcome of long-term diagnostic predictions yields significant type I and type II errors, as schematized in **Fig. 3**. One potential solution to this long-term prediction problem is to develop a computational disease model to extract vector information from baseline scalar data in order to estimate the disease progression rate. The addition of a disease vector feature would not only provide patient-specific prediction of future mental state, but also make it possible to enroll future ground-truth data as new training variables fed into machine learning algorithms.

Before detailing our computational strategy, it is worthwhile to recollect some key aspects of the AD pathological progression and its cumulative effect on brain morphometry and clinical staging.

AD pathology and the network diffusion model

According to Braak model (Braak and Braak 1996), the progression of Alzheimer's Disease (AD) is highly stereotyped, which could be explained by transsynaptic or transneuronal spread of misfolded b-amyloid and tau proteins, starting from entorhinal cortex and hippocampus until isocortical association areas. Morphological changes accompanying this pathological progression has solid support from MRI evidence, including both cross-sectional and longitudinal morphometric studies. (Fischl et al. 2002; Klauschen et al. 2009; Smith et al. 2004). Which is of concern, several longitudinal studies also confirm that such progression follows vulnerable fiber pathways rather than spatial proximity, closely mirroring Braak pathological stages (Apostolova and Thompson 2008; Apostolova et al. 2007; Thompson et al. 2003; Whitwell et al. 2007).

Standing on these findings, a network diffusion model (NDM) was recently proposed by our group to mathematically predict future patterns of regional atrophy and metabolism resulting from trans-neuronal transmission on the brain's connectivity network (Raj et al. 2015). This model accurately predicts time-of-interest future regional atrophy starting from subject's baseline sMRI. Rooted in this time-sensitive model connecting pathological accumulation with morphometric change, as a logical assumption, if we can construct the correlation between MCI/AD subject's neuroimaging morphometric progression and relative cognition feature change, it would open the possibility predicting the time window when AD spectrum conversion would happen for long-term prognosis.

In the current study, we focus on predicting future diagnosis within ADD spectrum, i.e. HC, MCI, and AD, at time-of-interest (1, 3, and 5 years) from baseline by using artificial neural network trained by both baseline and future-time-point cognition features extracted from ADNI data. One human connectome template is used to construct NDM in aid of exploiting progression-rate information from baseline sMRI and further helping predict future cognition features.

2. Methods

Data collection and grouping

The data involved in this study was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>). ADNI is a public-private, large multisite longitudinal study with the goal of tracking AD biomarkers and accelerate prevention and treatment of the disease. Over 2,000 subjects have been recruited from more than 50 intuitions

across the U.S. and Canada during 4 project phases from ADNI-1, through ADNI-Go and ADNI-2, to the on-going ADNI-3.

The current study utilizes 9,960 diagnostic records from 2,004 subjects from all phases available until the time of publication and is mostly based on the ADNIMERGE dataset ². This dataset also contains subjects’ demographic information, Apolipoprotein E4 (APOE4) genetic test outcome and records of clinical neuropsychological tests. See **Table 2** for details.

Table 2: Clinical predictor variables required for cognitive feature-extraction

Demographic data	Age Gender Marriage Education years
Cognitive tests	ADAS11 ADAS13 APOE4 CDR FAQ MMSE RAVLT_forgetting RAVLT_immediate RAVLT_learning RAVLT_perc_forgetting

We limit our dataset initially to the 1729 subjects with at least one diagnosed visit with post-processed sMRI information, which include voxel count of 86 brain regions (parcellated according to Desikan-Killiany atlas) and the corresponding scanning dates. The original source was

² contributed by *Michael C. Donohue* and *Chung-Kai Sun* from UC San Diego.

contributed and uploaded by Miriam Hartig et al. from the University of California, San Francisco. Finally, only sMRI data that passed their quality-control was used in this study. Since diagnosis are based on the evidence presented on the latest visit alone, we treat each visit and its corresponding clinical record as an independent observation.

For the purposes of constructing different computational models, the 1729-subject observation pool was divided into five training datasets: (i) cognition feature-extraction, (ii) sMRI feature-extraction, (iii) future cognition feature-regression, (iv) future sMRI feature-regression, and (v) future diagnosis-prediction. Specificities for each model are summarized in **Table 3**.

Table 3: Five computational models with different candidate algorithms and data recruitments

Target Computational Model	Candidate Algorithms	Data recruitments for baseline visit	Data recruitments for future visit
Cognition feature-extraction	Multinomial logistic regression	With complete cognitive tests from diagnostically steady subjects	-
sMRI feature-extraction	Network Diffusion Model + autoencoder neural network	With post-processed sMRI	-
Regression of future cognition-feature	Linear regression or regression neural network	With post-processed sMRI and complete cognitive tests	With complete cognitive tests
Regression of future sMRI-feature	Regression neural network	With post-processed sMRI and complete cognitive tests	With post-processed sMRI and complete cognitive tests
Prediction of future diagnosis	Classification neural network	With post-processed sMRI and complete cognitive tests	With post-processed sMRI and complete cognitive tests

Cognition feature-extraction

In the case of cognition feature-extraction, we want such a feature to be easily extracted as can sturdily reflect the relationship between available cognition indices and corresponding diagnosis and can be easily handled for the regression of future progression. Starting from this motivation, multinomial logistic regression is considered as our first choice, whose output is only two log-odds telling which diagnosis could be made with the highest probability, i.e. $\ln(P_{MCI}/P_{HC})$ and $\ln(P_{AD}/P_{MCI})$ in this study. Because we want to combine the diagnostic powers from the most essential clinical tests and demographic information, the enrolled observations should be equipped with complete cognitive tests and demographic records, totally 14 entries as shown in **Table 2.**

Meanwhile, in order to get rid of the noise from fluctuated clinical performance of diagnosis-converted subjects, only subjects holding the same diagnosis across the whole follow-up is recruited into the queue of logistic regression. One experimental logistic regression was run firstly with 14 entries mentioned above, for the purpose of determining the essential variables best at telling AD from HC or MCI, whose p-value should be less than 0.05 for at least one log-odds in this study. As long as these significant clinical predictors decided, another logistic regression was operated for the extraction of cognition-feature. The outcome logistic regression model and two log-odds of each subject would be saved for standby application.

Construction of atrophy-factor vector and diffusion-rate vector

All sMRI data used in this study were collected from ADNI database, post-processed by Miriam Hartig et al. using FreeSurfer version 5.1. ADNI-1 has both 1.5T and 3T sMRI scans while

ADNI-GO and ADNI-2 only have 3T data. Most sMRI data were acquired by using non-accelerated scan while about 8.8% data acquired via accelerated scan. In consideration of variable control, neither ADNI-1 1.5T nor accelerated scan was included in this study. No ADNI-3 processed sMRI data is available online by the time of publication.

Beginning with this dataset, brain atrophy factor was calculated for each of the 86 brain regions determined by Desikan-Killiany atlas. First, in order to cancel the influence from different subject brain sizes, each subject's 86-size brain volume vector was divided by its whole intracranial volume, leading to one 86-size normalized volume vector. Afterwards, all subjects with steady HC diagnosis across their entire follow-up were selected from sMRI feature-extraction dataset, therefore the average normalized volume and corresponding standard deviation could be calculated for each brain region based on normal controls. Given these 86 means and standard deviations, one region-wise normalization was operated followed by one sigmoid transform, leading to one 86-size atrophy factor vector for each observation.

Considering atrophy factor is merely one scalar feature of one subject's brain atrophy at one certain time point, NDM was used here to estimate atrophy diffusion rates among different brain regions, which leads to one 86-size diffusion rate vector for every practical observation. Within one such a diffusion rate vector, each value represents the summation of atrophy factors transmitted to one certain brain region, setting off from all other connected brain regions in a connectome-weighted way.

