



# Integrative analysis of multi-dimensional imaging genomics data for Alzheimer's disease prediction

Ziming Zhang<sup>1</sup>, Heng Huang<sup>2\*</sup>, Dinggang Shen<sup>1,3\*</sup> and  
The Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup> Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>2</sup> Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA

<sup>3</sup> Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

## Edited by:

Manuel Menéndez-González,  
Hospital Álvarez-Buylla, Spain

## Reviewed by:

Ziv Radisavljevic, Brigham and  
Women's Hospital, Harvard Medical  
School, USA

Dinesh Sriramulu, Shres  
Consultancy (Life Sciences), India

## \*Correspondence:

Dinggang Shen, Department of  
Radiology and Biomedical Research  
Imaging Center, University of North  
Carolina at Chapel Hill, MRI Building,  
CB #7513, 130 Mason Farm Road,  
Chapel Hill, NC 27599, USA

e-mail: dgshen@med.unc.edu;

Heng Huang, Department of  
Computer Science, University of  
Texas at Arlington, 533 ERB  
Building, 500 UTA Blvd., Arlington,  
TX 76019, USA

e-mail: heng@uta.edu

In this paper, we explore the effects of integrating multi-dimensional imaging genomics data for Alzheimer's disease (AD) prediction using machine learning approaches. Precisely, we compare our three recent proposed feature selection methods [i.e., multiple kernel learning (MKL), high-order graph matching based feature selection (HGM-FS), sparse multimodal learning (SMML)] using four widely-used modalities [i.e., magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF), and genetic modality single-nucleotide polymorphism (SNP)]. This study demonstrates the performance of each method using these modalities individually or integratively, and may be valuable to clinical tests in practice. Our experimental results suggest that for AD prediction, in general, (1) in terms of accuracy, PET is the best modality; (2) Even though the discriminant power of genetic SNP features is weak, adding this modality to other modalities does help improve the classification accuracy; (3) HGM-FS works best among the three feature selection methods; (4) Some of the selected features are shared by all the feature selection methods, which may have high correlation with the disease. Using all the modalities on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, the best accuracies, described as (mean  $\pm$  standard deviation)%, among the three methods are (76.2  $\pm$  11.3)% for AD vs. MCI, (94.8  $\pm$  7.3)% for AD vs. HC, (76.5  $\pm$  11.1)% for MCI vs. HC, and (71.0  $\pm$  8.4)% for AD vs. MCI vs. HC, respectively.

**Keywords: Alzheimer's disease prediction, modality integration, imaging genomics data, feature selection, binary and multiclass classification**

## 1. INTRODUCTION

Alzheimer's disease (AD) is a complex chronically progressive neurodegenerative disease and the most common form of dementia in elderly people worldwide. As reported in Wimo et al. (1997), the prevalence of clinically manifest AD is about 2% at the age of 65 years but increases to about 30% at the age of 85 years. Recent research (Brookmeyer et al., 2007) suggested the number of people with AD to be double within the next 20 years, and 1 in 85 people will be affected by 2050. With the increase of human's life expectancy, more and more elderly people will suffer from AD, and accordingly it will cause a heavy socioeconomic burden. Unfortunately, there is no treatment to cure or even slow the progression of this disorder currently (Weiner et al., 2012). Huge effort has been put on the better understanding of the disease for more effective treatment (Hardy and Selkoe, 2002; Jack et al., 2010; Weiner et al., 2010, 2012).

There are two more labels related to AD, that is, Healthy Control (HC) and Mild Cognitive Impairment (MCI) (Wee et al., 2011, 2012; Zhou et al., 2011; Zhang et al., 2012a). People in HC are actually not AD patients, while MCI can be considered as the early stage of AD (Dubois et al., 2010), in which people show mildly impaired in memory with relative preservation of other cognitive domains and functional activities and do not meet the

criteria for dementia (Petersen et al., 2009). Also in Petersen et al. (2009) it was showed that each year 10–15% of MCI patients progressed to AD. As we see, the disease is developed gradually from HC to MCI, and eventually to AD. Between different statuses, there are no clear rules for defining the status of the disease. Therefore, accurate prediction of disease status (i.e., HC, MCI, or AD) becomes very difficult and important for early treatment of the disease.

To predict AD, a variety of biomarkers have been found and proposed (some of them are referred to by the sampling techniques), such as magnetic resonance imaging (MRI) Fan et al. (2007), positron emission tomography (PET), FDG-PET, and cerebrospinal fluid (CSF) (Hampel et al., 2008). Amount of works (Vemuri et al., 2009; Kohannim et al., 2010; Salas-Gonzalez et al., 2010; Hinrichs et al., 2011; Wolz et al., 2011; Zhang et al., 2011a; Ewers et al., 2012) have focused on how to utilize these biomarkers to classify AD, and they suggest that combining them for prediction is better than using any of them independently. Besides these existing biomarkers, recently genetic information is explored and studied in some very interesting works (Wang et al., 2012a,b; Nho et al., 2013a,b) for AD prediction.

In particular, machine learning community has provided powerful classification tools that are used for AD prediction (Zhang

et al., 2011b, 2012b; Li et al., 2012; Liu et al., 2012; Zhang and Shen, 2012). Multiple kernel learning (MKL) framework (Rakotomamonjy et al., 2008; Zhang et al., 2010, 2011a) is one of the examples which can integrate different sources of information automatically using convex optimization. Feature selection (Liu et al., 2013) is another good example, where the first stage performs feature selection method, and in the second stage the selected features are fed into a classifier for training. In Wang et al. (2012a,b) the authors proposed group-sparse learning algorithms for regression and feature selection based on MRI, PET, and single-nucleotide polymorphism (SNP). The basic idea behind these algorithms is to select a small subset of features (i.e., feature selection) that will be commonly shared by different regression tasks.

The main contributions of this paper are two-fold:

**(1) Our first contribution is to compare the performances of three recent proposed feature selection methods from machine learning community in the same experimental environment.**

These methods are multiple kernel learning (MKL) (Zhang et al., 2011a), high-order graph matching based feature selection (HGM-FS) (Liu et al., 2013), and sparse multimodel learning (SMML) (Wang et al., 2013). Feature selection for AD prediction has been attracting more and more attention. With very limited data samples and high dimensional data representations, it is reasonable to assume that all the data samples actually lie in a low dimensional representation space/manifold where classifiers can achieve better generalization, and feature selection is used to find such low dimensional manifold. However, in the literature such feature selection methods are developed independently using their own experimental settings, making it difficult to tell which is the best in terms of classification accuracy and how they behave on the data. Understanding these factors will be very useful for clinical usage to predict AD by choosing proper methods.

