Task Balanced Multimodal Feature Selection to Predict the Progression of Alzheimer's Disease

Lodewijk Brand
Department of Computer Science
Colorado School of Mines
Golden, CO 80401, USA
lbrand@mymail.mines.edu

Braedon O'Callaghan
Department of Computer Science
Colorado School of Mines
Golden, CO 80401, USA
bocallaghan@mymail.mines.edu

Anthony Sun
Department of Computer Science
Colorado School of Mines
Golden, CO 80401, USA
sun@mymail.mines.edu

Hua Wang
Department of Computer Science
Colorado School of Mines
Golden, CO 80401, USA
huawangcs@gmail.com

for the Alzheimer's Disease Neuroimaging Initiative

Abstract—**The social and financial costs associated with Alzheimer's disease (AD) result in significant burdens on our society. In order to understand the causes of this disease, public-private partnerships such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) release data into the scientific community. These data are organized into various modalities (genetic, brain-imaging, cognitive scores, diagnoses, etc.) for analysis. Many statistical learning approaches used in medical image analysis do not explicitly take advantage of this multimodal data structure. In this work we propose a novel objective function and optimization algorithm that is designed to handle multimodal information for the prediction and analysis of AD. Our approach relies on robust matrix-factorization and row-wise sparsity provided by the $\ell_{2,1}$-norm in order to integrate multimodal data provided by the ADNI. These techniques are jointly optimized with a classification task to guide the feature selection in our proposed *Task Balanced Multimodal Feature Selection* method. Our results, when compared against some widely used machine learning algorithms, show improved balanced accuracies, precision, and Matthew's correlation coefficients for identifying cognitive decline. In addition to the improved prediction performance, our method is able to identify brain and genetic biomarkers that are of interest to the clinical research community. Our experiments validate existing brain biomarkers and single nucleotide polymorphisms located on chromosome 11 and detail novel polymorphisms on chromosome 10 that, to the best of the authors' knowledge, have not previously been reported. We anticipate that our method will be of interest to the greater research community and have released our method's code online.[1]**

Index Terms—**Alzheimer's disease, multimodal, classification, alternating direction method of multipliers, biomarker identification**

I. Introduction

Alzheimer's Disease (AD) is a chronic neurodegenerative condition that has significant health impacts on affected patients and imparts significant financial burden on society. AD is a progressive disease characterized by loss of memory and essential mental function. AD affects the neurons in the brain involved in thinking, learning and memory. Cognitive decline manifests itself in a patient when brain cells are damaged or destroyed by AD. By 2030, according to the Alzheimer's Association [12], the number of people worldwide living with AD is estimated to to rise to nearly 76 million people. Given the massive social and financial costs associated with AD it is critical that we develop strategies for the early-diagnosis and treatment of the disease.

As of 2019, none of the pharmacological treatments are able to stop or slow down the disease. The medications that are currently available are designed to temporarily alleviate symptoms associated with AD, not cure the disease. To address this gap researchers are working towards the development of AD treatments that are able to slow or stop the progression of the disease. In order to make progress on this goal the research community has tried to address two main issues; first, identify the underlying mechanisms of the disease and second, develop novel treatments that can halt progression of the disease.

The underlying mechanisms behind the development of the disease are not well understood. In a recent study [14] it was shown that damage to brain cells in the precuneus, a brain region related to cognitive function, can occur 20 years before any AD-related symptoms are observed. It is hypothesized that the folding of amyloid-β and tau proteins leads to the neurodegeneration that can lead to a future diagnosis although it is unclear exactly what mechanism causes this folding. In addition, many clinical AD trials have failed [26] to produce significant results in slowing or halting the progression of AD. It is possible, given that irreversible neural damage can occur years before any symptoms are observed, that patients included in these clinical studies had significant damage before treatment began during the trial. Nonetheless, both of these

[1] Code is provided at: https://github.com/minds-mines/TBMFS.jl

2471-7819/20/$31.00 ©2020 IEEE
DOI 10.1109/BIBE50027.2020.00040

issues has led to a concerted effort in identifying AD-relevant biomarkers that are predictive of a future AD diagnosis.

Organizations [19] such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) provide clinical data to researchers to analyze and understand the disease. These clinical data sources are inherently multimodal, meaning that a single patient may have data associated with multiple clinical tests and tools. Recent works [6], [8], [21], [33], [38], [39] have shown promise in diagnosing AD with machine learning approaches, although, many of these algorithms do not explicitly take into account the multimodal structures associated with the data provided. Recent multimodal deep learning approaches [35], have used various neural architectures to extract latent features from complex multimodal data. Once these latent features are extracted, they are concatenated together for the final classification task; this two-stage approach does not allow for the available labeled data to be effectively utilized during the multimodal combination.

In this work, we develop a novel method to combine multimodal neuroimaging and genetic data which is *jointly optimized* with a classification task; namely, identifying the cognitive status of patients in the ADNI cohort. Our approach, optimized by the alternating direction method of multipliers [5], works to balance multimodal feature selection with classification to simultaneously identify which features are important for an AD diagnosis. We present the following scientific contributions:

- A novel objective function that balances feature selection and classification to fuse multimodal data available through the ADNI.
- An algorithm derivation, using the multi-block alternating direction method of multipliers framework, to optimize the proposed *Task Balanced Multimodal Feature Selection* objective.
- Improved classification performance against an array of machine learning algorithms that have been widely used in AD classification and multimodal data integration.
- A validation of existing biomarkers reported in AD literature and an identification of a novel collection of genetic single nucleotide polymorphism (SNPs), specifically on chromosome 10, that warrant further investigation.

II. Methods

In this section we present the justification behind our proposed *Task Balanced Multimodal Feature Selection* method, build an associated objective function and propose an optimization algorithm. For the remainder of this manuscript we represent the rows and columns of the matrix $\mathbf{X}$ as $\mathbf{x}^i$ and $\mathbf{x}_i$ respectively.

