

Early prediction of progression to Alzheimer's disease using multi-modality neuroimages by a novel ordinal learning model ADPacer

Lujia Wang, Zhiyang Zheng, Yi Su, Kewei Chen, David Weidman, Teresa Wu, ShihChung Lo, Fleming Lure, Jing Li, for the Alzheimer's Disease Neuroimaging Initiative & the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing

To cite this article: Lujia Wang, Zhiyang Zheng, Yi Su, Kewei Chen, David Weidman, Teresa Wu, ShihChung Lo, Fleming Lure, Jing Li, for the Alzheimer's Disease Neuroimaging Initiative & the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (29 Aug 2023): Early prediction of progression to Alzheimer's disease using multi-modality neuroimages by a novel ordinal learning model ADPacer, IISE Transactions on Healthcare Systems Engineering, DOI: [10.1080/24725579.2023.2249487](https://doi.org/10.1080/24725579.2023.2249487)

To link to this article: <https://doi.org/10.1080/24725579.2023.2249487>



Published online: 29 Aug 2023.



Submit your article to this journal [↗](#)



Article views: 30



View related articles [↗](#)



View Crossmark data [↗](#)



Early prediction of progression to Alzheimer's disease using multi-modality neuroimages by a novel ordinal learning model ADPacer

Lujia Wang^a, Zhiyang Zheng^a, Yi Su^b, Kewei Chen^b, David Weidman^b, Teresa Wu^c, ShihChung Lo^d, Fleming Lure^d, Jing Li^a, for the Alzheimer's Disease Neuroimaging Initiative*, and the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing**

^aH. Hilton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Georgia, GA, USA; ^bBanner Alzheimer's Institute, Tucson, AZ, USA; ^cSchool of Computing and Augmented Intelligence, Arizona State University, Arizona, AZ, USA; ^dMS Technologies Corp, Rockville, MD, USA

ABSTRACT

Machine learning has shown great promise for integrating multi-modality neuroimaging datasets to predict the risk of progression/conversion to Alzheimer's Disease (AD) for individuals with Mild Cognitive Impairment (MCI). Most existing work aims to classify MCI patients into converters versus non-converters using a pre-defined timeframe. The limitation is a lack of granularity in differentiating MCI patients who convert at different paces. Progression pace prediction has important clinical values, which allow for more personalized interventional strategies, better preparation of patients and their caregivers, and facilitation of patient selection in clinical trials. We proposed a novel ADPacer model which formulated the pace prediction into an ordinal learning problem with a unique capability of leveraging training samples with label ambiguity to augment the training set. This capability differentiates ADPacer from existing ordinal learning algorithms. We applied ADPacer to MCI patient cohorts from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL), and demonstrated the superior performance of ADPacer compared to existing ordinal learning algorithms. We also integrated the SHapley Additive exPlanations (SHAP) method with ADPacer to assess the contributions from different modalities to the model prediction. The findings are consistent with the AD literature.

KEYWORDS

Machine learning; ordinal learning; Alzheimer's disease; mild cognitive impairment; label ambiguity

1. Introduction

Alzheimer's Disease (AD) is a devastating neurodegenerative disorder that currently affects 6.5 million people aged 65 and older in the U.S. (Alzheimer Association, 2022). The symptoms typically start with mild memory loss and cognitive decline, and inevitably progress with gradual deterioration of other brain functions. There is currently no cure for AD. It is widely accepted that disease-modifying treatments will have greater potential when given at early stages of the disease.

Mild Cognitive Impairment (MCI) is the prodromal phase of the disease when patients show noticeable signs of memory loss and cognitive decline, but their symptoms are not severe enough to disrupt ability to carry out daily activities independently. Individuals with MCI progress to AD at different paces. Also, in part because the MCI symptoms could be caused by other underlying diseases, some individuals with MCI may not eventually progress to AD dementia.

It is important to predict the likelihood of progression to AD for MCI patients, since interventions may have greater potential to slow down the disease progression, when given early, before neuronal damage is extensive.

Neuroimaging has shown great promise to predict MCI progression to AD. Especially, images of different types/modalities measure different aspects of the brain affected by the disease. Combining data from different neuroimaging modalities has demonstrated improved prediction power than using a single modality alone. Two commonly used neuroimaging modalities are volumetric magnetic resonance imaging (MRI) and positron emission tomography (PET), which measure brain structure and function, respectively. There is an abundance of research for predicting MCI progression to AD by integrating MRI and FDG-PET—a type of PET image that measures cerebral glucose metabolism, together with some non-imaging data. The existing research typically formulates the progression prediction into a

CONTACT Jing Li  jl3175@gatech.edu  H. Hilton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Georgia, GA USA.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

**Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

classification problem, i.e., to classify an MCI subject as a converter if the subject progresses/converts to AD within a pre-defined timeframe, and a non-converter otherwise. Next, we provide a brief review of this area, with a focus on methods developed in the recent few years.

Liu et al. (2017) proposed a view-aligned hypergraph learning method, which divided data into views based on different combinations of modalities and generated sparse representation to construct a hypergraph in each view space. Coherence among the views was obtained by using a view-aligned regularizer to generate class probability scores. The final classification was achieved by assembling the scores through a multi-view label fusion method. Zhou et al. (2019) proposed a stage-wise deep neural network, in which the latent representations of different combinations of modalities were learned at each stage of the neural network by taking the learned representation from the stage before. Zhou et al. (2019) proposed a latent representation learning framework which learned a common latent representation from multi-modality data and modality-specific latent representations from each modality. The latent representations were projected jointly to the label space for classification. Zhou et al. (2020) further proposed a latent representation learning method combined with ensemble support vector machine (SVM). Zhang and Shi (2020) proposed a deep multi-modal fusion network which utilized an attention mechanism to extract features from neuroimages and acquire relevant information. Based on the importance of the data, the fusion rate of each modality was assigned automatically and a hierarchical fusion method was adopted to fuse the multiple modalities. Shen et al. (2021) proposed a heterogeneous data fusion method to predict MCI conversion, which included data of healthy controls and patients with AD as auxiliary data to enhance the learning capacity of the classifier.

While most existing research, including all aforementioned works, has focused on integrating FDG-PET with MRI, more recently, amyloid-PET has been introduced to study MCI conversion. Pathologically, AD is characterized by amyloid plaques and neurofibrillary tangles (Holtzman et al., 2011). Amyloid-PET imaging measures the accumulation of amyloid plaques in the brain, which holds great promise for predicting MCI conversion to AD, especially when combined with structural MRI data (Rosenberg et al., 2013; Schwarz et al., 2016). Some recent works developed multi-modality models to classify MCI converters vs. non-converters based on MRI and amyloid-PET. Xu et al. (2016) proposed a weighted multi-modality sparse representation method, in which the classification was done by minimizing the weighted sum of mean-squared-errors of the predictions by multiple modalities. Zhu et al. (2019) proposed a self-paced multi-kernel learning method, in which a multi-kernel linear regression with low rank constraints on the regression coefficients was used to fuse heterogeneous modalities for classification. Liu et al. (2021) proposed an incomplete-multimodality transfer learning (IMTL) model, which built predictive models for different combinations of modalities and coupled the model estimation processes of different

combinations to allow for transfer learning. An Expectation-Maximization (EM) algorithm was utilized to estimate logistic regression parameters of IMTL and extended to a collaborative learning paradigm for patient privacy preservation.