To answer such questions, in this paper we did comprehensive comparison between the three feature selection methods above on the Alzheimer's Disease Neuroimaging Initiative (ADNI)<sup>1</sup> data set, which is designed to characterize clinical, genetic, imaging, and biochemical biomarkers of AD and identify the relationships between them over the course of disease progression from normal cognition to MCI to dementia. Using linear support vector machines (SVMs) as our classifiers, we report the classification accuracy, sensitivity, and specificity for both binary and multiclass classification tasks, based on 10-fold cross validation. Our experimental results suggests (1) *HGM-FS works best among the three feature selection methods, and (2) all the feature selection methods select some shared features.*

**(2) Our second contribution is to explore the effects of integrating multi-dimensional imaging genomics data for AD prediction based on the three feature selection methods above.**

Developing discriminative features are always very important for AD prediction. Particularly, in this paper we focus on the performances of four widely-used modalities, namely MRI, PET, CSF, and SNP, with feature selection. This is very important and valuable for clinical purpose, because the cost of each modality is very different, and if the cheap ones can give us satisfactory results,

we can utilize them first rather than utilizing costly modalities at the beginning. Our experimental results suggest that *in general, (1) PET is the best modality in terms of accuracy, and (2) adding SNP to other modalities does improve the classification accuracy, even though its discriminant power is weak.*

The rest of the paper is organized as follows. In Section 2 the materials used in our experiments are explained. In Section 3 the details of our classification methods are provided, including the preprocessing on genetic data, and our three recent proposed feature selection methods. In Section 4 our comparison results are listed and discussed. Finally, we conclude the paper in Section 5.

## 2. MATERIALS

Our dataset is a subset from ADNI, where each subject can be represented using either imaging or genetic information. In total, we use 189 subjects from this dataset: 49 patients with AD, 93 patients with MCI, and 47 HC. Image preprocessing is performed separately for magnetic resonance imaging (MRI) and Fluorodeoxyglucose (FDG) Positron-Emission Tomography (PET) data. The preprocessing steps of MRI data include skull-stripping (Wang et al., 2011), dura removal, intensity inhomogeneity correction, cerebellum removal, spatial segmentation, and registration. We then parcellate the preprocessed images into 93 regions according to the template in Kabani et al. (1998). Only gray matter volume of these 93 regions-of-interest (ROI) is used in the experiments. For the preprocessing of PET images, we align the PET image of each subject to its corresponding MRI image using a rigid transformation and the average intensity of each ROI is calculated as a feature. CSF data were collected in the morning after an overnight fast using a 20- or 24-gauge spinal needle, frozen within 1 h of collection, and transported on dry ice to the ADNI Biomarker Core laboratory. In this study, CSF A $\beta$ 42, CSF *t*-tau, and CSF *p*-tau are used as features.

The single-nucleotide polymorphism (SNP) data (Saykin et al., 2010) were genotyped using the Human 610-Quad BeadChip. Among all SNPs, only SNPs, belonging to the top AD candidate genes listed on the AlzGene database ([www.alzgene.org](http://www.alzgene.org)) as of June 10, 2010, were selected after the standard quality control (QC) and imputation steps. The QC criteria for the SNP data include (1) call rate check per subject and per SNP marker, (2) gender check, (3) sibling pair identification, (4) the Hardy-Weinberg equilibrium test, (5) marker removal by the minor allele frequency, and (6) population stratification. As the second preprocessing step, the quality-controlled SNPs were imputed using the MaCH software to estimate the missing genotypes. After that, the Illumina annotation information based on the Genome build 36.2 was used to select a subset of SNPs, belonging to the top 135 AD candidate genes (Bertram et al., 2007). The above procedure yielded 5677 SNPs from 135 genes. Because the dimensionality of SNPs is much higher than the ones of other neuroimaging features, we use the unsupervised feature selection method to reduce the dimensionality of SNPs to the similar level of other types of features.

## 3. METHODOLOGY

Our AD prediction framework is simple: Given the feature vectors for individuals, feature selection methods are applied first to

<sup>1</sup>This dataset can be downloaded from <http://www.loni.ucla.edu/ADNI>.

select discriminative features. Then by performing element-wise product between the selected features and the learned weights by feature selection methods, the new feature vectors are fed into linear support vector machines (SVMs) to train the predictors. During testing, the selected features, learned weights, and predictors are fixed, and each test sample is classified into one of the three labels (i.e., AD, MCI, and HC) whose predicted score is the maximum.

In this section, we begin with introducing the preprocessing on genetic data, which attempts to reduce the computational time of each feature selection method beforehand by reducing the dimensionality of genetic data. Then we summarize our three recent proposed feature selection methods for readers to better understand our prediction framework.

### 3.1. PREPROCESSING ON GENETIC DATA

Since the dimensionality of SNP features is so high that the computational time for the feature selection methods used in our experiments is very long, we perform a simple unsupervised dimension reduction method on the genetic data before applying those complicated feature selection in our experiments to reduce the computational time.

One commonly used unsupervised feature selection criterion is Laplacian Score (He et al., 2006), which aims to select features that can best preserve the local manifold structure. He et al. (2006) argued that: in many classification tasks, the local structure of data is expected to be more important than global structure.

Motivated by such observation, Laplacian Score was proposed to capture the local structure of data using graph Laplacian. Given a set of  $N$  training data  $\{\mathbf{x}_i\}_{i=1,\dots,N}$  where  $\forall i, \mathbf{x}_i \in \mathbb{R}^d$  and the corresponding data matrix  $\mathbf{X} \in \mathbb{R}^{d \times N}$ , a data similarity matrix  $\mathbf{W} \in \mathbb{R}^{N \times N}$  can be calculated using the heat kernel. By summing up the elements in  $\mathbf{W}$  along each row, we can further create a diagonal matrix  $\mathbf{D} \in \mathbb{R}^{N \times N}$ . The Laplacian Score of a feature is defined as  $s(k) = \frac{\mathbf{X}_k \mathbf{L} \mathbf{X}_k^T}{\mathbf{X}_k \mathbf{D} \mathbf{X}_k^T}$ , where  $\mathbf{X}_k$  is the  $k$ th row in matrix  $\mathbf{X}$  consisting of the values for the  $k$ th feature in all the data samples,  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  is the graph Laplacian matrix, and  $(\cdot)^T$  denotes the vector transpose operator.

One drawback of the Laplacian Score strategy is that: it is very sensitive to the heat kernel parameter  $\sigma$  which is used to construct similarity matrix. Thus, a huge amount of computation is needed to tune the parameter  $\sigma$ . On the other hand, constructing the similarity matrix itself is also time consuming [with the time complexity of at least  $\mathcal{O}(dN^2)$ ].

The Laplacian score only considers the local structure. In order to utilize both local and global structures, we adopt an unsupervised local and global discriminative (LGD) feature selection criterion. The score of each feature is defined as the ratio between global variance and local variance:

$$s(k) = \frac{\sum_i (\mathbf{x}_i(k) - \bar{\mathbf{x}}(k))^2}{\sum_j \sum_{\mathbf{x}_i(k) \in o(\mathbf{x}_j(k))} (\mathbf{x}_i(k) - \bar{\mathbf{x}}_j(k))^2}, \quad (1)$$

where  $\mathbf{x}_i(k)$  is the value for the  $k$ th feature of the  $i$ th sample,  $\bar{\mathbf{x}}(k)$  is the mean value for the  $k$ th feature in the data samples,  $o(\mathbf{x}_j(k))$  is

the set of neighbor points of the  $j$ th sample.  $\bar{\mathbf{x}}_j(k)$  is the mean computed from the set of neighbor points as  $\bar{\mathbf{x}}_j(k) = \frac{\sum_{\mathbf{x}_i(k) \in o(\mathbf{x}_j(k))} \mathbf{x}_i(k)}{|o(\mathbf{x}_j(k))|}$ , where  $|o(\mathbf{x}_j(k))|$  is the cardinality of the neighbor set. Because the neighbor points of point  $\mathbf{x}_i(k)$  can be efficiently computed by sorting the values  $\{\mathbf{x}_i(k)\}_{i=1,\dots,N}$ , the computational complexity of our method is  $\mathcal{O}(dN \log N)$ .

