A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data

Sidra Minhas, Aasia Khanum, Farhan Riaz, Atif Alvi, Shoab Ahmed Khan, and for the Alzheimer's Disease Neuroimaging Initiative

***Abstract*—The goal of this study is to introduce a nonparametric technique for predicting conversion from Mild Cognitive impairment (MCI)-to-Alzheimer's disease (AD). Progression of a slowly progressing disease such as AD benefits from the use of longitudinal data; however, research till now is limited due to the insufficient patient data and short follow-up time. A small dataset size invalidates the estimation of underlying disease progression model; hence, a supervised nonparametric method is proposed. While depicting a real-world setting, longitudinal data of three years are employed for training, whereas only the baseline visit's data is used for validation. The train set is preprocessed for extraction of two dense clusters representing the subjects who remain stable at MCI or progress to AD after three years of the baseline visit. Similarity between these clusters and the test point is calculated in Euclidean space. Multiple features from two modalities of biomarkers, i.e., neuropsychological measures (NM) and structural magnetic resonance imaging (MRI) morphometry are also analyzed. Due to the limited MCI dataset size (NM: 145, MRI: 52, NM+MRI: 29), leave-one-out cross validation setup is employed for performance evaluation. The algorithm performance is noted for both unimodal case and bimodal cases. Superior performance (accuracy: 89.66%, sensitivity: 87.50%, specificity: 92.31%, precision: 93.33%) is delivered by multivariate predictors. Three notable conclusions of this study are: 1) Longitudinal data are more powerful than the temporal data, 2) MRI is a better predictor of MCI-to-AD conversion than NM, and 3) multivariate predictors outperform single predictor models.**

***Index Terms*—Alzheimer's disease neuroimaging initiative (ADNI), classification, Euclidean space, feature ranking, linear regression, longitudinal data, mild cognitive impairment (MCI), nonparametric.**

I. Introduction

ALZHIEMER'S disease (AD) is the most prevalent form of dementia amongst the elderly and it is the sixth leading cause of death in USA [1]. AD is a slowly progressing brain disorder which cannot be prevented, cured, or slowed as of now. AD can be broadly segregated into three stages. Changes in the brain may begin 20 years or more before the diagnosis of AD, which is categorized as the preclinical stage [2]. The second stage or mild cognitive impairment (MCI) is reached once the internal brain changes start reflecting upon memory and general cognitive function [3]. The last stage is when dementia due to AD has established and is evident through memory loss, reserved cognition, and impaired daily activities [4].

The MCI stage and onslaught of AD was initially captured using the commonly known 1984 NINCDS-ADRAD criteria [5]. According to these guidelines, memory impairment and cognitive function decline were the imperative indicatives of dementia. However, in 2011, the National Institute of Ageing and the Alzheimer's Association recommended the use of advanced imaging like magnetic resonance imaging (MRI) for quantifying structural atrophy of the brain, positron emission tomography (PET) for reckoning metabolic alterations, and cerebrospinal fluid (CSF) for measuring pathological amyloid depositions. Over the past several years, a variety of high-dimensional pattern classification techniques have been designed for identifying MCI versus AD patients based on different modalities of biomarkers either individually [6]–[12] or combined [13]–[15].

However, over the last few years, the researchers have been more interested in identifying the MCI patients who are at a

Manuscript received March 18, 2016; revised June 14, 2016 and August 1, 2016; accepted September 6, 2016. Date of publication September 13, 2016; date of current version September 1, 2017. Data collection and sharing for this work was supported by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense Award Number W81XWH-12-2-0012). ADNI is supported by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through 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 and Development, LLC; Johnson & Johnson Pharmaceutical Research & Development, LLC; Lumosity; and 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 supported ADNI clinical sites in Canada.

S. Minhas is with the Computer Engineering Department, College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan, and also with the Computer Science Department, Forman Christian College, Lahore 54600, 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: aasiakhanum@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 and Technology, Islamabad 44000, Pakistan (e-mail: farhan.riaz@gmail.com; shoabak@gmail.com).

Digital Object Identifier 10.1109/JBHI.2016.2608998

higher risk of developing AD in the future rather than mere segregation of the diseased and the nondiseased. Manly *et al.* [16] calculated that 10% of the MCI patients develop dementia over the course of time. In other individuals, MCI either reverts to normal cognition or remains stable. Efforts are being made to identify the MCI patients at risk of progressing to AD (MCIp) and patients who retain a stable diagnosis of MCI (MCIs). This prediction of whether an MCI subject will develop AD or not reduces to a classification of MCIp versus MCIs. Early realization of the disease may result in superior quality of life for the patients and also benefit in designing of clinical trials in order to find a cure. In the literature, statistical approaches [17], [18] and hazard models [19], [20] have been suggested for risk determination of MCI-to-AD conversion, but these methods are greatly influenced by the sample size available thus limiting their application toward MCI subgroup identification. Numerous high-dimensional and multimodal classification and regression techniques have also been proposed to perform the conversion prediction task. Some of the previous studies include [21]–[24] which are limited in terms of accuracy, and, thus, do not present reliable results for effective usage of the respective methods in the prediction of the said conversion. Various modalities of data have also been verified for their predictive abilities. Moradi *et al.* [25], [26] resort to using brain morphometric information embedded in MRI, Davatzikosa [27] employ both MRI and CSF biomarkers, whereas Gomar *et al.* [28] use a combination of MRI, CSF biomarkers, and cognitive scores to predict a change in diagnostic status of MCI patients. Yet no consensus has been reached about the best predictors for MCI-to-AD conversion prediction task.