Most of the existing studies, including all that have been reviewed previously, formulated the prediction of MCI progression/conversion to AD into a binary classification problem. Using a pre-defined timeframe, these studies aimed to classify each MCI patient into a converter if the patient progresses/converts to AD dementia within the timeframe, and a non-converter otherwise. The limitation of these existing studies is a lack of granularity in differentiating MCI patients who convert at different paces. The capability of predicting the pace of conversion for each MCI patient has important clinical value. For example, more personalized interventional strategies can be given to patients depending on if an individual will convert to AD very fast, fast, moderate, or slow. Patients can better prepare themselves and their family members, planning in a timely manner to address future care needs, by knowing their pace of conversion. Additionally, this may help clinical trials select the appropriate patient cohort.

In this paper, we formulate the prediction of the pace of conversion into an ordinal learning problem, in which the response variable has C classes with a natural order, i.e., class 1, 2, ..., C represent different fast-to-slow paces of conversion. Ordinal learning is a subfield in machine learning (ML). Various ordinal classification models have been developed such as ordinal SVM (Chu and Keerthi, 2007), ordinal Gaussian Process (Chu et al., 2005), and ordinal logit regression (Harrell, 2015). Most of the existing algorithms can only use precisely-labeled samples in training with each sample belonging to one and only one ordinal class. However, in some applications, it is difficult to obtain precisely-labeled training samples due to cost, availability, and other practical constraints, whereas it is common to have samples with label ambiguity, i.e., we know that the sample is from a range of several ordinal classes but do not know which precise class the sample is from. Next, we give an example to further illustrate this situation.

Consider an example of predicting four different paces of MCI conversion to AD: conversion within the first year after the MCI diagnosis (class 1), conversion between the first and second year (class 2), conversion between the third and the fifth year (class 3), and conversion beyond the fifth year (class 4). One reason causing label ambiguity in the training data is that the assessment times of different patients may not perfectly align. For example, some MCI patients might not be assessed within the first year but in the second year when they were found to have converted to AD. In this case, we can only know that these patients belong to class 1 or 2, but not the precise class of these patients. As another example, some patients might be known to have not converted to AD within the first year but then they dropped out of the study. In this case, we can only know that these patients do not belong to class 1, i.e., they belong to class 2, 3, or 4, but we do not know which of the last three classes they precisely belong to. If conventional ordinal learning

models were used, they would have to exclude these ambiguously-labeled patients from the training set, which would greatly reduce the training sample size. Motivated by this gap in the existing research, we propose a new model called ADPacer to integrate both ambiguously- and precisely-labeled samples for training a robust ordinal classifier to predict the pace of conversion to AD for each patient with MCI based on multi-modality neuroimaging (MRI and PET) and non-imaging data.

The contributions of this paper are summarized as follows:

- Our study focused on predicting the pace of MCI progression/conversion to AD by integrating multi-modality neuroimaging and non-imaging datasets. This complements the existing studies in literature that mainly focus on binary classification of converters and non-converters. Prediction of progression pace has multifold benefits as it would allow for individually tailored intervention, better preparation of patients and caregivers, and more nuanced patient selection strategies in clinical trials.
- We proposed a novel ADPacer model to leverage training samples with label ambiguity to augment the training set with precisely-labeled samples. This capability differentiates ADPacer from existing ordinal learning algorithms.
- We applied ADPacer to MCI patient cohorts from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL), and demonstrated the superior performance of ADPacer compared to existing ordinal learning algorithms. We also integrated the SHapley Additive exPlanations (SHAP) method with ADPacer to assess the contributions from different modalities to the model prediction. The findings are consistent with the AD literature.

2. Method

2.1. Data description

This study used ADNI data as the primary source for developing and demonstrating ADPacer. As a supplementary, we also included some data from AIBL. Due to the small sample size of the AIBL data, its role in this study is secondary, limited to only providing some preliminary validation of the model.

2.1.1. Primary data source from ADNI

Introduction to ADNI. ADNI (<http://adni.loni.ucla.edu>) was launched in 2003 by the NIH, FDA, private pharmaceutical companies, and nonprofit organizations, as a \$60 000 000, 5-year public-private partnership. The primary goal of ADNI has been to test whether MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and

clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California-San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. For up-to-date information, please see <http://www.adni-info.org/>.

Patient cohort. Pathologically, AD is characterized by amyloid plaques and neurofibrillary tangles (Holtzman et al., 2011). Using amyloid-PET imaging, MCI patients can be divided into two subgroups according to published guidelines (Fleisher et al., 2011): amyloid-positive and amyloid-negative. MCI patients in the amyloid-positive subgroup are at an elevated risk of progressing to AD. Predicting the different paces of progression/conversion for patients in this subgroup has important clinical value, which is the goal of this study.

We present different fast-to-slow paces of conversion by four ordinal classes. Class 1-4 represents conversion to AD within one year, between one and two years, between three and five years, and beyond five years from the time of a clinical visit, t , when the patient is assessed as having amyloid-positive MCI. t could correspond to either the initial visit when the patient is first diagnosed with MCI, or any follow-up visit when the patient's condition remains as MCI. t is also the time when ADPacer can be deployed. In other words, ADPacer serves as a predictive tool that can be deployed at any clinical visit when the patient is assessed as having amyloid-positive MCI, and informs both the patient and clinician about the future course of the disease. As a final note about the definition of the classes, class 4 is relatively more heterogeneous than the other classes because it includes patients who eventually convert to AD but beyond five years, as well as those whose MCI is not due to AD but some other underlying conditions.

The dataset used in this study included 282 samples from 209 participants in ADNI. Each sample corresponds to a clinical visit. A participant could have one or multiple samples if the corresponding clinical visit(s) meet the following inclusion criteria: 1) the patient is assessed as having amyloid-positive MCI at the visit, which can be either the initial visit when the MCI diagnosis is first made or a follow-up visit when the MCI status remains unchanged; 2) T1 MRI and florbetapir-PET (a type of amyloid-PET) were collected during the visit. According to the forementioned definitions of the four ordinal classes, we found that among the 282 samples, 167 samples have precise class labels (46, 41, 45, and 35 for class 1-4, respectively), whereas 115 samples have label ambiguity (46 samples in class 2, 3, or 4; 69 samples in class 3 or 4).