A. Our Objective

The goal of our work is to design an algorithm that is able to integrate multiple sources of data, reduce their feature space, improve the classification accuracy, all while maintaining model interpretability. We begin with the multi-task feature learning objective motivated by Liu *et al.* [25]

$$\min_{\mathbf{W}} \ \|\mathbf{Y} - \mathbf{W}\mathbf{X}\|_F^2 + \gamma \|\mathbf{W}\|_{2,1} \ , \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{d \times N}, \mathbf{W} \in \mathbb{R}^{c \times d}$ and $\mathbf{Y} \in \mathbb{R}^{c \times N}$ are the input, regression coefficient and output matrices, $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{n} \left\|\mathbf{x}^i\right\|_2$ is the $\ell_{2,1}$-norm, and γ is a hyperparameter designed to control the row-sparsity of $\mathbf{W}$. The problem in Eq. (1) aims to learn a multi-target regression model to jointly predict c-related regression targets. It is worth noting that if c equals one than Eq. (1) is equivalent to lasso [34] regression. We aim to use the feature selection property of the $\ell_{2,1}$-norm to identify important features in the input data $\mathbf{X}$ via matrix-factorization. Thus, we can rewrite the optimization in Eq. (1) as

$$\min_{\mathbf{B}, \mathbf{Z}\mathbf{Z}^T = \mathbf{I}} \ \|\mathbf{X} - \mathbf{B}\mathbf{Z}\|_F^2 + \gamma \|\mathbf{B}\|_{2,1} \ , \tag{2}$$

where $\mathbf{B} \in \mathbb{R}^{d \times r}$, $\mathbf{Z} \in \mathbb{R}^{r \times n}$, and r a hyperparameter. Note that for any pair $\{\mathbf{B}, \mathbf{Z}\}$ a corresponding pair $\{\mathbf{B}/\alpha, \alpha\mathbf{Z}\}$ with $\alpha > 1$ has a smaller objective value of Eq. (2); this will force α to go to infinity. To handle this issue we carefully include an orthogonal constraint on $\mathbf{Z}\mathbf{Z}^T$. Equation (2), due to the squared Frobenius norm in the first term, is known to be sensitive to outliers in the input data $\mathbf{X}$. Following many existing works in statistical learning and data mining [11], [36], [37], we replace the squared Frobenius norm with the $\ell_{2,1}$-norm to optimize

$$\min_{\mathbf{B}, \mathbf{Z}\mathbf{Z}^T = \mathbf{I}} \ \|\mathbf{X} - \mathbf{B}\mathbf{Z}\|_{2,1} + \gamma \|\mathbf{B}\|_{2,1} \ . \tag{3}$$

Then, inspired by recent work from Ghosal *et al.* [13], we generalize the formulation in Eq. (3) to account for M modalities by

$$\min_{\mathbf{B}_m, \mathbf{Z}\mathbf{Z}^T = \mathbf{I}} \ \left[\alpha_m \|\mathbf{X}_m - \mathbf{B}_m\mathbf{Z}\|_{2,1} + \gamma_m \|\mathbf{B}_m\|_{2,1}\right] \ , \tag{4}$$

where each α_m and γ_m balance modality reconstruction and feature selection. In Eq. (4) we aim to learn a latent space representation, $\mathbf{Z}$, constructed via the multimodal features identified by each $\mathbf{B}_m \in \mathbb{R}^{d_m \times r}$. Finally, in order to identify features in each $\mathbf{X}_m$ that are predictive of a particular diagnosis we balance the feature selection terms in Eq. (4) with a classification task. To guide the latent space discovery we incorporate a hinge-loss support vector machine (SVM) into our objective

$$\begin{aligned} \min_{\substack{\mathbf{w}_k, b_k, \mathbf{B}_m, \\ \mathbf{Z}\mathbf{Z}^T = \mathbf{I}}} \ & \sum_{m=1}^{M} \left[\alpha_m \|\mathbf{X}_m - \mathbf{B}_m\mathbf{Z}\|_{2,1} + \gamma_m \|\mathbf{B}_m\|_{2,1}\right] \\ & + \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2 + C \sum_{i=1}^{N}\sum_{k=1}^{K} \left[\left(1 - \left(\mathbf{w}_k^T \mathbf{z}_i + b_k\right) y_{ik}\right)_+\right] \ , \end{aligned} \tag{5}$$

where $C > 0$ is a regularization parameter, $y_{ik} \in \{-1, 1\}$ are the multi-class labels associated with the i-th patient and $(\cdot)_+$ is defined as $(a)_+ = \max(0, a)$. Note that the classification takes as input the latent space $\mathbf{Z}$ instead of