## 3.2. FEATURE SELECTION METHODS

In general, feature selection methods can be considered as the combination of (1) data fitting terms and (2) sparse-induced regularization terms. Data fitting terms guarantee that the learned models are suitable for prediction within certain variations, while sparse-induced regularization terms make sure that only the “important” features have non-zero weights. By designing different combinations of these two terms, different feature selection methods are developed. For instance, hinge-loss or least-square loss can be used as data fitting terms, and  $\ell_1$  norm, trace norm, or group sparsity can be used as sparse-induced regularization terms. Typically, different combinations result in selecting different features, and we briefly summarize the details of MKL, HGM-FS, and SMML to show how they developed for feature selection.

### 3.2.1. Multiple kernel learning (MKL)

The basic idea of MKL is to build an optimal kernel for a specific task (e.g., classification) based on a set of basis kernels. In our experiments, we employ SimpleMKL (Rakotomamonjy et al., 2008) as the MKL competitor. Given training samples  $\{(\mathbf{x}_i, y_i)\}_{i=1,\dots,N}$ , where  $\forall i, \mathbf{x}_i \in \mathbb{R}^d$  is an input data vector and  $y_i \in \{1, -1\}$  is its binary label, and  $M$  feature mapping functions  $\{\phi_m : \mathbb{R}^d \rightarrow \mathbb{R}^{D_m}\}_{m=1,\dots,M}$ , SimpleMKL formulates the MKL problem with the hinge loss function for binary classification as follows:

$$\begin{aligned} \min_{\beta, \mathbf{w}, b} \sum_m \frac{\|\mathbf{w}_m\|_2^2}{2\beta_m} + C \sum_i \ell(\mathbf{x}_i, y_i; \mathbf{w}, b) \\ \text{s.t. } \forall m, \beta_m \geq 0, \|\beta\|_1 \leq 1 \end{aligned} \quad (2)$$

where  $\forall m, \beta_m$  denotes the weight for kernel  $m$  induced by feature mapping function  $\phi_m$ ,  $\beta$  denotes the kernel weight vector,  $\mathbf{w} = \{\mathbf{w}_m\}$  and  $b$  denote the classifier parameters,  $C \geq 0$  is a pre-defined regularization parameter,  $\|\cdot\|_1$  denotes the  $\ell_1$  norm of a vector,  $\forall i, \ell(\mathbf{x}_i, y_i; \mathbf{w}, b) = \max\{0, 1 - y_i [\sum_m \mathbf{w}_m^T \phi_m(\mathbf{x}_i) + b]\}$  denotes the hinge loss function, which is used in all the optimization problems in this paper, and  $(\cdot)^T$  in  $\ell$  denotes the matrix transpose operator. We adopt the method in Xu et al. (2010) to solve MKL in Equation 2.

### 3.2.2. High-order graph matching based feature selection (HGM-FS)

(Liu et al., 2013) This method extends the traditional LASSO (Tibshirani, 1994) by adding two regularizers which capture the geometrical relations (i.e., high-order statistics) between the predicted vectors and the target vectors (e.g., class label vectors). The underlying assumption of the method is that the predicted vectors are not only close to the target vectors in the target space, but also distributed similarly to the target vectors.

Given a set of training data  $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1, \dots, N}$  where  $\forall i, \mathbf{x}_i \in \mathbb{R}^d$  is a  $d$ -dimensional feature vector, and  $\mathbf{y}_i \in \mathbb{R}^{|\mathcal{C}|}$  is its associated  $|\mathcal{C}|$ -dimensional target vector, we denote  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$  as the feature matrix, and  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{|\mathcal{C}| \times N}$  as the target matrix. Letting  $\mathbf{W} \in \mathbb{R}^{d \times |\mathcal{C}|}$  be the regression coefficient matrix, and  $\|\cdot\|_F$  denote the Frobenius norm, the two new regularizers are defined as  $B = \sum_{i,j=1}^N \|\mathbf{y}_i - \mathbf{y}_j - \mathbf{W}^T(\mathbf{x}_i - \mathbf{x}_j)\|_F^2$  and  $T = \sum_{i,j,k=1}^N \|\mathbf{y}_i - \mathbf{y}_j\|^T(\mathbf{y}_j - \mathbf{y}_k) - (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W} \mathbf{W}^T (\mathbf{x}_j - \mathbf{x}_k)\|_F^2$ , where  $B$  and  $T$  capture the pair-wise and triplet-wise geometrical relations in the target space between the predicted vectors and the target vectors, respectively. Therefore, the final optimization formulation for HGM-FS can be written down as follows:

$$L(\mathbf{W}) = \min_{\mathbf{W}} \|\mathbf{W}^T \mathbf{X} - \mathbf{Y}\|_F^2 + \lambda_1 \|\mathbf{W}\|_1 + \lambda_2 B + \lambda_3 T \quad (3)$$

where  $\lambda_1 \geq 0, \lambda_2 \geq 0$ , and  $\lambda_3 \geq 0$  are regularization parameters. In theory, it is possible to add any higher-order graph matching information into the objective function above, and the features with non-zero regression coefficients from the original feature space are selected for final classification.

### 3.2.3. Sparse multimodal learning (SMML)

(Wang et al., 2013) SMML was proposed to integrate heterogeneous features from different modalities by using the joint structured sparsity regularizations to learn the feature importance from both group-wise and individual point of views.

Let  $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1, \dots, N}$  be  $N$  training samples, where each input vector  $\forall i, \mathbf{x}_i = \{(\mathbf{x}_i^1)^T, \dots, (\mathbf{x}_i^K)^T\} \in \mathbb{R}^d$  has  $d = \sum_{j=1}^K d_j$  dimensions containing all features from  $K$  modalities in total, each modality  $j$  has  $d_j$  dimensional feature vector,  $\mathbf{y}_i \in \mathbb{R}^{|\mathcal{C}|}$  is its associated  $|\mathcal{C}|$  dimensional class label vector, and  $|\mathcal{C}|$  is the number of classes. Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$  and  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{|\mathcal{C}| \times N}$ . We denote the classification coefficient matrix as  $\mathbf{W} = [\mathbf{w}_1^1, \dots, \mathbf{w}_{|\mathcal{C}|}^1; \dots; \mathbf{w}_1^K, \dots, \mathbf{w}_{|\mathcal{C}|}^K] \in \mathbb{R}^{d \times |\mathcal{C}|}$ , where  $\mathbf{w}_p^q \in \mathbb{R}^{d_q}$  is the weights of all features from the  $q$ th modality in the classification decision function of the  $p$ th class.