Feature-extraction of atrophy-factor and diffusion-rate

Because of limited sample size in terms of long-term follow-up, in order to prevent could-be overfitting it is necessary to condense the predictor vectors of both atrophy factor and diffusion rate via feature-extraction. In addition, one purified feature vector is more handy and sturdy for future regression, which would do much favor in estimation of future atrophy pattern.

Owing to these reasons, one sparse autoencoder was operated to extract atrophy feature and diffusion feature from each subject's atrophy factor and diffusion rate. Basically, autoencoder is one single-hidden-layer fully connected neural network with one predictor vector as its both input and output layer and the size of its hidden-layer determines the size of feature vector extracted from the input predictor.

Three autoencoder neural networks were designed with different hidden layer sizes, i.e. 86, 20, and 5, and thus their encoded feature vectors are referred to as 86-size, 20-size, and 5-size feature vector in turn. Some essential hyperparameters of autoencoder used in this study are listed in **Appx. Table 1**. Which should be noted, every inferior feature vector was encoded from its immediate superior feature vector. For example, one 86-size atrophy-feature was extracted from the subject's 86-size atrophy-factor, while the corresponding 20-size atrophy-feature was extracted from the 86-size atrophy-feature vector instead of the 86-size atrophy-factor.

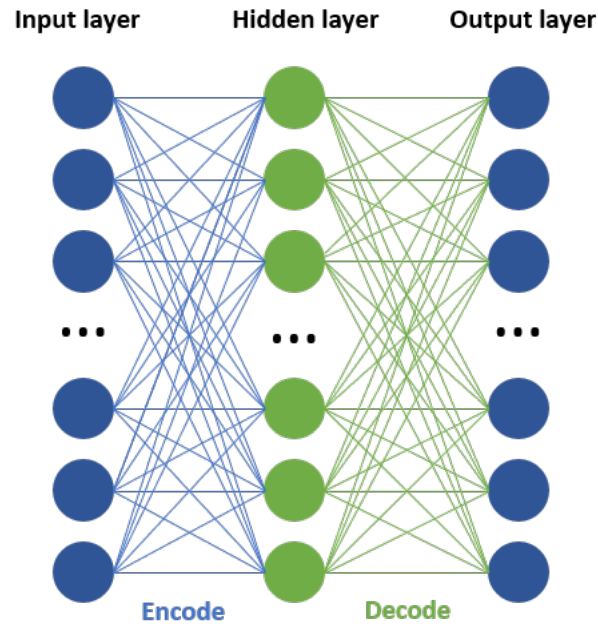


Figure 4: Architecture of autoencoder neural network. The input layer and output layer have equal size, but the hidden layer size is not necessarily the same.

Regression of future cognition-feature

The 2 log-odds resulted from the logistic regression before are considered as the cognition features of each observation. In order to estimate patient-specific cognition change pattern in 1 year, 3 years, and 5 years post baseline, a group of observation pairs have been enrolled provided that the involving subjects have comprehensive cognition-feature, atrophy-feature, and diffusion-feature for baseline observation along with cognition-feature for one future time-of-interest (1,3, or 5 years).

Two candidate regression models, i.e. linear regression and regression neural network, were tested for their performance in future prediction of cognition-feature, which shared the same input as baseline cognition-feature and baseline atrophy-feature with/without diffusion-feature.

The target variable is one of the 2 log-odds at future-time-point, for each of which one independent regression would be constructed.

With respect of linear regression strategy, the adopted algorithm is stepwise regression, which is a systematic method adding and removing predictors from a linear model according to their statistical significance in explaining the target variable. In this study, for each step or iteration of regression model fitting, one F-test was operated to calculate the change in the sum of squared error by adding one significant term (if $p < 0.05$) or removing one redundant term (if $p > 0.1$). Only linear terms were explicitly included in this linear regression model, excluding cross-product or higher order terms.

In terms of regression neural network, the input layer size depends on which kind of feature vector is used. For instance, if 20-size atrophy-feature is called upon without using of 20-size diffusion-feature, the input layer size would be 22 (20 baseline atrophy-feature + 2 baseline log-odds). Totally 2 hidden layers are created for this log-odds regression neural network, both of which hold the same size of input layer. The output layer size is 1, corresponding to one of the future 2 log-odds.

Regression of future atrophy-factor and atrophy-feature

In the current study, one subject's brain atrophy pattern is determined by its 86-size atrophy factor vector or different sizes of atrophy feature vectors (86, 20, or 5-size). We assume that the knowledge of both baseline and future-time-point atrophy outlines is going to provide a more specific simulation of one subject's disease progression than merely knowing its baseline

situation. Thus, another regression neural network was constructed for the estimation of each subject's future brain atrophy starting from the input of baseline atrophy-state as well as atrophy-diffusion originating from NDM.

The architecture's input layer is regulated by the feature vector fed in, which generally is composed of one baseline atrophy factor/feature vector and one corresponding diffusion rate/feature vector in equal length. The output layer is the time-of-interest (1, 3, or 5 years) atrophy factor/feature of the same subject, whose size remains the identical to input atrophy factor/feature vector. As for two hidden layers included, the first one is in paired size of the input layer, which is also the relationship between the second hidden layer and output layer.

To validate that NDM can bring forth more beneficial clues to future atrophy simulation, another collateral neural network has been tested with diffusion rate/feature vector removed from the input layer while saving other components and hyperparameters untouched.

Prediction of future diagnosis

At very first, we want to evaluate the diagnostic performance of the 2 predicted log-odds deriving from future cognition feature-regression. Principally, this sort of diagnosis is decided by 3 probable combinations of signs of the 2 log-odds, as shown in **Table 4**. Two sets of predicted log-odds (linear-regression-based and neural-network-based) have been tested separately.

Table 4: Three combinations of signs of 2 log-odds and corresponding diagnoses

	HC	MCI	AD
1st log-odds	-	+	+
2nd log-odds	-	-	+

Subsequently, we are more interested in introducing both baseline and future-time-point disease features into one classification neural network, identifying different cognition progression patterns within the disease spectrum, HC, MCI, or AD. Within this neural network architecture, the input layer consists of (i) baseline atrophy-feature, (ii) baseline cognition-feature, (iii) future-time-point atrophy-feature, and (iv) future-time-point cognition-feature. The size of the first hidden layer is the same of input while the second and also the last hidden layer is in half size of its sister. The output layer is one 3-bin logical vector with one 1-entry indicating the future diagnosis and another two 0-entries, i.e. HC is represented by $[1\ 0\ 0]$ while MCI by $[0\ 1\ 0]$ and AD by $[0\ 0\ 1]$.

This classification neural network was trained by ground-truth baseline and future-time-point features with 70% of dataset for training and another 20% for validation. The rest 10% of dataset was saved for testing in a 10-fold way, i.e. for each iteration out of ten, 90% of dataset was used for training and validation while 10% for testing. The final performance was defined as the average of these 10 outcomes. Meanwhile, which is more of clinical interest, the diagnostic performance of this classification model has also been 10-fold tested by an identical dataset except replacing the ground-truth future atrophy-feature and cognition-feature with regression-

based features. Based on such a test, the ensemble performance of (i) feature-extraction, (ii) future feature-regression and (iii) future diagnosis-prediction could be evaluated.

Another collateral neural network has also been constructed with all future-time-point data removed from the input layer, in the interest of identifying how many benefits this diagnosis prediction model derives from future information.

