Gene Expressions, Hippocampal Volume Loss and MMSE Scores in Computation of Progression and Pharmacologic Therapy Effects for Alzheimer's Disease

Aydin Saribudak, *Member, IEEE,* Adarsha A. Subick, *Member, IEEE,* Na Hyun Kim, Joshua A. Rutta, M. Ümit Uyar, *Fellow, IEEE* and the Alzheimer's Disease Neuroimaging Initiative †

Abstract—We build personalized relevance parameterization method (PReP-AD) based on artificial intelligence (AI) techniques to compute Alzheimer's disease (AD) progression for patients at mild cognitive impairment (MCI) stage. Expressions of AD related genes, mini mental state examination (MMSE) scores and hippocampal volume measurements of MCI patients are obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In evaluation of cognitive changes under pharmacological therapies, patients are grouped based on available clinical measurements and the type of therapy administered, namely donepezil monotherapy and polytherapy of donepezil with memantine. Average leave one out cross validation (LOOCV) error rates are calculated for PReP-AD results as less than 8% when MMSE scores are used to compute disease progression for a 60 month period, and 3% with hippocampal volume measurements for 12 months. Statistical significance is calculated as $p = 0.003$ for using AD related genes in disease progression and as $p < 0.05$ for the results computed by PReP-AD. These relatively small average LOOCV errors and p-values suggest that our PReP-AD methods employing gene expressions, MMSE scores and hippocampal volume loss measurements can be useful in supporting pharmacologic therapy decisions during early stages of AD.

Index Terms—Alzheimer's disease, mild cognitive impairment, AD biomarkers, gene expressions, cholinesterase inhibitors, donepezil, memantine, hippocampal volume loss, mini mental state examination, MMSE scores, artificial intelligence

1 Introduction

Alzheimer's disease (AD) is a dementia type neurodegenerative disorder associated with cognitive impairment, alterations in functional abilities and progressive brain tissue changes such as hippocampal volume loss [1], [2]. AD progression is also characterized with a number of histopathological findings, specifically the presence of extracellular amyloid plaques and intraneuronal neurofibrillary tangles causing loss of functional connectivity between distinct brain regions [3], [4]. Transition period from normal cognition to AD is called mild cognitive impairment (MCI) stage during which initial symptoms of AD such as short-term memory loss due to a decline in cognitive functions can be observed [5], [6].

To slow down cognitive decline caused by glutamatergic and cholinergic signalling impairments, US Food and Drug Administration (FDA) approved a number of drugs for pharmacologic therapies [7]. For example, cholinesterase inhibitors (ChEI) such as donepezil, galantamine and rivastigmine decelerate disease progression for mild to moderate stages, while memantine, an N-methyl-D-aspartate (NMDA) receptor antagonist, targets glutamate for patients with moderate to severe disease [8], [9]. A polytherapy of donepezil and memantine can also be administered to patients in mild to severe stages [7], [10]. However, late administration of pharmacologic therapies in the course of neuropathological processes decrease the effectiveness of therapies [11]. To stabilize cognitive and functional deterioration for elderly individuals with AD symptoms, treatments applied at earlier stages of the disease (e.g., MCI) is crucially important [12].

For diagnosis of AD and drug efficacy analysis, cognitive impairment is measured with a number of assessment tools [13]. A 30-point questionnaire, called Mini-Mental State Examination (MMSE), is widely used to assess disease symptoms including disorientation, loss of memory and attention, and eroded language abilities [14], [15]. The decline in learning and memory skills is also associated with a volume atrophy in hippocampus, where neurofibrillary tangle formation begins at early stages of AD [6], [16]. Repeated 3D magnetic resonance imaging (MRI) scans of whole brain recorded for tracing the rate of hippocampal volume loss can also assess disease progression [17].

The most well-characterized and validated protein biomarkers of AD, $A\beta_{1-42}$ (42 aminoacid form of β-amyloid plaques) and tau proteins (t-tau and its phosphorylated form p-tau), are indicators of tangle formations in cerebrospinal fluid (CSF) [2]. However, the invasive nature of obtaining CSF (i.e., lumbar puncture) limits the repeated use of these biomarkers [18]. To identify biological pathways associated with AD, alterations of gene expressions and their complex interactions are studied with gene expression profiling [11], [19]. Blood-based biomarkers are also identified to evaluate the levels of $A\beta_{1-42}$ and tau proteins [15].

- *A. Saribudak, A. A. Subick, N. H. Kim, J. A. Rutta and M. Ü. Uyar are with the Department of Electrical Engineering, the City College of the City University of New York, New York, NY, 10031. E-mail: uyar@ccny.cuny.edu*
- *† 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*

Impact of protein biomarkers, such as ApoE $\varepsilon 4$ alleles, on AD advancement is studied using information from clinical measurements of a number of research studies [2], [20]. Clinical measurements are used to design linear and non-linear models [13], [21] and curve fitting functions [22] to study the cognitive decline. In addition, therapy effects on AD and the interactions among AD drugs are examined to explore their underlying mechanisms [7], [10], [23]. In this paper, we analyze effect of gene expressions (obtained from blood-based biomarkers) and their correlations on AD progression at early stages of the disease under different drug therapies. We build our personalized relevance parameterization (PReP-AD) methods based on artificial intelligence (AI) techniques from our earlier research [24], [25] to compute impact of genes on cognitive decline. Our contribution to AD research aims to provide support for clinical decision makers in their therapy decisions. Our methods may help a clinician either to postpone administration of a drug therapy if a stable progression is expected for an MCI patient, or schedule and select drug regimen if worsening conditions are expected.

For a number of AD patients in MCI stage, we make use of expressions of a set of genes reported to be related to AD in literature [2], [15], [16], [19], MMSE scores and hippocampal volume measurements stored in Alzheimer's Disease Neuroimaging Initiative (ADNI) database [26]. We compute the impact of individual genes on AD advancement to parameterize disease progression by implementing an algorithm based on differential evolution (DE), a well-known evolutionary algorithm in AI, where a population of coefficient arrays comprising information for the gene impact on AD are evolved through a number of generations to converge to their fittest values over a multi-dimensional space. We introduce two methods, PReP-AD-MMSE employing MMSE scores and PReP-AD-HVL which uses hippocampal volume loss measurements in computation of progression parameters. To assess cognitive changes for MCI patients over time, we classify the patients into six groups based on their clinical measurements recorded in ADNI and the type of therapy they received. Patients given donepezil monotherapy, and polytherapy of donepezil with memantine are in MMSE-MONO and MMSE-POLY groups, respectively, whereas patients with no drug therapy but with MMSE scores are placed in MMSE-NT group. Similarly, HVL-NT, HVL-MONO and HVL-POLY groups are defined for the patients with hippocampal volume measurements. The patients with MMSE scores and volume measurements but have not received any therapy are assigned to both MMSE-NT and HVL-NT groups, and it is likewise for the patients in the therapy groups.

Our PReP-AD-MMSE method utilizes MMSE scores and AD gene expressions to compute the progression parameters, while hippocampal volumes are used with the expressions of AD genes for PReP-AD-HVL. We implement a validation algorithm using leave-one-out-cross-validation (LOOCV) technique [27] to calculate the error rates of the disease progression computed by PReP-AD-MMSE and PReP-AD-HVL methods. For a 60-month period, PReP-AD-MMSE computes parameters with an average LOOCV error rate of 4.8% for MMSE-NT, while average errors are obtained as 6.24% for MMSE-MONO, and 7.75% for MMSE-POLY. The errors for PReP-AD-HVL computed parameters are 1.63% for HVL-NT, 2.66% for HVL-MONO and 2.83% for HVL-POLY, for a 12 month period. Cumulative distributions of LOOCV error rates for both PReP-AD-MMSE and PReP-AD-HVL methods indicate high probabilities of obtaining small error rates.

We calculate a p-value for each of the six patient groups, where expressions of randomly selected genes are used in the null hypothesis, as opposed to using AD related gene expressions, as the inputs to PREP-AD. The p-values calculated for six patient groups are all smaller than 0.05. In addition, we build *cliques* [28] for gene expressions to evaluate statistical significance of correlation among the 51 AD related genes. A clique is defined as a set of genes whose pairwise correlations are larger than a threshold value that is determined based on experimental results. p-value is calculated as 0.003 using the correlations of genes randomly selected from a pool of 20,062 genes. We observe that having a larger average clique size for AD genes results in a smaller number of cliques. Relatively small averages for LOOCV errors and p-values suggest that our PReP-AD methods employing gene expressions, MMSE scores and hippocampal volume loss measurements can be useful in supporting pharmacologic therapy decisions during early stages of AD.

In the initial phases of our research [29], [30], we computed disease progression based on gene expressions for patients with MMSE scores, hippocampal volumes, with or without pharmacologic therapy. In this paper, we first extend our research on pharmacological therapies by using hippocampal volume losses. PReP-HVL-MONO and PReP-HVL-POLY methods are built as extensions of the initial four PReP methods. In addition, we repeat the computation process for several re-sampled AD related gene pools to evaluate the change in error rates as the number of genes increases. We analyze the error rate trend for the re-sampled AD patient pools. We make a comprehensive analysis of our overall findings on computational methods for AD.

The rest of this article is organized as follows. In Sec. 2 outlines related work reported in literature. MMSE and hippocampal volume loss measurements, pharmacologic therapy effects and AD related genes are in Sec. 3. Our algorithms for PReP-AD methods are presented in Sec. 4. Analytical results and errors are in Sec. 5. Statistical significance of using AD related genes is in Sec. 6. A brief discussion and evaluation of the results are in Sec. 7.

2 Related Work

AD is a neurodegenerative disease characterized by the formation of amyloid beta plaques and neurofibrillary tangles [1], [2], [3], [4]. A decrease in $A\beta_{1-42}$ and an increase in t-tau and p-tau proteins measured in CSF are reported as molecular-based biomarkers for AD [18].ApoE allele $\varepsilon 4$ gene, which is associated with a late onset of AD, is also widely accepted as a genetic risk factor [20].As an alternative blood-based biomarkers and patterns in microarray data are analyzed for obtaining gene expressions [15], [19].

In a comprehensive review [1], 68 clinical rating scales were identified to measure domains of AD including cognition, behavior, and quality of life. In assessment of the cognitive stage of a patient (e.g., memory recall and orientation to space and time), MMSE is the most common rating tool, which is a 30-point test requiring 5 to 10 minutes to complete [14].A quadratic and logarithmic curve fitting method [22], a trilinear model [21], and a model with nonlinear functions [13] are introduced in literature to evaluate disease progression based on MMSE scores. In many studies, analysis of AD patients with volumetric MRI has shown consistently greater hippocampal atrophy rates compared to aging control patients [31]. A functional connectivity between hippocampus and a set of disrupted brain regions for AD patients is reported in [32]. In [2], analyzing its relationship with protein

biomarkers, hippocampal volume loss is reported as an indicator of AD pathology and a marker for efficient therapy interventions.

The effect of donepezil, a popular ChEI drug, on social functioning is studied in [23] by evaluating ratings graded based on caregiving time and stress. In [10], researchers investigate the benefits of initiating a memantine therapy for the patients with moderate to severe disease and currently under treatment with donepezil. ChEI and memantine polytherapy is reported as safe and well tolerated based on clinical observations [7]. Therapies administered during early stages of AD, when they are the most effective in slowing cognitive decline, are discussed in [5], [33].