Based on these notations, SMML can be formulated as follows:

$$\begin{aligned} & \min_{\mathbf{W}} \sum_{i=1}^{|\mathcal{C}|} \sum_{j=1}^N \left(1 - y_{ji}(\mathbf{w}_i^T \mathbf{x}_j + b_i)\right)_+ \\ & + 2\gamma_1 \sum_{i=1}^{|\mathcal{C}|} \sum_{j=1}^K \|\mathbf{w}_i^j\|_2 + 2\gamma_2 \|\mathbf{W}\|_{2,1} \end{aligned} \quad (4)$$

where  $y_{ji} \in \{-1, 1\}$  denotes the binary class label for data  $\mathbf{x}_i$ ,  $\mathbf{w}_i \in \mathbb{R}^d$  denotes the classification coefficient vector for class  $i$ ,  $\mathbf{w}_i^j \in \mathbb{R}^{d_j}$  denotes the classification coefficient vector for class  $i$  and modality  $j$ ,  $\gamma_1 \geq 0$  and  $\gamma_2 \geq 0$  are two predefined regularization parameters,  $(\cdot)_+ = \max\{0, \cdot\}$ , and  $\|\cdot\|_{2,1}$  denotes the  $\ell_{2,1}$  norm of a matrix. Both the regularizers try to capture the essential structure in the classification coefficient matrix to improve the performance. In Wang et al. (2013), a very efficient algorithm has been proposed to solve Equation 4, which is guaranteed to converge to a global optimal solution.

## 4. RESULTS

In order to verify the effect of integration of both imaging and genetic information on the AD prediction accuracy, we perform two classification tasks separately: (1) binary classification, i.e., AD vs. MCI, AD vs. HC, and MCI vs. HC; (2) multiclass classification, i.e., AD vs. MCI vs. HC. Ten-fold cross validation is utilized for evaluating each method. LIBSVM (Chang and Lin, 2011) is employed as our linear SVM solver. The parameters for each method is determined using grid search. Best performance of each method is reported.

### 4.1. DATA PROCESS

For MRI, PET, and CSF, we utilize all the features, that is,  $93 + 93 + 3 = 189$  in total. For genetic SNP data, we apply the pre-processing method to rank the 5677 SNPs in a descending order, and finally preserve the top 189 SNPs as the final features for the genetic data without fully tuning, because we would like to make the data from both imaging and genetic information sources balanced.

For MRI, PET, and CSF, each feature is normalized by subtracting the mean and then divided by the standard deviation, where the mean and the standard deviation are calculated from training data. That is, given a set of training data  $\{\mathbf{x}_i\}_{i=1, \dots, N}$ , every single data sample  $\tilde{\mathbf{x}}$  needs to be normalized as  $\forall j, \tilde{\mathbf{x}}(j) \leftarrow \frac{\tilde{\mathbf{x}}(j) - \bar{\mathbf{x}}(j)}{\sqrt{\frac{1}{N} \sum_{i=1}^N [\mathbf{x}_i(j) - \bar{\mathbf{x}}(j)]^2 + \epsilon}}$ , where  $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$  denotes the mean of the training data,  $j$  denotes the  $j$ th feature in each feature vector, and  $\epsilon$  is a very small positive constant to avoid the case of the dominator equal to zero.

For genetic SNP features, each feature is normalized by subtracting the minimum value along the dimension and then divided by the difference between the maximum and the minimum along the dimension as well, calculated from training data. This normalization process can be written as  $\forall j, \tilde{\mathbf{x}}(j) \leftarrow \frac{\tilde{\mathbf{x}}(j) - \min_{i=1, \dots, N} \mathbf{x}_i(j)}{\max_{i=1, \dots, N} \mathbf{x}_i(j) - \min_{i=1, \dots, N} \mathbf{x}_i(j) + \epsilon}$ .

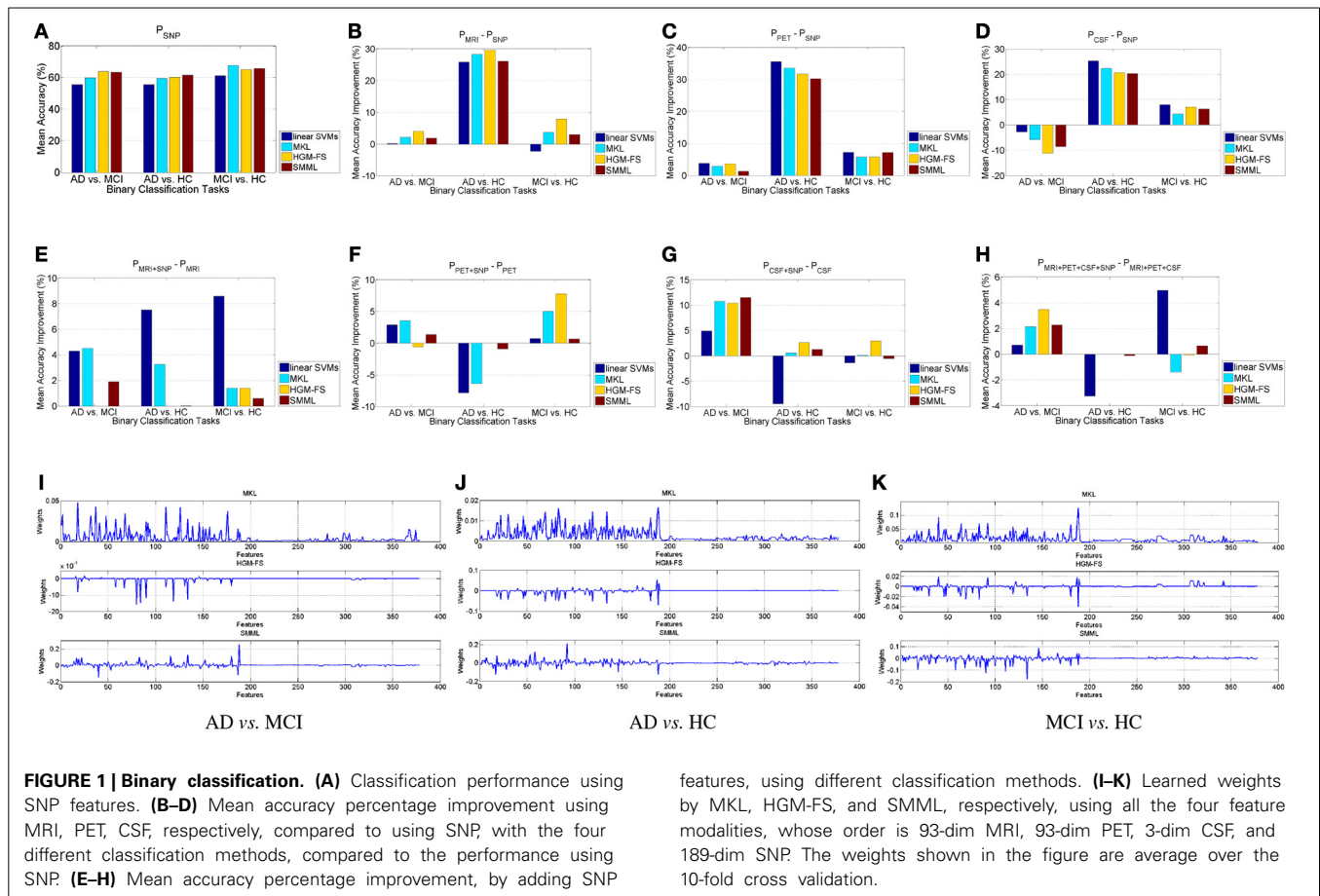
To generate the final feature vectors as the input for each feature selection method, each feature is further normalized to the  $\ell_2$ -norm unit ball based on the training data. This normalization can avoid the scaling bias in each feature. Again, this normalization process can be written as  $\forall j, \tilde{\mathbf{x}}(j) \leftarrow \frac{\tilde{\mathbf{x}}(j)}{\sqrt{\sum_{i=1}^N \mathbf{x}_i(j)^2 + \epsilon}}$ .