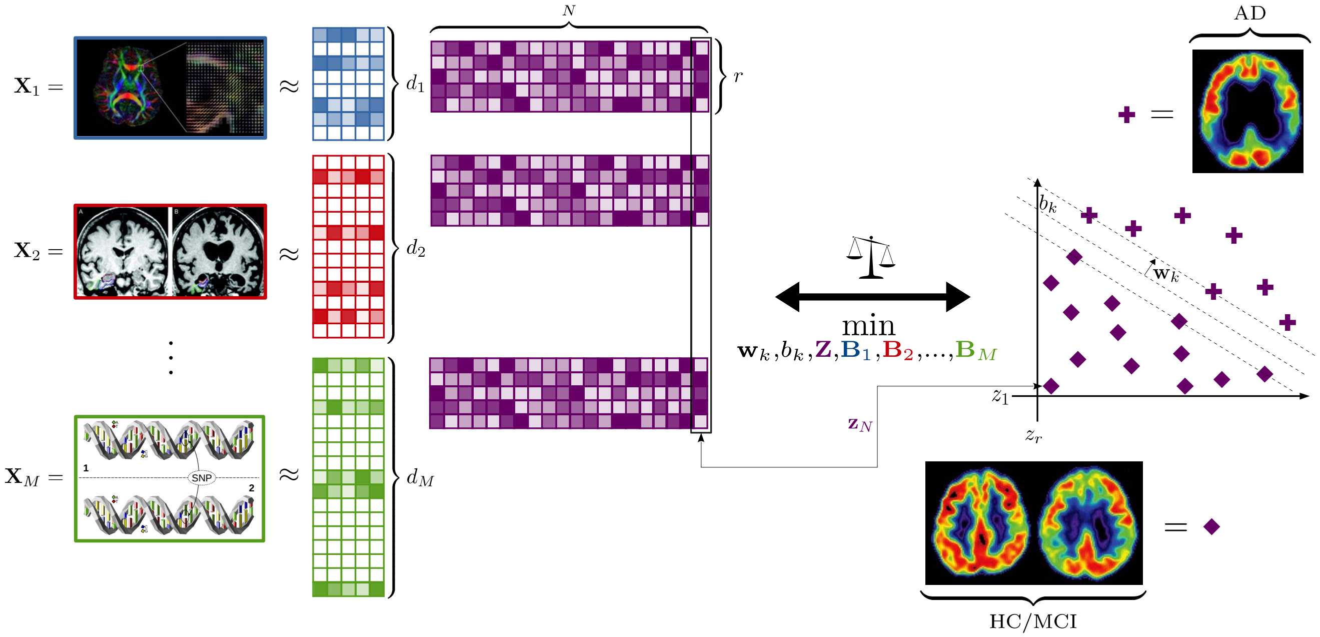


Fig. 1: Visualization of the *Task Balanced Multimodal Feature Selection* method. Our model takes as input M modalities, $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_M \in \mathbb{R}^{d_m \times N}$, and discovers a latent representation $\mathbf{Z} \in \mathbb{R}^{r \times N}$ (in purple) by way of simultaneous matrix factorization with $\mathbf{B}_1, \mathbf{B}_2, \ldots \mathbf{B}_M \in \mathbb{R}^{d_m \times r}$ (in blue, red, ..., and green). The M factorizations are jointly optimized with a classifier, specifically a support vector machine. Note that the N-th column of $\mathbf{Z}$ is shared across the M factorizations and is jointly dependent on the classification task. This work incorporates robust matrix-factorization and row-wise sparsity on each $\mathbf{B}_m$ by way of the $\ell_{2,1}$-norm to improve the joint classification task and identify biomarkers that are predictive of Alzheimer's disease. (Viewed best in color)

the raw multimodal data $\mathbf{X}_m$; this introduces an additional coupling between the first and last terms. We call Eq. (5) the *Task Balanced Multimodal Feature Selection* objective. A visual representation of our proposed method is provided in Figure 1. While the objective of our new method in Eq. (5) is clearly and reasonably motivated, the terms are dependent on one another, making it difficult to optimize this objective in general. To solve the proposed objective we derive an efficient iterative algorithm using the multi-block extension [16] of the alternating direction method of multipliers (ADMM).

B. Alternating Direction Method of Multipliers

The ADMM has been widely used to solve problems in bioinformatics, signal processing, and many other application areas across statistical learning [5]. The ADMM aims to decouple a larger and more difficult problem into a series of smaller sub-problems that are easier to solve. An extension to the ADMM, known as the multi-block ADMM [16], is designed to extend the ADMM framework to optimize functions of the following form:

$$\begin{aligned} \min_{x_i} \quad & f_1(x_1) + f_2(x_2) + \cdots + f_K(x_K) \ , \\ \text{subject to} \quad & \mathbf{E}_1 x_1 + \mathbf{E}_2 x_2 + \cdots + \mathbf{E}_K x_K = c \ . \end{aligned} \tag{6}$$

Equation. (6) can be solved by minimizing the following unconstrained objective:

$$\mathcal{L}(x_1, \ldots, x_k, \lambda) = \sum_{k=1}^{K} f(x_k) + \frac{\mu}{2} \left\| \sum_{k=1}^{K} \mathbf{E}_k x_k - c + \frac{1}{\mu}\lambda \right\|_2^2 \ , \tag{7}$$

where λ is a Lagrangian multiplier and $\mu > 0$ is a penalty parameter. The objective in Eq. (7) can be solved by the following iterative procedure that updates each x_k (primal) and the Lagrangian variable λ (dual):

$$\begin{cases} x_1^{t+1} \leftarrow \arg\min_{x_1} \mathcal{L}(x_1^t, x_2^t, \cdot\cdot, x_K^t, \lambda^t) \ , \\ x_2^{t+1} \leftarrow \arg\min_{x_2} \mathcal{L}(x_1^{t+1}, x_2^t, \cdot\cdot, x_K^t, \lambda^t) \ , \\ \qquad \cdots \\ x_K^{t+1} \leftarrow \arg\min_{x_K} \mathcal{L}(x_1^{t+1}, x_2^{t+1}, \ldots, x_K^t, \lambda^t) \ , \\ \lambda^{t+1} = \lambda^t + \mu \Big(\sum_{k=1}^{K} \mathbf{E}_k x_k - c \Big) \ , \\ \mu^{t+1} = \rho \mu^t \ , \end{cases} \tag{8}$$

where $\rho > 1$ is a constant. The process described above in Eq. (8) is repeated until the algorithm converges.

C. Algorithm Derivation

Since the terms in Eq. (5) are coupled across the predictors $\mathbf{Z}$, $\mathbf{B}_m$, and $\mathbf{W}$ and includes the non-smooth $\ell_{2,1}$-norm, it is difficult to optimize in general. To decouple the objective we introduce four sets of constraints $\hat{\mathbf{Z}} = \mathbf{Z}$, $\mathbf{F}_m = \mathbf{X}_m - \mathbf{B}_m\mathbf{Z}$, $\hat{\mathbf{B}}_m = \mathbf{B}_m$, and $e_{ik} = y_{ik} - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right)$. Since $y_{ik} \in \{-1, 1\}$, it follows that $1 - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right) y_{ik} = y_{ik}y_{ik} - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right) y_{ik} = y_{ik}\left(y_{ik} - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right)\right)$ [27]. This allows us to derive the SVM updates in the primal instead of the dual. Then, following the multi-block ADMM framework described above, we systematically incorporate these constraints into the objective

$$\begin{aligned}\min_{\substack{e_{ik},\mathbf{w}_k,b_k,\mathbf{Z},\\ \hat{\mathbf{Z}},\mathbf{F}_m,\mathbf{B}_m,\hat{\mathbf{B}}_m}} \mathcal{L} = \sum_{m=1}^{M} \Bigg[& \alpha_m \|\mathbf{F}_m\|_{2,1} + \frac{\mu}{2}\|\mathbf{F}_m - \mathbf{L}_m\|_F^2 \\ & + \gamma_m \left\|\hat{\mathbf{B}}_m\right\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{B}_m - \hat{\mathbf{B}}_m + \frac{1}{\mu}\mathbf{\Lambda}_m\right\|_F^2 \Bigg] \\ & + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{w}_k\|_2^2 + C\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_{ik}e_{ik}\right)_+ \\ & + \frac{\mu}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(e_{ik} - \left(y_{ik} - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right)\right) + \frac{1}{\mu}\eta_{ik}\right)^2 \\ & + \frac{\mu}{2}\left\|\mathbf{Z} - \hat{\mathbf{Z}} + \frac{1}{\mu}\mathbf{\Omega}\right\|_F^2 \quad \text{subject to} \quad \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T = \mathbf{I}\ , \end{aligned} \tag{9}$$

where $\mathbf{L}_m = (\mathbf{X}_m - \mathbf{B}_m\mathbf{Z}) - \frac{1}{\mu}\mathbf{\Theta}_m$, $\mathbf{\Theta}_m$, $\mathbf{\Lambda}_m$, $\mathbf{\Omega}$, and η_{ik} are Lagrange multipliers and $\mu > 0$ is a penalty parameter. For the remainder of this section, we derive the multi-block ADMM update steps for minimizing Eq. (9).