In our research, we develop AI based methods to compute progression of cancer and neurodegenerative diseases using clinical biomarkers such as gene expressions and morphological features of pathology samples. In [34], tumor growth parameters are computed for breast cancer patients using their gene expressions and MRI scans. We study the growth delay and shrinkage effect of anti-cancer compounds on breast cancer cell lines implanted in xenograft mice models in [35], [36]. In [25], [37], to evaluate kidney cancer tumor growth behavior, we compute morphological features of tumor specimen extracted from xenograft mice models using spatial pattern analysis and Voronoi tessellations. In this paper, we compute progression of AD using gene expressions, MMSE scores and hippocampal volume measurements of MCI patients to provide insight for building effective treatment strategies at the early stages of AD.

3 Alzheimer's Disease

Concern over AD continues to deepen, as it has been forecasted that the prevalence of the disease will only increase, with a predicted 1 in 85 persons worldwide living with AD by 2050 [38]. In response to this prediction of dramatic increase, there has been pressure to find better indicators to improve understanding of the disease and to evaluate prospective treatments. Current neuropsychological tests used to monitor the disease and evaluate their efficacy have several limitations [5]. They have a poor test-retest reliability with intraclass correlation coefficients of 0.5 to 0.8 as opposed to being very close to 1. They are often unable to distinguish between treatments that modify the disease progression and the ones that treat only the symptoms (e.g., drugs that enhance cognition without treating neurodegenerative process).

ADNI [26] is a longitudinal multisite study combining genetic information with clinical measurements to find better disease indicators. ADNI1, ADNI GO and ADNI2 are the three phases of the study. During each phase, the disease advancement is tracked over time for hundreds of participants who are recruited across North America. The research data is then shared through USC Image and Data Archive repository (IDA) to assist research studies. IDA also provides data from several other dementia studies such as DOD-ADNI conducted to study the effects of traumatic brain injury and post-traumatic stress disorder on AD, and AIBL study (Australian Imaging Biomarkers and Lifestyle Study of Aging).

The initial phase ADNI1 supports researchers with MRI and PET data, and clinical cognitive and biomarker data for the development of novel methods. The data in the repository is helpful to describe longitudinal changes in brain structure and metabolism. Second phase ADNI GO and third phase ADNI2 are providing clinical measurements and 3-T MRI data from newly enrolled and follow-up patients from ADNI1. In addition, the blood samples (for DNA and RNA extraction) and CSF biomarkers are collected to extract genetic and biochemical information. Within three phases, ADNI helps to determine the relationships among clinical and genetic biomarker characteristics as the disease pathology evolves [39]. For up-to-date information, see www.adni-info.org.

3.1 Clinical Measurements

MMSE is a widely used scale designed to assess cognitive items correlating with underlying neuropathology of AD [22]. To measure cognitive functions such as memory recall and orientation skills to time and space, a 30-point MMSE test with 11 questions in 7 different categories is given to patients [14]. Using MMSE scores collected at distinct time points, patients can be classified as having normal cognition, and mild and severe cognitive impairment.In addition, assessing the rate of change in MMSE over time at regular intervals, patterns of progression can be monitored [13].

Hippocampal volume can be an effective metric for diagnosing AD and predicting the disease progression [31]. The ADNI database records hippocampal volume measurements of AD patients for early and late MCI patients, as well as control patients. 3-Tesla MRI scanners have been used to produce high resolution baseline and follow-up MRI scans. To exemplify MRI imaging process in ADNI, we present sample MRI images in Fig. 1 for a female patient where the shrinkage in hippocampus volume and expansion of the ventricular cavity in the brain from ages 74 to 81 are discernible. The rate of change in hippocampal volume between visits is used as a metric for disease progression.

3.2 Pharmacologic Therapies

Activity of acetylcholinesterase (AChE) enzyme lowered levels of acetylcholine, a chemical acting as a neurotransmitter at the nerve endings and across synapses in the central nervous system [41]. Cholinesterase inhibitors (ChEI) donepezil, galantamine and rivastigmine are administered to AD patients to prevent breakdown of acetylcholine. The neurotransmitter glutamate is related to learning and memory processes by stimulating NMDA receptors [42]. Memantine, an NMDA receptor antagonist, addresses dysfunction in glutamatergic transmission and used especially for the patients with severe disease [8]. Efficacy of memantine and ChEI monotherapy in AD treatment has been widely studied [8], [41], [43]. It is also clinically observed that a polytherapy of ChEI donepezil and memantine is beneficial due to their complementary mechanisms [7]. Predicting drug therapy response for AD patients improves the cost-effectiveness of pharmacologic therapies [33].

In our study, we define six patient groups based on clinical measurements and the therapy administered. MMSE-MONO and MMSE-POLY groups include patients given donepezil monotherapy, and polytherapy of donepezil with memantine, respectively. The patients receiving no drug therapy are put into MMSE-NT. All patients in these three groups have also their MMSE scores. Similarly, HVL-NT, HVL-MONO and HVL-POLY groups are defined for the patients with hippocampal volume measurements. There are overlaps among the groups since the patients with both measurements are registered for both MMSE and HVL groups. For example, 87 patients are in both MMSE-NT and HVL-NT groups.

3.3 Alzheimer's Disease Related Genes

Table 1 lists the genes with their Human Genome Organisation (HUGO) symbol names that are reported in the literature to be related to AD [2], [15], [16], [19]. In our computations, we use

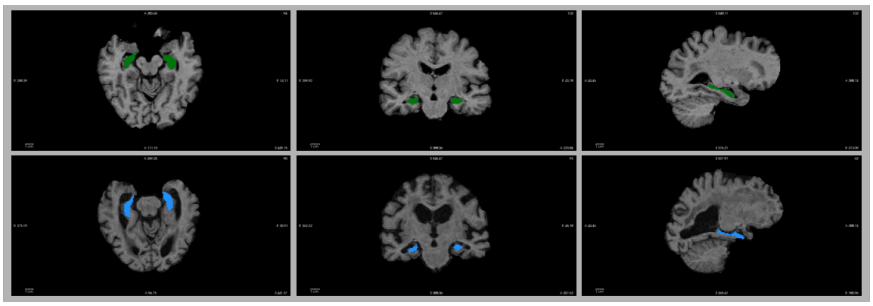

Fig. 1: MRI brain scan of a female MCI patient (ID: 023-S-0887). The skull has been digitally removed and the hippocampus volume has been isolated using the FSL library. The top row has images from the baseline scan (aged 74), whereas the bottom row includes the latest recorded scan (aged 81). From left to right, the figure shows the axial, coronal, and sagittal views of the brain, where hippocampus regions are highlighted in green and blue. (In this paper, brain images are presented using FreeView application of FreeSurfer [40].)

the mean values of the expressions for the genes either belonging to the same family or encoding a member of the same family of protein molecules such as enzymes and receptors. For example, we take into account the mean values of expressions for ACSF2 and ACSF3, which encode a member of Acyl-CoA synthetase family of enzymes. Note that, for simplicity, we use the term gene to refer either to a gene or gene family in the remainder of this paper.

TABLE 1: HUGO SYMBOLS OF AD RELATED GENES

58 genes reported in [15]					
ACSF2	CSF1R	EGFR	LEPROTL1	PDGFD	TNFAIP8
ACSF3	CSF2	HBEGF	MEGF6	PDGFRA	TNFAIP8L1
ADIPOQ	CSF2RA	ICAM1	MEGF8	PDGFRB	TNFAIP8L2
ADIPOR1	CSF2RB	LEP	MEGF9	PDGFRL	TNFAIP8L3
ADIPOR2	CSF3	LEPR	MEGF10	SELPLG	VEGFA
APP	CSF3R	LEPRE1	MEGF11	TF	VEGFB
BDNF	EGFL6	LEPREL1	OLR1	TNFAIP1	VEGFC
CCL5	EGFL7	LEPREL2	PDGFA	TNFAIP2	VWF
CCL22	EGFL8	LEPREL4	PDGFB	TNFAIP3	
CSF1	EGFLAM	LEPROT	PDGFC	TNFAIP6	
33 genes reported in [19]					
ADD3	CHN2	EEF1A1	INHBB	PCSK1N	RAP1GDS1
AGT	CLU	GSTM2	KRT8	PLEKHB1	RPL31
APLP1	CLUH	GSTM2P1	LIMS2	PLP1	TSPAN3
C4BPA	CLUL1	HBB	LRRC4B	PRDX1	
C4BPB	DMPK	HBG2	MT1G	PTS	
CD81	DTNA	IGHMBP2	OSBPL3	RANGAP1	
ApoE (tau activator) and MAPT (microtubule-associated tau) are reported in [2]					
SOD1, SOD2 and SOD3 are reported in [16]					

4 PREP-AD METHOD

In our research, we have studied the effect of gene expressions on predicting tumor mass proliferation. For example, for breast cancer patients, we compute exponential-linear model tumor growth parameters [34], and examined tumor shrinkage and growth delay effects of anti-cancer drugs [35], [36]. In this paper, we build our personalized relevance parameterization methods, namely PREP-AD-MMSE using MMSE scores and PREP-AD-HVL based on hippocampal volume loss measurements. Both methods aim to compute AD advancement to make pharmacological therapy analysis at the early stages of AD.

Deterministic methods are useful to solve problems defined in a multi-dimensional space if the algorithm running time does not exceed a polynomial bound. On the other hand, many well-known evolutionary algorithms in AI converge to the fittest candidate solution within specific constraints in typically low-cost tractable time by means of evolutionary mechanisms such as selection, cross-over and mutation. In this paper, our goal is to bring an efficient approach to compute AD progression based on gene expressions at early stages of the disease. Size of the solution space for this problem grows with a relatively large number of disease related genes (as presented in Sec. 3.3). For this reason, we implement a DE based evolutionary method to compute in polynomial time the parameters reflecting AD advancement and the pharmacological therapy effects. To identify individual gene impact on AD progress, our algorithm explores and exploits the solution space using pre-determined cross-over and mutation parameters and converges to the fittest weight coefficient array.

In Figure 2, we outline PREP-AD workflow. The shapes marked in blue are only for PREP-AD-HVL, green lines are for PREP-AD-MMSE, and orange shapes are for both methods. Set M is defined for the patients with MMSE scores, V is for those with volume measurements, $\overline{V}$ is for patients without volumes, and set $H = V \cup \overline{V}$, where $V \cap \overline{V} = \emptyset$, $M \cap H \neq \emptyset$. n_M is the number of patients in set M, and n_H and n_V are likewise.

In our workflow, the experimental data; MMSE scores of the patients from set M and volume measurements of set V patients, are processed to calculate progression parameters (gray box on the left pane). As shown in mid pane, Algorithm 2 computes parameters for set H patients using HVL parameters of set V patients and gene expressions of set H patients. HVL parameters of set H and MMSE parameters of set M patients are input to Algorithm 3 together with their gene expressions. Algorithm 1 computes the genetic accordance matrix for the training set of patients using their parameters and gene expressions and Algorithm 3 computes the parameters and LOOCV errors for the test patient. This procedure, shown in the box on the right pane, is repeated for all patients.

Note that the MMSE scores and volume measurements recorded during the MCI stage of the disease for all patients are used to calculate the progression parameters. The training progression parameter dataset is then used in PREP-AD algorithms which are implemented based on AI methods to compute disease progression for new patients and the corresponding error rates.

4.1 Genetic Data Vector

We construct genetic data matrix Θ for both PREP-AD methods, using expressions of AD related genes and their correlations from all MCI patients. Genetic data vector Θ_{β_u}, which is a column vector of matrix Θ, can be stated for a patient β_u as

$$\Theta_{\beta_u} = < c, e(g_1), \cdots, e(g_r), \theta_1, ..., \theta_s > \quad (1)$$

where $e(g_i)$ is the expression of gene g_i, θ_k is the k^{th} correlation parameter, for integers $i \in [1\ r]$ and $k \in [1\ s]$. The bias term, c, is a constant to fit a training data into the equation set of many optimization and learning algorithms (e.g., perceptron, gradient descent, and others) with a smaller mean squared error rate [44].