### 4.2. BINARY CLASSIFICATION

In this experiment, we perform three binary classification tasks: AD vs. MCI, AD vs. HC, and MCI vs. HC, respectively. For each task, the former is the positive class, and latter is the negative class. We also test 9 different configurations of modalities as features for feature selection methods: (1) MRI only, (2) PET only, (3) CSF only, (4) SNP only, (5) MRI+SNP, (6) PET+SNP, (7) CSF+SNP, (8) MRI+PET+CSF, and (9) MRI+PET+CSF+SNP. Linear SVMs with the final features are utilized as the baseline method, because there is no feature selection involved.

We first analyze the performance using each individual modality, compared to the performance using SNP. Figure 1A shows the performance using SNP features, and Figures 1B–D show our comparison results. From Figure 1A we can see that using SNP feature selection based classifiers outperform the baseline





method consistently, but among them there is no winner for all the binary classification tasks, and their performances are similar, in general. From **Figures 1B–D** we can see that, overall, MRI and PET outperform SNP significantly for all the methods and all the binary classification tasks, especially for AD vs. HC. For CSF, only for AD vs. MCI, its performances of all the methods are worse than those SNP correspondingly. For the other two classification tasks, the performances are still much better than using SNP. Notice that the dimensionality of CSF features is only 3, which leaves little room for selecting important features, while the dimensionality of SNP features are 189. However, it seems that the discriminative power of SNP features are very weak, which results in that even the dimensionality of the features is higher, its performance is still worse than others. General speaking, for AD prediction on the binary classification tasks, the discriminative power of different modalities can be ordered as  $PET > MRI > CSF > SNP$ .

Next we compare the performance with/without SNP features using MRI, PET, CSF, respectively, and MRI+PET+CSF. **Figures 1E–H** shows our comparison results. For AD vs. HC, by adding the SNP features the performance using each method improves little, in general, especially for each feature selection based classifiers. This is mainly because the performance using the rest modalities without SNP are almost saturated, and under this circumstance the SNP features will be considered as noisy features, which has little contribution to the performance. Another

reason is that AD vs. HC is the easiest task due to their differences. For the other two binary classification tasks, especially for AD vs. MCI, the performance improvement by adding SNP features is significant for all the methods. This is mainly because the disease statuses in each task have no clear differentiation, which makes the discriminative power of the rest modalities quite weak as well, and in this case the SNP features can provide complementary information to help differentiate the disease statuses. Overall, even though the discriminative power of SNP features for AD prediction are rather weak, these features do help other modalities improve the performance for each feature selection based classifier, especially on AD vs. MCI.

**Table 1** shows the classification performances of different methods using all the four modalities. As we see, all the feature selection based classifiers outperform the baseline method significantly, but still there is no winner for all the binary classification tasks. Among the three feature selection based classifiers, HGM-FS works the best, which has achieved the state-of-the-art performance.

Finally we analyze the learned weights by each feature selection method using all the four modalities, as shown in **Figures 1I–K**. As we see, the weights for SNP features are relatively smaller than those for the other modalities, on average, and most of them are zeros. This observation demonstrates that the discriminative power of SNP features is much weaker than that of the other three modalities, which is consistent with the observation in **Figure 1**.

However, these small non-zero weights are still very important for improving the performance for AD prediction. Also, for these methods, some of the selected features are shared, which may be very important for the prediction of the disease status. Among the three methods, the selected features by HGM-FS are sparsest for every binary classification task.

**4.3. MULTICLASS CLASSIFICATION**

For multiclass classification, we conduct three “one-vs-the-rest” binary classification tasks instead, i.e., HC vs. non-HC, MCI vs. non-MCI, and AD vs. non-AD. Then a test data sample

is classified to the class whose decision value is the maximum among the three binary classification tasks. Notice that multiclass classification is more difficult than binary classification in general.

As we did for the binary classification, we perform similar analysis for the multiclass classification as well. **Figure 2A** shows the performance of each method using SNP features, which is slightly worse than the corresponding performance in **Figures 1A** and **2B** shows the performance improvement using MRI, PET, CSF, respectively, compared to the performance using SNP. Still MRI and PET work better than SNP. However, different from binary classification above, CSF works worse than SNP, especially for feature selection based classifiers. This is because that the much lower dimensionality of CSF features makes it much more difficult to distinguish multiple classes simultaneously. To summarize, for AD prediction using multiclass classification, MRI and PET are better than both CSF and SNP, and PET again works best.

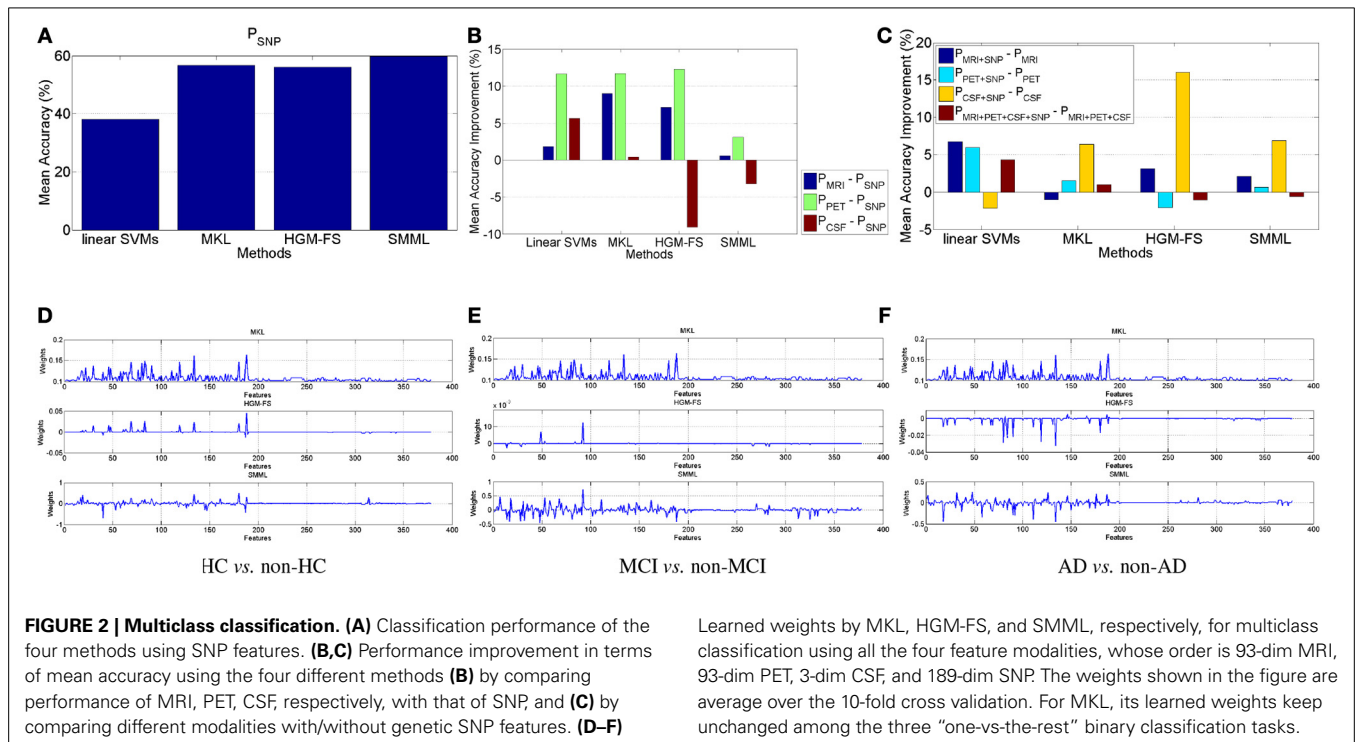
**Figure 2C** shows the performance improvements for all the four methods by comparing the performances using different modalities with/without SNP. Similar to **Figures 1E–H**, in general, adding SNP features can improve multiclass classification performance. And **Table 2** lists the performances of the four methods using all the modalities. Still HGM-FS outperforms the rest methods, and has achieved the state-of-the-art result on AD prediction using multiclass classification.