To prevent the risk of overfilling by including samples from the same patient in both the training and validation sets, we cautiously designed the cross validation (CV) scheme in model evaluation by splitting CV folds among patients not samples. In this way, samples from the same

patient will be included in either the training or validation set, but not in both, thus avoiding overfitting. Similar strategies have been adopted by other papers (Zhou et al., 2019). This CV scheme was followed in our case study in Section 3.

2.1.2. Secondary data source from AIBL

AIBL was launched in 2006 in Australia, with purposes of discovering biomarkers, cognitive characteristics, and health and lifestyle factors that determine subsequent development of symptomatic AD. Data was collected by the AIBL study group. AIBL study methodology has been reported previously (AIBL Research Group, 2009). For more information, please see <https://aibl.csiro.au/>. Our dataset includes 33 amyloid-positive MCI samples. Among them, 0, 9, 8, and 2 samples belong to class 1-4, respectively; 9 samples belong to class 2, 3, or 4; 5 samples belong to class 3 or 4. It is important to note that the acquisition of amyloid-PET for these samples involved the use of the Pittsburgh Compound B (PIB), a different tracer than ADNI. Due to this discrepancy, the inclusion of AIBL data in our study only intends to serve the purpose of some preliminary validation for the ADPacer model trained using ADNI data. Further validation is needed based on larger and more compatible datasets.

2.2. Image processing and feature computation

T1 MRI was processed by FreeSurfer v7.1 (Fischl, 2012) to obtain volumetric and cortical thickness measures following standard, published procedures (Schwarz et al., 2016). Amyloid-PET was processed by a PET Unified Pipeline to obtain regional standardized uptake value ratios (SUVR) measurements for FreeSurfer defined regions (Su et al., 2013, 2015).

156 features extracted from MRI were included: volumetric measures for 68 cortical regions of interest (ROIs), 14 sub-cortical structures, and 6 ventricle structures; cortical thickness measures for 68 ROIs. 151 features extracted from amyloid-PET were included: SUVRs of 68 cortical ROIs, 14 sub-cortical structures, and 68 white matter structures; a mean cortical SUVR feature. Additionally, in the ADNI dataset, we included non-imaging features including basic demographics such as gender, age and education level; scores from commonly used cognitive and clinical assessments such as the Mini-Mental State Examination (MMSE) total score and the Clinical Dementia Rating Scale (CDR) global score and sum-of-boxes score; status of the $\epsilon 4$ allele of apolipoprotein E (APOE) gene, which is a major genetic risk factor of AD. In the AIBL dataset, some of the aforementioned non-imaging features are missing, such as education and CDR sum-of-boxes score.

2.3. Proposed model: ADPacer

Consider a training dataset that consists of n MCI patients. Let x_i denote the feature vector of the i -th patient, which includes image features from MRI and PET as well as non-

imaging variables, $i = 1, \dots, n$. Let C denote the total number of ordinal classes corresponding to different paces of progression to AD. For conventional ordinal classification models, each sample in the training set should belong to one and only one ordinal class, i.e., each patient's pace of progression should be precisely-labeled. For the proposed ADPacer model, each sample in the training set can belong to a range of classes, $[Y_i^l, Y_i^r]$. When $Y_i^l < Y_i^r$, the range includes more than one class, i.e., the patient has label ambiguity. Note that the notation $[Y_i^l, Y_i^r]$ includes the precise label as a special case, i.e., when $Y_i^l = Y_i^r$, the sample is precisely labeled. The goal of ADPacer is to leverage both precisely- and ambiguously-labeled samples in training in order to build a robust ordinal classifier to differentiate the different paces of progression for MCI patients.

A typical ordinal classifier includes a set of ranking functions, $f_k, k = 1, \dots, C - 1$, which satisfy the constraint of $f_1 \leq \dots \leq f_{C-1}$ (Chu and Keerthi, 2007; Chu et al., 2005; Harrell, 2015). To predict the label for a sample x , one can compute the outputs from the ranking functions for this sample, $f_1(x), \dots, f_{C-1}(x)$. Then, the sample is assigned to class k if $f_k(x)$ is the first one being non-negative in the series of the ranking functions. For example, if all ranking functions are non-negative, the sample is assigned to class 1. If the $f_1(x) < 0$ and all $f_2(x), \dots, f_{C-1}(x) \geq 0$, then the sample is assigned to class 2. If all ranking functions are negative, the sample is assigned to class C . To train such ranking functions that compose an ordinal classifier, conventional algorithms can only use precisely-labeled samples (Chu and Keerthi, 2007; Chu et al., 2005; Harrell, 2015). ADPacer aims to include both precisely- and ambiguously-labeled samples in training. To achieve this, we first propose an alternative view of the ranking functions as a set of binary classifiers, i.e.,

$$\begin{aligned}
 f_1 : x_i &\rightarrow \text{class 1 vs. class } [2, \dots, C] \\
 &\vdots \\
 f_k : x_i &\rightarrow \text{class } [1, \dots, k] \text{ vs. class } [k+1, \dots, C] \\
 &\vdots \\
 f_{C-1} : x_i &\rightarrow \text{class } [1, \dots, C-1] \text{ vs. class } C \\
 f_1 &\leq \dots \leq f_{C-1}
 \end{aligned} \tag{1}$$

Each f_k is responsible for classifying a sample into the “left” or “right” side of k , where the left side includes classes ordered before k (including k) whereas the right side includes classes ordered after k . It can be proved that using this set of binary classifiers will guarantee to classify a test sample into one and only one classes, i.e., it can achieve the same goal of any conventional ordinal classifier at the test stage when the trained model is deployed, whereas the advantage of the proposed binary view is that it allows the training stage to accommodate samples with label ambiguity.

To train the proposed set of binary classifiers, we first identify a subset of samples in the training set, D_k , which can be used to train each binary classifier f_k . D_k should include samples whose range of classes, $[Y_i^l, Y_i^r]$, is either on

the left side of k , which means $Y_i^r \leq k$, or on the right side of k , which means $k < Y_i^l$. That is, D_k can include ambiguously-labeled samples as long as the range of classes for those samples are completely on the left or right side of k , i.e., D_k excludes samples whose range of classes include k . Mathematically, we can denote D_k as

$$\begin{aligned} D_k &= D_k^l \cup D_k^r \\ &= \left\{ (x_i, [Y_i^l, Y_i^r]) : Y_i^r \leq k, i = 1, \dots, n \right\} \\ &\quad \cup \left\{ (x_i, [Y_i^l, Y_i^r]) : k < Y_i^l, i = 1, \dots, n \right\}. \end{aligned} \quad (2)$$