To compute the correlation parameter θ_k, we built an $n\ x\ r$ dimensional E matrix using expressions for r AD related genes from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

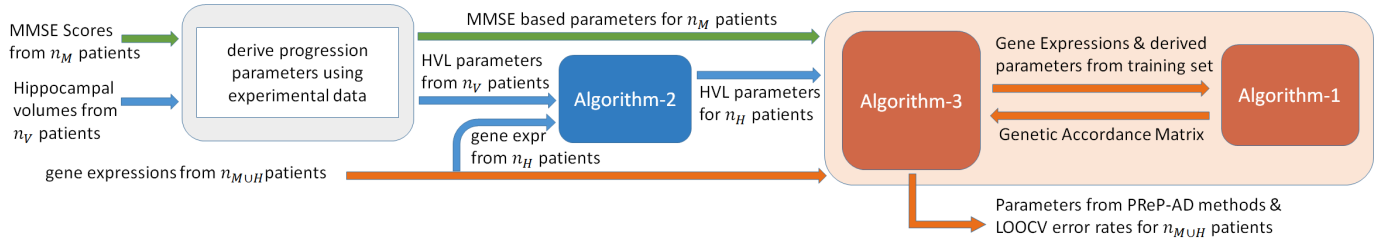

Fig. 2: Progression parameter computation flowchart (blue labelled lines and boxes are for PReP-AD-HVL, green is for PReP-AD-MMSE, orange is for both)

n MCI patients. The i^{th} column vector E_i consists of expression values $e(g_i)$ from n patients. The correlated gene set $\mathfrak{C}_k$, which is used to calculate parameter θ_k, can be expressed as

$$\mathfrak{C}_k = \{g_i \in G: \quad C(E_i,E_j) \geq \tau, \ \forall g_j \in \mathfrak{C}_k \backslash g_i\} \tag{2}$$

where G is the set of genes, $C(E_i,E_j)$ is the correlation between genes g_i and g_j, and τ is the threshold parameter, and correlated gene set $\mathfrak{C}_k$ is a unique set. As an example, let us suppose that, for a sample gene set of $G_{sample} = \{g_1,\ldots,g_{10}\}$, only the pairwise correlations of $C(E_1,E_3)$, $C(E_1,E_5)$, $C(E_1,E_6)$, $C(E_1,E_{10})$, $C(E_5,E_{10})$, and $C(E_6,E_{10})$ are greater than τ. Then, the correlation sets are calculated as $\mathfrak{C}_1 = \{g_1,g_3\}$, $\mathfrak{C}_2 = \{g_1,g_5,g_{10}\}$, and $\mathfrak{C}_3 = \{g_1,g_6,g_{10}\}$. Note that the set $\mathfrak{C} = \{g_1,g_3,g_5\}$ does not exist since $C(E_3,E_5) < \tau$. Once all correlation sets are calculated, the correlation parameter θ_k can be computed as

$$\theta_k = \prod_{g_i \in \mathfrak{C}_k} e(g_i) \tag{3}$$

In this study, the size of all $\mathfrak{C}_k$ are set to 2 since the correlation parameters θ_k calculated for larger sets are negligible when the threshold τ is small (e.g., 0.1). For example, a set consisting of four genes, each with an expression value of 0.01, generates a correlation parameter of 10^{-8}. For this purpose, pairwise correlations are calculated for sets with more than two genes. For the above example, instead of set $\mathfrak{C}_2 = \{g_1,g_5,g_{10}\}$, we prefer to define three sets as $\mathfrak{C}_2 = \{g_1,g_5\}$, $\mathfrak{C}_3 = \{g_1,g_{10}\}$, and $\mathfrak{C}_4 = \{g_5,g_{10}\}$. The threshold constant τ is determined to limit the number of correlation parameters within a predetermined range. By experimenting with many runs, the threshold values giving the most related results are identified: $\tau = 0.7$ was used in PReP-AD-MMSE method, whereas $\tau = 0.6$ for PReP-AD-HVL.

After genetic data vector Θ_{β_u} is built for patient β_u, we compute progression parameters as follows

$$P_{\beta_u} = \mathfrak{A}_{AD} \cdot \Theta_{\beta_u} \tag{4}$$

In Eq. (4), $\mathfrak{A}_{AD} = [a_{ij}]$ is the $\mu \times m$ dimensional genetic accordance matrix, where μ is the number of progression parameters and m is the number of elements of genetic data vector Θ_{β_u}. The relationship between parameters can be expressed as $m = r+s+1$. In our study, to calculate LOOCV errors for all patient, we exclude one test patient at each iteration and build PReP-AD method by making use of gene expressions and clinical information from the remaining patients (i.e., training set). This process generates as number of patients as many methods for each group. The m parameters that we calculate for all patients vary from 52 to 58, and smaller than the size of the the training sets. This results in an overdetermined system of equations at each iteration. In Table 2, we present the number of patients for the given m values computed by PReP-AD for six patient groups. For example, $m = 53$ for 112 patients and $m = 54$ for the remaining 3, all from MMSE-NT group.

TABLE 2: The number of patients from six patient groups for the given m values computed by PReP-AD methods

	Computed m-value						
	52	53	54	55	56	57	58
Patient Group	Number of Patients						
MMSE-NT	-	112	3	-	-	-	-
MMSE-MONO	-	-	1	14	38	28	-
MMSE-POLY	2	11	50	7	-	-	-
HVL-NT	-	-	-	2	170	2	1
HVL-MONO	-	-	7	50	20	3	-
HVL-POLY	1	64	-	-	-	-	-

A linear decline is reported in disease level of AD patients as time progresses during MCI stage [13], [21]. For both PReP-AD methods, we compute linear decline parameters in MMSE scores and hippocampus volume measurements for the MCI patients whose gene expressions and clinical measurements are available in ADNI database. Vector P_{β_u} can be stated using the progression parameters from patient β_u as follows

$$P_{\beta_u} = < \mathfrak{p}_{1,u}, \mathfrak{p}_{2,u}, \cdots, \mathfrak{p}_{\mu,u} > \tag{5}$$

where $\mathfrak{p}_{\nu,u}$ is the ν^{th} parameter and integer $\nu \in [1\ \mu]$.

4.2 Genetic Accordance Matrix for AD

Genetic accordance matrix $\mathfrak{A}_{AD}$ is defined to identify the effects of gene expressions and their correlations on progression of AD. Using genetic information from n MCI patients, each row vector of matrix $\mathfrak{A}_{AD}$ can be computed separately as

$$\mathfrak{A}^{\nu}_{AD} = \widetilde{P}^{\nu} \cdot \begin{bmatrix} C \\ \Theta_D \end{bmatrix}^{-1} \tag{6}$$

where C is a row vector with a bias term c for n patients and Θ_D is the reduced form of genetic data matrix Θ. $\widetilde{P}^{\nu}$ contains progression parameters from n patients

$$\widetilde{P}^{\nu} = (\tilde{\mathfrak{p}}_{\nu,1},\ \tilde{\mathfrak{p}}_{\nu,2},\ \cdots,\ \tilde{\mathfrak{p}}_{\nu,n}) \tag{7}$$

where $\tilde{\mathfrak{p}}_{\nu,u}$ is the estimated value of the ν^{th} progression parameter for patient β_u.

In this paper, we compute six different genetic accordance matrices, one for each patient group as defined in Sec. 3.2. Utilizing AI techniques from our earlier research [24], [25], we implement differential evolution (DE) (as shown in Algorithm 1) to generate the genetic accordance matrix. Inputs of Algorithm 1 include genetic data matrix Θ and progression parameter matrix P for n MCI patients. Elements of P are the rates of change in disease level and calculated by using MMSE scores and hippocampus volume

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

measurements which are collected multiple times at distinct time points. The output is a coefficient vector of the genetic accordance matrix.

In lines 2 to 21 of Algorithm 1, a row vector $\mathfrak{A}^{v}_{AD}$ of matrix $\mathfrak{A}_{AD}$ is computed. An initial population of candidate vectors $\vec{A}^{1}$ is

Algorithm 1 Computation of genetic accordance matrix based on gene expressions, MMSE scores and hippocampal volumes

Require: genetic data matrix Θ, progression parameter matrix P
Ensure: coefficients of genetic accordance matrix $\mathfrak{A}_{AD}$

1: **Start**
2: **for all** progression parameter $\mathfrak{p}_v$ **do**
3: random declaration of initial population $\vec{A}^{1}$;
4: compute initial fitness array $F(\vec{A}^{1})$;
5: $\vec{A}_{v,best} \leftarrow \vec{A}^{1}$; //initialize best pop.
6: $F(\vec{A}_{v,best}) \leftarrow F(\vec{A}^{1})$;
7: **for all** generation σ **do**
8: evolve crossover population $\vec{A}^{\sigma}$ from $\vec{A}_{v,best}$;
9: **for all** candidate i **do**
10: **for all** patient β_u **do**
11: compute parameter $\tilde{\mathfrak{p}}_{v,u}$ using Θ_{β_u};
12: **end for**
13: compute fitness $F(\vec{A}_i^{\sigma})$ of candidate vector;
14: **if** $F(\vec{A}_i^{\sigma}) < F(\vec{A}^{i}_{v,best})$ **then**
15: $\vec{A}^{i}_{v,best} \leftarrow \vec{A}_i^{\sigma}$; //update best pop.
16: and $F(\vec{A}^{i}_{v,best}) \leftarrow F(\vec{A}_i^{\sigma})$;
17: **end if**
18: **end for**
19: **end for**
20: compute $\mathfrak{A}^{v}_{AD}$ from $\vec{A}_{v,best}$; //fittest vector
21: **end for**
22: construct $\mathfrak{A}_{AD}$ matrix using $\mathfrak{A}^{v}_{AD}$ row vectors;
23: **End**

instantiated and the fitness of individuals are calculated as shown at lines 3 and 4. The candidates of initial population are generated randomly as DE based computation does not require an initial knowledge to track the solution space [45]. In lines 5 and 6, the fittest candidate array $\vec{A}_{v,best}$ and the corresponding fitness array $F(\vec{A}_{v,best})$ are initialized. In DE-based computation, a thorough search is applied in global solution space with the *exploration feature* of DE, whereas the local region around the current fittest solution is examined with its *exploitation feature* [45]. Initial population is evolved generation by generation (lines 7 to 19), where, at each generation, crossover population $\vec{A}^{\sigma}$ is calculated from the current best vector population $\vec{A}_{v,best}$. An individual candidate vector $\vec{A}_{\ell}^{\sigma}$ is computed as

$$\vec{A}_{\ell}^{\sigma} = \begin{cases} \vec{A}^{i}_{v,best} + f \cdot (\vec{A}^{j}_{v,best} - \vec{A}^{k}_{v,best}) & \text{if } rand \leq \text{CR} \\ \vec{A}^{\ell}_{v,best} + \varepsilon & \text{otherwise} \end{cases} \quad (8)$$

where i, j *and* k are randomly selected unique vector indices, $f \in [0\ 1]$ is the scaling factor, CR is the cross-over rate, $rand \in [0\ 1]$ is a random number, and $\varepsilon \approx 0$ is defined to exploit the neighborhood of the current best solution. In Fig. 3, the calculation of vector $\vec{A}_{\ell}^{\sigma}$ is represented in a 3-D space for $rand \leq$ CR case.