3. Results

Demographic characteristics of different training datasets

Totally eight training datasets have been structured for variable target computational models. Two datasets only focus on single-time-point visit data in consideration of feature-extraction, i.e. sMRI feature and cognition feature. The other six sub-pools are drawn for 2 types of regression models. Sample size for each training model is summarized in **Figure 5**.

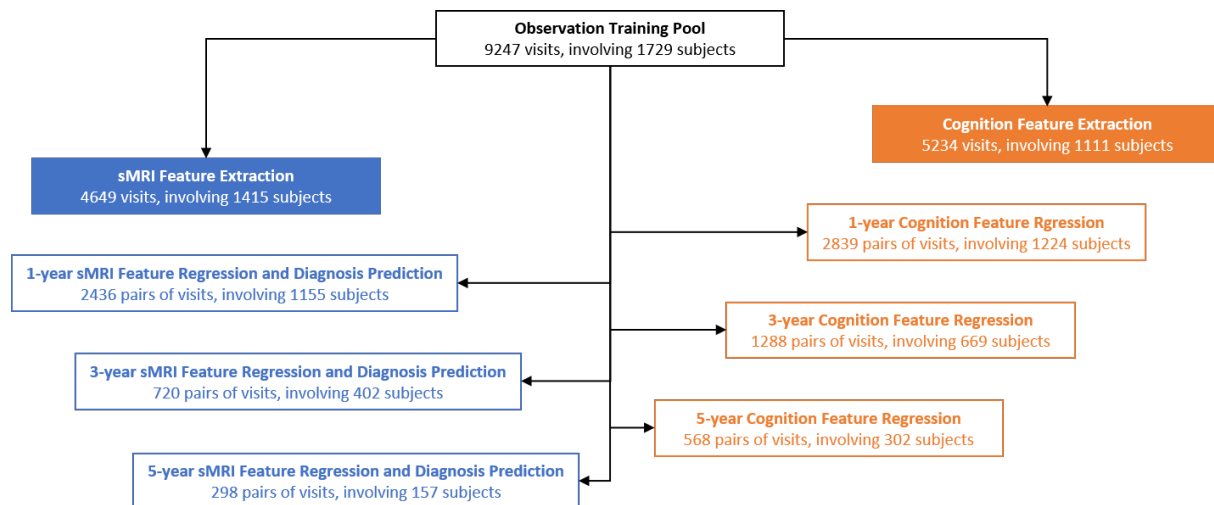


Figure 5: Data division to 2 feature-extraction datasets and 2 types of feature-regression datasets

As shown in **Appx. Table 2**, for these two feature-extraction datasets focusing on single-time-point observations, there is no significant difference ($p>0.05$) in diagnosis-proportion, age, or male-to-female ratio between sMRI feature-extraction and cognition feature-extraction datasets, using Friedman test.

In the case of double-time-point training datasets, there are 9 potential cognition progression patterns during follow-up, which include (i) 3 steady forms (steady-HC, steady-MCI, and steady-AD), (ii) 3 progressive conversions (HC-to-MCI, MCI-to-AD, and HC-to-MCI), and (iii) 3 anti-conversions (MCI-to-HC, AD-to-MCI, and AD-to-HC). Regarding so far there has been no effective treatment for AD-spectrum disease, the occurrence of anti-conversion is blamed on uncertain diagnosis or misdiagnosis in this study and thus its 3 subtypes are seen as one cognition progression pattern.

For each pair of these two feature-regression/diagnosis-prediction datasets with the same time-of-interest (1, 3, or 5 years), there is no significant difference in progression-pattern proportion, age, or male-to-female ratio ($p>0.05$, Friedman test). (Shown in **Appx. Table 3~5**).

Cognition feature-extraction

One experimental multinomial logistic regression has been employed using 14 demographic and cognitive variables, leading to 7 significant predictors of cognition feature with $p\text{-value} < 0.05$ for at least 1 log-odds, shown in **Appx. Table 6**.

Based on these 7 essential indexes, one following logistic regression gave rise to the cognition feature-extraction model. Thereafter, every observation's cognition state would only be represented by 2 log-odds out of this logistic regression model. Relevant coefficients and p-values are shown in **Appx. Table 7**.

Feature-extraction of atrophy-factor and diffusion-rate

After atrophy-factor and atrophy diffusion-rate calculated for each practical observation, they were thrown into a series of autoencoder neural networks, whose hidden-layer-sizes decrease from 86, to 20, and to 5 in turn. During this process, each observation's brain atrophy pattern and pathological diffusion model were encoded and saved within different hidden layers. After that, one backward decoding manipulation was performed to decide how much valuable information had been prevented in different sizes of feature vectors, by comparing the similarity between the reconstructed atrophy-factors or diffusion-rates and the original 86-size vectors. Correlation coefficients were calculated for such comparison, shown in **Appx. Table 8~9**. As we can see here, whatever for atrophy-feature or diffusion-feature, the deeper autoencoder network used, the smaller feature-vector started with, the harder it is to recover the original data completely.

Regression of future log-odds

Two strategies (stepwise linear regression and neural network) were tested for the regression of future-time-point 2 log-odds with the input of different sizes of atrophy-features with/without diffusion-features. In general, for long-term follow-up (3 and 5 years), the inclusion of diffusion-features always brings about better regression performance (R-squared) than merely using

atrophy-features (**Appx. Table 10~17**) ($p = 0.000$). However, when predicting log-odds 1 year after baseline, little improvement could be achieved by adding diffusion-features ($p = 0.527$).

When comparing the regression power from different sizes of feature vectors, the 86-size atrophy-factor and 86-size atrophy-feature have sturdier outcome in long-term cases (**Appx. Table 10~13**) than 20-size and 5-size feature vectors (**Appx. Table 14~17**) ($p = 0.000$).

Meanwhile, the application of neural network would provide these two 86-size vectors with further advancement in regression outcome versus using stepwise linear regression ($p = 0.033$).

To sum up, one optimal strategy for regression of future cognition-feature (log-odds) turns out to be making use of NDM and neural network architecture with the input including 86-size atrophy-factor or 86-size atrophy-feature. In consideration of simplifying variable control, only 86-size atrophy-factor was used for log-odds-regression purpose in the following study.

Regression of future atrophy-features

Because of limited sample size for future-atrophy-regression dataset, only 298 pairs of observations for 5-year follow-up (**Appx. Table 5**), smaller feature vectors are preferred to larger ones in order to prevent overfitting. At the same time, which is a similar story to future-log-odds-regression, the regression outcomes (R-value) took advantage from the input of diffusion-features, but only in the case of 86-size atrophy-factor and 86-size atrophy-feature other than those 2 smaller feature vectors (**Appx. Table 18~21**) ($p = 0.046$). It seems purified diffusion-feature cannot provide adequate information helpful for estimation of future atrophy pattern. As a result, there is a trade-off between small atrophy-features and large atrophy-features

when it comes to the prediction of atrophy progression; the final option among these features still needs testing by their performance in the following diagnosis-prediction later.

Prediction of future AD spectrum diagnosis

Quick impression of different diagnosis-prediction strategies

Aiming at a rough idea of the performance of different diagnosis-prediction strategies, 80% available observations were randomly selected for model construction, including intra-group validation (60% for training + 20% for validation), while the rest 20% for testing. No K-fold validation was included here for cheap test. The outcome is shown in **Table 5**.