Furthermore, to keep the intrinsic order of $f_1 \leq \dots \leq f_{C-1}$, we introduce intercepts/thresholds $b_1, \dots, b_{C-1}$ by enforcing $b_1 \leq \dots \leq b_{C-1}$, and let $f_k = h + b_k$, $k = 1, \dots, C-1$, where h is a shared function. We adopt the support vector formulation for h due to the success of SVM in various applications, and let $h(x) = w^T \phi(x)$, which ϕ contains non-linear transformations of the features. Note that we do not need to define the explicit form of the non-linear transformations and can use the kernel tricks similar to SVM (Chu and Keerthi, 2007). Finally, we propose the ADPacer model formulation as the following optimization problem:

$$\min_{w, \{b_k\}_{k=1}^{C-1}} \frac{1}{2} w^T w \quad (3)$$

subject to:

$$\left. \begin{aligned} w^T \phi(x_i) + b_k &\geq 1 - \xi_i^k, \quad i \in D_k^l \\ w^T \phi(x_j) + b_k &\leq -1 + \xi_j^k, \quad j \in D_k^r \\ \sum_{i \in D_k^l} \xi_i^k + \sum_{j \in D_k^r} \xi_j^k &\leq \epsilon^k \\ \xi_i^k &\geq 0, \quad i \in D_k^l; \quad \xi_j^k \geq 0, \quad j \in D_k^r \end{aligned} \right\} \text{for } k = 1, \dots, C-1 \quad (4)$$

$$b_1 \leq \dots \leq b_{C-1} \quad (5)$$

Among the $C-1$ sets of constraints in (4), each set corresponds to a binary classifier f_k with the purpose of achieving max-margin separation between *class* $[1, \dots, k]$ vs. *class* $[k+1, \dots, C]$. ξ_i^k and ξ_j^k are slack variables. ϵ^k is a tuning parameter.

To efficiently solve the ADPacer optimization, we derive the dual form of the primal problem in (3)–(5), which is summarized in Proposition 1. The proof is given in the Appendix.

Proposition 1: Let $|D_k^l|$ and $|D_k^r|$ denote the sample sizes of the training subsets, D_k^l and D_k^r , respectively, $k = 1, \dots, C-1$. Let Y denote a diagonal matrix, $Y = \text{diag}(\overbrace{1, \dots, 1}^{|D_1^l|}, \overbrace{-1, \dots, -1}^{|D_1^r|}, \overbrace{1, \dots, 1}^{|D_{C-1}^l|}, \overbrace{-1, \dots, -1}^{|D_{C-1}^r|})$. Furthermore, for any sample $i \in D_k = D_k^l \cup D_k^r$, introduce an indicator variable $y_i^k = 1$ or -1 to indicate if $x_i \in D_k^l$ or $x_i \in D_k^r$, respectively. Also, let K denote a covariance matrix with $K_{ij} = \phi(x_i)^T \phi(x_j) = k(x_i, x_j)$ that can be computed by a kernel

function defined on the feature space. Then, the dual form of the primal ADPacer optimization in (3)–(5) is:

$$\min_{\gamma} \frac{1}{2} \gamma^T Y K Y \gamma - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k,$$

subject to:

$$\sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k = 0, \quad (6)$$

$$\sum_{k=1}^L \sum_{i \in D_k} \alpha_i^k y_i^k \geq 0, \quad L = 1, \dots, C-2, \quad (7)$$

$$0 \leq \alpha_i^k \leq \lambda^k, \quad i \in D_k, \quad k = 1, \dots, C-1,$$

where $\gamma = \left(\left\{ \alpha_{i \in D_1}^1 \right\}, \dots, \left\{ \alpha_{i \in D_{C-1}}^{C-1} \right\} \right)^T$ are the Lagrange multipliers, and λ^k , $k = 1, \dots, C-1$, are tuning parameters.

The dual optimization is a convex quadratic programming problem, which can be solved by a standard solver such as CPLEX. After the optimal solution $\hat{\gamma}$ is obtained, we can obtain the optimal solution in the primal problem. Specifically, we can get:

$$\hat{w} = \sum_{k=1}^{C-1} \sum_{i \in D_k} \hat{\alpha}_i^k y_i^k \phi(x_i). \quad (8)$$

$\hat{b}_k$ can be estimated using any training sample that satisfies the KKT condition, i.e.,

$$\hat{b}_k = y_i^k - \hat{w}^T \phi(x_i), \quad (9)$$

with $x_i \in D^k$ and $0 < \alpha_i^k < \lambda^k$; $k = 1, \dots, C-1$. This concludes the training process of ADPacer.

To apply the trained ADPacer model to make a prediction for any new sample x , we can compute the ranking functions for the new sample as:

$$f_k(x) = \hat{w}^T \phi(x) + \hat{b}_k = \sum_{k=1}^{C-1} \sum_{i \in D_k} \hat{\alpha}_i^k y_i^k k(x_i, x) + \hat{b}_k. \quad (10)$$

Finally, based on the computed ranking functions, the new sample is assigned to class k if $f_k(x)$ is the first one being non-negative in the series of the ranking functions, i.e.,

$$\hat{k} = \arg \min_k \{k : f_k(x) \geq 0\}. \quad (11)$$

Please see Algorithm 1 for an overview of the entire training and application process of ADPacer.

Algorithm 1: ADPacer

Model Training Phase

Input: training set: $\mathcal{D} = \{(x_i, [Y_i^l, Y_i^r])\}_{i=1, \dots, n}$; setting of tuning parameters: λ^k , $k = 1, \dots, C-1$; form of the kernel function $k(\cdot)$

Output: estimated model parameters: $\hat{\gamma} = \left(\left\{ \hat{\alpha}_{i \in D_1}^1 \right\}, \dots, \left\{ \hat{\alpha}_{i \in D_{C-1}}^{C-1} \right\} \right)^T$, $\hat{w}$, $\hat{b}_k$, $k = 1, \dots, C-1$

Procedure:

1. Reformate the training set to fit the alternative view of ordinal classification as a set of binary classifiers in (1):

- a. For each binary classifier f_k , $k = 1, \dots, C - 1$, pull training samples into D_k using (2) and create the binary label as $y_i^k = 2I(x_i \in D_k^l) - 1$.
- b. Reformat the training set as $\mathcal{D}' = \{(x_i, y_i^k) : i \in D_k, k = 1, \dots, C\}$
2. Solve the dual ADPacer optimization in Proposition 1 using a convex quadratic solver and obtain the optimal solution $\hat{\gamma} = \left(\left\{ \hat{\alpha}_{i \in D_1}^1 \right\}, \dots, \left\{ \hat{\alpha}_{i \in D_{C-1}}^{C-1} \right\} \right)^T$.
3. Obtain the optimal solution for the primal ADPacer optimization in (3)–(5) using (8) and (9), i.e., $\hat{w} = \sum_{k=1}^{C-1} \sum_{i \in D_k} \hat{\alpha}_i^k y_i^k \phi(x_i)$; $\hat{b}_k = y_i^k - \hat{w}^T \phi(x_i)$ with $x_i \in D^k$ and $0 < \alpha_i^k < \lambda^k$; $k = 1, \dots, C - 1$.