$\mathbf{w}_k$ **Update:** Removing terms not dependent on $\mathbf{w}_k$ from Eq. (9) gives

$$\min_{\mathbf{w}_k} \frac{1}{2}\|\mathbf{w}_k\|_2^2 + \frac{\mu}{2}\sum_{i=1}^{N}\left(e_{ik} - y_{ik} + \mathbf{w}_k^T\mathbf{z}_i + b_k + \frac{1}{\mu}\eta_{ik}\right)^2 . \tag{10}$$

Taking the derivative of Eq. (10) with respect to $\mathbf{w}_k$, setting the result equal to zero, and solving for $\mathbf{w}_k$ gives

$$\begin{aligned}\mathbf{w}_k^T = & \sum_{i=1}^{N}\left[\left(y_{ik} - e_{ik} - b_k - \frac{1}{\mu}\eta_{ik}\right)\mathbf{z}_i^T\right] * \\ & \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i^T + \frac{1}{\mu}\mathbf{I}\right)^{-1} \quad , \end{aligned} \tag{11}$$

b_k **Update:** Taking the derivative of Eq. (10) with respect to b_k, setting the result equal to zero, and solving for b_k gives

$$b_k = \frac{\sum_{i=1}^{N}\left(y_{ik} - e_{ik} - \mathbf{w}_k^T\mathbf{z}_i - \frac{1}{\mu}\eta_{ik}\right)}{N} \quad . \tag{12}$$

e_{ik} **Update:** Removing terms not dependent on e_{ik} from Eq. (9) gives

$$\min_{e_{ik}} C\left(y_{ik}e_{ik}\right)_+ + \frac{\mu}{2}\left(e_{ik} - s_{ik}\right)^2 \quad , \tag{13}$$

```
Algorithm 1 ADMM algorithm to optimize Eq. (9)
Data: Multimodal data $\mathbf{X}_m$ for $m \in [1, M]$ and the $N \times K$ class labels $y_{ik} \in \mathbf{Y}$.
Hyperparameters: $C > 0$, $\alpha_m > 0$, $\gamma_m > 0$, $\mu > 0$, $\rho > 1$ and $r \in \mathbb{Z}_{\geq 1}$.
Initialize: $e_{ik}, \mathbf{w}_k, b_k, \mathbf{Z}, \hat{\mathbf{Z}}, \mathbf{F}_m, \mathbf{B}_m, \hat{\mathbf{B}}_m$, $\mathbf{\Theta}_m$, $\mathbf{\Lambda}_m$ and $\mathbf{\Omega}$.
while the objective in Eq. (9) not converged do
    for $k \in K$ do
        Update $\mathbf{w}_k$ by Eq. (11).
        Update $b_k$ by Eq. (12).
        for $i \in N$ do
            Update $e_{ik}$ by Eq. (14).
            Update $\eta_{ik} = \eta_{ik} + \mu(e_{ik} - y_{ik} + \mathbf{w}_k^T\mathbf{z}_i + b_k)$.
        end for
    end for
    Update $\mathbf{z}_i \in \mathbf{Z}$ by Eq. (16)
    Update $\hat{\mathbf{Z}}$ by Eq. (18)
    for $m \in M$ do
        Update $\mathbf{f}^i \in \mathbf{F}_m$ by Eq. (20).
        Update $\hat{\mathbf{b}}^i \in \hat{\mathbf{B}}_m$ by Eq. (22).
        Update $\mathbf{B}_m$ by Eq. (24).
        Update $\mathbf{\Theta}_m = \mathbf{\Theta}_m + \mu\left(\mathbf{F}_m - (\mathbf{X}_m - \mathbf{B}_m\mathbf{Z})\right)$.
        Update $\mathbf{\Lambda}_m = \mathbf{\Lambda}_m + \mu(\mathbf{B}_m - \hat{\mathbf{B}}_m)$.
    end for
    Update $\mathbf{\Omega} = \mathbf{\Omega} + \mu(\mathbf{Z} - \hat{\mathbf{Z}})$.
    Update $\mu = \rho\mu$
end while
```

where $s_{ik} = \left(y_{ik} - \left(\mathbf{w}_k^T\mathbf{z}_i + b_k\right)\right) - \frac{1}{\mu}\eta_{ik}$. Taking the derivative of Eq. (13) with respect to e_{ik}, setting the result equal to zero, and solving for each e_{ik} gives the closed-form updates

$$e_{ik} = \begin{cases} s_{ik} - \frac{C}{\mu}y_{ik} & \text{when } y_{ik}s_{ik} > \frac{C}{\mu}\ , \\ 0 & \text{when } 0 \leq y_{ik}s_{ik} \leq \frac{C}{\mu}\ , \\ s_{ik} & \text{when } y_{ik}s_{ik} < 0\ . \end{cases} \tag{14}$$