Table 5: Diagnostic performance of different computational models at 3 future-time-points

Strategy	Computational model	Total Accuracy					
		With NDM involved			Without NDM involved		
		1y	3y	5y	1y	3y	5y
Using predicted future log-odds only	Logistic regression + stepwise linear regression (86-size atrophy factor)	0.803	0.703	0.637	0.801	0.705	0.636
Using double-time-point log-odds and atrophy-feature	Neural network (86-size atrophy factor)	0.752	0.713	0.678	0.746	0.707	0.654
	Neural network (86-size atrophy feature)	0.828	0.708	0.691	0.821	0.675	0.661
	Neural network (20-size atrophy feature)	0.821	0.808	0.832	0.808	0.801	0.768
	Neural network (5-size atrophy feature)	0.830	0.778	0.732	0.831	0.750	0.728

The 2 future log-odds out of linear regression demonstrated surprisingly decent diagnostic accuracy when predicting the diagnosis 1 year later, but the performance retrogressed abruptly in long-term cases (3 and 5 years). Furthermore, the addition of diffusion-rate seems to provide no help but noise to linear regression of future log-odds.

As for classification neural network fed with double-time-point cognition-features and atrophy-features, the 20-size atrophy feature did the sturdiest job in diagnosis-prediction for all 3 future-time-points. To be reiterated here, all these neural networks were trained and validated with (i) ground-truth baseline features and (ii) ground-truth future-time-point features, while tested with (i) ground-truth baseline features but (ii) regression-based future cognition-feature and atrophy-feature. All future cognition-features resulted from another regression neural network mentioned in preceding text, inputted with baseline cognition-feature, 86-size atrophy-factor and 86-size diffusion rate. The final outcome suggests that neural network can make the best use of NDM for in improving the total prediction accuracy ($p = 0.04$).

Diagnosis-prediction using regression-based log-odds only

Both linear-regression-based and neural-network-based log-odds are predicted with the input of baseline log-odds, 86-size atrophy-factor and diffusion-rate. These two computational models have been tested for their diagnostic performance with 10-fold validation, and Student's t-test was applied to compare several important sensitivities and specificities between two strategies, especially in terms of progressive-MCI (MCI-to-AD). Anti-conversion cases will not be covered in this study.

For prediction of diagnosis 1 year after baseline, **Appx. Table 22~23**, the total diagnostic accuracy of stepwise-linear-regression (0.794 ± 0.044) is better than that of neural-network-regression (0.740 ± 0.087), $p = 0.001$. Neural-network-regression demonstrated quite lower sensitivity to steady HC ($p = 0.000$), suggesting it is more inclined to classify susceptible

subjects into higher diagnosis of dementia, which could be supported by its higher sensitivity to cases of MCI-to-AD, i.e. progressive-MCI ($p = 0.009$).

For 3-year-interval diagnosis-prediction, **Appx. Table 24~25**, the total accuracy of neural-network-regression (0.714 ± 0.119) caught up with stepwise-linear-regression (0.687 ± 0.080), $p = 0.249$. In fact, such a tie should be mainly explained by the precipitous decline of linear-regression performance. Neural-network-regression model has shown both higher sensitivity (0.754 ± 0.256 , $p = 0.007$) and higher specificity (0.802 ± 0.108 , $p = 0.016$) for progressive MCI than stepwise linear regression.

When using predicted future log-odds to predict mental state 5 years after baseline (**Appx. Table 26~27**), the outcome is expressly fairish. Notably, the total accuracy of neural-network-regression model (0.710 ± 0.093) is significantly higher ($p = 0.001$) than that of stepwise-linear-regression (0.632 ± 0.107). At the same time, neural-network-based regression of log-odds also presented better sensitivity (0.902 ± 0.190 , $p = 0.003$) to progressive MCI, with acceptable specificity (0.812 ± 0.244) at the same time. However, such a log-odds based classification model was not robust enough to handle different sets of testing data, with sensitivity and specificity performance fluctuating turbulently across 10-fold validation.

Diagnosis-prediction using double-time-point log-odds and atrophy-features

To overcome the performance flexibility of merely using log-odds for diagnosis-prediction, one classification neural network has been constructed and fed with the cognition-features and atrophy-features from both baseline and time-of-interest. Such a computational model was trained with ground-truth double-time-point data but tested with baseline variables and predicted

future feature vectors. The predicted future cognition-feature is the outcome of one regression neural network with input of baseline cognition-feature, 86-size atrophy-factor and 86-size diffusion rate; the predicted future atrophy-feature is in 20-size coming out of one regression neural network inputted with baseline 20-size atrophy-feature and 20-size diffusion-feature.

As one collateral control of this double-time-point prediction model, another single-time-point classification neural network was built up using the same architecture but removing time-of-interest information from the input layer. 10-fold validation was applied to estimate the performance of these two prediction models and Student's t-test was used for comparison. The ultimate outcome turned out that both strategies demonstrated comparable accuracy for short-term (1y) prediction while double-time-point neural network had significant advantage in long-term (3y and 5y) prognosis accuracy ($p = 0.431$ for 1y, $p = 0.000$ for 3y, $p = 0.000$ for 5y).

Regarding 1-year prognosis (**Appx. Table 28~29**), although single-time-point feature vectors promised similar prediction accuracy compared with double-time-point input (0.826 ± 0.061 vs. 0.834 ± 0.015), most of its point score came from steady-diagnosis, inert to the detection of disease progression. With respect to double-time-point prediction model, it had higher sensitivity (0.594 ± 0.047 , $p = 0.000$) and specificity (0.540 ± 0.034 , $p = 0.029$) in foresight of progressive-MCI, which could be explained by its innate inclination to observing disease progression. Nevertheless, neither of these two models has done a decent job in screening out susceptible HC subjects 1 year before their diagnostic conversion.

Concerning 3-year follow-up (**Appx. Table 30~31**), which is the span-of-interest for most earlier studies, double-time-point neural network exhibited satisfying sensitivity and specificity (0.844 ± 0.047 and 0.855 ± 0.070) in the detection of progressive-MCI, which were much higher than the performance of single-time-point prediction model ($p = 0.008$ and 0.009). Moreover, for the very first time in this study, double-time-point prediction model provided the possibility of detecting MCI-susceptible HC subjects from cognitively normal community with passable sensitivity and specificity (0.673 ± 0.102 and 0.428 ± 0.115).

As regards diagnosis-prediction 5 years after baseline (**Table 6~7**), double-time-point prediction model returns one excellent outcome in foreseeing conversion cases starting from either MCI or HC, shown in **Table 6**. The sensitivity and specificity of double-time-point neural network for progressive-MCI are significantly better than single-time-point ($p = 0.022$ and 0.007). With respect of HC-to-AD conversion screening, over 80% of sensitivity and 60% of specificity were offered by such a classification neural network fed with baseline and time-of-interest information.

Table 6: Diagnosis-prediction (5y) using double-time-point neural networks, total accuracy = 0.845 ± 0.039

	Sensitivity	Baseline diagnosis				Specificity	Baseline diagnosis		
		HC	MCI	AD			HC	MCI	AD
Future diagnosis	HC	0.848 ± 0.054	0.600 ± 0.350	NaN	HC	0.945 ± 0.019	0.327 ± 0.308	NaN	
	MCI	0.680 ± 0.119	0.852 ± 0.101	NaN	MCI	0.477 ± 0.065	0.885 ± 0.077	0.000 ± 0.000	
	AD	0.900 ± 0.074	0.890 ± 0.083	0.964 ± 0.094	AD	0.744 ± 0.154	0.923 ± 0.045	1.00 ± 0.000	

Table 7: Diagnosis-prediction (5y) using single-time-point neural networks, total accuracy = 0.708 ± 0.175

	Sensitivity	Baseline diagnosis				Specificity	Baseline diagnosis		
		HC	MCI	AD			HC	MCI	AD
Future diagnosis	HC	0.861 ± 0.152	0.167 ± 0.577	NaN	HC	0.840 ± 0.148	0.167 ± 0.816	NaN	
	MCI	0.167 ± 0.444	0.684 ± 0.412	NaN	MCI	0.185 ± 0.455	0.740 ± 0.378	0.000 ± 0.000	
	AD	0.458 ± 0.850	0.750 ± 0.402	0.771 ± 0.733	AD	0.542 ± 1.004	0.743 ± 0.411	1.000 ± 0.000	

4. Discussion

As a summary, by using the framework discussed above (**Figure 6**), the established computational model has proved decent performance in (i) detection of MCI-to-AD conversion 3 and 5 years earlier, and (ii) detection of HC-to-MCI-or-AD conversion 3 and 5 years earlier. To the best of our knowledge, this is one of the very first studies focusing on long-term prediction of AD spectrum conversion starting from both HC and MCI.