To compute fitness of candidate vectors, progression parameters are computed for each patient β_u in lines 10 to 12. At line 13, the fitness of the candidate solution $F(\vec{A}_i^{\sigma})$ is computed using square differences of estimated and measured progression

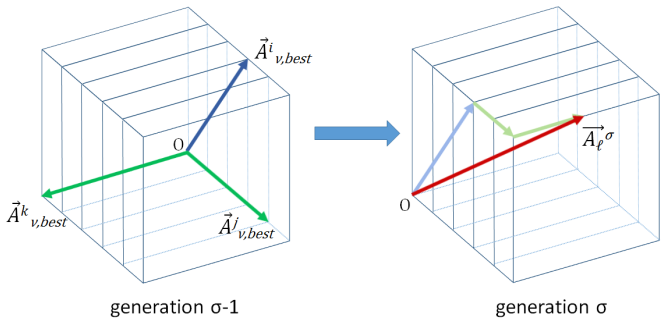

Fig. 3: A sample 3D representation for evolution of a cross-over candidate vector A_{ℓ}^{σ} at generation σ using three randomly selected distinct vectors from generation $\sigma - 1$

parameters, $\tilde{\mathfrak{p}}_{v,u}$ and $\mathfrak{p}_{v,u}$, respectively, for each patient β_u. The fitness function can be formulated as

$$F(\vec{A}_i^{\sigma}) = \frac{1}{n} \cdot \sqrt{\sum_{u=1}^{n} \left(\frac{\mathfrak{p}_{v,u} - \tilde{\mathfrak{p}}_{v,u}}{\mathfrak{p}_{v,u}} \right)^2} \quad (9)$$

Here, to decrease variation among error rates, we inversely weight the least square differences with the magnitude of measured progression parameter. In lines 14 to 17, each candidate vector $\vec{A}^{i}_{v,best}$ of best population and corresponding fitness $F(\vec{A}^{i}_{v,best})$ are updated. With DE based computation, the crossover population of next generation is evolved from the current best population array $\vec{A}_{v,best}$ to converge to the fittest solution. At line 20, the fittest candidate vector of the best population is computed as

$$\vec{\mathfrak{A}}^{v}_{AD} = \underset{\vec{A}_i^{\sigma} \in \vec{A}_{v,best}}{\operatorname{argmin}} \; F(\vec{A}_i^{\sigma}) \quad (10)$$

where $\vec{\mathfrak{A}}^{v}_{AD}$ is the row vector corresponding to parameter $\mathfrak{p}_v$. Genetic accordance matrix $\mathfrak{A}_{AD}$ is constructed with row vectors computed for each progression parameter at line 22. Using matrix $\mathfrak{A}_{AD}$ in Eq. (4), parameter $\mathfrak{p}_{v,u}$ of patient β_u is calculated as

$$\mathfrak{p}_{v,u} = c \cdot a_{v,0} + \sum_{i=1}^{r} a_{v,i} \cdot e(g_i) + \sum_{j=1}^{s} a_{v,(j+r)} \cdot \theta_j \quad (11)$$

where c is the bias term, $a_{v,i}$ is the i^{th} coefficient of v^{th} row of matrix $\mathfrak{A}_{AD}$, $e(g_i)$ and θ_j are the elements of genetic data vector Θ_{β_u} of patient β_u as given in Eq. (1).

4.3 Paired Difference Method

Although gene expressions and MMSE scores were available for most of the MCI patients in ADNI database, the hippocampal volume measurements were not recorded for some of them. We introduce Algorithm 2 exploiting genetic similarities among the patients to compute the hippocampal volume loss rate for the patients whose volume measurements were not available. *Manhattan distance* metric [46] was used to calculate genetic similarities among patients. The volume loss rates computed by Algorithm 2 are then input to Algorithm 1 to generate genetic accordance matrix for PREP-AD-HVL method.

In Algorithm 2, the patients with and without hippocampal volume measurements are placed into sets V and $\overline{V}$, respectively. We initialize gene expression matrix E and calculate progression parameters using volume information of ADNI for set V patients (the gray box in Fig. 2). Volume loss rate difference array L_{Δ}

and gene expression difference matrix E_Δ are generated in lines 7 to 12. At line 9, to generate array E_Δ^{ij} for patients β_i and β_j, the difference between expressions of gene g_ϕ is computed as $\left|e(g_{\phi,i}) - e(g_{\phi,j})\right|$. The rate difference L_Δ^{ij} is computed as $\left|\ell_{v,i} - \ell_{v,j}\right|$ where $\ell_{v,i}$ and $\ell_{v,j}$ are the loss rates for patients β_i and β_j. Then, with DE based methods [25], weight coefficients are computed for the elements of matrix E_Δ (lines 14 to 22). At line 14, we generate an initial population of weight vector $\vec{w}^{\,1}$ and the fitness of initial array $F(\vec{w}^{\,1})$ is computed at line 15. We evolve a crossover population $\vec{w}^{\,\sigma}$ at each generation σ from the current best population $\vec{w}_{best}$. The fitness is computed for $\vec{w}^{\,\sigma}$ at line 18, and each candidate $\vec{w}_{i,\,best}$ of best population is updated based on its fitness value $F(\vec{w}_{i,\,best})$ in lines 19 to 21. Here, DE is implemented for the relationship between gene expression and volume loss rate differences which can be formulated as

$$L_\Delta = \left[\ \Gamma \mid E_\Delta\ \right] \cdot \vec{w}_{fit} \tag{12}$$

where Γ is a column vector with negative bias terms, E_Δ is a $\binom{n}{2} \times r$ dimensional matrix for n patients and r genes. $\vec{w}_{fit}$ is the fittest weight vector and calculated as

$$\vec{w}_{fit} = \underset{\vec{w}_{i,\,best} \in \vec{w}_{best}}{\operatorname{argmin}} F(\vec{w}_{i,\,best}) \tag{13}$$

Finally, using the fittest weight vector $\vec{w}_{fit}$, we compute hippocampal volume loss rates for the patients from set $\overline{V}$ (lines 24 to 33). At line 24, we randomly select a patient $\beta_{\overline{u}}$ from set $\overline{V}$. Then, paired difference $f_\delta(\beta_{\overline{u}}, \beta_i)$ of gene expressions is computed for patient $\beta_{\overline{u}}$ and each patient β_i from set V in lines 26 to 28 as

$$f_\delta(\beta_{\overline{u}}, \beta_i) = w_0 \cdot c + \sum_{\phi=1}^{|G|} w_\phi \cdot |e(g_{\phi,\overline{u}}) - e(g_{\phi,i})| \tag{14}$$

where G is the set of AD related genes. At line 29, most genetically similar patient β_u from set V is found for patient $\beta_{\overline{u}}$ as

$$\beta_u = \underset{\beta_i \in V}{\operatorname{argmin}}\ f_\delta(\beta_{\overline{u}}, \beta_i) \tag{15}$$

At line 30, corresponding hippocampal volume loss rate $\ell_{v,\overline{u}}$ for patient $\beta_{\overline{u}}$ is calculated as

$$\ell_{v,\overline{u}} = \ell_{v,u} \cdot \begin{cases} [1 + \eta \cdot f_\delta(\beta_{\overline{u}}, \beta_u)] & \text{if } \gamma \leq 0.5 \\ [1 - \eta \cdot f_\delta(\beta_{\overline{u}}, \beta_u)] & \text{otherwise} \end{cases} \tag{16}$$

where $\gamma \in [0\ 1]$ is a random number and η is the normalization constant assigned to calculate rate $\ell_{v,\overline{u}}$ to assure that its difference with $\ell_{v,u}$ of the genetically similar patient β_u is in a pre-determined range (e.g., $(1 \pm 0.25) \cdot \ell_{v,u}$). Patient $\beta_{\overline{u}}$ is removed from set $\overline{V}$ at line 31 and inserted into V at line 32. The process is repeated until the rates are calculated for all patients in $\overline{V}$.

4.4 Cross Validation for PReP-AD

Progression parameters are computed using hippocampal volume measurements in Algorithms 1 and 2 for PREP-AD-HVL and with MMSE scores in Algorithm 1 for PREP-AD-MMSE. We implement a validation algorithm based on LOOCV method [27] to compare the progression parameters computed by PREP-AD methods with the clinical measurements reported in ADNI database.

In Algorithm 3, genetic data matrix Θ and progression parameter matrix P are initialized in lines 2 and 3. For a test patient β_u, genetic data vector Θ_{β_u} and parameter vector P_{β_u} are constructed in lines 6 and 7, respectively. At line 9, using genetic data vectors from remaining $(n-1)$ patients, reduced genetic data matrix Θ_R is computed as a training set. Similarly, reduced progression parameter matrix P_R of training set is computed at line 10. At line 12, using matrices Θ_R and P_R in Algorithm 1, genetic accordance matrix $\mathfrak{A}_{AD}$ is computed for the training set patients. Estimated parameter vector, $P_{\beta_u}^{est}$, is computed for patient β_u at line 13. At line 14, disease progression is calculated for PREP-AD-MMSE using estimated MMSE scores or for PREP-AD-HVL method using estimated volume measurements. In lines 16 to 18, average LOOCV error rate $\varepsilon(t_i)$ is computed for each time point t_i, which can be formulated as

$$\varepsilon(t_i) = \frac{1}{n} \cdot \sum_{u=1}^{n} \left| 1 - \frac{\widetilde{\Delta}_{\beta_u}(t_i)}{\Delta_{\beta_u}(t_i)} \right| \tag{17}$$

In Eq. (17), for MMSE scores, $\Delta_{\beta_u}(t_i)$ is the amount of change at time point t_i recorded in ADNI database for a patient β_u with respect to the baseline measurement. $\widetilde{\Delta}_{\beta_u}(t_i)$ is the difference between the score computed by PREP-AD-MMSE. Similarly, for hippocampal volume measurements, $\Delta_{\beta_u}(t_i)$ is the change at time point t_i taken from the database (or from Algorithm 2) compared

Algorithm 2 Genetic similarity based computation of hippocampal volume loss rate for patients without measurements

Require: gene expressions and hippocampal volumes from ADNI
Ensure: estimated hippocampal volume loss rate

1: **Start**
2: //initialization
3: initialize sets V and $\overline{V}$;
4: initialize gene expression matrix E for MCI patients;
5: calculate parameter $\mathfrak{p}_{v,u}$ for each patient $\beta_u \in V$;
6: //construct L_Δ array and E_Δ matrix
7: **for** $i = 1$ to $|V| - 1$ **do**
8: **for** $j = i + 1$ to $|V|$ **do**
9: calculate difference array E_Δ^{ij} for gene pair;
10: calculate parameter difference L_Δ^{ij};
11: **end for**
12: **end for**
13: //differential evolution process
14: random declaration of initial population $\vec{w}^{\,1}$;
15: compute fitness array $F(\vec{w}^{\,1})$ for initial population;
16: **for all** generation σ **do**
17: evolve crossover population $\vec{w}^{\,\sigma}$ from $\vec{w}_{best}$;
18: compute fitness array $F(\vec{w}^{\,\sigma})$ for candidate vectors;
19: **for all** individual i **do**
20: calculate $\vec{w}_{i,\,best}$ and $F(\vec{w}_{i,\,best})$;
21: **end for**
22: **end for**
23: //paired difference computation for $\overline{V}$
24: **repeat**
25: get patient $\beta_{\overline{u}} \in \overline{V}$; //random selection
26: **for** $i = 1$ to $|V|$ **do**
27: calculate paired difference $f_\delta(\beta_{\overline{u}}, \beta_i)$ using $\vec{w}_{fit}$;
28: **end for**
29: calculate patient $\beta_u \in V$ genetically similar to $\beta_{\overline{u}}$;
30: calculate hippocampal loss rate $\ell_{v,\overline{u}}$ for patient $\beta_{\overline{u}}$;
31: $\overline{V} \leftarrow \overline{V} - \{\beta_{\overline{u}}\}$;
32: $V \leftarrow V \cup \{\beta_{\overline{u}}\}$;
33: **until** $|\overline{V}| = 0$
34: **End**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Algorithm 3 Cross validation algorithm for MCI patients

Require: prog. parameters and gene expressions from n patients
Ensure: LOOCV error rates for n patients

Start
  initialization of matrix Θ with genetic data vectors;
  initialization of matrix P with progression parameters;
  for $u = 1$ to n **do**
    //define for test patient β_u
    construct genetic data vector Θ_{β_u};
    construct parameter vector P_{β_u};
    //define for training set of patients
    compute matrix with gene expressions: $\Theta_R = \Theta - \Theta_{\beta_u}$;
    compute parameter vector: $P_R = P - P_{\beta_u}$;
    //compute disease progression
    compute $\mathfrak{A}_{AD}$ matrix using Θ_R and P_R in Algorithm 1;
    compute estimated vector $P^{est}_{\beta_u}$ using $\mathfrak{A}_{AD}$ in Eq. (4);
    compute disease progression using vector $P^{est}_{\beta_u}$;
  end for
  for all time point $t_i \in$ T **do**
    calculate average LOOCV error rate $\varepsilon(t_i)$;
  end for
End

to its baseline value. For this case, $\widetilde{\Delta}_{\beta_u}(t_i)$ represents the volume difference computed by PREP-AD-HVL.