Most of the previous studies performed prediction using baseline (BL) data only. Only a few researchers have exploited the longitudinal data for prognostic purposes [27], [29], [30]. Zhang *et al.* [31] and Fan *et al.* [32], [33] have predicted future clinical scores and assigned the prediction label and reported the accuracy from 60% to 80%. On the contrary, we state that predicting future clinical scores and using them for classification is highly restricted due to small number of data and time readings for a slowly progressing disease such as AD. Resultantly, assumptions about underlying longitudinal data distributions, models, and parameters become inaccurate.

In order to achieve individual classification, we propose a supervised nonparametric method to predict the development of AD in the forthcoming years using only the BL readings of an MCI patient. The contributions of this paper are as follows.

1) Comprehensive study of feature selection amongst the neuropsychological measures (NM) and MRI-derived volumetric measures for early AD prediction.
2) Leveraging upon the three year longitudinal training data.
3) Handling of nonconsistent dimensionality of train and test sets.
4) Nonparametric classification of MCIp and MCIs.

Section II defines the ADNI data used in this study. Section III describes the methodology followed. Section IV presents the detailed results of the proposed method, and finally conclusions are presented in Section V.

II. Materials

A. Participants

Data used in the preparation of this study were obtained from the ADNI database.[1] A detailed information regarding the ADNI study and patient inclusion/exclusion criteria is provided in the ADNI General Procedures Manual [34]. Data used in this study were downloaded on December 8th, 2015.

In this study, we restricted our analysis to the MCI subjects recruited in ADNI-1 followed for a period of three years. Description of MCI and health subjects are described elsewhere [28] and also on the website.[2] In short, MCI patients had minimental state exam (MMSE) scores between 24 and 30 (inclusive), a memory complaint, objective memory loss, a clinical dementia rating score of 0.5, absence of significant impairment in other cognitive domains, and preserved activities of daily living. Fourhundred MCI patients were enrolled in ADNI1. Follow ups for these patients were conducted at 6th month (M06), 12th month (M12), 18th month (M18), 24th month (M24), and 36th month (M36), after the BL visit. Over a period of three years, 200 of the MCI patients were reported to be in MCIp group, whereas 100 of them remained stable as MCIs. The remaining 100 patients had unstable diagnosis, thus invalidating their use in this particular study. In an attempt to validate our method, we used the same MCIp and MCIs subjects as identified by Moradi *et al.* [25].

B. Feature Set

1) Biomarker Selection: In ADNI-1, five types of biomarkers were noted for each MCI patient, i.e., NM, clinical function scores, MR images, FDG-PET scans, and biochemical readings via lumbar puncture. A study regarding the dynamics of disease biomarkers in ADNI presented by Caroli *et al.* [35] clearly present that major deterioration of biochemical readings can be observed prior to MCI development, whereas the clinical function is the most differentiating once dementia has developed. This fact is further consolidated in a figure at the ADNI website[3] which advocates that NM and brain imaging (MRI and PET) best capture the transition of MCI-to-AD. Henceforth, in this study, we considered only NM- and MRI-derived morphometric measures to identify cognitive decline and brain atrophy, respectively, for early prediction of AD. Neurodegeneration measured via FDG-PET are excluded from this study due to limited data.

2) Feature Selection: The NM used in this study is selected from cognitive assessment measures explained in ADNI General Procedures Manual [34]. The NM feature set comprised of Alzheimer's disease assessment score (ADAS) and scores of Rey auditory verbal learning test (RAVLT), clock drawing test (CDT), clock copying test (CCT), logical memory immediate recall (LIMM), MMSE, trail making test A (TRAA), and trail making test B (TRAB).

[1] www.loni.ucla.edu/ADNI
[2] www.adni-info.org
[3] http://adni.loni.usc.edu/study-design/background-rationale/

Apart from cognitive scores, MRI volumetric data provided by ADNI were downloaded.[4] These measures were calculated by the University of California, San Francisco, Memory and Ageing Center using the T1 weighted MPRAGE MR scans acquired from a 1.5-T Siemens scanner (dimensions 1 mm × 1 mm × 1 mm, TR: 20 ms, TE: 5 ms). These images were preprocessed for gradient warping, scaling, B1 correction, and N3 inhomogeneity correction by Mayo Clinic [36]–[38]. Cortical reconstruction and volumetric segmentation was then performed using FreeSurfer image analysis suite version 4.3 which is based on the framework provided by Reuter *et al.* [39]. In this framework, an intersubject template was generated by iteratively aligning all input images to a median image using a registration method described in [40]. Using the median image for template generation caters for the within subject bias. Cortical volumetric measures were obtained after iterative topology correction, nonlinear atlas registration, and nonlinear spherical surface registration. The segmented images were passed through an intensive quality control process [41] by Mayo Clinic and provided for use on ADNI website. For this study, five MRI-derived measures selected were the volumes of entorhinal, fusiform, hippocampus, middle temporal lobe (MTL), and the ventricles. These volumes were normalized by intracranial volume to remove interpatient variances.

3) Feature Preprocessing and Analysis: Longitudinal studies are inherently faced with the problem of missed follow-up readings. Currently only those MCI patients "*i*" are included in the study which have some feature value "*f*" for "*j*th" feature over all time points "*t*" and have confirm belonging to either of the *c* classes: MCIp (labeled 1) or MCIs (labeled 0). Henceforward, each feature is referred to as f_{ji_c}. "*t*" is fixed from BL ($t = 0$) to three years ($t = 3$) with an equal interval of one year by combining M06 and M12 for the first year, M18 and M24 for the second year, and M36 for the third year. In this study, missing feature readings are not imputed or calculated in order to avoid induction of bias in the results. Therefore, the dataset size reduced considerably due to one or more missed follow-up readings. Table I lists the demographics of the subjects used in this study stratified according to different modalities of biomarkers under consideration.