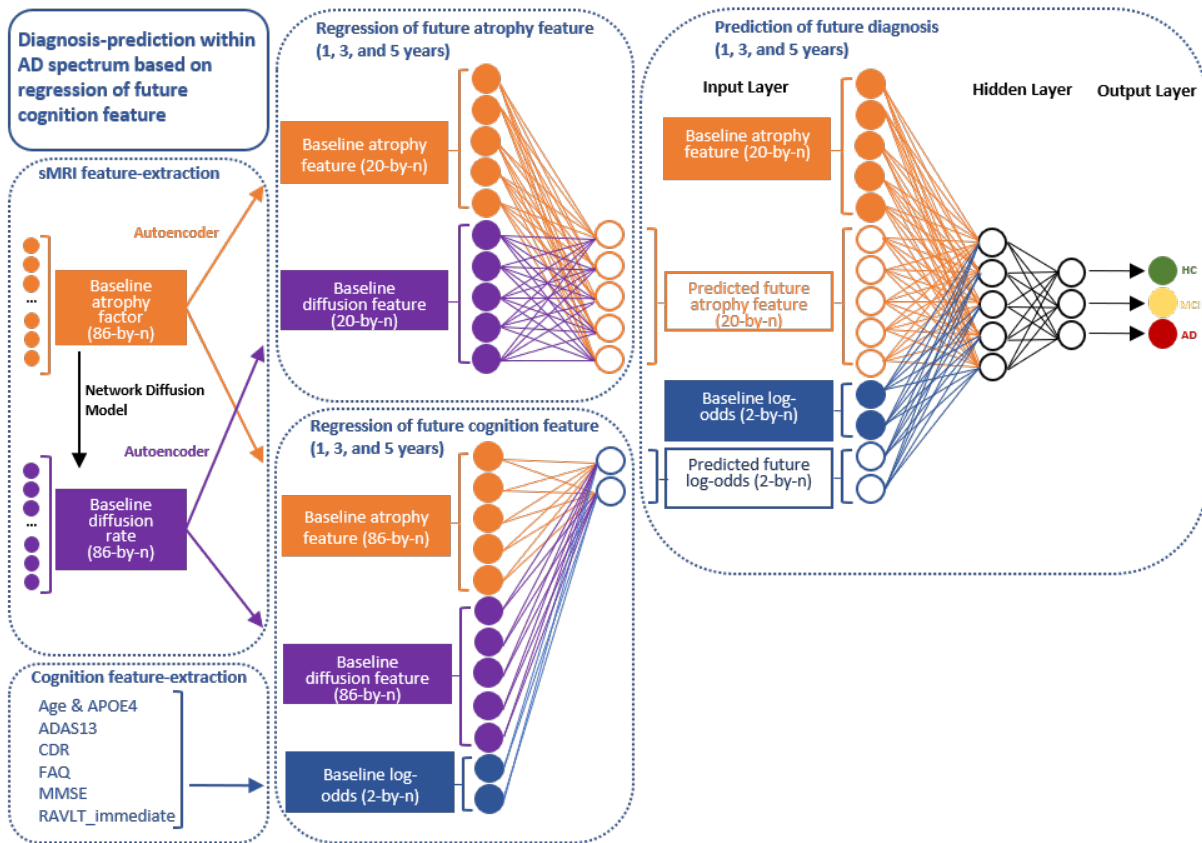


Figure 6: Diagram of framework constructed for diagnosis-prediction within AD spectrum based on regression of future cognition features

The same workflow has also been tested with future predictors removed from the final classification neural network, whose outcome was inferior to double-time-point input but still presented with comparable performance like earlier studies. Such a result, once again, suggests that single-time-point cross-sectional clinical data is not adequate enough for an accurate estimation of patient-specific disease progression. For all that, owing to the speed-extraction function of NDM, we can determine the pathological diffusion pattern among each subject's brain regions and further helping estimate its future clinical state.

On the other hand, the current study is nothing more than taking two cross-sections of one subject's cognitive change and is far from a time-sensitive model in outlining comprehensive disease progression. One of many reasonable solutions is turning to multimodality neuroimaging methods, like perfusion-weighted MRI and amyloid-PET, for multi-dimensions of disease description. Besides, although neural network is competent in regression-task, in the field of classifier it is usually overrun by support vector machine (SVM), which explicitly determines optimal decision boundaries between hyperplanes. In the future, several candidate machine learning algorithms, like SVM or random forest, would be made use of for further research.

References

1. Apostolova, L.G., Steiner, C.A., Akopyan, G.G., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., and Thompson, P.M. (2007). Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. *Arch. Neurol.* 64, 1489–1495.
2. Apostolova, L.G., and Thompson P.M. (2008). Mapping progressive brain structural changes in early Alzheimer's disease and mild cognitive impairment. *Neuropsychologia* 46, 1597–1612.
3. Batmanghelich KN, Ye DH, Pohl KM, Taskar B, Davatzikos C. Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE; 2011. Disease classification and prediction via semi-supervised dimensionality reduction; pp. 1086–1090.
4. Braak H., and Braak E. (1996). Evolution of the neuropathology of Alzheimer's disease. *Acta Neurol. Scand. Suppl.* 165, 3–12.
5. Cheng B, Zhang D, Shen D. Domain transfer learning for MCI conversion prediction. *MICCAI.* 2012;2012:82–90.
6. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging.* 2011;32(12):2322–e19.
7. Dubois B., Hampel H., Feldman H. H., Scheltens P., Aisen P., Andrieu S., et al.. (2016). Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimers Dement.* 12, 292–323. 10.1016/j.jalz.2016.02.002
8. Ewers M., Brendel M., Rizk-Jackson A., Rominger A., Bartenstein P., Schuff N., et al.. (2014). Reduced FDG-PET brain metabolism and executive function predict clinical progression in elderly healthy subjects. *Neuroimage Clin.* 4, 45–52. 10.1016/j.nicl.2013.10.018
9. Feldman HH, Ferris S, Winblad B, et al. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. *Lancet Neurol.* 2007; 6:501–12.