LOOCV based Algorithm 3 helps avoiding over-fitting of results computed to the existing patient data because it assigns a training set for each patient at each iteration for both PREP-AD methods. The accuracy of disease progression results can be further improved with larger number of MCI patients.

5 Analytical Results

With our PREP-AD methods, we have computed disease progression for MCI patients whose gene expressions and clinical measurements are provided in ADNI database. Six patient groups, namely MMSE-NT, MMSE-MONO, MMSE-POLY, HVL-NT, HVL-MONO and HVL-POLY, are defined for two PREP-AD methods based on clinical measurements and the type of therapy administered. The patients with MMSE scores available in ADNI are placed in one of the three MMSE- groups based on their therapy, whereas the patients with hippocampal volume measurements are in one of the three HVL- groups. MMSE-NT group contains 115 patients without drug administration, MMSE-MONO group has 81 patients who were given monotherapy of ChEI donepezil, and MMSE-POLY has 70 patients who received polytherapy of donepezil and memantine. 175 patients given no therapy are in HVL-NT, whereas HVL-MONO consists of donepezil administered 80 patients and HVL-POLY has 65 patients under polytherapy. Note that there is some overlap among the groups since some patients have both MMSE scores and hippocampal volume measurements available. 87 patients are in both MMSE-NT and HVL-NT groups, while 77 patients are common for MMSE-MONO and HVL-MONO, and HVL-POLY is a subset of MMSE-POLY.

Patients who received AD related drugs have followed a consistent schedule. For MMSE-MONO and HVL-MONO, all patients except one were administered donepezil at a daily dose of 5 to 10 mg before 36^{th} month of the therapy (remaining one patient started at month 60). For MMSE-POLY and HVL-POLY, all patients except five were given donepezil with a dose of 5 to 20 mg/day starting before month 36, and memantine with a dose of 10 to 30 mg/day starting before 48^{th} month. The same five patients are the outliers for both polytherapy groups with the following therapy schedules: one patient received memantine at a dose of 5 mg/day (too small compared to other patients) while four started their therapy after 48^{th} month, which is outside the timeframe of our study.

TABLE 3: Number of patients with increasing, stable, moderately and significantly decreasing cognition computed by PREP-AD

Disease Progress	MMSE-NT	MMSE-MONO	MMSE-POLY
Increasing	3	13	11
Stable	76	32	15
Decreasing	36	36	44
	HVL-NT	HVL-MONO	HVL-POLY
Increasing	3	15	6
Stable	29	26	14
Decreasing	98	22	16
Significant Decr.	45	17	29

At the early stages of AD, a linear decline in disease progression has been reported in several studies [20], [21]. In this paper, we compute the amount of linear change to reflect disease progression during the MCI stage based on clinical measurements from the ADNI. Since volume measurements are not recorded in ADNI for all MCI patients, we first compute the volume based progression parameters with Algorithm 2 for each HVL- group. Then, genetic accordance matrix $\mathfrak{A}_{AD}$ is computed for all six groups by Algorithm 1, and corresponding LOOCV error rates by Algorithm 3. With PREP-AD-MMSE, for a 60 month period of disease progression, we obtain an average LOOCV error rate of 4.8% for MMSE-NT, 6.24% for MMSE-MONO, and 7.75% for MMSE-POLY groups. PREP-AD-HVL results have error rates of 1.63%, 2.66% and 2.83% for HVL-NT, HVL-MONO and HVL-POLY groups, respectively, for a 12 month of disease progression.

5.1 Disease Progression by PReP-AD-MMSE

PREP-AD-MMSE method computes disease progression of MCI patients using their gene expressions and MMSE scores. For the patients in the three MMSE- groups, either a decreasing, stable or an increasing progress is computed in cognitive behavior, where a stable progression is defined as a change in MMSE scores in the range of ± 2 points for a 60 month period. Table 3 lists the number of patients for each of the disease progress categories computed by both PREP-AD methods. We observe that, for a larger percentage of patients from MMSE-MONO and MMSE-POLY groups, our PREP-AD-MMSE computes a decrease in cognition compared to the no-therapy group of MMSE-NT. This is likely because the patients from these groups were administered therapy as a result of an increase in their deteriorating condition. However, PREP-AD-MMSE computes an increase in MMSE scores for some patients under therapy (13 from MMSE-MONO, 11 from MMSE-POLY), whereas it is stable for MMSE-NT, which may be due to a positive response to the drug therapy.

The boxplot graphs in Fig. 4 illustrate the distribution of error rates, where the interquartile range (IQR) is shown with a box, the red horizontal line denoting the median value, horizontal black lines for maximum and minimum values and red crosses for the outliers. The distribution of LOOCV error rates for the results from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 4: LOOCV error rate statistics for PREP-AD-MMSE results

	PREP-AD-MMSE Statistics				
Month	12	24	36	48	60
	MMSE-NT Group				
No of Patients	115	98	88	69	33
Mean	3.47	4.32	4.25	5.49	6.25
Median	2.59	2.97	3.17	4.53	5.05
Std Dev	2.82	4.28	3.83	4.55	4.28
IQR	3.05	4.89	4.32	5.37	5.11
No of Outliers	6	7	3	3	2
	MMSE-MONO Group				
No of Patients	78	70	62	48	27
Mean	4.28	5.46	6.99	8.40	9.62
Median	2.79	4.54	5.79	6.79	8.01
Std Dev	3.72	4.50	5.93	6.88	6.94
IQR	5.23	4.71	8.87	11.29	9.42
No of Outliers	2	3	-	-	-
	MMSE-POLY Group				
No of Patients	70	68	56	38	27
Mean	4.45	6.83	9.01	10.71	12.84
Median	3.64	4.73	8.72	10.66	11.60
Std Dev	3.69	5.58	6.63	7.81	8.67
IQR	4.22	6.55	10.72	14.04	9.42
No of Outliers	2	2	-	-	-

PREP-AD-MMSE in first row of Fig. 4 is for a 60 month period. The statistics of error rates are given in Table 4, where the median, IQR and the number of outliers correspond to the boxplot graphs. Although error rates increase as time progresses for all patient groups, we observe that median error rate is lower than 10%, except for months 48 and 60 in MMSE-POLY group.

Fig. 5 presents disease progression for six selected patients computed by PREP-AD-MMSE (shown in blue) and the corresponding MMSE scores reported in ADNI database (shown in brown). These patients are selected to represent stable and decreasing disease progression. For example, stable group patients 114-S-1118 and 018-S-2133 from MMSE-NT group have average error rates of 3.34% and 8.15%, respectively. For MMSE-MONO group, we selected patient 014-S-0563 with an average error rate of 4.84% and 003-S-1057 with 8.13%. Selected patients for MMSE-POLY group are patient 011-S-2274 with an average error rate of 3.49% and 011-S-1080 with 8.64%.

5.2 Disease Progression by PReP-AD-HVL

Disease progression for MCI patients based on their gene expressions and hippocampal volume measurements is computed with PREP-AD-HVL method for a 12 month period. Volume measurements are available for 26 of 175 patients from HVL-NT, for 17 of 80 patients from HVL-MONO, and for 24 of 65 patients from HVL-POLY group. For the patients without hippocampal volumes recorded in ADNI, Algorithm 2 computes the volumes based on genetic similarities. Once the volume measurements are obtained for all patients, Algorithms 1 and 3 compute disease progression and the corresponding LOOCV error rates.

In Table 3, the number of patients with increasing, stable, moderately and significantly decreasing cognition computed by PREP-AD-HVL are listed. A *stable* progression is defined for a total hippocampal volume change of 30 mm^3 or less within a 12 month period, while a *moderate* decrease is smaller than 80 mm^3 for the same period. Similar to the results for MMSE-MONO and MMSE-POLY groups, there are more patients with a decreasing cognition compared to the ones with stable or increasing behavior in HVL groups. As can be seen in Table 3, the number of patients with increasing cognition computed by PREP-AD-HVL is the largest for HVL-MONO (15 out of 80 patients) and HVL-POLY (6 out of 63 patients) groups compared to HVL-NT (3 out of 175 patients), which is most likely as a result of the drug therapy.

TABLE 5: LOOCV error rate statistics for PREP-AD-HVL results

	PREP-AD-HVL Statistics					
Month	No Pat.	Mean	Median	Std	IQR	No Out.
	HVL-NT Group					
6	175	0.95	0.46	1.03	0.87	21
12	175	1.63	0.91	1.81	1.50	21
	HVL-MONO Group					
6	80	1.65	1.18	1.67	1.58	6
12	80	2.66	2.07	2.35	2.80	4
	HVL-POLY Group					
6	65	1.94	1.42	1.90	1.76	4
12	65	2.83	2.47	2.31	2.46	5

Error rate distributions computed for the PREP-AD-HVL results for months 6 and 12 for patients from HVL-NT, HVL-MONO and HVL-POLY groups are given in Fig. 4. The statistics calculated from boxplot graphs, and the mean and standard deviation values are given in Table 5. We observe in Table 5 that HVL-NT group error rates are smaller than those for therapy groups. This is because variation in disease progression for patients without therapy is the smallest (i.e., disease progression results are more similar). The larger range of volume measurements (i.e., approximately 1,500 - 3,000 mm^3) compared to the smaller scale of MMSE scores (i.e., 0 to 30 points) contribute to smaller HVL group errors.

For selected six patients representing stable and decreasing cognition in each group, disease progression results for PREP-AD-HVL are shown in the third and fourth columns of Fig. 5, where the computed results are in blue and the volumes reported in ADNI in carmine. For patients 128-S-0770 and 137-S-0800 from HVL-NT, the average error rates are calculated as 2.37% and 3.74%, respectively. For HVL-MONO, the error is 1.66% for 027-S-0644 and 2.42% for 073-S-0746. It is 2.07% for 037-S-0501 and 0.72% for 027-S-0408, both from HVL-POLY group.