Table II summarizes the group wise statistics of all "*N*" features used in this study. The measurement $F_{j_c}(t)$, of each feature of *i* instances belonging to one of the *c* classes at various times *t* is calculated by

$$F_{j_c}(t) = \frac{\sum_{i=1}^{n} f_{ji_c}(t)}{n} \quad \text{for } t = 0-3,\ j = 1-N,\ i = 1-n,\ c = 0/1. \quad (1)$$

The standard deviation of $F_{j_c}(t)$ is also mentioned in Table II. *p* value is used to demonstrate the significance of each feature at various time points. It is worth noting that some features like CDT and volume of MTL demonstrate low significance at initial times but are highly differentiating as time proceeds, hence reinforcing the pertinence of longitudinal data for conversion prediction.

[4]www.ida.loni.usc.edu

TABLE I
SUBJECT INFORMATION

NM	MCIp ($n = 78$)	MCIs ($n = 67$)
Age	74.7 ± 6.88	74.72 ± 6.87
Male/Female	45/33	40/27
Education	16 ± 2.78	15.43 ± 3.1
MRI	MCIp ($n = 27$)	MCIs ($n = 25$)
Age	71.24 ± 7.16	75.22 ± 7
Male/Female	17/10	20/5
Education	16 ± 3.19	14.8 ± 3.29N
MRI+NM	MCIp ($n = 16$)	MCIs ($n = 13$)
Age	71.43 ± 7.2	75.1 ± 6.68
Male/Female	9/7	6/7
Education	15.9 ± 2.5	14.9 ± 3.1

NM = Neurophysiological measures, MCI = mild cognitive impairment, MCIp = MCI subjects progressing to AD in three years, MCIs = MCI subjects staying stable as MCI in three years, and n = number of subjects in each category.

III. METHODS

Fig. 1 presents the overview of the nonparametric AD conversion prediction system presented in this paper. The individual modules are detailed in the following sections.

A. Feature Ranking

The first step in the proposed method was to sift the features according to their ranks in contribution toward MCIp versus MCIs segregation. For this we performed the two sampled student's t-test on normalized BL feature readings whose *p* values indicate about significance of a particular features toward effective diagnostics. We designed a wrapper-based system, where the number of features is incrementally increased to observe the effect on performance.

B. Classification Setup

Given the limited amount of data which is available for performing the experiments, we used leave-one-out cross validation where $n - 1$ data were used for training and 1 instance was used for validation. Three sets of experiments were designed which considered different features for the conversion prediction task. In the first two experiments, each of the biomarkers (NM and MRI) was considered separately, whereas in the third experiment, all features from both biomarkers were combined. For performance analysis and comparison accuracy, sensitivity, specificity, and precision values were recorded.

There are two types of classification methods: Online classification and offline classification which are described as follows.

1) *Online Classification:* In which all features are combined to determine the class label. Combination of multivariate features requires normalization or mapping techniques to cater for bias and scaling issues.
2) *Offline Classification:* In which each feature is considered separately to determine a class label, and later majority voting algorithm (MVA) is employed for final class label.

In this study, we opted for an offline classification. The MVA formulated for this study is defined later. The feature set initially comprised of the highest ranked feature, and later more features were added incrementally according to their ranks.

TABLE II
STATISTICS OF LONGITUDINAL FEATURES AT THREE TIME POINTS

	F(0)			F(1)			F(2)			F(3)		
NM	MCIp	MCIs	p val	MCIp	MCIs	p val	MCIp	MCIs	p val	MCIp	MCIs	p val
ADAS	19.85 ± 5.1	15.6 ± 5.5	<0.0001	21.8 ± 5.16	14 ± 7.1	<0.0001	25 ± 6.0	16 ± 5.9	<0.0001	29.3 ± 8.2	15.9 ± 6	<0.0001
RAVLT	6.6 ± 1.2	8.9 ± 2.7	<0.0001	6.30 ± 1.9	8.7 ± 2.4	<0.0001	5.5 ± 1.8	8.6 ± 2.8	<0.0001	4.8 ± 1.9	7.7 ± 2.8	<0.0001
CDT	4.04 ± 1.2	4.5 ± 0.7	0.0104	4.14 ± 1.0	4.5 ± 0.6	0.0159	3.9 ± 1.1	4.5 ± 0.6	<0.0001	3.7 ± 1.2	4.5 ± 0.8	<0.0001
CCT	4.62 ± 0.6	4.7 ± 0.6	0.3344	4.68 ± 0.6	4.8 ± 0.5	0.1144	4.6 ± 0.6	4.7 ± 0.6	0.3034	4.4 ± 0.9	4.7 ± 0.6	0.0123
LIMM	7.26 ± 2.9	8.24 ± 2.7	0.0433	6.60 ± 3.0	9.9 ± 3.4	<0.0001	5.7 ± 3.1	9.8 ± 3.6	<0.0001	4.4 ± 3.0	9.9 ± 4.3	<0.0001
MMSE	27 ± 1.7	27.8 ± 1.7	0.0046	25.8 ± 2.4	28 ± 1.6	<0.0001	24 ± 2.7	28 ± 2.0	<0.0001	23 ± 3.3	28 ± 2.1	<0.0001
TRAA	42 ± 19	39.6 ± 13	0.3924	45.2 ± 22	41 ± 17	0.2275	49 ± 26	41 ± 24	0.0508	59 ± 37	42 ± 18	0.0004
TRAB	118 ± 59	112 ± 61	0.5384	142 ± 81	105 ± 54	0.0016	162 ± 91	105 ± 53	<0.0001	194 ± 130	121 ± 67	<0.0001
MRI	MCIp ($\times 10^3$)	MCIs ($\times 10^3$)	p val	MCIp ($\times 10^3$)	MCIs ($\times 10^3$)	p val	MCIp ($\times 10^3$)	MCIs ($\times 10^3$)	p val	MCIp ($\times 10^3$)	MCIs ($\times 10^3$)	p val
Entor.	2.9 ± 6.2	3.8 ± 0.54	<0.0001	2.8 ± 0.47	3.8 ± 0.7	<0.0001	2.6 ± 0.45	3.6 ± 0.6	<0.0001	2.5 ± 0.45	3.7 ± 0.7	<0.0001
Fusiform	15.6 ± 1.8	16.8 ± 1.5	0.0142	15.7 ± 2.0	17 ± 1.7	0.0370	15 ± 1.8	16 ± 1.6	0.0028	15 ± 2.0	16.7 ± 2	0.0043
Hipp	6 ± 0.8	6.9 ± 0.98	0.0012	5.9 ± 0.83	6.8 ± 0.9	0.0007	5.8 ± 0.87	6.7 ± 0.9	0.0005	5.7 ± 0.9	6.7 ± 0.9	0.0003
MT L	17.4 ± 1.7	19.4 ± 1.9	0.0002	17.1 ± 1.7	19. ± 1.9	0.0001	17 ± 1.8	19 ± 1.9	<0.0001	16 ± 2.1	19 ± 2.2	<0.0001
Ventricles	45.6 ± 22	45 ± 24.7	0.9286	51.3 ± 25.4	47 ± 27	0.5495	57 ± 28.1	45 ± 28	0.1529	55 ± 28	52.7 ± 29	0.7821

