

Predicting Progression From Mild Cognitive Impairment to Alzheimer's Disease Using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers

Sidra Minhas¹, Aasia Khanum, Farhan Riaz, Shoab A. Khan, and Atif Alvi

Abstract—Mild cognitive impairment is a preclinical stage of Alzheimer's disease (AD). For effective treatment of AD, it is important to identify mild cognitive impairment (MCI) patients who are at a high risk of developing AD over the course of time. In this study, autoregressive modelling of multiple heterogeneous predictors of Alzheimer's disease is performed to capture their evolution over time. The models are trained using three different arrangements of longitudinal data. These models are then used to estimate future biomarker readings of individual test subjects. Finally, standard support vector machine classifier is employed for detecting MCI patients at risk of developing AD over the coming years. The proposed models are thoroughly evaluated for their predictive capability using both cognitive scores and MRI-derived measures. In a stratified

five-fold cross validation setup, our proposed methodology delivered highest AUC of 88.93% (Accuracy = 84.29%) and 88.13% (Accuracy = 83.26%) for 1 year and 2 year ahead AD conversion prediction, respectively, on the most widely used Alzheimer's disease neuroimaging initiative data. The notable conclusions of this study are: 1) Clinical changes in MRI-derived measures can be better forecasted than cognitive scores, 2) Multiple predictor models deliver better conversion prediction than single biomarker models, 3) Cognitive score boosted by MRI-derived measures delivers better short-term ahead conversion prediction, and 4) Neuropsychological scores alone can deliver good accuracy for long-term conversion prediction.

Index Terms—Autoregressive, biomarkers, classification, correlation, linear, longitudinal, Mild cognitive impairment, prediction.

Manuscript received October 17, 2016; revised January 31, 2017 and April 1, 2017; accepted May 3, 2017. Date of publication May 15, 2017; date of current version May 3, 2018. Data collection and sharing for this project was supported in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and in part by the Department of Defense ADNI (Department of Defense award number W81XWH-12-2-0012). The work of ADNI was supported in part by the National Institute on Aging, in part by the National Institute of Biomedical Imaging and Bioengineering, and in part by the generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; Euroimmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). (Corresponding author: Sidra Minhas.)

M. Sidra is with the Computer Engineering Department, College of Electrical and Mechanical Engineering, National University of Sciences & Technology, Islamabad 44000, Pakistan (e-mail: sidra_m21@hotmail.com).

A. Khanum and A. Alvi are with the Computer Science Department, Forman Christian College, Lahore 54600, Pakistan (e-mail: aasiahanum@fccollege.edu.pk; atifalvi@fccollege.edu.pk).

F. Riaz and S. A. Khan are with the Computer Engineering Department, College of Electrical and Mechanical Engineering, National University of Sciences & Technology, Islamabad 44000, Pakistan (e-mail: farhan.riaz@gmail.com; shoabak@gmail.com).

Digital Object Identifier 10.1109/JBHI.2017.2703918

I. INTRODUCTION

TODAY, over 30 million people live with dementia due to Alzheimer's disease (AD) worldwide. This number is estimated to increase three fold by 2050 [1]. AD is a progressive neurodegenerative disorder that starts as memory impairment, followed by severe cognitive decline and eventually complete loss of function. The course of disease is generally divided into three stages. During the first stage, which is mostly pre-symptomatic, degenerative pathological changes happen in the form of β -amyloid ($A\beta$) plaques deposition in the brain [2]. After a patient-specific interval, the second stage called Mild Cognitive Impairment (MCI) sets in. At this stage, neuronal degeneration and neuronal dysfunction accelerates and expresses behaviourally as mild decline in cognitive abilities along with memory and thinking problems [3]. The last and final stage in evolution of the disease is dementia wherein the brain damage becomes so extensive that the patient becomes completely debilitated, with the effects often leading to death [4].

Various biomarkers and clinical symptoms are employed for evaluating AD progression depending upon condition of the patient and stage of the disease [4]. According to the hypothetical model presented by Sperling *et al.* [5], Cerebrospinal Fluid (CSF) bio-specimens and Positron Emission Tomography (PET) $A\beta$ imaging are considered appropriate for determining

early pathological degenerations. Neurodegeneration and neuronal dysfunction is better understood with structural Magnetic Resonance Imaging (MRI) and Fluorodeoxyglucose PET (FDG-PET). Neuropsychological Measures (NM) are used to measure the level of cognitive status of the patient at any stage.

At present, AD is diagnosed if the patient is clinically declared to have dementia [4]. On the other hand, MCI is considered as an intermediate stage between normal aging and dementia which may or may not lead to AD [6]. It has been observed that around 10–15 percent of MCI patients develop AD annually, whereas a significant number remains stable and some may even return to normal [7]. Accurate prediction of MCI course is highly desirable for early control and management of AD. Research into accurate prediction of MCI course is on the rise. MCI patients can be widely segregated into two groups i.e. those who progress to AD (MCIp) and those who stay stable as MCI (MCIs) over the coming years. Hence, the MCI-to-AD conversion prediction task reduces to that of binary classification of the MCI group into MCIp and MCIs.