5.3 Error Rate Analysis

5.3.1 Cumulative Distribution of Error Rates

To examine the accuracy of AD progression results from PREP-AD methods, we generate cumulative distribution functions (CDFs) of LOOCV error rates. For a given cut-off error rate of ε, fractional form of CDF $F_c(\varepsilon)$ is $F_c(\varepsilon) = \sum_{u=0}^{\varepsilon} Pr(u)$, where $Pr(u)$ is the probability of error rate u. Cumulative Weibull distribution function $W_c(\varepsilon|\lambda,\omega)$, an approximation for fractional form of a cumulative distribution, is $W_c(\varepsilon|\lambda,\omega) = 1 - e^{-(\frac{\varepsilon}{\lambda})^{\omega}}$, where ω is the shape parameter and λ is of scale (i.e., dispersion of probability distribution) [47]. A Weibull curve similar to a fractional function implies the significance of empirical observations as they fit into a normal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

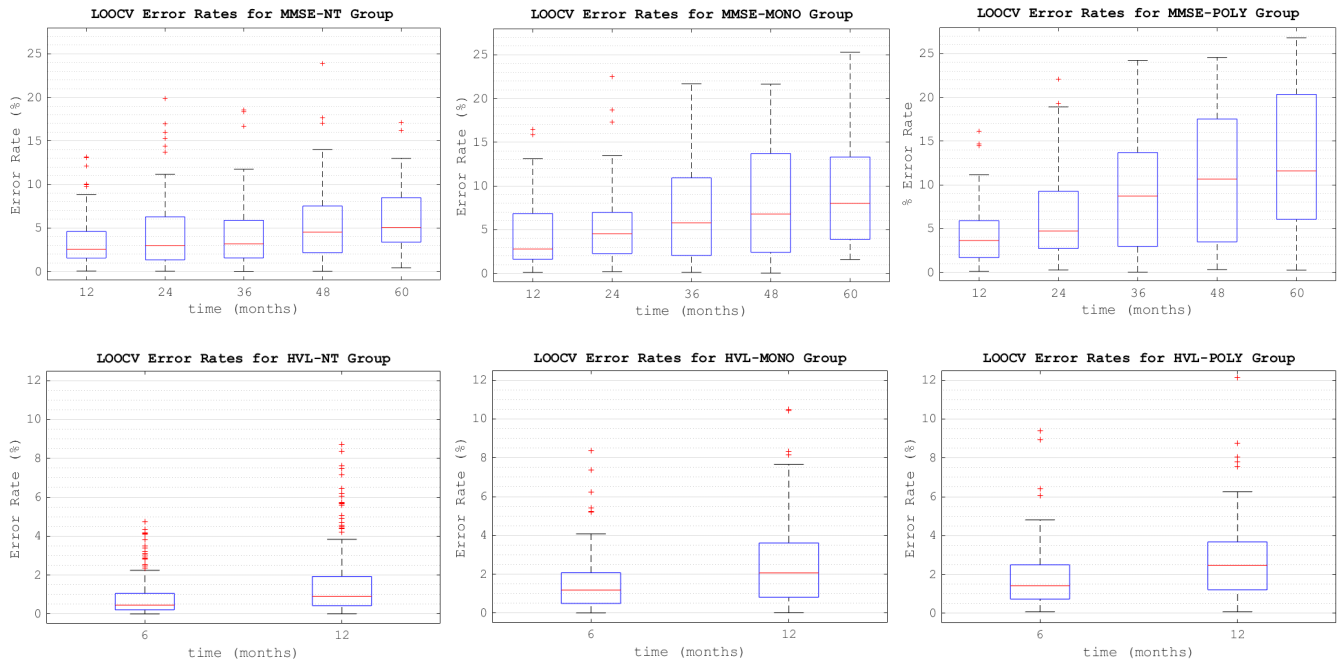

Fig. 4: LOOCV errors computed for PREP-AD-MMSE results (months 12 to 60), and for PREP-AD-HVL results (months 6 and 12)

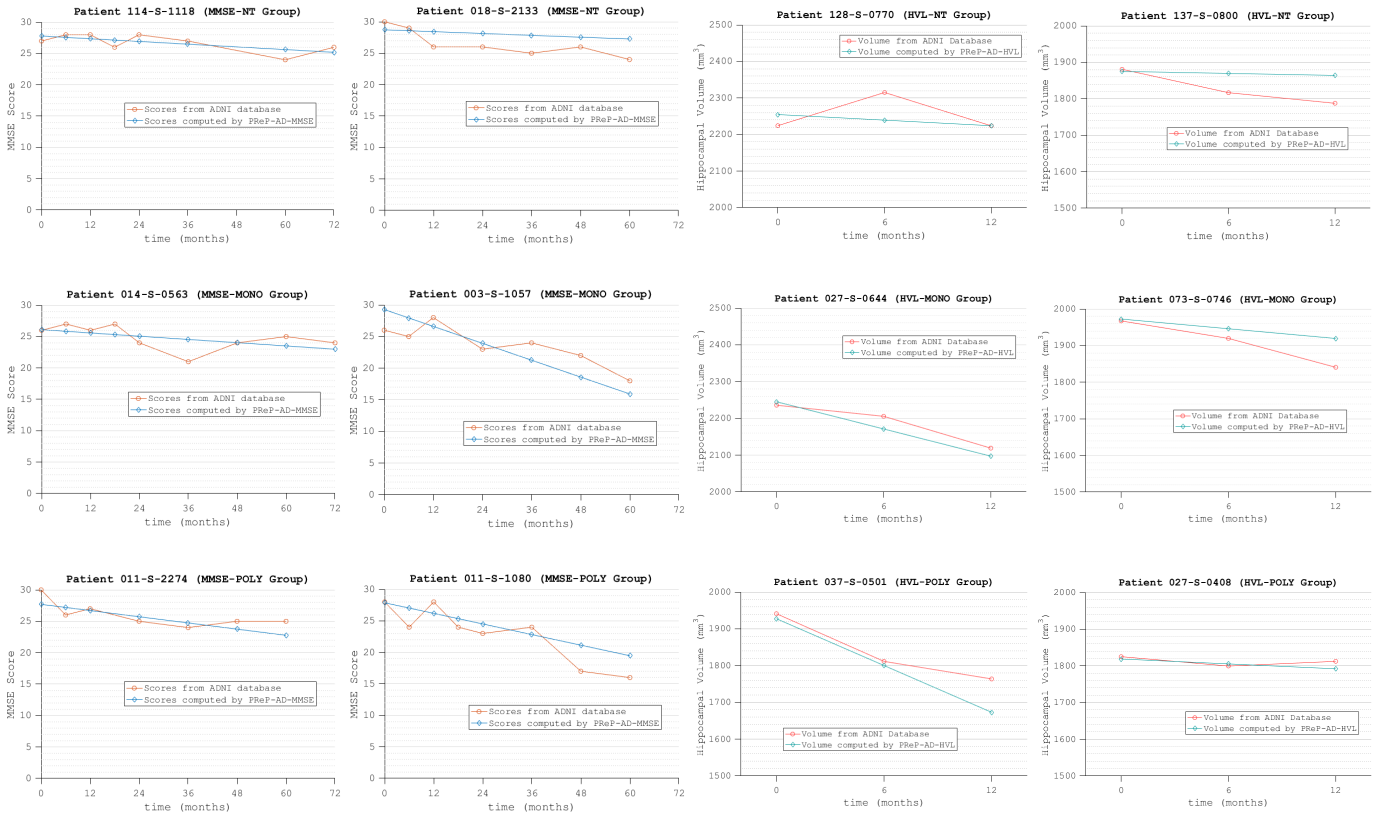

Fig. 5: Examples of AD progress computed by PREP-AD and reported in ADNI with average errors of 3.34% for 114-S-1118, 8.15% for 018-S-2133, 2.37% for 128-S-0770, 3.74% for 137-S-0800, 4.84% for 014-S-0563, 8.13% for 003-S-1057, 1.66% for 027-S-0644, 2.42% for 073-S-0746, 3.49% for 011-S-2274, 8.64% for 011-S-1080, 2.07% for 037-S-0501, and 0.72% for 027-S-0408

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

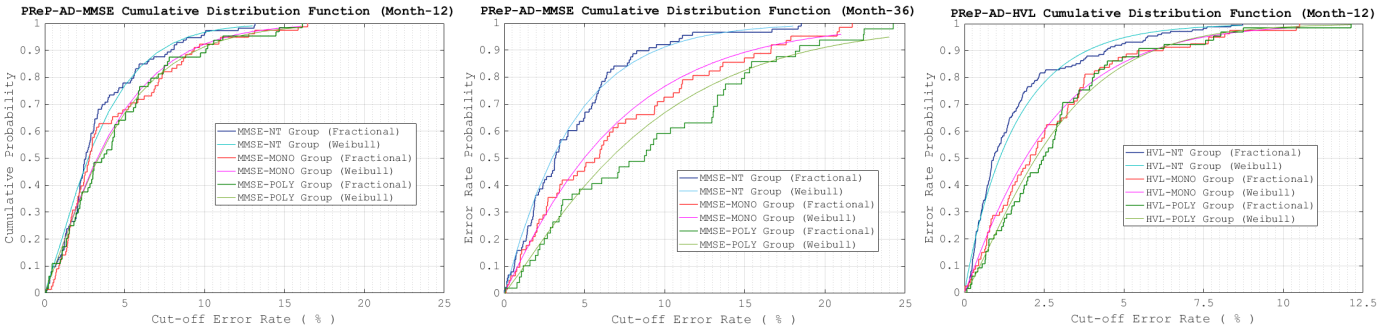

Fig. 6: Fractional cumulative distribution and Weibull approximations of LOOCV error rates: month 12 (left pane) and month 36 (mid pane) of all MMSE groups, and month 12 (right pane) of all HVL groups

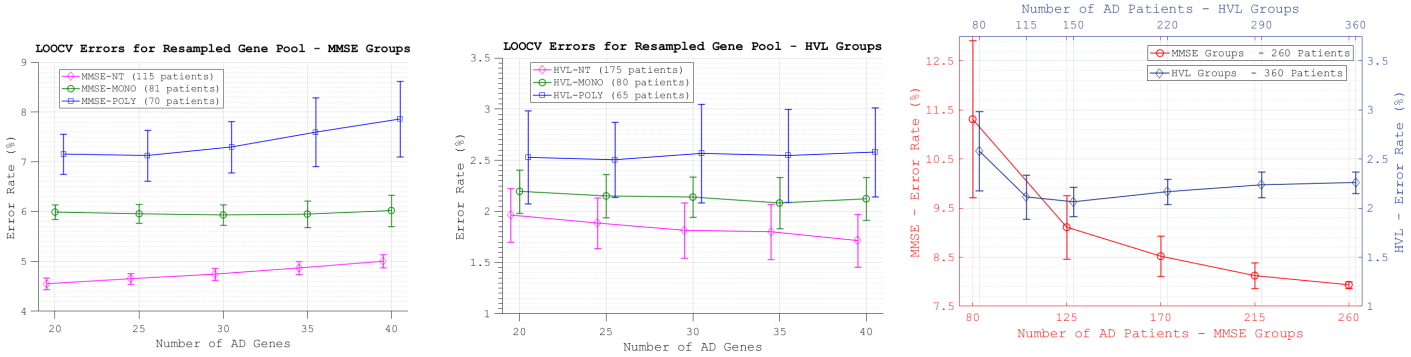

Fig. 7: LOOCV errors calculated using distinct number of randomly selected **AD genes** for MMSE groups (left pane), and for HVL groups (mid pane); LOOCV errors calculated for distinct number of randomly selected **AD patients** from all MMSE groups and from all HVL groups (right pane)

distribution. Fractional and Weibull distributions of LOOCV errors for PREP-AD-HVL for selected time points are in Fig. 6.