ADAS: Alzheimer's disease assessment score, RAVLT: Reys auditory verbal test, CDT: clock drawing test, CCT: clock copying test, LIMM: immediate recall total score, MMSE: minimental state examination, TRAA: trail making test A, TRAB: trail making test B, Entor: entorhinal, Hipp: hippocampus, and MTL: midtemporal lobe.

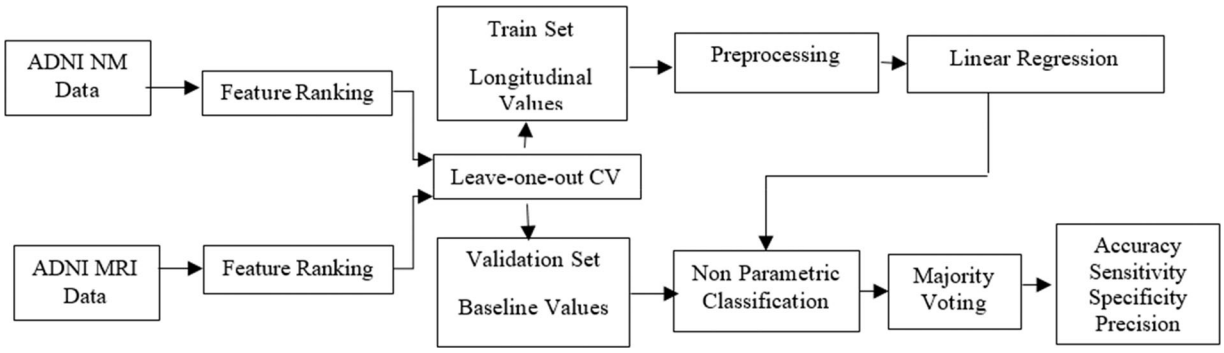

Fig. 1. Overview of the proposed nonparametric technique.

1) Longitudinal Train Set Preprocessing: Many measurements in longitudinal clinical research are based on clinicians' observations, are therefore prone to error and nonuniformity, and, hence, call for denoising measures. In this study, the train set for each feature j and each class c is represented as T_{jc} defined by

$$T_{jc} = \{[f_{ji_c}(0), f_{ji_c}(1), f_{ji_c}(2), f_{ji_c}(3)]\} \quad \text{for } j = 1-N,\ i = 1-n,\ c = 0-1 \tag{2}$$

where T_{jc} is a set of trajectories of the jth feature values of i training instances over three years. For serving the purpose of noise removal, more cohesive trajectory clusters for each class are extracted from within T_{jc} of the respective groups. Let $M_{jc} = f_{jc}(0),\ f_{jc}(1),\ f_{jc}(2),\ f_{jc}(3)$ be the mean feature values of the jth feature of all instances of c class at each time point. The resultant training trajectory cluster for a class c was the set of "l" *nearest* trajectories to M_{jc} in terms of some metric. Closeness metric used in our case was the pairwise Euclidean distance. The selection of "l" was subject to provide sufficient denoising for both majority and minority classes.

2) Train Set Modeling: The training trajectories contained in the local area defined by the previous step are modeled assuming a linear time trend. Linear regression is performed to model each training feature f_{ji_c} versus time t, in the resultant reduced training set according to

$$f_{ji_c} = \beta_o + \beta_1(t) + \epsilon, \text{ for } j = 1-N,\ i = 1-n,\ c-0/1. \tag{3}$$

Here, β_o is the y-intercept, β_1 is the regression coefficient, and ϵ is the error term. The error is measured as the difference in actual value and the modeled value and is reduced in a least squares sense to obtain best estimates for β_0 and β_1.