Nonetheless, there is no consensus about which biomarker or combination of biomarkers coupled with classification methodology is most accurate in predicting MCI to AD conversion. Some of the reported results rely on individual modality tests e.g. MRI/PET imaging [8], [9], Diffusion Tensor (DT) imaging [10], florbetapir-F18 positron emission tomography (FBP PET) [11], [12], CSF biomarker [13], [14], while others use a combination thereof. For instance, [15]–[18], employ a combination of CSF, Functional imaging (FDG-PET), and structural imaging (MRI) biomarkers for performing the classification. However, studies pertaining to modelling the disease progression over time have been limited. In [19] logistic regression and in [20] regularized linear regression are used to model the MCI-to-AD conversion and predictions are made with MRI and cognitive data as independent variables. However, the model accuracy remains below 70%. In [21] a Disease State Index (DSI) is developed by combining multiple AD predictors to capture disease progression over time and differentiate between MCIp and MCIs. Hall *et al.* [22] discusses the generalizability of this DSI over multiple MCI cohorts, however the absence of biomarker selection strategies from amongst the pool of biomarkers makes the technique less accurate. In [23] a novel MRI biomarker is proposed and used in a semi-supervised Machine Learning framework for AD prediction. In [24] a statistical Disease State Fingerprint (DSF) is created to assess a patient's disease state on basis of similarity with previous patients' biomarker data, nonetheless such statistical techniques are prone to sensitivity towards dataset size.

Most of the above mentioned approaches including [20], [25], [26] and [23] rely only on baseline data for building their predictive models, therefore resulting in limited accuracy ranging from 65–84%. However, AD is a gradually progressing disorder and therefore a lot of predictive power can be offered by temporally collected longitudinal data encompassing the dynamics of the disease. Very few papers have based their predictions on longitudinal data so far. Some example approaches include the image based minimal set classifier [27] and Spatial Pattern of Abnormalities for Recognition of Early AD score (SPARE-AD)

[28], [29], yet the performance remains limited. The main aim of this paper is to employ the latent information in longitudinal biomarker values to predict MCI-to-AD progression. Contributions of this paper are as follows:

- 1) Forecasting future clinical changes in the biomarker values using only the baseline and first annual follow-up biomarker readings. Autoregressive modelling of the multimodal, multivariate predictors is performed using three different setups employing single and multiple biomarkers
- 2) Classification of individual MCI trajectories into MCIp and MCIs using an optimized classifier.
- 3) Identifying most robust predictors of conversion in unimodal and multimodal settings.
- 4) Use of the publically available AD dataset provided by the Alzheimer's disease Neuroimaging Initiative (ADNI) for validation purposes.

Rest of the paper is organized in four sections. Section II entails the materials used, Section III describes the methodology followed, Section IV quantifies the results and mentions a brief comparison with previously published work. Section V presents the conclusion.

II. MATERIALS

A. ADNI Biomarkers

The data used in this study were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) website¹ on 30th May, 2015. ADNI is a five year public-private partnership which recruited 800 adults, ages 55 to 90, to participate in the research – approximately 200 Cognitively Normal (CN) older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information see www.adni-info.org. ADNI is engaged in gathering AD related biomarkers from Neuropsychological measures (NM), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) scans, biochemical readings via lumbar puncture and genetic alleles. According to the AD disease dynamics model presented at² it is visible that brain structure captured via MR Images and cognition status encapsulated by cognitive tests deteriorate most when the disease progresses from MCI to AD. Hence, only these two modalities are selected for the said classification task.

For the present analysis, we use the longitudinal MRI morphometric measures provided by the University of California, San Francisco, Memory and Ageing Centre on the ADNI website. The complete details of extracting the MRI derived measures are provided in [30]. The current study includes the information from those images only which passed the overall QC process. The MRI biomarkers consist of volumes of brain regions obtained after cortical parcellation and white matter parcellation, surface area of the brain regions and cortical thickness of the brain regions provided in the UCSFFSL file on ADNI website.

¹adni.loni.edu

² <http://adni.loni.usc.edu/study-design/background-rationale/>

TABLE I
SUBJECT INFORMATION

2 Year Data		
	MCIP (n = 37)	MCIs (n = 65)
Age	74.1 + 7.4	74.7 + 7.5
Education	16 + 3.1	15.2 + 2.9
Gender (M/F)	20/17	36/29
3 Year Data		
	MCIP (n = 54)	MCIs (n = 65)
Age	74.7 + 7	74.7 + 7.5
Education	15.8 + 3	15.2 + 2.9
Gender (M/F)	24/23	36/29

In addition to the MRI measures, the non-invasive, cheap and easy to obtain Neuropsychological Measures (NM) are also employed. The NM selected in present study are the scores of the AD Assessment Scale (ADAS), Rey Auditory Verbal Test (AVLT), Clock Drawing Test (CDT), Clock Copying Test (CCT), Immediate Recall Total Score (LIMM), Mini Mental State Examination (MMSE), Trail Making Test A (TRA) and Trail Making Test B (TRAB). The details of the test procedures and scoring criterion is given in the ADNI General Procedures Manual [30]. For an in-depth analysis, both MRI and NM biomarkers were first analysed individually and then combined to observe the effect on classification performance.