**Figures 2D–F** shows the learned weights by MKL, HGM-FS, and SMML, respectively, for the three “one-vs-the-rest” binary classification tasks using all the four modalities. Again, on each binary classification task, the learned weights for SNP features are relatively smaller than those for the rest modalities, some of the features with non-zero weights are shared as well, and among the

**Table 1 | Performance comparison (%) among different methods using all the four modalities (i.e., MRI, PET, CSF, and SNP) in terms of “mean ± standard deviation” for binary classification tasks (i.e., positive class vs. negative class).**

		Linear SVMs	MKL	HGM-FS	SMML
AD vs. MCI	Accuracy	65.0 ± 8.7	67.8 ± 9.3	<b>76.2 ± 11.3</b>	69.1 ± 14.8
	Sensitivity	52.0 ± 23.5	50.0 ± 27.1	<b>74.0 ± 25.0</b>	69.5 ± 23.4
	Specificity	72.3 ± 11.4	<b>77.8 ± 16.2</b>	77.8 ± 16.7	69.0 ± 18.6
AD vs. HC	Accuracy	87.5 ± 14.1	91.9 ± 13.2	92.9 ± 9.5	<b>94.8 ± 7.3</b>
	Sensitivity	90.0 ± 14.1	94.0 ± 13.5	<b>94.0 ± 9.7</b>	<b>94.0 ± 9.7</b>
	Specificity	85.0 ± 17.0	89.5 ± 17.4	91.5 ± 14.5	<b>95.5 ± 9.6</b>
MCI vs. HC	Accuracy	71.2 ± 14.6	<b>76.5 ± 11.1</b>	76.4 ± 10.7	75.6 ± 10.8
	Sensitivity	76.4 ± 15.8	76.2 ± 11.7	<b>78.8 ± 9.1</b>	76.3 ± 12.1
	Specificity	62.0 ± 26.3	<b>76.5 ± 16.3</b>	72.0 ± 19.8	73.5 ± 18.3

For better comparison, we highlight the best performances using red color.



**FIGURE 2 | Multiclass classification. (A)** Classification performance of the four methods using SNP features. **(B,C)** Performance improvement in terms of mean accuracy using the four different methods **(B)** by comparing performance of MRI, PET, CSF, respectively, with that of SNP, and **(C)** by comparing different modalities with/without genetic SNP features. **(D–F)**

Learned weights by MKL, HGM-FS, and SMML, respectively, for multiclass classification using all the four feature modalities, whose order is 93-dim MRI, 93-dim PET, 3-dim CSF, and 189-dim SNP. The weights shown in the figure are average over the 10-fold cross validation. For MKL, its learned weights keep unchanged among the three “one-vs-the-rest” binary classification tasks.

three methods, HGM-FS produces the sparsest selected features for multiclass classification.

**4.4. TOP SELECTED GENETIC BASIS**

The top selected SNPs by three different methods in binary classifications are plotted in **Figure 3A** (HGM-FS method), **Figure 3B** (MKL method), and **Figure 3C** (SMML method). In all figures,

the color map is used to show the weights of different features in classifications. If the weight is large, the feature is important to the corresponding class.

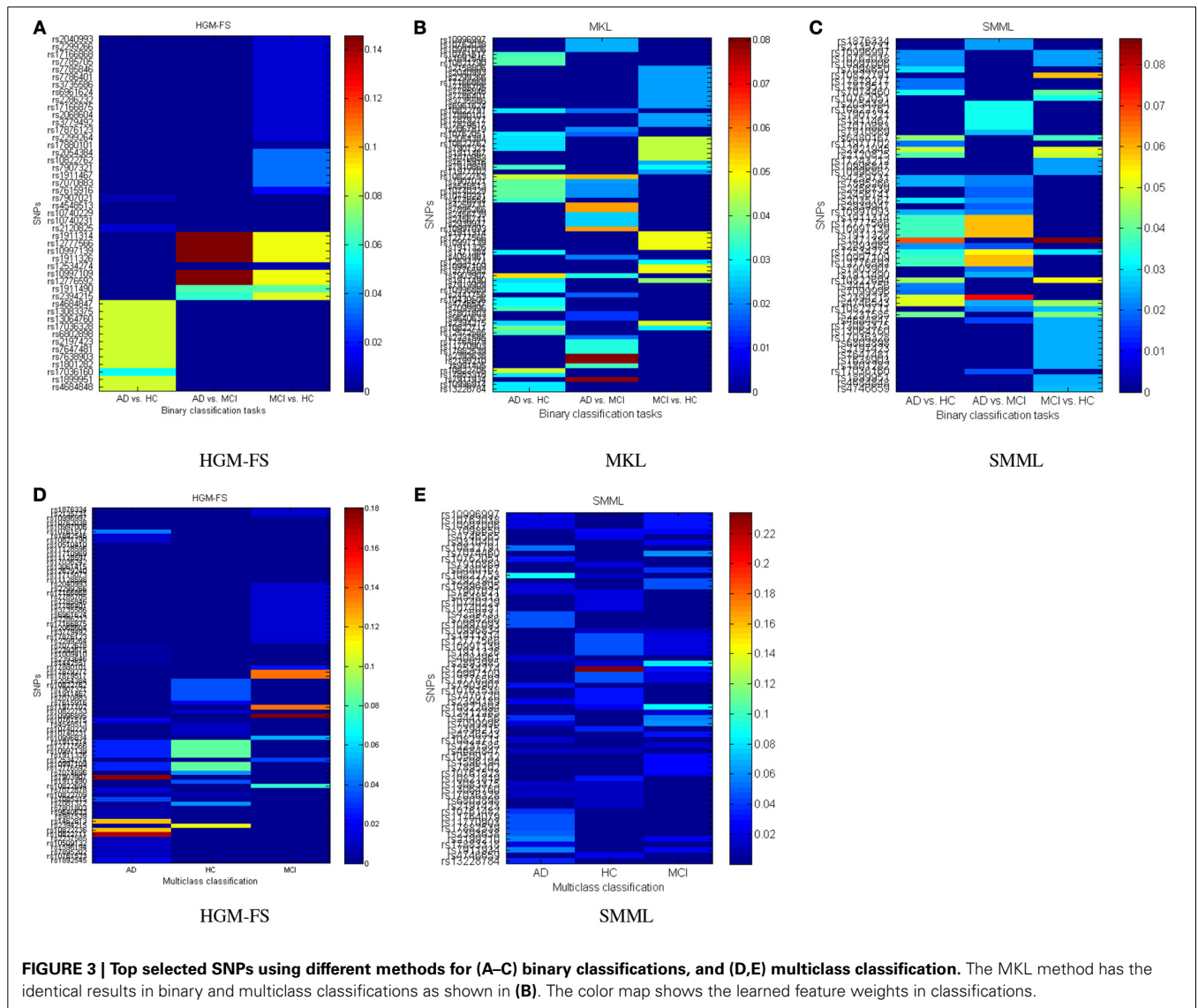
The top selected SNPs by two different methods in multiclass classifications are plotted in **Figure 3D** (HGM-FS method) and **Figure 3E** (SMML method). The MKL method has the identical results in binary and multiclass classifications, thus the selected SNPs by MKL method are the same as **Figure 3B**.