3) Nonparametric Classification: Most of the sophisticated classification algorithms require the train and test sets to have equal dimensions. Because of this restriction, many of the previous studies have either used one time point data only [21], [22], [25] or they have adopted various algorithms for predicting future feature values for test instances and then used existing classifiers [29]–[31]. Contrarily in this study, longitudinal time point readings are used for training the classifier, while only the BL readings are used validation. Under this setting, a three-year ahead prediction of MCI-to-AD conversion is delivered. The validation point V_j is represented as $V_j = \{f_{jw}(0)\}$, i.e., jth feature value of validation instance w at $t = 0$. The AD conversion label c assigned to the validation point V_j according to

each feature "*j*" is determined by the classification rule

$$c(v_j) = \min\left(d(V_j, T_{j0}), d(V_j, T_{j1})\right)$$

where

$$d(V_j, T_{jc}) = \overline{\|V_j - T_{jc}\|} \quad \text{for } j = 1 - N, i = 1 - n \tag{4}$$

where $d(V_j, T_{jc})$ is the average Euclidean distance between a validation point at BL V_j and the linear training trajectories T_{jc} of each class c [42]. The classification label *c(Vj)* for the *j*th validation feature is assigned as 1 if the validation point is nearer to T_{j1} and 0 otherwise.

4) Majority Voting Algorithm: For determining the final conversion prediction label ω for the validation point *V,* the process above is repeated for each of the *N* features. A vector *C* is maintained as follows:

$$C = [c(V_1), c(V_2), \ldots .., \ c(V_N)] \tag{5}$$

where $c(V_j)$ is the class label assigned to validation point V_j according to (4). Another vector *D* of length *N* is maintained according to

$$D = \ [\min(d(V_1, T_{10}), d(V_1, T_{11})), \ldots \min(d(V_N, T_{N0}), d(V_N, T_{N1}))] \tag{6}$$

where each element of *D* corresponds to the smallest value of the average distance between the validation point and training trajectories of each class. The two vectors *C* and *D* are input to the MVA described in Algorithm 1, *where* $n(C == 1)$ represents the number of 1's contained in the labels vector *C* and vice versa. If more features predict progression, i.e., label 1, the final label ω is set to 1 and vice versa. If an equal number of features predict conversion and stability for a particular validation point, the final decision is based on the distance values calculated in (4). The label corresponding to the minimum value in *D* is selected for ω.

IV. Results

The features ranks assigned to the features under the three experiments are enumerated in Table III. RAVLT was found to be the most discriminating NM, whereas Entorhinal was the top ranked MRI volumetric measure. When both biomarkers are considered collectively, CDT took the first place and TRAB the last.

Fig. 2 pictorially represents the details of one of the longitudinal features considered in this study. Fig. 2(a) shows a simple spaghetti plot for the longitudinal ADAS values of individuals of the two groups. Fig. 2(b) shows the denoised and reduced trajectories resulting from *l nearest trajectory* technique for one of the cross validation folds. It can be visualized that some noisy instances have been deleted and more cohesive trajectories represent both groups. Extensive experimentation was performed to select the value of "*l*." In case of NM, the upper limit for "*l*" was set to 67, i.e., minority dataset size, while other values resulted from 25%, 15%, and 5% reduction in minority dataset size. Hence, the resultant set of value for "*l*" was {50, 57, 63, 67}. Similar approach for MRI-derived measures resulted in a set of {19, 21, 23, 25} for "*l*." However, in case of bimodal data, dataset reduction would result in loss of information; hence "*l*" was set as the minority dataset size, i.e., 13. For benchmarking, results without preprocessing were also recorded. Fig. 3 graphically displays the recorded accuracy measures with respect to varying values of "*l*." It was observed that in case of NM, 15% and 25% data reduction provided better accuracy measures owing to effective noise removal. Thus, $l = 57$ was selected. However, for MRI volumetric measures, better accuracy was obtained when no data removal was performed or when the dataset size was set equal to the minority class. 5%, 15%, and 25% reduction resulted in loss of viable information, and, hence, a lower accuracy measure. "*l*" was set to 25 in this case. For bimodal data, accuracy measure was observed to be same with or without preprocessing.

For independent evaluation of NM and MRI measures, the performance measures for the proposed classifier are listed in Tables IV and V, respectively. The variance of each measure over the cross validation folds is also stated. The effect of incrementally increasing feature set size according to feature ranks can be witnessed also. In case of NM, the most discriminating feature, i.e., RAVLT alone is able to deliver the best performance in terms of accuracy, i.e., 73.79%. However,

TABLE III
Results of Feature Ranking

	Feature Name	Unimodal Ranks	Bimodal Ranks
NM	ADAS	2	6
	RAVLT	1	5
	CDT	6	1
	CCT	3	8
	LIMM	5	2
	MMSE	4	3
	TRAA	7	7
	TRAB	8	13
MRI	Entorhinal	1	4
	Fusiform	4	12
	Hippocampus	3	10
	MTL	2	11
	Ventricles	5	9

ADAS: Alzheimer's disease assessment score, RAVLT: Reys auditory verbal test, CDT: clock drawing test, CCT: clock copying test, LIMM: immediate recall total score, MMSE: minimental state examination, TRAA: trail making test A, TRAB: trail making test B, and MTL: midtemporal Lobe.