10. Filipovych R, Davatzikos C. Semi-supervised pattern classification of medical images: application to mild cognitive impairment (MCI) *Neuroimage*. 2011;55(3):1109–1119.
11. Fischl B., Salat D.H., Busa E., Albert M., Dieterich M., Haselgrove C., van der Kouwe A., Killiany R., Kennedy D., Klaveness S., et al.. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
12. Fox N. C., Warrington E. K., Rossor M. N. (1999). Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease. *Lancet* 353, 2125–2125. 10.1016/S0140-6736(99)00496-1
13. Johnson K. A., Fox N. C., Sperling R. A., Klunk W. E. (2012). Brain imaging in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2:a006213. 10.1101/cshperspect.a006213
14. Klauschen F., Goldman A., Barra V., Meyer-Lindenberg A., and Lundervold A. (2009). Evaluation of automated brain MR image segmentation and volume- try methods. *Hum. Brain Mapp.* 30, 1310–1327.
15. Mintun M. A., Larossa G. N., Sheline Y. I., Dence C. S., Lee S. Y., Mach R. H., et al.. (2006). [11C]PIB in a nondemented population: potential antecedent marker of Alzheimer disease. *Neurology* 67, 446–452. 10.1212/01.wnl.0000228230.26044.a4
16. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative A.D.N. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage*. 2015;104:398–412.
17. Mosconi L., Tsui W. H., Herholz K., Pupi A., Drzezga A., Lucignani G., et al.. (2008). Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J. Nucl. Med.* 49, 390–398. 10.2967/jnumed.107.045385
18. Petersen RC, Thomas RG, Grundman M, et al. Alzheimer's Disease Cooperative Study Group. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med.* 2005; 352:2379–88.
19. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prima M. World Alzheimer Report 2015 The global impact of dementia analysis of prevalence, incidence, cost and trends.

20. Rahul C. Deo. Machine Learning in Medicine. *Circulation*. 2015;132:1920-1930.
21. Raj A, LoCastro E, Kuceyeski A, Tosun D, Relkin N, and Weiner M. (2015). Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer's Disease. *Cell Reports* 10, 359–369.
22. Reisberg B, Ferris S, de Leon MJ, et al. Stage-specific behavioral, cognitive, and in vivo changes in community residing subjects with age-associated memory impairment and primary degenerative dementia of the Alzheimer type. *Drug Dev Res*. 1988; 15:101–14.
23. Shaffer JL, Petrella JR, Sheldon FC, Choudhury KR, Calhoun VD, Coleman RE, Doraiswamy PM. Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology*. 2013;266(2):583–591.
24. Smith S.M., Jenkinson M., Woolrich M.W., Beckmann C.F., Behrens T.E.J., Johansen-Berg H., Bannister P.R., De Luca M., Drobnjak I., Flitney D.E., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 (1), S208–S219.
25. Thal LJ, Ferris SH, Kirby L, et al. Rofecoxib Protocol 078 study group. A randomized, double-blind, study of rofecoxib in patients with mild cognitive impairment. *Neuropsychopharmacology*. 2005; 30:1204–15.
26. Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Dordrell, D.M., and Toga, A.W. (2003). Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci*. 23, 994–1005.
27. Whitwell J.L., Przybelski S.A., Weigand S.D., Knopman D.S., Boeve B.F., Petersen R.C., and Jack C.R. Jr. (2007). 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain* 130, 1777–1786.
28. Winblad B, Gauthier S, Scinto L, et al. GAL-INT-11/18 Study Group. Safety and efficacy of galantamine in subjects with mild cognitive impairment. *Neurology*. 2008; 70:2024–35.
29. Ye DH, Pohl KM, Davatzikos C. Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on. IEEE; 2011. Semi-supervised pattern classification: application to structural MRI of Alzheimer's disease; pp. 1–4.

Appendix

Table 1: Hyperparameters of autoencoder for sMRI feature-extraction

	86-size feature vector	20-size feature vector	5-size feature vector
Hidden layer size	86	20	5
L2 Regularization	0.001	0.001	0.001
Sparsity Regularization	1	1	1
Sparsity Proportion	0.25	0.25	0.25
Input-hidden layer transfer function	log-sigmoid	log-sigmoid	log-sigmoid
Hidden-output layer transfer function	pure linear	pure linear	pure linear
Maximum number of training epochs	2000	2000	2000

Table 2: no significant difference ($p > 0.05$, Friedman test) in demographic characteristics

between two feature-extraction datasets

Target Computational Model	Characteristics of dataset	HC	MCI	AD	Number of involving subjects
sMRI feature-extraction (autoencoder)	Number of visits	1526(32.82%)	2192(47.15%)	931(20.03%)	776 males + 639 females
	Age	75.26±11.66	73.36±14.97	74.84±13.03	
	Males/Females	743/783	1300/783	496/435	
Cognition feature-extraction (logistic regression)	Number of visits	1923(36.74%)	2263(43.34%)	1048(20.02%)	608 males + 503 females
	Age	76.06±12.40	74.52±15.49	75.71±14.70	
	Males/Females	923/1000	1363/900	604/444	

Table 3: No significant difference ($p>0.05$, Friedman test) in demographic characteristics between two 1-year feature-regression/diagnosis-prediction datasets

Target Computational Model	Characteristics of dataset	steady HC	steady MCI	steady AD	HC-to-MCI	MCI-to-AD	HC-to-AD	anti-conversion	Number of involving subjects
Regression of future cognition-feature	Number of visit pairs	830(29.24%)	1244(43.82%)	473(16.66%)	40(1.41%)	199(7.01%)	1(0.00%)	52(1.83%)	681 males + 543 females
	Baseline Age	75.59±11.44	73.55±14.78	74.63±14.56	75.32±12.78	74.38±14.88	85.21	69.01±15.29	
	Males/Females	406/424	748/496	265/208	27/13	109/90	0/1	31/21	
Regression of future sMRI-feature and diagnosis-prediction	Number of visit pairs	737(30.25%)	1054(43.27%)	397(16.30%)	32(1.31%)	177(7.27%)	1(0.04%)	38(1.56%)	639 males + 516 females
	Baseline Age	75.46±11.31	73.43±14.65	74.43±14.66	75.85±13.29	74.03±14.89	85.21	68.69±14.46	
	Males/Females	359/378	636/418	223/174	23/9	96/81	0/1	24/14	

Table 4: No significant difference ($p>0.05$, Friedman test) in demographic characteristics between two 3-year feature-regression/diagnosis-prediction datasets

Target Computational Model	Characteristics of dataset	steady HC	steady MCI	steady AD	HC-to-MCI	MCI-to-AD	HC-to-AD	anti-conversion	Number of involving subjects
Regression of future cognition-feature	Number of visit pairs	399(30.98%)	486(37.73%)	78(6.06%)	62(4.81%)	186(14.44%)	10(0.78%)	67(5.20%)	363 males + 306 females
	Baseline Age	75.63±10.52	72.35±14.47	72.91±13.22	77.33±10.12	74.28±13.79	79.19±9.26	68.39±15.32	
	Males/Females	182/217	303/183	42/36	30/32	111/75	7/3	32/35	
Regression of Future sMRI-feature and diagnosis-prediction	Number of visit pairs	287(39.86%)	219(30.42%)	43(5.97%)	33(4.58%)	107(14.86%)	4(0.50%)	27(3.75%)	232 males + 170 females
	Baseline Age	75.97±9.78	73.10±14.59	72.01±12.18	77.49±9.78	73.76±14.07	82.17±9.27	68.96±14.82	
	Males/Females	130/157	147/72	24/29	17/16	65/42	3/1	21/6	

Table 5: No significant difference ($p > 0.05$, Friedman test) in demographic characteristics between two 5-year feature-regression/diagnosis-prediction datasets