Model Application Phase

Input: estimated model parameters from training: $\hat{\gamma}$, $\hat{w}$, $\hat{b}_k$, $k = 1, \dots, C - 1$; reformatted training set $\mathcal{D}'$; a new sample x

Output: Predicted ordinal class $\hat{k}$ for the new sample

Procedure:

1. Compute the ranking functions for the new sample x using (10): $f_k(x) = \hat{w}^T \phi(x) + \hat{b}_k = \sum_{k=1}^{C-1} \sum_{i \in D_k} \hat{\alpha}_i^k y_i^k k(x_i, x) + \hat{b}_k$, $k = 1, \dots, C - 1$.
2. Predict the ordinal class using (11): $\hat{k} = \arg \min_k \{k : f_k(x) \geq 0\}$

Implementation details. Previously we have discussed the general ADPacer model. Next, we will present the detailed implementation of ADPacer to the specific dataset described in Section 2.1 and 2.2. Our dataset has four ordinal classes corresponding to fast-to-slow paces of progression, which means that there are three tuning parameters in the dual optimization of ADPacer, $\lambda^1, \lambda^2, \lambda^3$. Our dataset is fairly balanced in each class. Thus, we set $\lambda^1 = \lambda^2 = \lambda^3 = \lambda$. We used CV based on patients not samples to avoid overfitting (see more discussion in Section 2.1), and selected the λ with the highest CV accuracy. A final detail to mention is that because of the high-dimensional features used to train ADPacer, we embedded a feature weighting scheme within the CV. Specifically, the conventional kernel is computed by weighting all features equally. For example, in the commonly used radial basis function (RBF) kernel, the kernel between two samples x_i and x_j is $\exp(-\theta \sum_{l=1}^p (x_{i,l} - x_{j,l})^2)$. To estimate feature-specific weights θ_l , $l = 1, \dots, p$, we adopted a simple idea of pre-training a linear ADPacer (i.e., ADPacer with a linear kernel), from which we could obtain the feature weights as $\theta_l = \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k x_{i,l}$. Then, we used these weights in the RBF kernel to re-train a nonlinear ADPacer. The nonlinear ADPacer was used as the final ordinal classifier to leverage the flexibility provided by a nonlinear model. The pre-trained linear ADPacer served the purpose of estimating the feature-specific weights to inform

the nonlinear kernel. This simple strategy turned out to work quite well in our case study.

2.4. Integration of ADPacer and SHAP for model interpretation

To help interpret the ADPacer model, we integrated a popular, model-agnostic method called SHapley Additive exPlanations (SHAP) (34) with ADPacer, with purpose of quantifying the contributions of different data modalities to the pace prediction. Specifically, we used SHAP to estimate the contribution score of each of the three data modalities (e.g., MRI, amyloid-PET, and non-imaging data), namely the SHAP value, by computing the difference in the ADPacer's prediction when the modality is included vs. excluded. The higher the absolute SHAP value of a data modality, the greater contribution of the modality.

3. Results

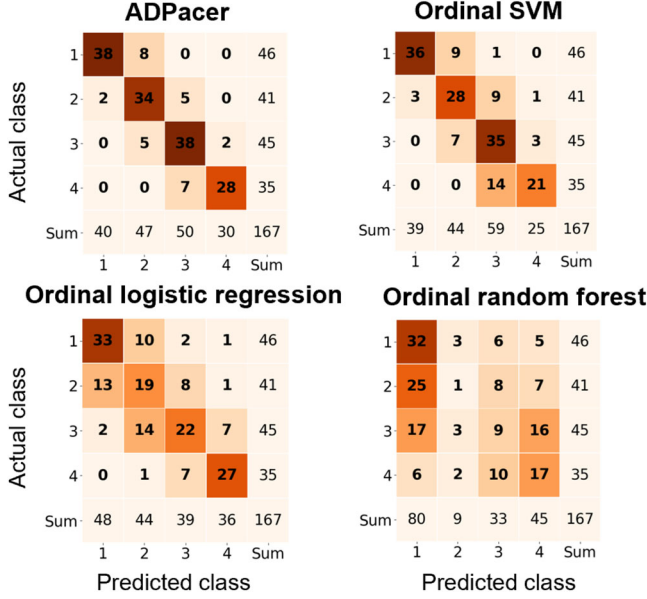
We applied ADPacer to the ADNI dataset described in Sections 2.1 and 2.2. Specifically, we used a 30-fold CV scheme in which the precisely-labeled samples were divided into 30 folds. One fold was left out as the validation set, whereas the other folds were combined with the ambiguously-labeled samples to form the training set. The rationale behind employing a 30-fold CV is similar to the use of leave-one-out CV in studies with limited samples size, both aiming to allocate more of the available data in training to build a more reliable model. It is important to note that in this CV scheme, ambiguously-labeled samples were only included to help model training. When the trained model is in use, it should and will classify each sample into a precise label. This is why only precisely-labeled samples were included in the left-out validation fold, such that the CV accuracy reflects the performance of the model for classifying each sample into the precise label it belongs to.

A special caution in the training-validation split was that we avoided putting samples from the same patient in both training and validation, which may cause overfitting, as mentioned in Section 2.1. ADPacer was trained using the training set and the model was then applied to classify samples in the validation set. This process was iterated over the 30 folds. We reported the overall accuracy and class-specific accuracy based on the CV in Table 1. We also composed a confusion matrix to show the distribution of correctly and wrongly classified samples in Figure 1.

In comparison, we applied three existing ordinal learning algorithms to the same dataset, including ordinal SVM (Chu and Keerthi, 2007), ordinal logistic regression (Harrell, 2015), and ordinal random forest (Hornung, 2020). The same 30-fold CV scheme as ADPacer was used, except that these existing methods could not incorporate ambiguously-labeled samples in training. We reported the overall accuracy, class-specific accuracy, and confusion matrix of each method in Table 1 and Figure 1 to compare with ADPacer. It is clear from the results that ADPacer achieved the highest overall accuracy and class-specific accuracy.

Table 1. Accuracies of ADPacer and existing ordinal learning methods based on CV.