Because the feature selection mechanisms are different in three methods, the top selected SNPs are different. However, after our careful investigations, we found that many of the most top selected SNPs among three methods come from the same several genes. For example, the most top selected SNPs in HGM-FS method in multiclass classifications come from gene “CTNNA3,” which is a protein-coding gene and is associated to late-onset Alzheimer’s disease (Miyashita et al., 2007). The most top selected SNPs in SMML method in multiclass classifications come from gene “PON2,” which encodes paraoxonase-2 gene and

**Table 2 | Performance comparison (%) among different methods using all the four modalities (i.e., MRI, PET, CSF, and SNP) in terms of “mean ± standard deviation” for the multiclass classification task (i.e., AD vs. MCI vs. HC).**

	Linear SVMs	MKL	HGM-FS	SMML
Accuracy	53.0 ± 14.0	68.4 ± 11.8	71.0 ± 8.4	66.8 ± 11.3

For better comparison, we highlight the best performances using red color.



is associated with apolipoprotein E4 allele in both Alzheimer's and vascular dementias (Janka et al., 2001). The most top selected SNPs in MKL method in multiclass classifications also come from gene "CTNNA3" and gene "PON2." Although the well-known APOE SNP is not the top one SNP in our list, it still appears in the top rank list. Because our studies are data-driven integrative multi-variate studies, our results are consistent with the existing GWAS results but also show the difference. The top selected SNPs in our studies reveal more interactions between genotypes and phenotypes.

## 5. DISCUSSION AND CONCLUSIONS

In this paper, we conduct a comprehensive study on modality integration for Alzheimer's disease (AD) prediction using the ADNI dataset. We employ four widely-used modalities (i.e., MRI, PET, CSF, and SNP), and compare three state-of-the-art feature selection based linear classifiers (i.e., MKL/HGM-FS/SMML + linear SVMs) with the baseline classifier (i.e., linear SVMs without feature selection). In our experiments, we perform both binary classification (i.e., AD vs. MCI, AD vs. HC, and MCI vs. HC) and multiclass classification (i.e., AD vs. MCI vs. HC), and analyze the results, respectively, based on 10-fold cross validation.

The key observations from our experimental results are: (1) Among all the compared methods, MRI and PET perform better than CSF and SNP in terms of prediction accuracy, and PET is the best among all the four modalities; (2) In general, SNP performs worst, but it is still helpful to improve the performance with other modalities together; (3) Among the three feature selection based classifiers, HGM-FS with linear SVMs performs best, and using all the four modalities, it has achieved the state-of-the-art performance for either binary or multiclass classification for AD prediction; (4) The selected features by each method share some common parts, the learned weights for SNP features are relatively smaller than those for the others, and HGM-FS produces the sparsest features among the three methods.

Our results are also very useful for clinical usage. For instance, considering the prediction performance, PET should be preferred, which is consistent with some recent evidences in the diagnosis of AD (Sperling and Johnson, 2013). For AD vs. HC, imaging modalities are highly recommended, and adding genomics data will, generally speaking, harm the classification accuracy. However, for AD vs. MCI, considering all the imaging and genomics data is highly recommended, since these modalities contain complementary information and integration of these information for diagnosis will be very helpful. Among the compared feature selection methods, one should be chosen according to the classification task, which will maximize the accuracy, e.g., for AD vs. HC, SMML is preferred rather than HGM-FS. Several imaging and genetic features are commonly selected by the feature selection methods in our experiments. These features may have high correlation with AD, and understanding why and how these features change may provide useful evidence to understand AD.

## ACKNOWLEDGMENT

This work was supported in part by NIH grants EB006733, EB008374, EB009634, AG041721, MH100217, and AG042599.

## REFERENCES

- Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., and Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the alzgene database. *Nat. Genet.* 39, 17–23. doi: 10.1038/ng1934
- Brookmeyer, R., Johnson, E., Ziegler-Graham, K., and Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement.* 3, 186–191. doi: 10.1016/j.jalz.2007.04.381
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Int. Syst. Technol.* 2, 27:1–27:27. doi: 10.1145/1961189.1961199
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., et al. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 9, 1118–1127. doi: 10.1016/S1474-4422(10)70223-4
- Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack, C. R., et al. (2012). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol. Aging* 33, 1203–1214.e2. doi: 10.1016/j.neurobiolaging.2010.10.019
- Fan, Y., Rao, H., Hurt, H., Giannetta, J., Korczykowski, M., Shera, D., et al. (2007). Multivariate examination of brain abnormality using both structural and functional MRI. *NeuroImage* 36, 1189–1199. doi: 10.1016/j.neuroimage.2007.04.009
- Hampel, H., Brgr, K., Teipel, S. J., Bokke, A. L., Zetterberg, H., and Blennow, K. (2008). Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's Dement.* 4, 38–48. doi: 10.1016/j.jalz.2007.08.006
- Hardy, J., and Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297, 353–356. doi: 10.1126/science.1072994
- He, X., Cai, D., and Niyogi, P. (2006). "Laplacian score for feature selection," in *proceeding of Advances in Neural Information Processing Systems*, Vol. 18 (Vancouver, BC), 507–514.
- Hinrichs, C., Singh, V., Xu, G., and Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 55, 574–589. doi: 10.1016/j.neuroimage.2010.10.081
- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 9, 119–128. doi: 10.1016/S1474-4422(09)70299-6
- Janka, Z., Juhasz, A., Rimanoczy, A., Boda, K., Marki-Zay, J., and Kalman, J. (2001). Codon 311 (cys— > ser) polymorphism of paraoxonase-2 gene is associated with apolipoprotein e4 allele in both Alzheimer's and vascular dementias. *Mol. Psychiatry* 7, 110–112. doi: 10.1038/sj/mp/4000916
- Kabani, N. J., Macdonald, D. J., Holmes, C. J., and Evans, A. C. (1998). "3D anatomical atlas of the human brain," in *20th Annual Meeting of the Organization for Human Brain Mapping* (Montreal, QC).
- Kohannim, O., Hua, X., Hibar, D. P., Lee, S., Chou, Y.-Y., Toga, A. W., et al. (2010). Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiol. Aging* 31, 1429–1442. doi: 10.1016/j.neurobiolaging.2010.04.022
- Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., et al. (2012). Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol. Aging* 33, 427.e15–e30. doi: 10.1016/j.neurobiolaging.2010.11.008
- Liu, F., Suk, H.-I., Wee, C.-Y., Chen, H., and Shen, D. (2013). "High-order graph matching based feature selection for Alzheimer's disease identification," in *proceedings of 2013 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)* (Nagoya).
- Liu, M., Zhang, D., and Shen, D. (2012). Ensemble sparse classification of Alzheimer's disease. *NeuroImage* 60, 1106–1116. doi: 10.1016/j.neuroimage.2012.01.055
- Miyashita, A., Arai, H., Asada, T., Imagawa, M., Matsubara, E., Shoji, M., et al. (2007). Genetic association of CTNNA3 with late-onset alzheimer's disease in females. *Hum. Mol. Genet.* 16, 2854–2869. doi: 10.1093/hmg/ddm244
- Nho, K., Corneveaux, J. J., Kim, S., Lin, H., Risacher, S. L., Shen, L., et al. (2013a). Identification of functional variants from whole-exome sequencing, combined with neuroimaging genetics. *Mol. Psychiatry* 18, 739–739. doi: 10.1038/mp.2013.81
- Nho, K., Corneveaux, J. J., Kim, S., Lin, H., Risacher, S. L., Shen, L., et al. (2013b). Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Mol. Psychiatry* 18, 781–787. doi: 10.1038/mp.2013.24



- Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., et al. (2009). Mild cognitive impairment: ten years later. *Arch. Neurol.* 66, 1447–1455. doi: 10.1001/archneurol.2009.266
- Rakotomamonjy, A., Rouen, U. D., Bach, F., Canu, S., and Grandvalet, Y. (2008). SimpleMKL. *J. Mach. Learn. Res.* 9, 2491–2521.
- Salas-Gonzalez, D., Grriz, J., Ramirez, J., Illán, I., López, M., Segovia, F., et al. (2010). Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images. *Med. Phys.* 37, 6084–6095. doi: 10.1118/1.3488894
- Saykin, A. J., Shen, L., Foroud, T. M., Potkin, S. G., Swaminathan, S., Kim, S., et al. (2010). Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimer's Dement.* 6, 265–273. doi: 10.1016/j.jalz.2010.03.013
- Sperling, R., and Johnson, K. (2013). Biomarkers of Alzheimer disease: current and future applications to diagnostic criteria. *Continuum (Minneapolis Minn.)* 19, 325–338. doi: 10.1212/01.CON.0000429181.60095.99
- Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Ser. B* 58, 267–288.
- Vemuri, P., Wiste, H., Weigand, S., Shaw, L., Trojanowski, J., Weiner, M., et al. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations. *Neurology* 73, 287–293. doi: 10.1212/WNL.0b013e3181af79e5
- Wang, H., Nie, F., Huang, H., and Ding, C. H. Q. (2013). “Heterogeneous visual features fusion via sparse multimodal machine,” in *The 26th IEEE Conference on Computer Vision and Pattern Recognition* (Portland, OR), 3097–3102.
- Wang, H., Nie, F., Huang, H., Kim, S., Nho, K., Risacher, S., et al. (2012a). Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* 28, 229–237. doi: 10.1093/bioinformatics/btr649
- Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, A., and Shen, L. (2012b). Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* 28, i127–i136. doi: 10.1093/bioinformatics/bts228
- Wang, Y., Nie, J., Yap, P.-T., Shi, F., Guo, L., and Shen, D. (2011). “Robust deformable-surface-based skull-stripping for large-scale studies,” in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, eds G. Fichtinger, A. L. Martel, and T. M. Peters (Toronto, ON: Springer), 635–642.
- Wee, C.-Y., Yap, P.-T., Li, W., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2011). Enriched white matter connectivity networks for accurate identification of MCI patients. *NeuroImage* 54, 1812–1822. doi: 10.1016/j.neuroimage.2010.10.026
- Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2012). Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage* 59, 2045–2056. doi: 10.1016/j.neuroimage.2011.10.015
- Weiner, M. W., Aisen, P. S., Jack, C. R. Jr., Jagust, W. J., Trojanowski, J. Q., Shaw, L., et al. (2010). The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's Dement.* 6, 202–211.e7. doi: 10.1016/j.jalz.2010.03.007
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2012). The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's Dement.* 8, S1–S68. doi: 10.1016/j.jalz.2011.09.172
- Wimo, A., Ljunggren, G., and Winblad, B. (1997). Costs of dementia and dementia care: a review. *Int. J. Geriatr. Psychiatry* 12, 841–856. doi: 10.1002/(SICI)1099-1166(199708)12:8<841::AID-GPS652>3.0.CO;2-R
- Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D., Rueckert, D., et al. (2011). Multi-method analysis of mri images in early diagnostics of Alzheimer's disease. *PLoS ONE* 6:e25446. doi: 10.1371/journal.pone.0025446
- Xu, Z., Jin, R., Yang, H., King, L., and Lyu, M. R. (2010). “Simple and efficient multiple kernel learning by group lasso,” in *Proceedings of the 27th International Conference on Machine Learning* (Haifa), 1175–1182.
- Zhang, D., and Shen, D. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage* 59, 895–907. doi: 10.1016/j.neuroimage.2011.09.069
- Zhang, D., Shen, D., and Alzheimer's Disease Neuroimaging, I. (2012a). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS ONE* 7:e33182. doi: 10.1371/journal.pone.0033182
- Zhang, Z., Sturgess, P., Sengupta, S., Crook, N., and Torr, P. H. S. (2012b). “Efficient discriminative learning of parametric nearest neighbor classifiers,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Providence), 2232–2239. doi: 10.1109/CVPR.2012.6247932
- Zhang, Z., Ladicky, L., Torr, P. H. S., and Saffari, A. (2011b). “Learning anchor planes for classification,” in *Advances in Neural Information Processing Systems 24* (Granada), 1611–1619.
- Zhang, Z., Li, Z.-N., and Drew, M. S. (2010). “AdaMKL: A novel biconvex multiple kernel learning approach,” in *ICPR* (Istanbul), 2126–2129. doi: 10.1109/ICPR.2010.521
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zhou, L., Wang, Y., Li, Y., Yap, P.-T., Shen, D., and Alzheimer's Disease Neuroimaging, I. (2011). Hierarchical anatomical brain networks for MCI prediction: revisiting volumetric measures. *PLoS ONE* 6:e21935. doi: 10.1371/journal.pone.0021935

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 July 2014; accepted: 10 September 2014; published online: 17 October 2014.

Citation: Zhang Z, Huang H, Shen D and The Alzheimer's Disease Neuroimaging Initiative (2014) Integrative analysis of multi-dimensional imaging genomics data for Alzheimer's disease prediction. *Front. Aging Neurosci.* 6:260. doi: 10.3389/fnagi.2014.00260

This article was submitted to the journal *Frontiers in Aging Neuroscience*.

Copyright © 2014 Zhang, Huang, Shen and The Alzheimer's Disease Neuroimaging Initiative. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.