B. Subjects

For the predictive modelling framework presented in this paper, biomarker readings recorded at regular annual intervals after the baseline (BL) reading were considered i.e. 12th month (M12), 24th month (M24) and 36th month (36M). The missing follow-up readings were imputed using Last Observation Carried Forward (LOCF) method [31]. This study includes the MCI patients who retained the diagnosis of MCI at BL and all follow-up visits as MCIs class. The MCI patients that retained MCI diagnosis at BL and M12 but converted to AD at M24 or M36 visit as MCIP class. For analysis standardization the subjects with complete readings for both NM and MRI measures available were kept. The detailed demographic information of the subjects is mentioned in **Table I**. It can be seen that both groups were well matched in terms of number, age, gender and education in all cases.

III. METHODS

The overview of the machine learning framework for MCIP-to-AD conversion prediction task is shown in **Fig. 1**. The MCI population is divided into training and test sets using a stratified 5 fold cross validation scheme according to which 80% of the instances are used for training while the remaining instances are used for testing. The training data is used for marker ranking and selection and is further divided in Leave-one-out (LOO) cross validation setup to generate and validate the autoregressive (AR) parameters. The AR parameters obtained are used to

forecast future biomarker values of the test data. Once complete time domain trajectories are available, they are embedded in a Support Vector Machine (SVM) classifier to segregate MCIP and MCIs. This process is repeated again for increasing number of features i.e. model ranks and performance metrics are noted. The individual modules are detailed in the following pages.

A. Training Phase

As stated earlier, each of the training fold is further divided in LOO setup in which 1 instance is kept for validation while all others are used for generating the AR parameters. In each loop biomarker ranking is performed and the selected biomarkers are used to calculate the AR parameters. The average of AR parameters resulting from each loop are kept for future biomarker value forecasting.

1) Biomarker Ranking and Subset Selection: The current biomarker set consists of 264 MRI and 8 NM biomarkers, of which not all contribute effectively towards MCIP vs. MCIs segregation. Hence, biomarker selection strategies are called for. In this study, a wrapper based approach is adopted for selecting significant biomarkers. The biomarkers are sifted according to their ranks in impact towards MCIP vs. MCIs discrimination. For this the two sampled student's t-test is performed on baseline biomarker readings. The p values of the t-test indicate the significance of a particular biomarker towards effective diagnostics. Later, biomarker subsets are formed by incrementally adding one biomarker at a time according to its significance. The number of biomarkers in each subset corresponds to the model rank.

2) AR Modelling: Future biomarker value is forecasted using the models generated by the longitudinal training data. The variable trajectories are assumed to erupt from an autoregressive model where the next value depends linearly on its own previous values. Hence, we put forward three least squares linear prediction models to estimate next in sequence data. Let X be the matrix of baseline and M12 biomarker readings of the n training instances and P be the vector of biomarker values of the training instances at the next time stamp .i.e. M24. Note that M24 biomarker readings of the validation instances are not known, and need to be determined using the proposed models. The linear relation between X and P is modelled by (1).

$$X.w = P \quad (1)$$

Where w is the vector of linear autoregressive prediction coefficients. The weights w are calculated in a least squares sense using (2)

$$w = (X^T X)^{-1} X^T P \quad (2)$$

Where $(X^T X)^{-1} X^T$ is the pseudo inverse of the matrix X . The known M12 and calculated M24 values are later used to forecast the M36 reading. The above stated model is trained in different ways in order to determine the effect of atomic and combined biomarkers towards future value estimation of a particular biomarker. The following variations are proposed.

a) Single Predictor Model (SPM): This model assumes each biomarker to be independent and uncorrelated with