Algorithm 1: Majority Voting Algorithm.

```
procedure: MVA
Input: C, D
if n(C == 1) > n(C == 0),
    ω = 1
else if n(C == 1) < n(C == 0),
    ω = 0
else if n(C == 1) == n(C == 0),
    ω = C(1, (D == min(D))
end if
    return ω
    end procedure
```

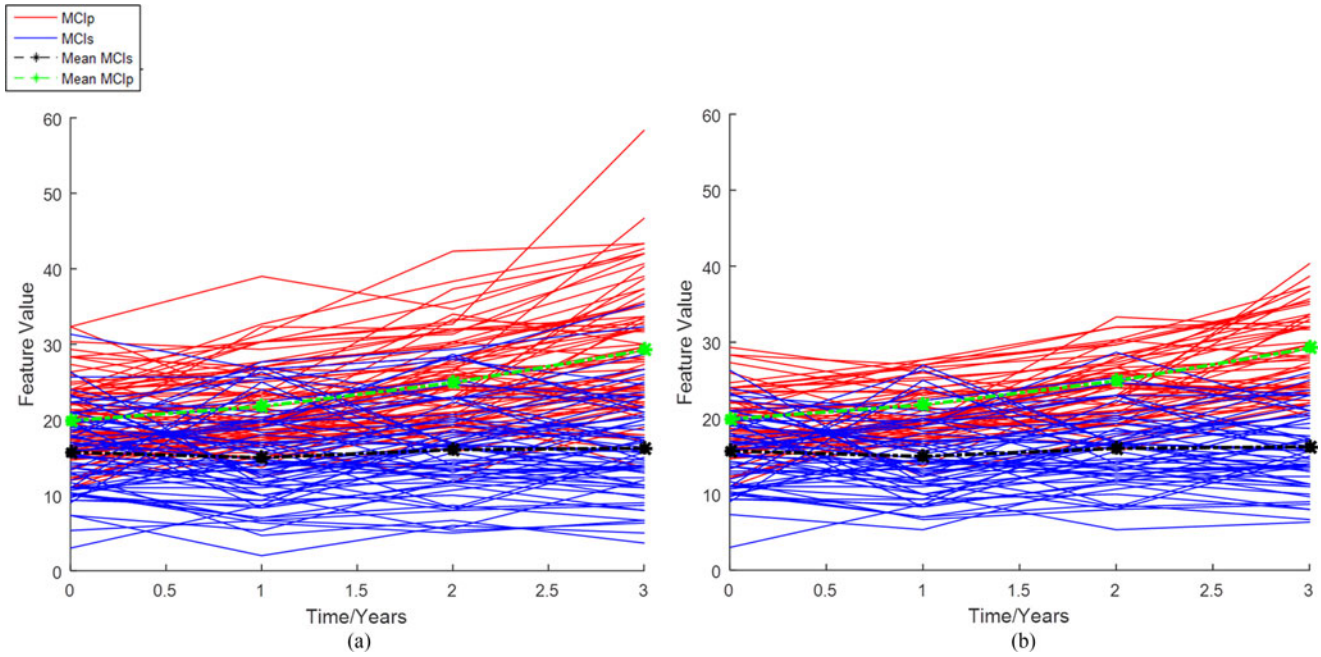

Fig. 2. (a) Spaghetti plot of ADAS versus time. (b) Spaghetti plot of reduced denoised ADAS versus time.

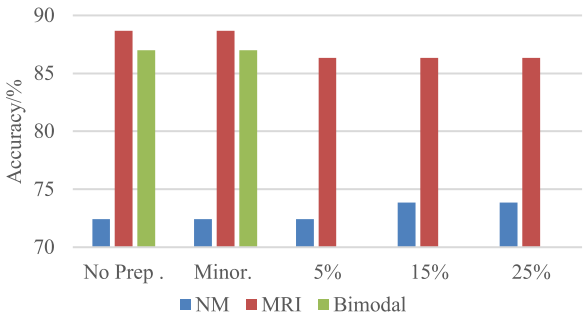

Fig. 3. Accuracy comparison for selection of "*l*" for multimodality data: First with no preprocessing, and then l = minority class size, and 5%, 10%, 15% reduction from the minority class size.

TABLE IV
PERFORMANCE METRICS FOR NM ($l = 57$)

# of Features	Acc (%)	Var.	Sen (%)	Var.	Spe (%)	Var.	Pre	Var
1	73.79	19.47	76.92	17.98	70.15	21.26	75.00	18.99
2	72.41	20.11	76.92	17.98	67.16	22.39	73.17	19.87
3	63.45	23.35	48.72	25.31	80.60	15.88	74.51	19.37
4	66.21	22.53	47.44	25.26	88.06	10.67	82.22	14.95
5	68.97	21.55	61.54	23.98	77.61	17.64	76.19	18.43
6	63.45	23.35	46.15	25.17	83.58	13.93	76.60	18.32
7	63.45	23.35	55.13	25.06	73.13	19.95	70.49	21.15
8	58.62	24.43	46.15	25.17	73.13	19.95	66.67	22.64

Acc = accuracy, Sen = sensitivity, Spe = specificity, Pre = precision, Var = variance.

TABLE V
PERFORMANCE METRICS FOR MRI ($l = 25$)

# of Features	Acc (%)	Var.	Sen (%)	Var.	Spe (%)	Var.	Pre	Var
1	84.62	13.27	85.19	13.11	84.00	14.00	85.19	13.11
2	84.62	13.27	85.19	13.11	84.00	14.00	85.19	13.11
3	78.85	17.01	81.48	15.67	76.00	19.00	78.57	17.46
4	84.62	13.27	88.89	10.26	80.00	16.67	82.76	14.78
5	78.85	17.01	92.59	7.12	64.00	24.00	73.53	20.05

Acc = accuracy, Sen = sensitivity, Spe = specificity, Pre = precision, Var = variance.

by combining other features, a better estimate for specificity and precision was observed but considerably low measures of sensitivity were obtained. A good balance of sensitivity and specificity was observed with RAVLT as conversion predictor.

On the other hand, using one, two, and four top ranked MRI volumetric features delivered the same accuracy of 84.62%. Using single and two top ranked features deliver increased precision and specificity; thus, we conclude that a single feature, i.e., volume of Entorhinal is able to deliver the best performance under the framework presented in this paper.