Z Update: Removing terms not dependent on $\mathbf{Z}$ from Eq. (9) and optimizing each column of $\mathbf{Z}$ individually gives the following N minimizations

$$\begin{aligned}\min_{\mathbf{z}_i} & \sum_{m=1}^{M}\|\mathbf{B}_m\mathbf{z}_i - \mathbf{t}_{im}\|_2^2 + \sum_{k=1}^{K}\left(\mathbf{w}_k^T\mathbf{z}_i - u_{ik}\right)^2 \\ & + \left\|\mathbf{z}_i - \hat{\mathbf{z}}_i + \frac{1}{\mu}\omega_i\right\|_2^2 \quad , \end{aligned} \tag{15}$$

where ω_i are the columns of $\mathbf{\Omega}$, $\mathbf{t}_{im} = \mathbf{x}_{im} - \mathbf{f}_{im} - \frac{1}{\mu}\theta_{im}$ and $u_{ik} = y_{ik} - e_{ik} - b_k - \frac{1}{\mu}\eta_{ik}$. Taking the derivative of Eq. (15) with respect to $\mathbf{z}_i$, setting it equal to zero, and solving for $\mathbf{z}_i$ gives the closed-form update

$$\begin{aligned}\mathbf{z}_i = & \left(\sum_{m=1}^{M}\mathbf{B}_m^T\mathbf{B}_m + \sum_{k=1}^{K}\mathbf{w}_k\mathbf{w}_k^T + \mathbf{I}\right)^{-1} * \\ & \left[\sum_{m=1}^{M}\mathbf{B}_m^T\mathbf{t}_{im} + \sum_{k=1}^{K}\mathbf{w}_k u_{ik} + \hat{\mathbf{z}}_i - \frac{1}{\mu}\omega_i\right] . \end{aligned} \tag{16}$$

$\hat{\mathbf{Z}}$ Update: Removing terms not dependent on $\hat{\mathbf{Z}}$ from Eq. (9) gives

$$\min_{\hat{\mathbf{Z}}} \left\| \mathbf{Z} - \hat{\mathbf{Z}} + \frac{1}{\mu}\mathbf{\Omega} \right\|_F^2 \quad \text{subject to} \quad \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T = \mathbf{I} \ , \qquad (17)$$

which is an instance of the orthogonal Procustes problem [31]:

$$\hat{\mathbf{Z}} = \mathbf{U}\mathbf{V}^T \text{ where } \left\{\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^T\right\} = svd(\mathbf{Z} + \frac{1}{\mu}\mathbf{\Omega}) \ . \qquad (18)$$

$\mathbf{F}_m$ Update: Removing all terms from Eq. (9) that do not include $\mathbf{F}_m$ gives

$$\min_{\mathbf{F}_m} \alpha_m \|\mathbf{F}_m\|_{2,1} + \frac{\mu}{2}\|\mathbf{F}_m - \mathbf{L}_m\|_F^2 \ . \qquad (19)$$

We can decouple Eq. (19) by row and use the results derived in [23] to update each row of $\mathbf{f}^i$ in a give $\mathbf{F}_m$ by

$$\mathbf{f}^i = \mathbf{l}^i \left(1 - \alpha_m/(\mu \left\|\mathbf{l}^i\right\|_2)\right)_+ \ , \qquad (20)$$

where $(x)_+ = \max(0, x)$. This procedure is repeated for $m \in [1, M]$.

$\hat{\mathbf{B}}_m$ Update: Dropping all terms without a $\hat{\mathbf{B}}_m$ in Eq. (9) gives

$$\min_{\hat{\mathbf{B}}_m} \gamma_m \left\|\hat{\mathbf{B}}_m\right\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{O} - \hat{\mathbf{B}}_m\right\|_F^2 \ , \qquad (21)$$

where $\mathbf{O}_m = \mathbf{B}_m + \frac{1}{\mu}\mathbf{\Lambda}_m$. Similar to the update for $\mathbf{F}_m$, we can decouple Eq. (21) by row-by-row and derive an update for each $\hat{\mathbf{b}}^i \in \mathbf{B}_m$ by

$$\hat{\mathbf{b}}^i = \mathbf{o}^i \left(1 - \gamma_m/(\mu \left\|\mathbf{o}^i\right\|_2)\right)_+ \ . \qquad (22)$$

$\mathbf{B}_m$ Update: Keeping all terms from Eq. (9) that contain $\mathbf{B}_m$ gives

$$\min_{\mathbf{B}_m} \|\mathbf{B}_m\mathbf{Z} + \mathbf{M}_m\|_F^2 + \|\mathbf{B}_m + \mathbf{N}_m\|_F^2 \ , \qquad (23)$$

where $\mathbf{M}_m = \mathbf{F}_m - \mathbf{X}_m + \frac{1}{\mu}\mathbf{\Theta}_m$ and $\mathbf{N}_m = -\hat{\mathbf{B}}_m + \frac{1}{\mu}\mathbf{\Lambda}_m$. Taking the derivative of Eq. (23), setting the result equal to zero, and solving for $\mathbf{B}_m$ gives

$$\mathbf{B}_m = \left(-\mathbf{M}_m\mathbf{Z}^T - \mathbf{N}_m\right)\left(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}\right)^{-1} \ . \qquad (24)$$

The final sequence of primal and dual updates designed to minimize Eq. (9) are summarized in Algorithm 1.

III. Experiments

A. Experimental Data

The baseline magnetic resonance imaging (MRI) scans, single nucleotide polymorphism (SNP) arrays, and demographic information for 821 ADNI-1 participants were obtained from the ADNI website. We performed FreeSurfer automated parcellation on the MRI data and extracted voxel-based morphometry (VBM) measures for 90 target regions of interest by following steps detailed in Risacher *et al.* [29]. For the SNP data the quality control steps discussed in Shen *et al.* [32] were followed. The labels, Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy control (HC), were used as diagnostic classification groups. Participants with no missing MRI, SNP, or diagnostic information were included, providing a set of 723 subjects (170 AD, 352 MCI, 201 HC) across the FreeSurfer, VBM, and SNP modalities.

B. Experimental Settings

In the following experiments we take, as input, the multimodal data described above and perform the binary classification task to predict AD *vs*. HC/MCI. The performance results from the binary classification experiments are reported from a repeated-k-fold cross validation scheme where the input and output data are shuffled in-between each k-fold cross validation experiment. The hyperparameter settings for our method are $C = 1, \alpha_1 = 100, \alpha_2 = 100, \alpha_3 = 0.01, \gamma_1 = 100, \gamma_2 = 100, \gamma_3 = 0.1, r = 5, \mu = 0.01, \rho = 1.1$ where FreeSurfer, VBM, and SNP are the first, second, and third modalities.