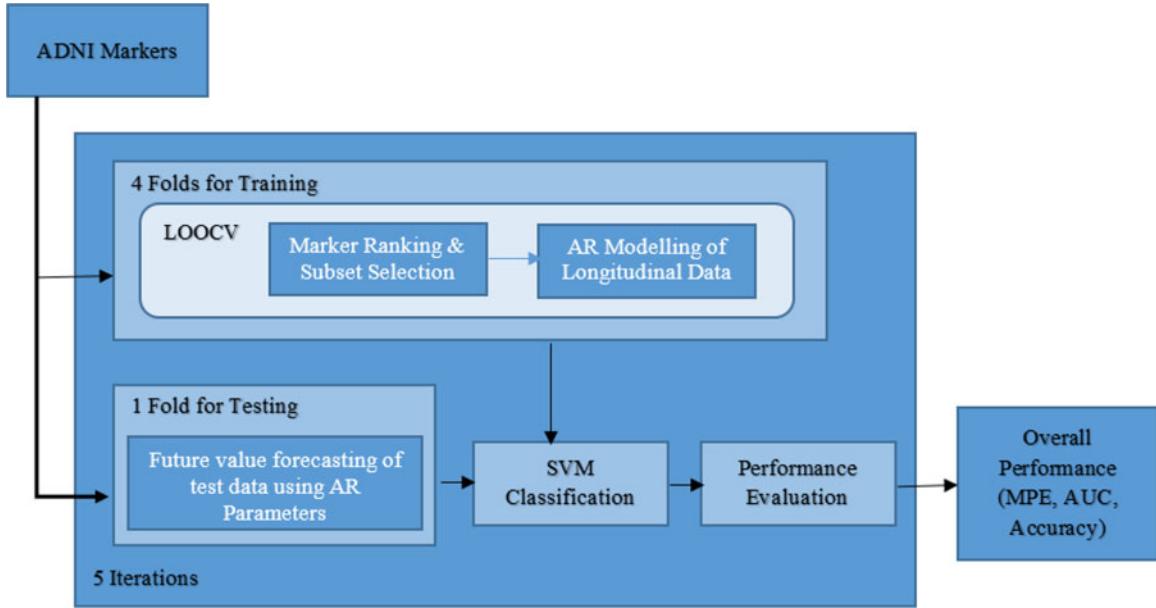


Fig. 1. System overview.

other biomarkers. Then for a particular biomarker m , let the training matrix X_m be:

$$X_m = \begin{bmatrix} x_i(t) & x_i(t+1) \\ \vdots & \vdots \\ x_n(t) & x_n(t+1) \end{bmatrix}, \text{ for } m = 1, \dots, N, i = 1, \dots, n \quad (3)$$

Where X_m is a matrix of baseline and M12 readings of m^{th} biomarker of n training instances. N corresponds to the model rank. Let P be the vector of next time point readings of the training data represented by (4).

$$P_m = \begin{bmatrix} x_i(t+2) \\ \vdots \\ x_n(t+2) \end{bmatrix}, \text{ for } m = 1, \dots, N, i = 1, \dots, n \quad (4)$$

The values of X_m and P_m are substituted as X and P , respectively, in (2) and the linear prediction coefficients w are calculated.

b) Multiple Predictor Timestamp Model (MPTM): It is rather important to quantify the effect of multiple previous time point readings onto future readings separately. For the n training instances, let $X(t)$ and $X(t+1)$ be the matrixes of baseline and M12 readings of the m biomarkers respectively as shown by (5).

$$X(t) = \begin{bmatrix} x_{1i}(t) & \dots & x_{mi}(t) \\ \vdots & \ddots & \vdots \\ x_{1n}(t) & \dots & x_{mn}(t) \end{bmatrix}$$

$$X(t+1) = \begin{bmatrix} x_{1i}(t+1) & \dots & x_{mi}(t+1) \\ \vdots & \ddots & \vdots \\ x_{1n}(t+1) & \dots & x_{mn}(t+1) \end{bmatrix} \quad (5)$$

For $m = 1, \dots, N, i = 1, \dots, n$

Then for linear models depicting mapping of respective time stamp readings onto future values, we employ (6)

$$\begin{aligned} X(t).w_t &= P_m \\ X(t+1).w_{t+1} &= P_m \end{aligned} \quad (6)$$

Where P is the vector of values of m^{th} biomarker at M24 of n training instances as shown in (4) while w_t and w_{t+1} are the linear prediction coefficients calculated in least squares sense using (2) by substituting the values of $X(t)$ and $X(t+1)$ respectively.

B. Testing Phase

1) Longitudinal Future Value Forecasting: Once the linear prediction coefficients have been obtained, they are used to forecast the future marker readings of the separated test instances. For estimating the next time stamp values of the test instances $Y(t+2)$ under SPM method, the matrix Y is populated with previously known values of $Y(t)$ and $Y(t+1)$ according to (3). Later Y is multiplied by the coefficients w to obtain $Y(t+2)$ as shown in (7).

$$Y(t+2) = Y.w \quad (7)$$

Likewise, for MPTM method (8) is used to forecast the next in sequence value.

$$Y(t+2) = [Y(t).w_t + Y(t+1).w_{t+1}] / 2 \quad (8)$$

Where $Y(t)$ and $Y(t+1)$ are analogous to $X(t)$ and $X(t+1)$ in (5). $Y(t+2)$ is dependent upon previous time point readings of all N biomarkers contained in vectors $Y(t)$ and $Y(t+1)$. It is assumed that both baseline and M12 readings equally impact the future value at M24, thus averaging is performed. Similar calculations are performed to calculate $Y(t+3)$ i.e. M36 biomarker values.