Target Computational Model	Characteristics of dataset	steady HC	steady MCI	steady AD	HC-to-MCI	MCI-to-AD	HC-to-AD	anti-conversion	Number of involving subjects
Regression of future cognition-feature	Number of visit pairs	212(37.32%)	160(28.17%)	22(3.87%)	44(7.75%)	81(14.26%)	20(3.52%)	29(5.11%)	172 males + 130 females
	Baseline Age	75.37±9.31	72.55±15.27	71.44±10.83	76.88±8.24	73.63±13.44	77.28±6.49	69.49±15.48	
	Males/Females	99/113	109/51	10/12	25/19	54/27	10/10	17/12	
Regression of Future sMRI-feature and diagnosis-prediction	Number of visit pairs	135(45.30%)	62(20.81%)	11(3.69%)	25(8.39%)	48(16.11%)	13(4.36%)	4(1.34%)	92 males + 65 females
	Baseline Age	75.35±8.70	74.38±15.27	70.84±6.24	75.48±6.39	73.39±13.34	77.92±6.80	79.24±6.51	
	Males/Females	60/75	47/15	5/6	15/10	32/16	7/6	4/0	

Table 6: Logistic regression of baseline diagnosis using 14 demographic and cognitive predictors

Total accuracy = 86.81%	Age	Gender	Years of Education	Marriage	APOE4	ADAS11	ADAS13	RAVLT immediate	RAVLT learning	RAVLT forgetting	RAVLT perc_forgetting	FAQ	CDR	MMSE
p-value for log-odds 1	0.000	0.818	0.160	0.640	0.003	0.531	0.000	0.001	0.133	0.803	0.160	0.176	0.000	0.168
p-value for log-odds 2	0.001	0.073	0.570	0.917	0.042	0.009	0.000	0.012	0.129	0.138	0.475	0.000	0.000	0.000

Table 7: Logistic regression of baseline diagnosis using 7 essential demographic and cognitive predictors

Training Accuracy = 86.53% Testing Accuracy = 87.33%	Age	APOE4	ADAS13	RAVLT_immediate	FAQ	CDR	MMSE
p-value for log-odds 1	0.000	0.004	0.000	0.000	0.195	0.000	0.258
p-value for log-odds 2	0.000	0.006	0.000	0.225	0.000	0.000	0.000
coefficients for log-odds 1	0.568	-0.356	-0.531	0.346	-0.060	-1.401	0.071
coefficients for log-odds 2	0.2202	-0.344	-0.430	0.157	-0.094	-0.244	0.278

Table 8: Original atrophy factor vs. reconstructed atrophy factor using autoencoder of 86, 20, or 5-size hidden layer

	86-size hidden layer		20-size hidden layer		5-size hidden layer	
	R-value	p-value	R-value	p-value	R-value	p-value
Training	0.978	0.000	0.817	0.000	0.673	0.000
Testing	0.973	0.000	0.796	0.000	0.639	0.000

Table 9: original diffusion rate vs. reconstructed diffusion rate using autoencoder of 86, 20, or 5-size hidden layer

	86-size hidden layer		20-size hidden layer		5-size hidden layer	
	R-value	p-value	R-value	p-value	R-value	p-value
Training	0.983	0.000	0.838	0.000	0.711	0.000
Testing	0.981	0.000	0.832	0.000	0.705	0.000

Table 10: Stepwise linear regression of future log odds with 86-size atrophy factor

	R-squared for log-odds 1		R-squared for log-odds 2	
	86-size atrophy factor + 86-size diffusion rate + baseline 2 log-odds	86-size atrophy factor + baseline 2 log-odds	86-size atrophy factor + 86-size diffusion rate + baseline 2 log-odds	86-size atrophy factor + baseline 2 log-odds
1 year	0.865	0.865	0.907	0.907
3 years	0.744	0.742	0.787	0.787
5 years	0.691	0.683	0.719	0.718

Table 11: Neural network regression of future log odds with 86-size atrophy factor

	R squared for log-odds 1		R squared for log-odds 2	
	86-size atrophy factor + 86-size diffusion rate + baseline 2 log-odds	86-size atrophy factor + baseline 2 log-odds	86-size atrophy factor + 86-size diffusion rate + baseline 2 log-odds	86-size atrophy factor + baseline 2 log-odds
1 year	0.865	0.849	0.906	0.874
3 years	0.807	0.703	0.834	0.783
5 years	0.811	0.748	0.708	0.817

Table 12: Stepwise linear regression of future log odds with 86-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	86-size atrophy feature + 86-size diffusion feature + baseline 2 log-odds	86-size atrophy rate + baseline 2 log-odds	86-size atrophy feature + 86-size diffusion feature + baseline 2 log-odds	86-size atrophy feature + baseline 2 log-odds
1 year	0.867	0.865	0.908	0.908
3 years	0.747	0.745	0.795	0.788
5 years	0.704	0.688	0.730	0.718

Table 13: Neural network regression of future log odds with 86-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	86-size atrophy feature + 86-size diffusion feature + baseline 2 log-odds	86-size atrophy feature + baseline 2 log-odds	86-size atrophy feature + 86-size diffusion feature + baseline 2 log-odds	86-size atrophy feature + baseline 2 log-odds
1 year	0.838	0.859	0.899	0.873
3 years	0.787	0.768	0.835	0.788
5 years	0.774	0.777	0.811	0.739

Table 14: Stepwise linear regression of future log odds with 20-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	20-size atrophy feature + 20-size diffusion feature + baseline 2 log-odds	20-size atrophy feature + baseline 2 log-odds	20-size atrophy feature + 20-size diffusion feature + baseline 2 log-odds	20-size atrophy feature + baseline 2 log-odds
1 year	0.865	0.864	0.906	0.906
3 years	0.740	0.737	0.784	0.783
5 years	0.676	0.663	0.696	0.687

Table 15: Neural Network regression of future log odds with 20-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	20-size atrophy feature + 20-size diffusion feature + baseline 2 log-odds	20-size atrophy feature + baseline 2 log-odds	20-size atrophy feature + 20-size diffusion feature + baseline 2 log-odds	20-size atrophy feature + baseline 2 log-odds
1 year	0.856	0.867	0.912	0.909
3 years	0.752	0.714	0.752	0.744
5 years	0.744	0.636	0.738	0.699

Table 16: Stepwise linear regression of future log odds with 5-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	5-size atrophy feature + 5-size diffusion feature + baseline 2 log-odds	5-size atrophy feature + baseline 2 log-odds	5-size atrophy feature + 5-size diffusion feature + baseline 2 log-odds	5-size atrophy feature + baseline 2 log-odds
1 year	0.862	0.862	0.904	0.904
3 years	0.729	0.726	0.774	0.771
5 years	0.643	0.639	0.672	0.672

Table 17: Neural Network regression of future log odds with 5-size atrophy feature

	R squared for log-odds 1		R squared for log-odds 2	
	5-size atrophy feature + 5-size diffusion feature + baseline 2 log-odds	5-size atrophy feature + baseline 2 log-odds	5-size atrophy feature + 5-size diffusion feature + baseline 2 log-odds	5-size atrophy feature + baseline 2 log-odds
1 year	0.825	0.862	0.868	0.893
3 years	0.741	0.739	0.800	0.764
5 years	0.676	0.611	0.712	0.711

Table 18: Neural network regression of future 86-size atrophy factor using baseline 86-size atrophy factor with/without 86-size diffusion rate

Baseline 86-size atrophy factor + 86-size diffusion rate	R-value	p-value	Baseline 86-size atrophy factor	R-value	p-value
1 year	0.907	0.00	1 year	0.906	0.00
3 years	0.872	0.00	3 years	0.855	0.00
5 years	0.835	0.00	5 years	0.756	0.00

Table 19: Neural network regression of future 86-size atrophy feature using baseline 86-size atrophy feature with/without 86-size diffusion feature

Baseline 86-size atrophy feature + 86-size diffusion feature	R-value	p-value	Baseline 86-size atrophy feature	R-value	p-value
1 year	0.788	0.00	1 year	0.888	0.00
3 years	0.755	0.00	3 years	0.601	0.00
5 years	0.552	0.00	5 years	0.548	0.00