We compare the proposed *Task Balanced Multimodal Feature Selection* method (*Ours*) against k-nearest neighbors (*k-NN*), support vector machines (*SVM*) via the LIBSVM library [9], ℓ_1-regularized logistic regression (*Logistic*), a multi-layer perceptron neural network (*MLP*), and two gradient boosted decision trees using the XGBoost [10] and LightGBM [22] libraries. The XGBoost [33], support vector machine [8], variants on logistic regression [21], and multi-layer perceptron neural network [38] methods have all been used in the past to identify AD *vs*. HC/MCI. The LightGBM method has been used [1] as the final classifier from the output of a deep neural architecture applied to multimodal data. For each of the compared algorithms, we concatenate features along the vertical dimension for each modality to construct the input matrix $\mathbf{X} \in \mathbb{R}^{(d_1+d_2+d_3)\times n}$. We test the performance of each compared method against the following metrics: Balanced Accuracy (BACC) [7], precision, recall, F_1-score $= 2 \times \frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, and Matthews Correlation Coefficient (MCC) [4]. Each of the compared methods have undergone extensive hyperparameter tuning to ensure a fair comparison. The logistic regression, k-nearest neighbors, multi-layer perceptron neural network algorithms, and all repeated-k-fold cross validation/hyperparameter-search results were implemented using the Flux [18] and MLJ [3] libraries.

C. Classification Performance

In Table I, we report the average performance and standard deviation results of our method compared against the aforementioned algorithms. The results presented in Table I show that our proposed algorithm outperforms the compared methods in terms of BACC, precision, and MCC. Our algorithm, within a standard deviation, is competitive against the other algorithm's F_1-scores. Our approach does not perform as well with regards to recall. Nonetheless, a high precision value indicates that our method produces fewer false positives when predicting an AD classification when compared to the other methods. This is likely due to the strong biomarker identification properties provided by the $\ell_{2,1}$-norm. This robust feature selection property may cause our approach to ignore features that contain subtle variations that may be important for a higher recall score.

Model	BACC	Precision	Recall	F_1-score	MCC
k-NN	0.515±0.031	0.508±0.016	**0.972±0.026**	0.885±0.022	0.070±0.136
SVM	0.598±0.060	0.566±0.044	0.876±0.050	0.859±0.036	0.213±0.128
Logistic	0.629±0.071	0.589±0.056	0.894±0.035	0.873±0.028	0.276±0.143
MLP	0.593±0.078	0.643±0.146	0.591±0.219	0.672±0.185	0.158±0.125
XGBoost	0.618±0.062	0.576±0.044	0.930±0.032	0.886±0.031	0.288±0.138
LightGBM	0.607±0.064	0.566±0.045	0.950±0.031	**0.893±0.031**	0.287±0.161
Ours	**0.728±0.074**	**0.751±0.096**	0.721±0.094	0.805±0.050	**0.372±0.129**

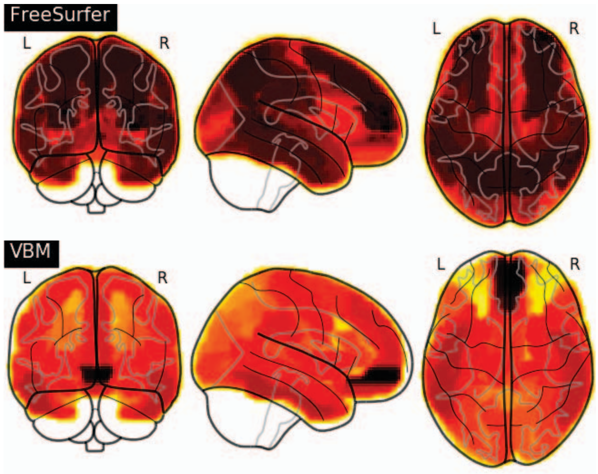
TABLE I: Ten repeated six-fold cross-validations and their standard deviations for identifying ADNI cohort participants with AD *vs*. HC/MCI. Each of the compared methods have undergone extensive hyperparameter tuning.

D. Biomarker Identification

In addition to the improved predictive performance reported in Table I, the *Task Balanced Multimodal Feature Selection* method can be analyzed to identify which biomarkers are most important for prediction. The key insight that reveals the interpretability of our model is that each learned $\mathbf{B}_m$ in Eq. (5) determines the construction of the latent representation $\mathbf{Z}$. Since the construction of $\mathbf{Z}$ is balanced via the classification task we expect that the features identified by each $\mathbf{B}_m$ can provide novel insight into AD-related biomarkers. In Figs. 2, 3 and 4, we analyze, rank, and plot each row-sum (reduced over r) of $\mathbf{B}_m$ for the FreeSurfer, VBM, and SNP modalities.

In Fig. 2 we provide the FreeSurfer and VBM biomarkers identified by our method. The top-5 areas of the brain identified by our method generally match up with the literature. For instance, atrophy of the precuneus [28] and inferior temporal gyrus [30] have been discovered in patients with AD-related dementias; these biomarkers are both ranked highly by our approach. Furthermore, Jacobs *et al*. [20] identified that increased connectivity in the parietal lobe is frequently observed in patients suffering from mild forms of AD. They argue that increased connectivity in the parietal lobe, an area ranked-highly by our method in the FreeSurfer modality, is a compensation mechanism designed to counteract mild AD symptoms; further study of these compensation mechanisms may be a promising path forward for future AD treatment.

The SNP results reported in Fig. 3 and Fig. 4 provide additional validation of our method's biomarker identification capacity. The ranked SNPs from rs2511175 through rs10899496 are associated with the GRB-2-associated-binding protein 2 (GAB2) which has been shown in multiple works [15], [17], to be associated with both early and late-onset AD. Hibar *et al*. [15] proposed that the polymorphisms associated with the GAB2 protein manifest themselves in observable changes to brain morphology. The other SNPs in the top-twenty occur on chromosome 10 and, to the best of the authors' knowledge, are not currently published in the literature. Nonetheless, we do see evidence, specifically in Bertram *et al*. [2] and Lendon *et al*. [24], that chromosome 10 could play a role in the pathology of AD. The collection of biomarkers identified in the FreeSurfer, VBM, and SNP modalities, provides substantial evidence that our approach is able identify biomarkers associated with cognitive decline.



Top-5, FreeSurfer	***Top-5, VBM***
left superior parietal lobule	*left inferior temporal gyrus*
right superior parietal lobule	*right inferior temporal gyrus*
left precuneus	*right gyrus rectus*
right precuneus	*left gyrus rectus*
right caudal middle frontal gyrus	*right medial superior frontal gyrus*

Fig. 2: FreeSurfer and VBM biomarkers identified by our method in the experiment reported in Table I. The top-5 identified brain biomarkers are listed for each modality.