2) SVM Classification: The training and completed test trajectories for the N biomarkers considered for modelling are

TABLE II
FEATURE RANKS

1 year ahead			
RANK	NM	MRI	NM+MRI
1	RAVLT	VRA	RAVLT
2	ADAS	CTLIC	ADAS
3	CDT	VLE	VRA
4	MMSE	CTRIC	CTLIC
5	TRAAC	CTRIF	VLE
2 year ahead			
	NM	MRI	NM+MRI
1	RAVLT	VLIT	RAVLT
2	ADAS	VRA	ADAS
3	MMSE	VLE	VLIT
4	CDT	SALCAC	VRA
5	LIMM	CTLPC	VLE

RAVLT: Reys Auditory Verbal Test, ADAS: Alzheimer's Disease Assessment Score, MMSE: Mini Mental State Examination, CDT: Clock Drawing Test, LIMM: Immediate Recall Total Score, CTRE: Cortical Thickness of Right Entorhinal, VRA: Volume of Right Amygdala, CTLIC: Cortical Thickness Average of Left Isthmus Cingulate, CTRIC: Cortical Thickness Average of Right Isthmus Cingulate, VLE: Volume of Left Entorhinal, CTRF: Cortical Thickness of Right Fusiform, VLIT: Volume Left Inferior Temporal, SALCAC: Surface Area of Left Caudal Anterior Cingulate, CTLPC: Cortical Thickness Average of Left Posterior Cingulate.

then fed to a standard SVM classifier. The classifier is trained with a Radial Basis Kernel Function and soft margin C is set to 1. For each validation instance, this classification results in a class label and a likelihood score that the label came from the particular class. Optimization of the SVM classifier employed will be studied in the future.

C. Performance Analysis

To assess the capability of the forecasting methods, Mean Percentage Error (MPE) between the original biomarker readings and forecasted biomarker readings is recorded. To quantify the classification results, AUC, Accuracy, Sensitivity and Specificity are recorded.

IV. RESULTS

The aim of this paper is to quantify the impact of longitudinal unimodal and multimodal biomarkers on MCI-to-AD conversion prediction by estimating the future course of the biomarkers. For consolidation of our hypothesis, the experiments were performed for both 1 year and 2 year ahead conversion prediction. For 1 year ahead conversion prediction, 2 year data was used for training whose baseline and M12 predictor values were used to forecast the M24 value. Likewise for 2 year ahead conversion prediction, 3 year data was used for training, and the known baseline and M12 values along with the forecasted M24 readings were collectively used to forecast the M36 value.

Table II describes the most frequent ranks assigned to each of the biomarkers under the unimodal and bimodal experiments. Due to space limitations, only top 5 ranked biomarkers are mentioned. Note that model rank 4 represents a model constructed with top 4 ranked biomarkers and so on. Similar trends in ranking have been reported in many studies including [20], [23],

[19]. **Fig. 2** presents a pictorial representation of top ranked NM and MRI biomarkers. By examining the mean trajectory of each group, it can be noticed that the baseline as well as longitudinal values of the NM marker differ significantly between the MCIP and MCIs group. A considerably sharper decline over 3 years is observed in the cognitive measure. This is contrasting to the mean trajectories for the MRI markers in which little change is observed in subsequent readings making them less useful in the current longitudinal setup. Resultantly, NM (RAVLT and ADAS) take top most ranks in multimodal experiments as well.

We first analyse the effect of single and multiple predictor data modelling. **Tables III** and **IV**, column 9 quantify the extent of similarity between originally observed and forecasted biomarker data. In all of the cases it was observed that SPM produced lower MPE as compared to multiple predictor MPTM models. When multiple predictors were forecasted the error generated accumulates resulting in a higher MPE. It was also detected that MRI measures are modelled most accurately (min. MPE: 2.38% and 2.78% for 1 year and 2 year ahead prediction respectively) as underlying biological changes are trending and irreversible. On the contrary, NM are modelled least accurately and produce highest error (min MPE: 10.02% and 10.81% % for 1 year and 2 year ahead prediction respectively). NM are prone to behavioural change and observer variances. However when NM are combined with MRI, the biomarker forecasting method becomes more precise (min MPE: 3.39% and 3.15% for 1 year and 2 year ahead prediction respectively).

Another aim of this work was to classify an MCI subject as MCIP or MCIs before the onset of dementia. For benchmarking, we first note the performance of our classification system by using the originally observed time domain biomarker trajectories. This provided us with the ground truth (GT) measures to compare the proposed forecasting techniques with. The ground truth metrics are mentioned in **Tables III** and **IV**, Row 2, 5 and 8 for NM only, MRI measure only and NM and MRI combined markers respectively. It was observed that MRI measures alone deliver minimum GT performance in both cases as it was witnessed earlier that MRI are less capable of monitoring longitudinal change. For 1 year ahead prediction (see **Table III**) it can be observed that combined NM and MRI measures deliver maximum GT AUC and Accuracy of 91.86% and 86.23% respectively using top ranked 43 features. A trade-off between sensitivity and specificity was observed due to class imbalance in 1 year ahead prediction data.

While looking at the performance of the proposed classification system using forecasted biomarker trajectories for 1 year ahead prediction (see **Table III**), it was noted that MPTM method using features from both NM and MRI outperform all other methods (AUC = 88.93%, Accuracy = 84.29%, Sensitivity = 70.36%, Specificity = 92.31%). Although the accuracy delivered by SPM and MPTM methods are comparable, a significantly higher AUC was witnessed in MPTM scheme with 45 multimodal biomarkers which is a more reliable metric for class imbalanced sets.