	Overall accuracy	Class 1 accuracy	Class 2 accuracy	Class 3 accuracy	Class 4 accuracy
ADPacer	0.83	0.83	0.83	0.84	0.79
Ordinal SVM	0.71	0.78	0.68	0.78	0.60
Ordinal logistic regression	0.60	0.72	0.46	0.49	0.76
Ordinal random forest	0.35	0.69	0.024	0.20	0.49

**Figure 1.** Confusion matrices of ADPacer and existing ordinal learning methods based on CV.

Additionally, we applied several multi-class classification algorithms including multi-class SVM, logistic regression, and random forest to the same dataset. These algorithms had low overall accuracy (range: 0.32-0.39), which was not only significantly worse than ADPacer but also worse than most existing ordinal learning algorithms. This result empirically justified the appropriateness of considering the four different paces of conversion to AD as ordinal classes.

Furthermore, as mentioned in Introduction, most existing research in predicting MCI conversion to AD focused on a binary classification of MCI patients into converters vs. non-converters based on a pre-defined timeframe. To facilitate a meaningful comparison between ADPacer and the reported binary classification performance in the literature, we evaluated the accuracies of ADPacer in three binary classification tasks: 1) predicting MCI conversion to AD within or beyond one year (class 1 vs. 2-4); 2) predicting MCI conversion within or beyond two years (class 1-2 vs. 3-4); 3) predicting MCI conversion within or beyond five years (class 1-3 vs. 4). It is worth noting that this evaluation did not require model re-training. ADPacer remained trained as an ordinal learning model, but we computed the accuracy of the model for each binary task. For instance, in task 1), if a sample from class 2, 3, or 4 is predicted to be in any of these classes, it is considered correctly classified, even if the true and predicted ordinal classes of the sample do not exactly match. In a similar way, we evaluated the accuracies of the existing ordinal learning methods in the three binary classification tasks. In addition, for binary classification tasks, it is natural to train models specifically designed for

Table 2. Accuracies of ADPacer, existing ordinal learning methods, and binary classifiers in binary classification tasks based on CV.

	Accuracy of classifying class 1 vs 2-4	Accuracy of classifying 1-2 vs 3-4	Accuracy of classifying 1-3 vs 4
ADPacer	0.94	0.94	0.95
Ordinal SVM	0.92	0.89	0.89
Ordinal logistic regression	0.83	0.82	0.89
Ordinal random forest	0.63	0.67	0.72
Binary SVM	0.67	0.78	0.30
Binary logistic regression	0.70	0.75	0.69
Binary random forest	0.64	0.66	0.66

binary classification. To this end, we trained three widely-used binary classification models including binary SVM, logistic regression, and random forest. Unlike existing ordinal learning models that cannot incorporate ambiguously-labeled samples in training, these binary classifiers can accommodate such samples. For instance, if a patient has not converted to AD within the first year and subsequently dropped out of the study, the data of this patient cannot be included in training the existing ordinal learning models but can be included in training binary classifiers in task 1). Therefore, in training binary classifiers, we included ambiguously-labeled samples whenever possible for each binary task.

Table 2 summarizes the results of ADPacer and existing ordinal learning models used for the binary classification tasks, as well as results of the binary classifiers. ADPacer achieved superior accuracies compared to all the other methods. Also, comparing Tables 1 and 2, we can observe that when a trained ordinal learning model is used to make a binary classification, its accuracy is higher than that when used to make an ordinal classification. This is expected because the former does not require an exact match between the predicted and true ordinal classes of samples. Additionally, it is interesting to observe that despite the capability of including ambiguously-labeled samples in training, binary classifiers performed worse than most ordinal learning algorithms. This may be due to the need to merge samples from multiple ordinal classes into a single class when training a binary classifier. The inherent heterogeneity present within these combined classes can make it challenging to identify an optimal binary classification boundary. This result reinforced the appropriateness of our focus on pace prediction using ordinal learning in this paper. ADPacer further expanded the capability of ordinal learning by incorporating ambiguously-labeled samples in training.

Also, we showed the contribution of each data modality using the SHAP method in Figure 2.

Finally, we applied the ADPacer model trained using ADNI data to AIBL samples. For this purpose, the model was re-trained using all the ADNI data under the optimal tuning parameters found under CV. Then, the trained

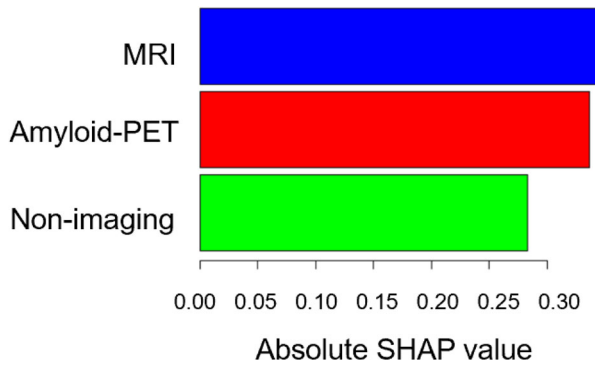


Figure 2. Modality contributions to ADPacer prediction.

model was applied to predict the ordinal class of each sample in AIBL. During this training-application process, two strategies were implemented to mitigate the discrepancies between the two datasets. First, to mitigate the impact of different tracers used to acquire amyloid-PET in the ADNI and AIBL datasets, the features extracted from amyloid-PET were standardized within each dataset separately. Second, to address the discrepancy that some non-imaging features in ADNI are unavailable in AIBL, ADPacer was trained using only the overlapping non-imaging features along with imaging features.

To evaluate the performance of ADPacer on AIBL data, it is unreliable to evaluate its ordinal classification accuracy because the precisely-labeled samples in each ordinal class are too small. Thus, we adopted an alternative approach by evaluating the accuracy of ADPacer used for binary classification tasks. Given the sample size constraint, our focus was narrowed down to two binary classification tasks: (a) class 1-2 vs. 3-4 (24 samples); (b) class 1-3 vs. 4 (19 samples). On the AIBL data, ADPacer achieved accuracies of 0.75 and 0.84 accuracies for task (a) and (b), respectively. In comparison, the highest accuracies achieved by the other ordinal learning models were 0.68 and 0.74, which were lower than ADPacer.

4. Discussion

We proposed an ADPacer model to predict the fast-to-slow paces of conversion to AD for MCI patients by integrating both ambiguously- and precisely-labeled samples in training. The capability of leveraging ambiguously-labeled samples differentiated ADPacer from existing ordinal learning models, and improved the training capability especially when precisely-labeled samples are limited. We applied ADPacer to predict the fast-to-slow paces of conversion to AD for MCI patients based on multi-modality neuroimaging and non-imaging data. While there is existing work using machine learning models to classify MCI patients into converters or non-converters based on a pre-defined time frame, the present study focused on conversion pace prediction, which greatly complemented the existing research. On the ADNI data, ADPacer achieved significantly higher accuracy (0.83 overall accuracy) than other ordinal learning models (the second best, ordinal SVM, has only 0.71 overall accuracy). The class-specific accuracy of ADPacer was also

fairly balanced, with 0.83-0.84 for class 1-3, and 0.79 for class 4. The reason for a slightly lower class-4 accuracy may be that this class, defined as MCI patients who have not converted to AD in five years, has substantial heterogeneity. That is, class 4 includes MCI patients who convert to AD in any number of years beyond five, as well as patients whose MCI symptoms may be due to other underlying conditions that are not AD-related so that they will not convert to AD dementia. Compared to ADPacer, the other ordinal learning models had more imbalanced class-specific accuracy and low overall accuracy. The reason is that these existing models can only include precisely-labeled samples in training. Since the precisely-labeled samples in each class are quite limited in this dataset whereas the feature dimension is high, these existing models may have suffered more from curse of dimensionality. In contrast, ADPacer was capable of additionally including ambiguous-labeled samples, so that the total training samples size nearly doubled that by the other models. This alleviation of curse of dimensionality led to more robust model training of ADPacer and higher accuracy.