Table 20: Neural network regression of future 20-size atrophy feature using baseline 20-size atrophy feature with/without 20-size diffusion feature

Baseline 20-size atrophy feature + 20-size diffusion feature	R-value	p-value	Baseline 20-size atrophy feature	R-value	p-value
1 year	0.893	0.00	1 year	0.906	0.00
3 years	0.857	0.00	3 years	0.877	0.00
5 years	0.775	0.00	5 years	0.792	0.00

Table 21: Neural network regression of future 5-size atrophy feature using baseline 5-size atrophy feature with/without 5-size diffusion feature

Baseline 5-size atrophy feature + 5-size diffusion feature	R value	p-value	Baseline 5-size atrophy feature	R value	p-value
1 year	0.933	0.00	1 year	0.933	0.00
3 years	0.885	0.00	3 years	0.803	0.00
5 years	0.827	0.00	5 years	0.813	0.00

Table 22: Diagnosis-prediction (1y) using future log-odds out of stepwise-linear-regression, total accuracy = 0.794 ± 0.044

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.819 ± 0.059	0.261 ± 0.294	NaN	HC	0.973 ± 0.029	0.096 ± 0.138	NaN
	MCI	0.522 ± 0.569	0.824 ± 0.073	0.500 ± 1.155	MCI	0.12 ± 0.170	0.879 ± 0.031	0.033 ± 0.141
	AD	0.00 ± 0.00	0.463 ± 0.097	0.899 ± 0.123	AD	NaN	0.484 ± 0.191	0.992 ± 0.036

Table 23: Diagnosis-prediction (1y) using future log-odds out of neural-network-regression, total accuracy = 0.740 ± 0.087

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.678 ± 0.197	0.479 ± 0.530	NaN	HC	0.974 ± 0.028	0.121 ± 0.149	NaN
	MCI	0.632 ± 0.453	0.785 ± 0.126	0.800 ± 0.894	MCI	0.091 ± 0.104	0.892 ± 0.052	0.124 ± 0.426
	AD	0.00 ± 0.00	0.527 ± 0.123	0.872 ± 0.180	AD	0.00 ± 0.00	0.497 ± 0.216	0.998 ± 0.015

Table 24: Diagnosis-prediction (3y) using future log-odds out of stepwise-linear-regression, total accuracy = 0.687 ± 0.080

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.641 ± 0.118	0.347 ± 0.484	NaN	HC	0.917 ± 0.119	0.257 ± 0.385	NaN
	MCI	0.643 ± 0.378	0.768 ± 0.137	0.667 ± 1.155	MCI	0.202 ± 0.164	0.773 ± 0.068	0.375 ± 0.957
	AD	0.000 ± 0.000	0.631 ± 0.131	0.962 ± 0.138	AD	NaN	0.696 ± 0.268	0.989 ± 0.070

Table 25: Diagnosis-prediction (3y) using future log-odds out of neural-network-regression, total accuracy = 0.714 ± 0.119

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.653 ± 0.169	0.524 ± 0.407	NaN	HC	0.910 ± 0.126	0.287 ± 0.238	0.000 ± 0.000
	MCI	0.670 ± 0.431	0.756 ± 0.176	1.000 ± 0.000	MCI	0.215 ± 0.146	0.838 ± 0.091	0.417 ± 0.983
	AD	0.143 ± 0.756	0.754 ± 0.256	0.916 ± 0.223	AD	0.5 ± 1.414	0.802 ± 0.108	1.00 ± 0.000

Table 26: Diagnosis-prediction (5y) using future log-odds out of stepwise-linear-regression, total accuracy = 0.632 ± 0.107

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.542 ± 0.142	0.200 ± 0.343	NaN	HC	0.911 ± 0.158	0.261 ± 0.670	NaN
	MCI	0.740 ± 0.456	0.703 ± 0.254	1.000 ± 0.000	MCI	0.222 ± 0.165	0.736 ± 0.232	1.000 ± 0.000
	AD	0.233 ± 0.689	0.776 ± 0.173	1.000 ± 0.000	AD	0.800 ± 0.894	0.726 ± 0.296	1.00 ± 0.000

Table 27: Diagnosis-prediction (5y) using future log-odds out of neural-network-regression, total accuracy = 0.710 ± 0.093

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.663 ± 0.175	0.524 ± 0.694	NaN	HC	0.921 ± 0.112	0.333 ± 0.510	0.000 ± 0.000
	MCI	0.623 ± 0.462	0.704 ± 0.218	0.000 ± 0.000	MCI	0.258 ± 0.232	0.843 ± 0.221	0.000 ± 0.000
	AD	0.567 ± 0.781	0.902 ± 0.190	0.917 ± 0.360	AD	0.722 ± 0.882	0.812 ± 0.244	1.00 ± 0.000

Table 28: Diagnosis-prediction (1y) using double-time-point neural network, total accuracy = 0.834 ± 0.015

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.861 ± 0.041	0.194 ± 0.122	NaN	HC	0.976 ± 0.002	0.090 ± 0.052	NaN
	MCI	0.513 ± 0.044	0.852 ± 0.014	0.500 ± 0.00	MCI	0.139 ± 0.035	0.901 ± 0.009	0.067 ± 0.017
	AD	0.800 ± 0.843	0.594 ± 0.047	0.929 ± 0.020	AD	1.00 ± 0.00	0.540 ± 0.034	0.995 ± 0.000

Table 29: Diagnosis-prediction (1y) using single-time-point neural network, total accuracy = 0.826 ± 0.061

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.902 ± 0.074	0.260 ± 0.725	NaN	HC	0.967 ± 0.051	0.074 ± 0.229	NaN
	MCI	0.327 ± 0.829	0.843 ± 0.061	0.500 ± 1.00	MCI	0.102 ± 0.264	0.880 ± 0.078	0.067 ± 0.281
	AD	0.000 ± 0.000	0.482 ± 0.158	0.898 ± 0.106	AD	0.00 ± 0.00	0.476 ± 0.197	0.995 ± 0.023

Table 30: Diagnosis-prediction (3y) using double-time-point neural network, total accuracy = 0.864 ± 0.026

	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.904 ± 0.050	0.641 ± 0.093	NaN	HC	0.963 ± 0.014	0.547 ± 0.114	NaN
	MCI	0.673 ± 0.102	0.865 ± 0.066	NaN	MCI	0.428 ± 0.115	0.879 ± 0.022	0.000 ± 0.000
	AD	0.250 ± 0.408	0.844 ± 0.047	0.986 ± 0.024	AD	0.500 ± 0.591	0.855 ± 0.070	1.000 ± 0.000

Table 31: Diagnosis-prediction (3y) using single-time-point neural network, total accuracy = 0.771 ± 0.079

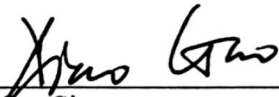
	Sensitivity	Baseline diagnosis			Specificity	Baseline diagnosis		
		HC	MCI	AD		HC	MCI	AD
Future diagnosis	HC	0.912 ± 0.144	0.432 ± 0.769	NaN	HC	0.916 ± 0.106	0.202 ± 0.298	NaN
	MCI	0.292 ± 0.663	0.702 ± 0.144	NaN	MCI	0.235 ± 0.431	0.788 ± 0.197	0.000 ± 0.000
	AD	0.000 ± 0.000	0.723 ± 0.283	0.933 ± 0.322	AD	0.000 ± 0.000	0.706 ± 0.358	1.000 ± 0.000

Publishing Agreement

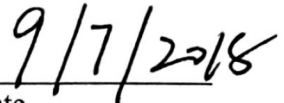
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date