Table IV tabulates the performance metrics for 2 year ahead conversion prediction. It can be detected that highest GT AUC

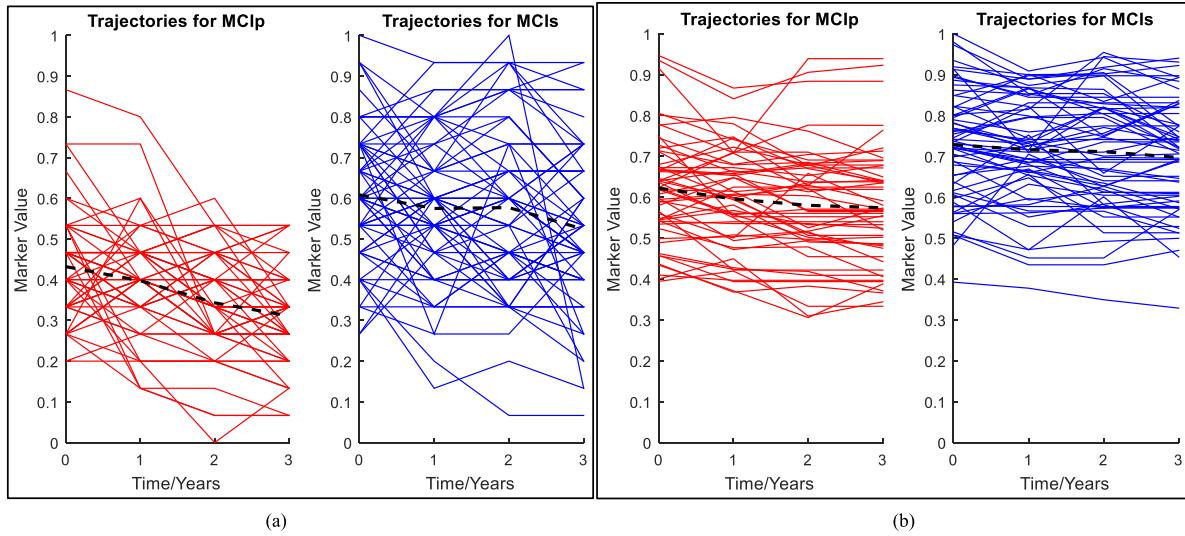


Fig. 2. Feature trajectories over 3 years. (a) Top ranked NM feature (RAVLT), (b) top ranked MRI feature (VLIT). Black dotted line represents the mean trajectory of each group.

TABLE III
1 YEAR AHEAD PREDICTION PERFORMANCE RESULTS

	Model	Rank	AUC	Accuracy	Sensitivity	Specificity	Precision	MPE
NM	Ground Truth	4	90.30	83.38	73.57	89.23	81.98	-
	SPM	6	83.54	79.33	67.50	86.15	76	10.02
	MPTM	6	84.01	78.33	70	83.09	72.33	10.46
MRI	Ground Truth	38	84.86	80.24	58.57	92.31	82.98	-
	SPM	38	81.98	78.38	59.29	89.23	78.48	2.38
	MPTM	38	81.13	77.48	53.93	89.23	76.29	3.09
NM+MRI	Ground Truth	43	91.86	86.23	72.85	93.84	88.57	-
	SPM	10	86.15	83.29	70.00	90.77	82.00	3.39
	MPTM	45	88.93	84.29	70.36	92.31	84.29	3.64

TABLE IV
2 YEAR AHEAD PREDICTION PERFORMANCE RESULTS

	Model	Rank	AUC	Accuracy	Sensitivity	Specificity	Precision	MPE
NM	Ground Truth	4	94.49	89.93	89.09	90.77	89.03	-
	SPM	4	86.90	80.72	85.45	76.92	75.6	10.81
	MPTM	8	88.13	83.26	85.45	81.54	79.41	11.26
MRI	Ground Truth	50	83.89	77.32	73.82	80.00	76.17	-
	SPM	42	79.93	71.41	68.36	73.85	69.17	2.78
	MPTM	42	79.96	68.08	62.91	72.31	65.33	3.28
NM+MRI	Ground Truth	43	91.34	82.36	75.64	87.69	85.35	-
	SPM	43	87.96	80.65	75.64	84.62	81.21	3.15
	MPTM	43	88.97	80.65	75.64	84.62	81.21	3.63

SPM: Single Predictor Model, MPTM: Multiple Predictor Timestamp Model, MPE: Mean Percentage Error.

(94.49%) and accuracy (89.93%) is delivered by using NM again by using a minimal of top 4 features. Similar trend was observed when the forecasting techniques were employed and NM alone delivered a high AUC of 88.13% and highest accuracy of 83.26% (sensitivity: 85.45%, specificity: 81.54%) using MPTM method and all eight markers. A similar conclusion has been drawn in a recent study [26].

The proposed pipeline was also tested for age, gender and education related confounding by using these demographic information as features. However, they were sifted down according

to the ranking criterion used in this study and hence not found to effect the performance.