For MMSE groups, in month 12, we observe a sharp increase in cumulative probability distributions for all patient groups, implying that a small error rate is obtained with a high probability. For example, the probability of an error rate of less than 10% is greater than 0.9 for all groups. The increase in error distribution functions for PREP-AD-MMSE results at month 36 is less sharp for the therapy groups compared to MMSE-NT group.Probability for obtaining a cut-off error of 10% for MMSE-NT group is 0.9, and slightly greater than 15% for MMSE-MONO and MMSE-POLY groups. The CDFs for PREP-AD-HVL results indicates low error rates since 90% of the results are with a cut-off error rate of approximately 5% for all groups.We also observe that the increase in distribution functions is more pronounced for the non-therapy group, which indicates smaller error rates because of similarity in disease progression for patients not receiving drug therapy.

5.3.2 *Resampling for Error Analysis*

To evaluate the effect of the number of genes on error rates, we rerun the analysis using 20, 25, 30, 35 and 40 randomly selected AD genes for all patients in each group. The average LOOCV error rates with the standard deviation, where each case is run 100 times, are illustrated on the left and middle panes of Fig. 7. For example, the errors calculated for 175 patients in HVL-NT are shown in blue square-shaped markers in the mid pane of Fig. 7 (obtained by $175 \times 100 \times 5 = 87,500$ runs). The total number of runs to study the effect of the number of AD genes on error is 293,000.

For MMSE-NT and MMSE-POLY groups, the average error increases for a larger number of AD genes. The average errors of MMSE-MONO group are close to each other for different number of genes. We obtain a decreasing trend for HVL-NT group, while it is stable for HVL-MONO, and increases for HVL-POLY as the number of AD genes increases. The changes in average errors is a result of the number of patients and the variance of disease progression in a group. For example, having 175 patients in HVL-NT results in a decreasing trend in average errors, while the average error increases as the number of genes increases for 115 patients in MMSE-NT. Similarly, while average errors are similar for MMSE-MONO, it has an increasing trend for MMSE-POLY, most likely because of the larger variance in progression results for MMSE-POLY patients. The standard deviation is smaller for MMSE groups (except for MMSE-POLY) compared to those of HVL groups. This difference is due to having volume measurements only at months 6 and 12, while collecting MMSE scores at five distinct time points in a 60-month period. A larger deviation is obtained for POLY group patients since their disease progression is with a larger variance.

We study the change in error rates for distinct number of patients by using overdetermined systems of equations in PREP-AD. We define two patient sets: a set of 260 patients from three MMSE groups, and a second set of 360 patients from three HVL groups. The average errors are calculated for randomly selected 80, 125, 170, 215 and 260 patients from MMSE set, and 80, 115, 150, 220, 290 and 360 patients from HVL set (100 runs for each patient in each case). We have a total of 206,500 runs to analyze error for distinct number of AD patients used in PREP-AD.

In the right pane of Fig. 7, the left and bottom axes show the average error and the number of patients for MMSE set, whereas they are scaled on the right and the top axes for HVL set. For

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2018.2870363, IEEE/ACM Transactions on Computational Biology and Bioinformatics

MMSE patients, the error exponentially decreases as the number of patients increases. For HVL set, a similar characteristic is observed until the number of patients is larger than 150 (with a small increase after that). This difference can be due to a larger number of HVL patients and only from month 12, whereas it is from a 60 month period for MMSE set. For both MMSE and HVL sets, the deviation decreases as the number of patients increases.

Consistent error rates for variances of input data points out the efficacy of PREP-AD methods in computing disease progression.

6 Statistical Significance

Correlation among the AD related genes is examined with weighted gene co-expression network (WGCN) [48], a non-directed graph, where each node corresponds to a gene and an edge between two nodes is the pairwise correlation between the expressions of those two genes. An edge in WGCN is placed between two nodes if their correlation is more than a given threshold value. A WGCN can also be presented as an adjacency matrix W, where an element w_{ij} is the pairwise correlation between expressions of genes i and j. Matrix W is symmetric and each element $w_{ij} \in [-1, 1]$. A *clique* is a subgraph of WGCN of which all nodes are connected to each other [28]. It consists of genes whose pairwise correlations of expressions are greater than a threshold. Many algorithms can be used to construct maximal clique of an adjacency matrix (see for example [49]).

We construct an adjacency matrix using expressions of AD related genes, and additional matrices whose elements are pairwise correlations of randomly selected gene expressions from ADNI database. We show that AD related genes form larger cliques compared to the randomly chosen ones and that the number of cliques formed using AD genes is smaller than those by random genes. For this comparison, we use 51 AD related and 51 randomly chosen genes out of 20,062 in the database, and repeat this process 10,000 times. In Fig. 8, CDFs of the total number of cliques in adjacency matrices for AD related and randomly selected genes are shown. The dotted lines in Fig. 8 represent the deviation among 10,000 random runs, where the lower and upper bounds are calculated as $(\mu \pm \sigma)$ for a mean value of μ and a standard deviation of σ. The average of cumulative distributions for random genes is marked with a continuous blue line, and the distribution for AD genes in yellow. The threshold is selected as the correlation value that results in the maximum difference of CDF results. The difference between the CDFs for AD related and randomly chosen genes shown in Fig. 8 implies a relationship among the AD genes.

For threshold value of 0.13, obtained by the maximum difference between CDFs (vertical dotted red line in Fig. 8), the total number of cliques is 691 with an average clique size of 10.59 for AD genes. However, for 10,000 sets of random genes, the average clique size is 6.69. p-value is obtained as 0.0036 for the average clique size for AD genes, which implies that AD related genes build bigger cliques since they are more interrelated with each other.

For statistical significance of PREP-AD results, we compute genetic accordance matrix $\mathfrak{A}_{AD}$ using randomly chosen genes. Since the magnitude of a weight coefficient in $\mathfrak{A}_{AD}$ represents the impact of a gene on disease progression, the deviation among coefficients is an indicator of significance. We first calculate the standard deviation among the $\mathfrak{A}_{AD}$ coefficients using AD-related genes for the six patient groups. Then we repeat this process with randomly selected genes by computing $\mathfrak{A}_{AD}$ for 1,000 times for each patient group to obtain the standard deviation of coefficient

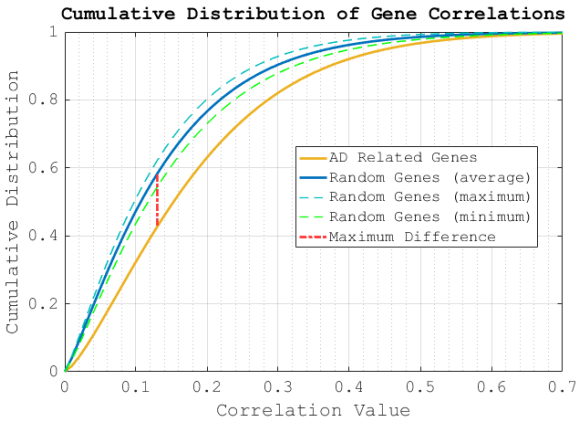

Fig. 8: CDFs of adjacency matrices for AD genes and 10,000 sets of randomly selected genes are given with upper and lower bounds of CDFs of random runs and the maximum difference between curves

values. We calculated the p-values for MMSE-NT, MMSE-MONO and MMSE-POLY as 0.0315, 0.0245 and 0.0288, respectively. For HVL-NT, HVL-MONO and HVL-POLY, they are calculated as 0.0239, 0.0440 and 0.0173 respectively. These relatively small p-values ($p \leq 0.05$) point out the significance of using AD related genes in computation of AD progression by PREP-AD methods.

We utilized supercomputers at the College of Staten Island of the CUNY for error analysis and statistical significance.

7 Discussion

The results in Sec. 5 show that PREP-AD methods compute a decrease in cognition for a larger percentage of patients in therapy groups. Since therapy were typically administered to the patients with already worsening conditions, a decreasing cognition can be expected. Table 3 row titled *Increasing Disease Progress* lists the number of patients with improved cognition, which may be attributed to a positive response to drug therapy.

Distribution of LOOCV error rates for PREP-AD results (boxplots in Fig. 4 and CDFs in Fig. 6) are presented in Sec. 5 together with their statistics (Tables 4 and 5). We observe from Fig. 4 that the errors in results for both PREP-AD methods are greater for therapy groups. This difference in error between the no-therapy and therapy groups is also apparent in their CDFs. As can be seen in Table 3, compared to therapy groups, there are more patients with stable progression in no-therapy groups. Therefore, a larger deviation among training set of patients for therapy groups is typical, which may be the reason for larger errors in therapy group computations. Moreover, having more patients in no-therapy groups may account for the smaller error rates obtained in the results. The error rates for PREP-AD increase with time for all patient groups, most likely because fewer patients continued to have follow-up measurements of their MMSE scores and hippocampal volumes at the later stages of their disease.

In Sec. 6, p-values are calculated for statistical significance of using AD related genes and for PREP-AD results. Small p-values ($p \leq 0.05$) indicate that AD related genes can be effectively used in our PREP-AD methods to compute disease progression. Error distribution graphs and the statistical significance computed for AD genes and PREP-AD methods imply that AI based techniques may be useful to compute progression of AD. An improvement in efficiency of drug therapies is expected by computing gene expression based progression results for different drug therapies.

8 Concluding Remarks

Personalized relevance parameterization methods are built based on AI techniques to compute disease progression at MCI stage of AD by evaluating gene expression values, MMSE scores and hippocampal volume loss over time. We define six patient groups to evaluate pharmacological therapy effects on disease progression based on their clinical measurements and administered drug therapy. Average LOOCV error rates are less than 8% in PReP-AD-MMSE results for a 60 month period and 3% in PReP-AD-HVL for 12 months. Statistical significance is evaluated for using AD related genes ($p = 0.003$) in PReP-AD methods based on clinical data and gene expressions ($p < 0.05$ for all patient groups).

Relatively small average for LOOCV errors and the corresponding p-values are encouraging to build AI methods based on gene expressions and clinical measurements, such as MMSE scores and MRI scans for hippocampal atrophy to support clinical studies toward pharmacologic therapy decisions at early stages of AD. We plan to extend our research to compute AD progression and drug efficacy with additional clinical biomarkers.

Acknowledgements

The initial research used in this work was supported by U.S. Army Communications-Electronics RD&E Center contracts W15P7T-09-CS021 and W15P7T-06-C-P217, and by the NSF grants ECCS-0421159, CNS-0619577 and IIP-1265265. The contents of this document represent the views of the authors and are not necessarily the official views of, or are endorsed by, the U.S. Government, DoD, Dept. of the Army or the U.S. Army CERDEC.

Data collection and sharing for this project was funded by the ADNI (NIH Grant U01 AG024904) and DOD ADNI (DoD award number W81XWH-12-2-0012). ADNI is funded 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.; Cogstate; 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). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the USC. ADNI data are disseminated by the Laboratory for Neuro Imaging at the USC.

This research was supported, in part, under National Science Foundation Grants CNS-0958379, CNS-0855217, ACI-1126113 and the City University of New York High Performance Computing Center at the College of Staten Island.

Conflict of Interest

The authors do not declare any conflict of interest.

References

[1] P Robert, S Ferris, S Gauthier, R Ihl, B Winblad, and F Tennigkeit. Review of alzheimers disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice. *Alzheimers Res Ther*, 2(4):24, 2010.

[2] N Schuff, N Woerner, L Boreta, T Kornfield, L M Shaw, J Q Trojanowski, P M Thompson, C R Jack, M W Weiner, Disease Neuroimaging Initiative, et al. Mri of hippocampal volume loss in early alzheimer's disease in relation to apoe genotype and biomarkers. *Brain*, 132(4):1067–1077, 2009.

[3] J J Kril, S Patel, A J Harding, and G M Halliday. Neuron loss from the hippocampus of alzheimer's disease exceeds extracellular neurofibrillary tangle formation. *Acta neuropathologica*, 103(4):370–376, 2002.