The performance scores obtained with NM and MRI measures combined are quantified in Table VI. Best results were delivered when the feature set comprised of all features from both biomarkers. Accuracy approximately increased by 15% and 6% from the case when NM and MRI were considered separately for classification, respectively. Under the proposed framework, sensitivity of 87.50%, specificity of 92.31%, and precision of 93.33% was noted. Low variance of performance metrics across the cross validation folds further strengthens the stability of our model.

From the above results, it can be concluded that NM alone cannot be selected for MCI conversion prediction. The cheap and easy to obtain and decipher NM are prone to observer variances. This can be confirmed by larger variance of performance metrics over the validation folds as compared to when MRI-derived measures are used. Brain MRI can record more accurately the underlying changes that occur when a patient transits from MCI-to-AD stage. However, our results show that

TABLE VI
PERFORMANCE METRICS FOR BIMODAL CASE ($l = 13$)

# of Features	Acc (%)	Var.	Sen (%)	Var.	Spe (%)	Var.	Pre	Var
1	62.07	24.38	31.25	22.92	100.00	0.00	100.00	0.00
2	68.97	22.17	62.50	25.00	76.92	19.23	76.92	19.23
3	79.31	17.00	62.50	25.00	100.00	0.00	100.00	0.00
4	75.86	18.97	68.75	22.92	84.62	14.10	84.62	14.10
5	82.76	14.78	75.00	20.00	92.31	7.69	92.31	7.69
6	82.76	14.78	81.25	16.25	84.62	14.10	86.67	12.38
7	86.21	12.32	81.25	16.25	92.31	7.69	92.86	7.14
8	75.86	18.97	68.75	22.92	84.62	14.10	84.62	14.10
9	75.86	18.97	62.50	25.00	92.31	7.69	90.91	9.09
10	86.21	12.32	81.25	16.25	92.31	7.69	92.86	7.14
11	82.76	14.78	75.00	20.00	92.31	7.69	92.31	7.69
12	86.21	12.32	81.25	16.25	92.31	7.69	92.86	7.14
13	89.66	9.61	87.50	11.67	92.31	7.69	93.33	6.67

Acc = accuracy, Sen = sensitivity, Spe = specificity, Pre = precision, Var = variance.

TABLE VII
COMPARISON WITH PREVIOUS TECHNIQUES

Author	Data	Validation method	Follow-up time	Accuracy
Zhang *et al.* [31]	NM, MRI, PET	Tenfold cv	0–24	78.40%
Misra *et al.* [29]	MRI	Leave-one-out cv	0–15	81.50%
Cui *et al.* [44]	NM, CSF, MRI	Tenfold cv	0–24	67.13%
Gomar *et al.* [28]	NM	–	0–48	78.00%
Hu *et al.* [45]	MRI	–	0–36	76.69%
Proposed	NM, MRI	Leave-one-out cv	0–36	89.66%

NM: Neuropsychological measures, CSF: cerebro spinal fluid, and MRI: magnetic resonance imaging.

a combination of multimodal biomarkers convey enhanced performance. Our findings are in accordance with various previous studies [22], [27], [31].

Although direct comparison amongst various MCI-to-AD conversion prediction techniques is bound on dataset size, feature set dimensions, cross validation folds, follow-up time, etc., Table VII presents a brief summary of competing techniques based on longitudinal data ADNI data.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a new approach for MCI-to-AD conversion prediction using the available longitudinal data. Our nonparametric classifier has no limitation for the train and test sets to have equal dimensions. For classifier training, feature values at different time points are employed; however, identification of MCI progressors is based on BL (single time point) readings only. The longitudinal feature variation is modeled using linear regression. The conversion label for an unknown test point is assigned by measuring similarity between the test point and the train trajectories in Euclidean space. Our presented technique out performs other state-of–the-art methods in terms of both accuracy (89.66%) and precision (93.33%) for detecting MCI patients who will develop AD within three years of their BL visit. Specifically, we have shown that 1) from amongst cognitive scores and brain structural atrophy captured via MRI, the later has better ability to detect MCI progressors, and 2) by combining multiple biomarkers, MCI population can be segregated into stables and progressors more accurately and precisely.

In continuation of this study and improving prediction performance, we aim to incorporate missing data computation to obtain a larger dataset size as done by [43]. We would also be validating other biomarkers like FDG-PET and biochemical readings to further enhance the disease forecasting.

ACKNOWLEDGMENT

ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. 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, CA, USA. 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 study. Private sector contributions are facilitated by the Foundation for the National Institutes of Health.

REFERENCES

[1] Alzheimer Association, "2015 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 11, no. 3, pp. 332–384, 2015.
[2] 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, J. Alzheimer's Assoc.*, vol. 7, no. 3, pp. 280–292, 2011.
[3] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the national institute on aging—Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia, J. Alzheimer's Assoc.*, vol. 7, no. 3, pp. 270–279, 2011.
[4] G. M. McKhann, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging—Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia, J. Alzheimer's Assoc.*, vol. 7, no. 3, pp. 263–269, 2011.
[5] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–944, 1984.
[6] 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.
[7] F. Li, L. Tran, K. Han Thung, S. Ji, D. Shen, and J. Li, "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.
[8] 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.
[9] A. Hake *et al.*, "Florbetapir positron emission tomography and cerebrospinal fluid biomarkers.," *Alzhiemers Dementia*, vol. 11, no. 8, pp. 986–993, 2015.
[10] 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.
[11] L. M. Shaw *et al.*, "Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging," *Ann. Neurol.*, vol. 65, pp. 403–413, 2009.
[12] N. Mattsson *et al.*, "CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment," *JAMA*, vol. 302, pp. 385–393, 2009.
[13] D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "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.