IV. Conclusion

In this work we present the *Task Balanced Multimodal Feature Selection* method to identify cognitive decline in the ADNI cohort. The proposed algorithm incorporates robust matrix-factorization and feature selection balanced with a classification task and shows promising performance when applied to predict AD when compared to other popular statistical learning methods. Our approach discovers existing, as well as novel, brain and genetic biomarkers associated with AD. In addition, we release the code associated with this method to the wider research community. In the future, we plan to extend this method to other multimodal AD datasets and design novel mechanisms for incorporating longitudinal and missing clinical data.

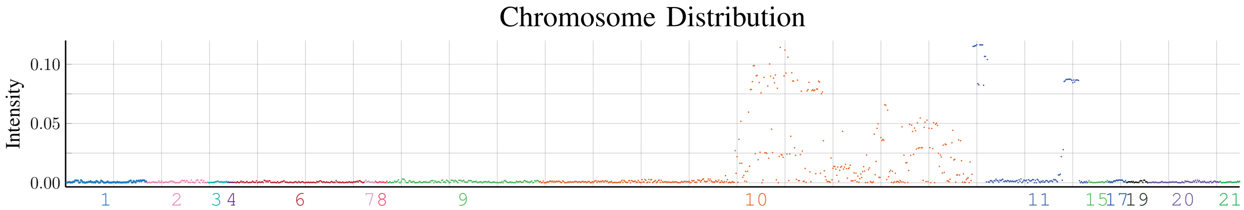


Fig. 3: SNP biomarkers identified by our method. Intensities, calculated via the absolute *row*-sum from the SNP feature selection matrix, color-coded by chromosome. The $\ell_{2,1}$-norm in the *Task Balanced Multimodal Feature Selection* objective clearly identifies a sparse set of SNPs located on chromosome 10 (orange) and chromosome 11 (blue). (Viewed best in color)

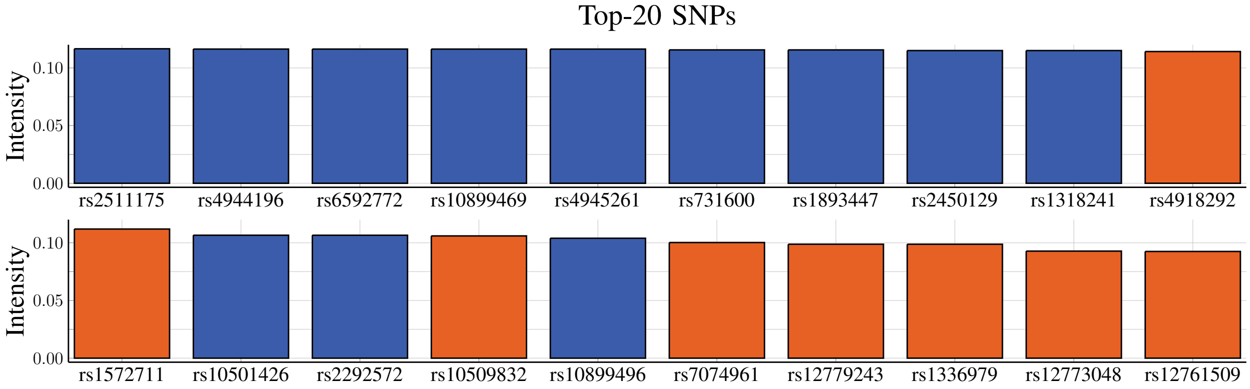


Fig. 4: Top-20 SNPs (sorted and named) identified by our method. The identified SNPs in this panel are colored by the chromosome on which they occur in Fig. 3.

Acknowledgment

Corresponding author: Hua Wang (huawangcs@gmail.com).

This work was supported in part by the National Science Foundation (NSF) under the grants of IIS 1652943, IIS 1849359, CNS 1932482 and CCF 2029543. Lodewijk Brand was supported in part by the Department of Defense (DoD) Science, Mathematics And Research for Transformation (SMART) Scholarship-for-Service program. We thank Dr. Heng Huang in the Department of Electrical and Computer Engineering of University of Pittsburgh and Dr. Li Shen in the Department of Biostatistics, Epidemiology and Informatics of University of Pennsylvania for providing with us the parsed ADNI data which are used in the experiments of this study.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DoD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is 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 University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] Akram Bakkour, John C Morris, David A Wolk, and Bradford C Dickerson. The effects of aging and alzheimer's disease on cerebral cortical anatomy: specificity and differential relationships with cognition. *Neuroimage*, 76:332–344, 2013.

[2] Lars Bertram, Deborah Blacker, Kristina Mullin, Devon Keeney, Jennifer Jones, Sanjay Basu, Stephen Yhu, Melvin G McInnis, Rodney CP Go, Konstantinos Vekrellis, et al. Evidence for genetic linkage of alzheimer's disease to chromosome 10q. *Science*, 290(5500):2302–2303, 2000.

[3] Anthony Blaom, Franz Kiraly, Thibaut Lienart, and Sebastian Vollmer. alan-turing-institute/mlj.jl: v0.5.3, November 2019.

[4] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PloS one*, 12(6), 2017.

[5] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[6] Lodewijk Brand, Kai Nichols, Hua Wang, Li Shen, and Heng Huang. Joint multi-modal longitudinal regression and classification for alzheimers disease prediction. *IEEE Transactions on Medical Imaging*, 39(6):1845–1855, 2019.

[7] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, pages 3121–3124. IEEE, 2010.

[8] Ramon Casanova, Fang-Chi Hsu, Mark A Espeland, Alzheimer's Disease Neuroimaging Initiative, et al. Classification of structural mri images in alzheimer's disease from the perspective of ill-posed problems. *PloS one*, 7(10), 2012.

[9] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

[10] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[11] Liang Du, Peng Zhou, Lei Shi, Hanmo Wang, Mingyu Fan, Wenjian Wang, and Yi-Dong Shen. Robust multiple kernel k-means using l21-norm. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[12] Joseph Gaugler, Bryan James, Tricia Johnson, Allison Marin, and Jennifer Weuve. 2019 alzheimer's disease facts and figures. *Alzheimers & Dementia*, 15(3):321–387, 2019.

[13] Sayan Ghosal, Qiang Chen, Aaron L Goldman, William Ulrich, Karen F Berman, Daniel R Weinberger, Venkata S Mattay, and Archana Venkataraman. Bridging imaging, genetics, and diagnosis in a coupled low-dimensional framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 647–655. Springer, 2019.