[4] L Wang, J S Swank, I E Glick, M H Gado, M I Miller, J C Morris, and J G Csernansky. Changes in hippocampal volume and shape across time distinguish dementia of the alzheimer type from healthy aging? *Neuroimage*, 20(2):667–682, 2003.

[5] S G Mueller, M W Weiner, L J Thal, R C Petersen, C R Jack, W Jagust, J Q Trojanowski, A W Toga, and L Beckett. Ways toward an early diagnosis in alzheimer's disease: The alzheimer's disease neuroimaging initiative (adni). *Alzheimers Dementia*, 1(1), 2005.

[6] C R Jack, R C Petersen, Y Xu, P C Obrien, G E Smith, R J Ivnik, B F Boeve, E G Tangalos, and E Kokmen. Rates of hippocampal atrophy correlate with change in clinical status in aging and ad. *Neurology*, 55(4):484–490, 2000.

[7] C G Parsons, W Danysz, A Dekundy, and I Pulte. Memantine and cholinesterase inhibitors: complementary mechanisms in the treatment of alzheimers disease. *Neurotoxicity research*, 24(3):358–369, 2013.

[8] R A Hansen, G Gartlehner, A P Webb, L C Morgan, C G Moore, and D E Jonas. Efficacy and safety of donepezil, galantamine, and rivastigmine for the treatment of alzheimers disease: a systematic review and meta-analysis. *Clinical interventions in aging*, 3(2):211, 2008.

[9] H Kavirajan and L S Schneider. Efficacy and adverse effects of cholinesterase inhibitors and memantine in vascular dementia: a meta-analysis of randomised controlled trials. *The Lancet Neurology*, 6(9):782–792, 2007.

[10] R Howard, R McShane, J Lindesay, C Ritchie, A Baldwin, R Barber, A Burns, T Dening, D Findlay, C Holmes, et al. Donepezil and memantine for moderate-to-severe alzheimer's disease. *New England Journal of Medicine*, 366(10):893–903, 2012.

[11] C Laske, H R Sohrabi, S M Frost, K López-de Ipiña, P Garrard, M Buscema, J Dauwels, S R Soekadar, S Mueller, C Linnemann, et al. Innovative diagnostic tools for early detection of alzheimer's disease. *Alzheimer's & Dementia*, 11(5):561–578, 2015.

[12] S Norton, F E Matthews, D E Barnes, K Yaffe, and C Brayne. Potential for primary prevention of alzheimer's disease: an analysis of population-based data. *The Lancet Neurology*, 13(8):788–794, 2014.

[13] Y Stern, X Liu, M Albert, J Brandt, D M Jacobs, C Del Castillo-Castaneda, K Marder, K Bell, M Sano, F Bylsma, et al. Application of a growth curve approach to modeling the progression of alzheimer's disease. *The J. of Gerontology Series A: Biological Sciences and Medical Sciences*, 51(4):M179–M184, 1996.

[14] M F Folstein, S E Folstein, and P R McHugh. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. of psychiatric research*, 12(3):189–198, 1975.

[15] C Laske, T Leyhe, E Stransky, N Hoffmann, A J Fallgatter, and J Dietzsch. Identification of a blood-based biomarker panel for classification of alzheimer's disease. *International Journal of Neuropsychopharmacology*, 14(9):1147–1155, 2011.

[16] T-T Huang, Y Zou, and R Corniola. Oxidative stress and adult neurogenesis effects of radiation and superoxide dismutase deficiency. In *Seminars in cell & developmental biology, V 23, No 7, pp 738-744*. Elsevier, 2012.

[17] P M Thompson, K M Hayashi, G I De Zubicaray, A L Janke, S E Rose, J Semple, M S Hong, D H Herman, D Gravano, D M Doddrell, and A W Toga. Mapping hippocampal and ventricular change in Alzheimer's Disease. *Neuroimage*, 22(4), 2004.

[18] A L Baird, S Westwood, and S Lovestone. Blood-based proteomic biomarkers of alzheimers disease pathology. *Frontiers in neurology*, 6, 2015.

[19] P R Walker, B Smith, Q Y Liu, A F Famili, J J Valdés, Z Liu, and B Lach. Data mining of gene expression changes in alzheimer brain. *Artificial intelligence in medicine*, 31(2):137–154, 2004.

[20] B D Hoyt, P J Massman, C Schatschneider, N Cooke, and R S Doody. Individual growth curve analysis of apoe ε4–associated cognitive decline in alzheimer disease. *Archives of neurology*, 62(3):454–459, 2005.

[21] J O Brooks, H C Kraemer, E D Tanke, and J A Yesavage. The methodology of studying decline in alzheimer's disease. *J. of the American Geriatrics Society*, 41(6):623–628, 1993.
[22] M S Mendiondo, JW Ashford, R J Kryscio, and F A Schmitt. Modelling mini mental state examination changes in alzheimer's disease. *Statistics in medicine*, 19(11-12):1607–1616, 2000.
[23] H Feldman, S Gauthier, J Hecker, B Vellas, B Emir, V Mastey, P Subbiah, and The Donepezil MSAD Study Investigators Group. Efficacy of donepezil on maintenance of activities of daily living in patients with moderate to severe alzheimer's disease and the effect on caregiver burden. *Journal of the American Geriatrics Society*, 51(6):737–744, 2003.
[24] J Kusyk, C S Sahin, M U Uyar, E Urrea, and S Gundry. Self organization of nodes in mobile ad hoc networks using evolutionary games and genetic algorithms. *J. of Advanced Research*, 2(3):253 – 264, 2011.
[25] A Saribudak, Y Dong, J Hsieh, and M U Uyar. Bio-inspired computation approach for tumor growth with spatial randomness analysis of kidney cancer xenograft pathology slides. In *9th EAI Conf. on Bio-inspired Information and Communications Technologies*, pages 1–8. IEEE, 2015.
[26] ADNI - Alzheimer's Disease Neuroimaging Initiative. In *National Institutes of Health, (up-to-date info., https://www.adni-info.org)*.
[27] S Arlot, A Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.
[28] S S Moy and J J Nadler. Advances in behavioral genetics: mouse models of autism. *Molecular psychiatry*, 13(1):4–26, 2008.
[29] A Saribudak, A A Subick, J A Rutta, M U Uyar, and ADNI. Gene expression based computation methods for alzheimer's disease progression using hippocampal volume loss and mmse scores. In *Proc. of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 165–174. ACM, 2016.
[30] A Saribudak, A A Subick, M U Uyar, and ADNI. Computation of pharmacologic therapy effects on cognitive abilities of alzheimers disease patients. In *Bioinformatics and Bioengineering (BIBE), 2016 IEEE 16th International Conference on*, pages 129–136. IEEE, 2016.
[31] MJ De Leon, S DeSanti, R Zinkowski, PD Mehta, D Pratico, S Segal, H Rusinek, J Li, W Tsui, LA Saint Louis, et al. Longitudinal csf and mri biomarkers improve the diagnosis of mild cognitive impairment. *Neurobiology of aging*, 27(3):394–401, 2006.
[32] L Wang, Y Zang, Y He, M Liang, X Zhang, L Tian, T Wu, T Jiang, and K Li. Changes in hippocampal connectivity in the early stages of alzheimer's disease: evidence from resting state fmri. *Neuroimage*, 31(2):496–504, 2006.
[33] C Green, J Shearer, C W Ritchie, and J P Zajicek. Model-based economic evaluation in alzheimer's disease: a review of the methods available to model alzheimer's disease progression. *Value in health*, 14(5):621–630, 2011.
[34] A Saribudak, E Ganic, J Zou, S Gundry, and M U Uyar. Toward genomic based personalized mathematical models for breast cancer tumor growth. In *IEEE 14th Int'l. Conf. on BioInf. & BioEng., pp. 115-119*, Nov. 2014.
[35] A Saribudak, S Gundry, J Zou, and MU Uyar. Genomic based personalized chemotherapy analysis to support decision systems for breast cancer. In *Medical Meas. and Appl., Int'l. Symp. on*, pages 495–500. IEEE, 2015.
[36] A Saribudak, S Gundry, J Zou, and M U Uyar. A gene expression-based mathematical modeling approach for breast cancer tumor growth and shrinkage. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 4(1):1–13, 2015.
[37] A Saribudak, Y Dong, S Gundry, J Hsieh, and M U Uyar. Mathematical models of tumor growth using voronoi tessellations in pathology slides of kidney cancer. In *Eng. in Medicine and Biology Society, 37th Annual Int'l. Conf.*, pages 4454–4457. IEEE, 2015.
[38] R Brookmeyer, E Johnson, K Ziegler-Graham, and H M Arrighi. Forecasting the global burden of alzheimers disease. *Alzheimer's & dementia*, 3(3):186–191, 2007.
[39] C R Jack, M A Bernstein, N C Fox, and et al. The alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: JMRI*, 27(4), 2008.
[40] Freesurfer software suite. *http://surfer.nmr.mgh.harvard.edu/*.
[41] E Giacobini. Cholinesterase inhibitors: new roles and therapeutic alternatives. *Pharmacological research*, 50(4):433–440, 2004.
[42] B Reisberg, R Doody, A Stöffler, F Schmitt, S Ferris, and H J Möbius. Memantine in moderate-to-severe alzheimer's disease. *New England J. of Medicine*, 348(14):1333–1341, 2003.
[43] P N Tariot, M R Farlow, G T Grossberg, S M Graham, S McDonald, I Gergel, Memantine Study Group, et al. Memantine treatment in patients with moderate to severe alzheimer disease already receiving donepezil: a randomized controlled trial. *Jama*, 291(3):317–324, 2004.
[44] Hal Daumé III. A course in machine learning. *Publisher, ciml. info*, pages 5–73, 2012.
[45] K Price, R M Storn, and J A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.
[46] J Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7:1–30, 2006.
[47] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.
[48] B Zhang, S Horvath, et al. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, 4(1):1128, 2005.
[49] C Bron and J Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

Aydin Saribudak received his B.S. degree in Electrical and Electronics Engineering from Middle East Technical University, Turkey, in 2005. He earned his M.Phil in 2017 and M.E. in 2018 from the City College of the CUNY, both in Electrical Engineering. He is currently working as a Data Scientist for Neustar, Inc. His interests include AI methods for the personalized progression and therapy analysis for human diseases. He is also working on machine learning techniques for the contextual marketing data analysis.

Adarsha A. Subick received his B.E. degree in 2014 and M.E. in 2017, both in Electrical Engineering Department of the City College of the City University of New York. He is currently a Solutions Architect with Amazon Web Services. His interests include artificial intelligence and their applications to the treatment of neurological disorders, high performance computing, and internet of things.

Na Hyun Kim received her B.E. degree in Electrical Engineering Department of the City College of the City University of New York in 2016. She is currently working as a Software Engineer at Walmart Jet.com. Her interests include biologically inspired computation algorithms and web applications for real-time tracking systems.

Joshua A. Rutta received his B.E. in Electrical Engineering at the Grove School of Engineering at the City College of New York in 2018. He is currently enrolled in the M.S. program in Electrical Engineering at Columbia University, doing a concentration in Data Driven Analysis and Computation. His interests include signal processing and machine learning.

M. Ümit Uyar, PhD, is a professor at the Electrical Engineering Department of the City College of the City University of New York (CUNY) and Computer Science Department of the Graduate Center of CUNY. His research interests include formal testing of complex computer systems and artificial intelligence techniques for cyber warfare, personalized anti-cancer therapy and clinical decision support systems. Prior to joining academia, he was a Distinguished Member of Technical Staff at AT&T Bell Labs. He has a B.S. degree from Istanbul Technical University, and M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, all in Electrical Engineering. Dr. Uyar is an IEEE Fellow and holds seven U.S. patents.