[14] D. Zhang, Y. Wanga, L. Zhou, H. Yuan, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
[15] B. Lei, D. Ni, and T. Wang, "Joint learning of multiple longitudinal prediction models," *Mach. Learn. Med. Imag.*, vol. 9352, pp. 330–337, 2015.
[16] J. J. Manly, M. X. Tang, N. Schupf, Y. Stern, J. P. Vonsattel, and R. Mayeux, "Frequency and course of mild cognitive impairment in a multiethinic community," *Ann. Neurol.*, vol. 63, pp. 494–506, 2008.
[17] P. Johnson, L. Vandewater, and W. Wilson *et al.*, "Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease," *BMC Bioinformat.*, vol. 15, no. 16, 2014, doi:10.1186/1471-2105-15-S16-S11.
[18] E. L. Abner, R. J. Kryscio, G. E. Cooper, and G.E. Fardo *et al.*, "Mild cognitive impairment: Statistical models of transition using longitudinal clinical data," *Int. J. Alzheimer's Dis.*, vol. 2012, 2012, doi.org/10.1155/2012/291920(ID 291920).
[19] A. Espinosa *et al.*, "A longitudinal follow-up of 550 mild cognitive impairment patients: Evidence for large conversion to dementia rates and detection of major risk factors involved," *J. Alzheimer's Dis.*, vol. 34, no. 3, pp. 769–780, 2015.
[20] M. Alegret *et al.*, "Cognitive, genetic, and brain perfusion factors associated with four year incidence of Alzheimer's disease from mild cognitive impairment," *J. Alzheimer's Dis.*, vol. 41, no. 3, pp. 739–748, 2014.
[21] R. Casanova *et al.*, "Alzheimer's disease risk assessment using large-scale machine learning methods," *Plos One*, vol. 8, no. 11, p. e77949, 2013.
[22] M. Ewers *et al.*, "Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance," *Neurobiol. Aging*, vol. 33, no. 7, pp. 1203–1214, 2012.
[23] H. Runtti, J. Mattila, M. van Gils, J. Koikkalainen, H. Soininen, and J. Lötjönen, "Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort," *J. Alzheimer's Dis.*, vol. 39, no. 1, pp. 49–61, 2014.
[24] 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, p. e0138866, 2016.
[25] E. Moradi, A. Pepe, C. Gaser, H. Huttunena, J. Tohka, and for the Alzheimer's Disease Neuroimaging Initiative, "Machine learning framework for early MRI-based Alzheimer's conversion," *NeuroImage*, vol. 104, pp. 398–412, 2015.
[26] C. Plant *et al.*, "Automated detection of brain atrophy patterns based on MRI for the prediction of AD," *NeuroImage*, vol. 50, pp. 162–174, 2010.
[27] C. Davatzikosa, P. Bhatta, L. M. Shawb, K. N. Batmanghelicha, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiol. Aging*, vol. 32, no. 12, pp. 2322–2327, 2011.
[28] J. J. Gomar, M. T. Bobes-Bascaran, C. Conejero-Goldberg, and P. Davies, "Utility of combinations of biomarkers, cognitive markers and risk factors to predict conversion from MCI toAD in patients in the ADNI," *Arch. Gen. Psych.*, vol. 68, pp. 961–969, 2011.
[29] C. Misra, Y. Fan, and C. Davatzikos, "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.
[30] C. Davatzikos, F. Xu, Y. An, Y. Fan, and S. M. Resnick, "Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: The SPARE-AD index," *Brain*, vol. 132, pp. 2026–2035, 2009.
[31] D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Predicting future clinical changes of MCI patients," *PLos One*, vol. 7, no. 3, p. 33182, 2012.
[32] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables," *NeuroImage*, vol. 50, pp. 1519–1535, 2010.
[33] Y. Fan, D. Kaufer, and D. Shen, "Joint estimation of multiple clinical variables of neurological diseases from imaging patterns," in *Proc. 2010 IEEE Int. Conf. Biomed. Imaging: From Nano Macro*, Rotterdam, The Netherlands, 2010, pp. 852–855.
[34] ADNI. (2016, Feb.), ADNI general procedures manual. [Online]. Available: https://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf
[35] A. Carolia and G. B. Frisonia, "The dynamics of Alzheimer's disease biomarkers in the Alzheimer's disease neuroimaging initiative cohort," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1263–1274, 2010.
[36] F. Segonne *et al.*, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage*, vol. 22, pp. 1060–1075, 2004.
[37] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, pp. 968–980, 2006.
[38] B. Fischl *et al.*, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, pp. 341–355, 2002.
[39] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *NeuroImage*, vol. 61, no. 4, pp. 1402–1418, 2012.
[40] M. Reuter, H. D. Rosas, and B. Fischl, "Highly accurate inverse consistent registration: A robust approach," *NeuroImage*, vol. 53, no. 4, pp. 1181–1196, 2010.
[41] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.
[42] E. W. Weisstein. (2016, Feb.), Point-line distance-2-dimensional. From mathworld—A wolfram web resource. [Online]. Available: http://mathworld.wolfram.com/Point-LineDistance2-Dimensional.html
[43] 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," *Intell. Data Eng. Autom. Learn.*, vol. 9375, pp. 424–432, 2015.
[44] Y. Cui *et al.*, "Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors," *PLos One*, vol. 6, no. 7, p. e21896, 2011.
[45] K. Hu *et al.*, "Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis," *Neurocomputing*, vol. 175, pp. 132–145, 2016.

Authors' photographs and biographies not available at the time of publication.