In the existing research that predicts MCI conversion as a binary classification problem, the reported accuracy ranges from 0.743-0.898 based on integrating MRI and PET with some studies additionally including non-imaging datasets (Liu et al., 2017, 2021; Schwarz et al., 2016; Shen et al., 2021; Zhu et al., 2019). This range can hardly be taken as a point of reference to our work because the existing studies focused on binary but not pace/ordinal class prediction. Also, the various studies differ in the dataset and imaging/non-imaging features used. Nevertheless, if we just take the face values of the reported accuracies in the various existing studies and compare with that in the present study, our accuracy is comparable to the existing studies especially considering that pace/ordinal classification is a more challenging task than binary classification. Furthermore, we demonstrated that ADPacer achieved superior accuracy (0.94-0.95) if used as a binary classifier.

Because we embedded a feature weighting scheme in training ADPacer (under “implementation details” in Section 2.3), we can naturally know the relative importance of features. We found that, among the top ranked features, regions such as hippocampus, superior_parietal, inferior_temporal, amygdala, and thalamus have been widely reported in the literature (Cai et al., 2017; Desikan et al., 2010; Jacobs et al., 2012; Pini et al., 2016). Furthermore, by using the method in Section 2.4, we can know the modality-level contribution to the model prediction. Specifically, according to absolute SHAP value of each modality in Figure 2, both MRI and amyloid-PET contributed more than non-imaging data. Neuroimaging data contains rich information about brain alteration associated with AD, which has been demonstrated in various studies (van Oostveen and de Lange, 2021). Between MRI and amyloid-PET, MRI contributed slightly more. Overall, all the included neuroimaging and non-imaging modalities more or less contributed to the pace prediction of MCI conversion to AD.

Recognizing the value of including a separate dataset to evaluate ML models, we applied the ADPacer model trained using ADNI data to the data from another public dataset, AIBL. Our results showed that ADPacer performed better than existing methods. On the other hand, we acknowledge the limitations associated with the AIBL dataset, which make it less than ideal as a validation set for ADPacer. These limitations include the limited sample size, the utilization of a different tracer for amyloid-PET compared to ADNI, the unavailability of certain non-imaging features, and the sub-optimal nature of our evaluation approach due to the constraint of the small sample size. Therefore, the inclusion of AIBL data in our study was only intended to serve the purpose of some preliminary validation for the ADPacer model. Further validation is needed based on larger and more compatible datasets.

This study has several limitations. First, our dataset has a relatively small sample size, which is inherent to the nature of the study that focused on amyloid-positive MCI patients. This focus has significant clinical value, as this patient subgroup has an elevated risk of progressing to AD and the ability of accurately predicting their paces of progression would greatly help the patient and clinician for decision making. However, this focus limits the sample size. Future research is needed to validate the findings in this paper using larger datasets. Second, this study includes volumetric MRI, amyloid-PET, and some demographic and clinical variables for pace prediction. There are other structural and functional neuroimages as well as non-imaging datasets such as CSF and genomics, which could provide additional value and help improve accuracy. Third, this study included four ordinal classes with relatively balanced sample sizes which may not accurately represent the population-level distribution of patients. Currently, there is a lack of population-level statistics in this regard, due to relative recency of the criteria for defining amyloid-positivity, limited access to amyloid-PET exams for patients, and clinical expertise required for image interpretation, which lead to limited sample sizes for conducting population studies (Klunk et al., 2015; Su et al., 2019; Zhang et al., 2014). In this study, we chose to include all available samples in the ADNI database according to the inclusion criteria stated in Section 2.1. In the future, refinement of class definition and sample inclusion can be done to align with population-level statistics and enhance result representativeness.

Last but not least, the ADPacer algorithms have some limitations that could be improved in future research. Specifically, ADPacer is kernel-based model. While we used the SHAP method for interpreting feature contributions, this occurs post-analysis. Regression and decision tree are well-known white-box methods, possessing inherent interpretability. Thus, it would be interesting to extend ADPacer by using regression or decision tree as the base model and modifying it to have the capability of incorporating both precisely- and ambiguously-labeled samples in training. Furthermore, while the case study has a relatively balanced sample size across different classes, this may not hold true in other applications of ADPacer. Thus, future research may

investigate how to optimize the performance of ADPacer in the presence of class imbalance, using strategies such as data resampling, class weighting and cost-sensitive training.

5. Conclusion

We developed an ADPacer model to integrate both ambiguously- and precisely-labeled samples for training a robust ordinal classifier to predict the fast-to-slow paces of MCI conversion to AD based on multi-modality neuroimaging and non-imaging datasets. ADPacer showed significantly better performance than existing ordinal learning models. Prediction of progression pace has multifold benefits in aspects of facilitating individually-tailored intervention, better preparation of patients and caregivers, and more nuanced patient selection strategies in clinical trials.

Acknowledgments

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 and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and 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. Data used in the preparation of this article was also obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI).

Disclosure statement

The authors report no conflict of interest.

Consent and approval statement

This study has been exempted from the requirement for approval by an institutional review board. The data corpus is publicly available.

Funding

Research reported in this publication was supported by NIH grant 2R42AG053149-02A1 and NSF grant DMS-2053170. This research was also supported by NIH grants R01AG069453 and P30AG072980, the State of Arizona, and Banner Alzheimer's Foundation.