[14] Brian A Gordon, Tyler M Blazey, Yi Su, Amrita Hari-Raj, Aylin Dincer, Shaney Flores, Jon Christensen, Eric McDade, Guoqiao Wang, Chengjie Xiong, et al. Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant alzheimer's disease: a longitudinal study. *The Lancet Neurology*, 17(3):241–250, 2018.

[15] Derrek P Hibar, Neda Jahanshad, Jason L Stein, Omid Kohannim, Arthur W Toga, Sarah E Medland, Narelle K Hansell, Katie L McMahon, Greig I de Zubicaray, Grant W Montgomery, et al. Alzheimer's disease risk gene, gab2, is associated with regional brain volume differences in 755 young healthy twins. *Twin Research and Human Genetics*, 15(3):286–295, 2012.

[16] Mingyi Hong and Zhi-Quan Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1-2):165–199, 2017.

[17] M Arfan Ikram, Fan Liu, Ben A Oostra, Albert Hofman, Cornelia M van Duijn, and Monique MB Breteler. The gab2 gene and the risk of alzheimer's disease: replication and meta-analysis. *Biological Psychiatry*, 65(11):995–999, 2009.

[18] Michael Innes, Elliot Saba, Keno Fischer, Dhairya Gandhi, Marco Concetto Rudilosso, Neethu Mariya Joy, Tejan Karmali, Avik Pal, and Viral Shah. Fashionable modelling with flux. *CoRR*, abs/1811.01457, 2018.

[19] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

[20] HIL Jacobs, MPJ Van Boxtel, A Heinecke, EHBM Gronenschild, WH Backes, IHGB Ramakers, J Jolles, and FRJ Verhey. Functional integration of parietal lobe activity in early alzheimer disease. *Neurology*, 78(5):352–360, 2012.

[21] Piers Johnson, Luke Vandewater, William Wilson, Paul Maruff, Greg Savage, Petra Graham, Lance S Macaulay, Kathryn A Ellis, Cassandra Szoeke, Ralph N Martins, et al. Genetic algorithm with logistic regression for prediction of progression to alzheimer's disease. *BMC bioinformatics*, 15(S16):S11, 2014.

[22] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, pages 3146–3154, 2017.

[23] Matthieu Kowalski. Sparse regression using mixed norms. *Applied and Computational Harmonic Analysis*, 27(3):303–324, 2009.

[24] Corinne Lendon and Nick Craddock. Susceptibility gene (s) for alzheimer's disease on chromosome 10. *Trends in neurosciences*, 24(10):557–559, 2001.

[25] Jun Liu, Shuiwang Ji, and Jieping Ye. Multi-task feature learning via efficient l2, 1-norm minimization. *arXiv preprint arXiv:1205.2631*, 2012.

[26] Dev Mehta, Robert Jackson, Gaurav Paul, Jiong Shi, and Marwan Sabbagh. Why do trials for alzheimers disease drugs keep failing? a discontinued drug perspective for 2010-2015. *Expert opinion on investigational drugs*, 26(6):735–739, 2017.

[27] Feiping Nie, Yizhen Huang, Xiaoqian Wang, and Heng Huang. New primal svm solver with linear computational cost for big data classifications. In *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32*, pages II–505, 2014.

[28] Gil D Rabinovici, Ansgar J Furst, Adi Alkalay, Caroline A Racine, James P ONeil, Mustafa Janabi, Suzanne L Baker, Neha Agarwal, Stephen J Bonasera, Elizabeth C Mormino, et al. Increased metabolic vulnerability in early-onset alzheimers disease is not related to amyloid burden. *Brain*, 133(2):512–528, 2010.

[29] Shannon L Risacher, Li Shen, John D West, Sungeun Kim, Brenna C McDonald, Laurel A Beckett, Danielle J Harvey, Clifford R Jack Jr, Michael W Weiner, Andrew J Saykin, et al. Longitudinal mri atrophy biomarkers: relationship to conversion in the adni cohort. *Neurobiology of aging*, 31(8):1401–1418, 2010.

[30] Stephen W Scheff, Douglas A Price, Frederick A Schmitt, Melissa A Scheff, and Elliott J Mufson. Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and alzheimer's disease. *Journal of Alzheimer's Disease*, 24(3):547–557, 2011.

[31] Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[32] Li Shen, Sungeun Kim, Shannon L Risacher, Kwangsik Nho, Shanker Swaminathan, John D West, Tatiana Foroud, Nathan Pankratz, Jason H Moore, Chantel D Sloan, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort. *Neuroimage*, 53(3):1051–1063, 2010.

[33] Daniel Stamate, Min Kim, Petroula Proitsi, Sarah Westwood, Alison Baird, Alejo Nevado-Holgado, Abdul Hye, Isabelle Bos, Stephanie JB Vos, Rik Vandenberghe, et al. A metabolite-based machine learning approach to diagnose alzheimer-type dementia in blood: Results from the european medical information framework for alzheimer disease biomarker discovery cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5(C):933–938, 2019.

[34] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[35] Aleksei Tiulpin, Stefan Klein, Sita MA Bierma-Zeinstra, Jérôme Thevenot, Esa Rahtu, Joyce van Meurs, Edwin HG Oei, and Simo Saarakkala. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Scientific reports*, 9(1):1–11, 2019.

[36] Hua Wang, Feiping Nie, and Heng Huang. Robust and discriminative distance for multi-instance learning. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2919–2924. IEEE, 2012.

[37] Hua Wang, Feiping Nie, Heng Huang, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics*, 28(12):i127–i136, 2012.

[38] Shui-Hua Wang, Yin Zhang, Yu-Jie Li, Wen-Juan Jia, Fang-Yuan Liu, Meng-Meng Yang, and Yu-Dong Zhang. Single slice based detection for alzheimers disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools and Applications*, 77(9):10393–10417, 2018.

[39] Jingwen Yan, Taiyong Li, Hua Wang, Heng Huang, Jing Wan, Kwangsik Nho, Sungeun Kim, Shannon L Risacher, Andrew J Saykin, Li Shen, et al. Cortical surface biomarkers for predicting cognitive outcomes using group $\ell_{2,1}$-norm. *Neurobiology of aging*, 36:S185–S193, 2015.