A few notable conclusions of this study are: 1) Multiple predictors together can deliver a more accurate estimate to future marker values, 2) MRI derived measures alone are not adequate for MCI-to-AD conversion prediction, 3) Short term prediction is better provided by incorporating both NM and MRI derived measures, 4) Long term ahead prediction is better provided by using the behavioural changes encapsulated by NM alone. Brooks and Loewenstein [32] also concluded that biomarkers

TABLE V
PERFORMANCE COMPARISON

Author	Predictors	Dataset size	Follow-up	Validation	AUC (%)	Accuracy (%)
Gomar, 2014 [33]	NM	150 MCIp, 168 MCIs	4 Years	5 Fold	78	-
Gomar, 2012 [25]	NM	116 MCIp, 204 MCIs	2 Years	5 Fold	80	-
Misra, 2009 [27]	MR Images	27 MCIp, 76 MCIs	1.5 Years	LOO	77	81.5
Davatzikos, 2011 [29]	MR Images	69 MCIp, 170 MCIs	1 Year	5 fold	73.4	55.8
Hirnichs, 2011 [35]	MRI	48 AD, 66 CN, 119 MCI	2 Year	LOO	79	72.3
Zhang, 2012 [16]	MR Images	38 MCIp, 50 MCIs	2 Year	10 Fold	76.8	78.4
Arco, 2016 [36]	MRI, NM	73 MCIp, 61 MCIs	1 Year	LOO	79.23	73.95
Proposed	NM, MRI	54 MCIp, 65 MCIs	2 Year	5 Fold	88.93	84.29
Proposed	NM	37 MCIp, 65 MCIs	3 Year	5 Fold	88.13	83.26

are most sensitive to early AD but not optimal for monitoring longitudinal change. A sharper decline in longitudinal cognitive measures than in other biological markers has also been observed by [25] and [33] enabling the selected NM to deliver higher predictive accuracy. [34] Also signifies that many older people may have AD pathology but do not manifest cognitive symptoms during life hence rendering pathological measures less effective. In spite of these literary citations, we suspect that the drop performance delivered by MRI alone while making 2 year ahead prediction may be because of the missing data imputation technique (LOCF) used in this study. MRI derived measures tend to follow a steady decline making LOCF incapable of capturing the changes in predictor values and hence resulting in an inaccurate longitudinal model for MRI measures.

Table V shows that our best prediction model for both 2 year (AUC: 88.93%, Accuracy: 84.29%) and 3 year (AUC: 88.13%, Accuracy: 83.26%) follow-up time performed very favourably with recently published models for ADNI data. Best possible comparison has been stated in terms of biomarkers used, dataset size, follow-up time and cross validation (CV) strategies employed. It can be seen that longitudinal NM and MRI biomarkers embedded in our proposed MPTM framework outperform other classification methods.

V. CONCLUSION

In summary, a machine learning framework for predicting progression from MCI to AD during two and three year period is proposed. Both MRI and NM biomarkers are employed for the said purpose. First, unknown longitudinal biomarker values are estimated from baseline and first follow-up readings of selected multimodal biomarkers using autoregressive parameters generated by longitudinal training data. Three methods for calculating the linear prediction coefficients are proposed which employ either individual predictors or multiple predictors. It is noted that combination of multiple predictors provides more accurate forecasting of future clinical changes. The marker trajectories obtained are then fed to the SVM classifier to segregate MCIp and MCIs. With the proposed framework, the best predictive AUCs of 89.93% and 88.13% are obtained for 2 year and 3 year follow up duration, respectively, which is superior to other recently proposed strategies. In the current study, it is observed that NM and MRI measures combined are better for short term ahead prediction whereas NM alone are more powerful long

term predictors of MCI progression. It is also concluded that MRI measures fail to perform well in a longitudinal setting.

A limitation of this study is the follow-up duration (3 years) which is relatively short to capture the changes in brain morphology due to slowly progressing disease such as AD. The proposed setup will benefit from data from longer follow-up periods. In the future, the performance of this framework will be optimized with feature selection strategies and boosting algorithms. Other missing value imputation strategies will also be incorporated to obtain a larger dataset for more robust modelling, and hence better conversion prediction.

ACKNOWLEDGMENT

Data used in preparation of this article were obtained from the Alzheimer's 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 listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

- [1] D. E. Barnes and K. Yaffe, "The projected effect of risk factor reduction on AD prevalence," *Lancet Neurol.*, vol. 10, no. 9, pp. 819–828, 2011.
- [2] H. Braak and E. Braak, "Neuropathological stageing of Alzheimer-related changes," *Acta Neuropathologica*, vol. 82, pp. 239–259, 1991.
- [3] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dementia*, vol. 7, no. 3, pp. 280–292, 2012.
- [4] C. R. Jack *et al.*, "Hypothetical model of dynamic biomarkers of the AD pathological cascade," *Lancet Neurol.*, vol. 9, no. 1, pp. 119–128, 2010.
- [5] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 280–292, 2011.
- [6] W. R. Markesberry, "Neurpathologic alterations in MCI: A review," *J. Alzheimer's Dis.*, vol. 19, no. 1, pp. 221–228, 2010.
- [7] R. C. Peterson *et al.*, "Mild cognitive impairment: Ten years later," *Arch. Neurol.*, vol. 66, no. 12, pp. 1447–1455, 2009.

- [8] F. Li *et al.*, "A robust deep model for improved classification of AD/MCI patients," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1610–1619, Sep. 2015.
- [9] B. Lei, D. Ni, and T. Wang, "Joint learning of multiple longitudinal prediction models," in *Machine Learning Medical Imaging*. New York, NY, USA: Springer, 2015, pp. 330–337.
- [10] S. Haller *et al.*, "Individual classification of mild cognitive impairment subtypes by support vector machine analysis of white matter DT," *Amer. J. Neuroradiol.*, vol. 34, pp. 283–291, 2013.
- [11] A. Hake *et al.*, "Florbetapir positron emission tomography and cerebrospinal fluid biomarkers," *Alzheimer's Dementia*, vol. 11, no. 8, pp. 986–993, 2015.
- [12] M. Pagani *et al.*, "Volume of interest-based [18F]fluorodeoxyglucose PET discriminates MCI converting to Alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study," *NeuroImage: Clin.*, vol. 7, pp. 34–42, 2015.
- [13] L. M. Shaw *et al.*, "Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging," *Ann. Neurol.*, vol. 65, pp. 403–413, 2009.
- [14] N. Mattson *et al.*, "CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment," *J. Amer. Med. Assoc.*, vol. 302, pp. 385–393, 2009.
- [15] H. I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," *Med. Image Comput. Comput. Assist. Intervention*, vol. 16, no. 2, pp. 583–590, 2013.
- [16] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.
- [17] D. Zhang *et al.*, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [18] Y. Liu *et al.*, "Predicting AD conversion: Comparison between prodromal AD guidelines and computer assisted predictAD tool," *PLoS One*, vol. 8, 2013, Art. no. e55246.
- [19] R. Casanova *et al.*, "Alzheimer's disease risk assessment using large-scale machine learning methods," *PLoS One*, vol. 8, no. 11, 2013, Art. no. e77949.
- [20] M. Ewers *et al.*, "Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance," *Neurobiol. Ageing*, vol. 33, no. 7, pp. 1203–1214, 2012.
- [21] H. Runtti *et al.*, "Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort," *J. Alzheimer's Dis.*, vol. 39, no. 1, pp. 49–61, 2014.
- [22] A. Hall *et al.*, "Generalizability of the disease state index prediction model for identifying patients progressing from MCI to AD," *J. Alzheimer's Dis.*, vol. 44, pp. 79–92, 2014.
- [23] E. Moradi *et al.*, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *NeuroImage*, vol. 104, pp. 398–412, 2015.
- [24] J. Mattila *et al.*, "A disease state fingerprint for evaluation of AD," *J. Alzheimer's Dis.*, vol. 27, pp. 163–176, 2011.
- [25] J. J. Gomar *et al.*, "Utility of combinations of biomarkers, cognitive markers and risk factors to predict conversion from MCI toAD in patients in the ADNI," *Arch. Gen. Psychiatry*, vol. 68, pp. 961–969, 2011.
- [26] I. O. Korolev *et al.*, "Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification," *PLoS One*, vol. 11, no. 2, 2016, Art. no. e0138866.
- [27] C. Misra *et al.*, "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI," *NeuroImage*, vol. 44, no. 4, pp. 1415–1422, 2009.
- [28] C. Davatzikos *et al.*, "Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: The SPARE-AD index," *Brain*, vol. 132, pp. 2026–2035, 2009.
- [29] C. Davatzikos *et al.*, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiol. Ageing*, vol. 32, no. 12, pp. 2322–2327, 2011.
- [30] ADNI, "ADNI General Procedures Manual," [Online]. Available: <https://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI-GeneralProceduresManual.pdf>. Accessed on: Jun. 23, 2015.
- [31] S Minhas, A. Khanum, F. Riaz, A. Alvi, and S. A. Khan, "Early alzheimer's disease prediction in machine learning setup: Empirical analysis with missing value computation," in *Proc. Intell. Data Eng. Autom. Learn.*, Wroclaw, Poland, 2015, pp. 424–432.
- [32] L. G. Brooks and D. A. Loewenstein, "Assessing the progression of mild cognitive impairment to Alzheimer's disease: Current trends and future directions," *Alzheimers Res. Therapy*, vol. 2, no. 5, 2010, Art. no. 28.
- [33] J. J. Gomar, C. C. Goldberg, P. Davies, and T. E. Goldberg, "Extension and refinement of the predictive value of different classes of markers in ADNI: Four year follow-up data," *Alzheimers Dementia*, vol. 10, no. 6, pp. 704–712, 2014.
- [34] C. C. Rowe *et al.*, "Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging," *Neurobiol. Ageing*, vol. 31, no. 8, pp. 1275–1283, 2010.
- [35] C. Hinrichs *et al.*, "Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in ADNI population," *NeuroImage*, vol. 55, pp. 574–589, 2011.
- [36] J. Arco *et al.*, "Short-term prediction of MCI to AD conversion based on longitudinal MRI analysis and neuropsychological tests," in *Innovation in Medicine and Healthcare 2015*. New York, NY, USA: Springer, 2016, pp. 385–394.