References

- Alzheimer Association. (2022). 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 18(4), 700–789. <https://doi.org/10.1002/alz.12638>
- Cai, K., Xu, H., Guan, H., Zhu, W., Jiang, J., Cui, Y., Zhang, J., Liu, T., & Wen, W. (2017). Identification of early-stage Alzheimer's disease using Sulcal morphology and other common neuroimaging indices. *PLoS One*, 12(1), e0170875. <https://doi.org/10.1371/journal.pone.0170875>
- Chu, W., & Ghahramani, Z. (2005). Gaussian processes for ordinal regression. *Journal of Machine Learning Research*, 6(35), 1019–1041.
- Chu, W., & Keerthi, S. S. (2007). Support vector ordinal regression. *Neural Computation*, 19(3), 792–815. <https://doi.org/10.1162/neco.2007.19.3.792>
- Desikan, R. S., Cabral, H. J., Settecase, F., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., Schmansky, N. J., Salat, D. H., & Fischl, B.; Alzheimer's Disease Neuroimaging Initiative. (2010). Automated MRI measures predict progression to Alzheimer's disease. *Neurobiology of Aging*, 31(8), 1364–1374. <https://doi.org/10.1016/j.neurobiolaging.2010.04.023>
- Doyle, O. M., Westman, E., Marquand, A. F., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Williams, S. C. R., & Simmons, A. (2014). Predicting progression of Alzheimer's disease using ordinal regression. *PLoS One*, 9(8), e105542. <https://doi.org/10.1371/journal.pone.0105542>
- Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoek, C., Taddei, K., Villemagne, V., Woodward, M., & Ames, D.; AIBL Research Group. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*, 21(4), 672–687. <https://doi.org/10.1017/S1041610209009405>
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Fleisher, A. S., Chen, K., Liu, X., Roontiva, A., Thiyyagura, P., Ayutyanont, N., Joshi, A. D., Clark, C. M., Mintun, M. A., Pontecorvo, M. J., Doraiswamy, P. M., Johnson, K. A., Skovronsky, D. M., & Reiman, E. M. (2011). Using positron emission tomography and florbetapir F 18 to image cortical amyloid in patients with mild cognitive impairment or dementia due to Alzheimer disease. *Archives of Neurology*, 68(11), 1404–1411. <https://doi.org/10.1001/archneurol.2011.150>
- Harrell, F. E. (2015). Ordinal logistic regression, in *Regression modeling strategies* (pp. 311–325). Springer.
- Holtzman, D. M., Morris, J. C., & Goate, A. M. (2011). Alzheimer's disease: The challenge of the second century. *Science Translational Medicine*, 3(77), 77sr1. <https://doi.org/10.1126/scitranslmed.3002369>
- Hornung, R. (2020). Ordinal forests. *Journal of Classification*, 37(1), 4–17. <https://doi.org/10.1007/s00357-018-9302-x>
- Jacobs, H. I. L., van Bostel, M. P. J., Jolles, J., Verhey, F. R. J., & Uylings, H. B. M. (2012). Parietal cortex matters in Alzheimer's disease: An overview of structural, functional and metabolic findings. *Neuroscience and Biobehavioral Reviews*, 36(1), 297–309. <https://doi.org/10.1016/j.neubiorev.2011.06.009>
- Klunk, W. E., Koeppe, R. A., Price, J. C., Benzinger, T. L., Devous, M. D., Jagust, W. J., Johnson, K. A., Mathis, C. A., Minhas, D., Pontecorvo, M. J., Rowe, C. C., Skovronsky, D. M., & Mintun, M. A. (2015). The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 11(1), 1–15. <https://doi.org/10.1016/j.jalz.2014.07.003>
- Liu, M., Zhang, J., Yap, P.-T., & Shen, D. (2017). View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. *Medical Image Analysis*, 36, 123–134. <https://doi.org/10.1016/j.media.2016.11.002>
- Liu, X., Chen, K., Weidman, D., Wu, T., Lure, F., & Li, J.; Alzheimer's Disease Neuroimaging Initiative. (2021). A novel transfer learning model for predictive analytics using incomplete multimodality data. *IJSE Transactions*, 53(9), 1010–1022. <https://doi.org/10.1080/24725854.2020.1798569>
- Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., & Frisoni, G. B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, 30, 25–48. <https://doi.org/10.1016/j.arr.2016.01.002>
- Rosenberg, P. B., Wong, D. F., Edell, S. L., Ross, J. S., Joshi, A. D., Brašić, J. R., Zhou, Y., Raymont, V., Kumar, A., Ravert, H. T., Dannals, R. F., Pontecorvo, M. J., Skovronsky, D. M., & Lyketsos, C. G. (2013). Cognition and amyloid load in Alzheimer disease imaged with florbetapir F 18 (AV-45) positron emission tomography. *The American Journal of Geriatric Psychiatry*, 21(3), 272–278. <https://doi.org/10.1016/j.jagp.2012.11.016>
- Schwarz, C. G., Gunter, J. L., Wiste, H. J., Przybelski, S. A., Weigand, S. D., Ward, C. P., Senjem, M. L., Vemuri, P., Murray, M. E., Dickson, D. W., Parisi, J. E., Kantarci, K., Weiner, M. W., Petersen, R. C., & Jack, C. R.; Alzheimer's Disease Neuroimaging Initiative. (2016). A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage: Clinical*, 11, 802–812. <https://doi.org/10.1016/j.nicl.2016.05.017>
- Shen, H. T., Zhu, X., Zhang, Z., Wang, S.-H., Chen, Y., Xu, X., & Shao, J. (2021). Heterogeneous data fusion for predicting mild cognitive impairment conversion. *Information Fusion*, 66, 54–63. <https://doi.org/10.1016/j.inffus.2020.08.023>
- Su, Y., Blazey, T. M., Snyder, A. Z., Raichle, M. E., Marcus, D. S., Ances, B. M., Bateman, R. J., Cairns, N. J., Aldea, P., Cash, L., Christensen, J. J., Friedrichsen, K., Hornbeck, R. C., Farrar, A. M., Owen, C. J., Mayeux, R., Brickman, A. M., Klunk, W., Price, J. C., ... Benzinger, T. L. S.; Dominantly Inherited Alzheimer Network. (2015). Partial volume correction in quantitative amyloid imaging. *Neuroimage*, 107, 55–64. <https://doi.org/10.1016/j.neuroimage.2014.11.058>
- Su, Y., D'Angelo, G. M., Vlassenko, A. G., Zhou, G., Snyder, A. Z., Marcus, D. S., Blazey, T. M., Christensen, J. J., Vora, S., Morris, J. C., Mintun, M. A., & Benzinger, T. L. S. (2013). Quantitative analysis of PiB-PET with freesurfer ROIs. *PLoS One*, 8(11), e73377. <https://doi.org/10.1371/journal.pone.0073377>
- Su, Y., Flores, S., Wang, G., Hornbeck, R. C., Speidel, B., Joseph-Mathurin, N., Vlassenko, A. G., Gordon, B. A., Koeppe, R. A., Klunk, W. E., Jack, C. R., Farlow, M. R., Salloway, S., Snider, B. J., Berman, S. B., Roberson, E. D., Brosch, J., Jimenez-Velazquez, I., van Dyck, C. H., ... Benzinger, T. L. S. (2019). Comparison of Pittsburgh compound B and florbetapir in cross-sectional and longitudinal studies. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 11(1), 180–190. <https://doi.org/10.1016/j.dadm.2018.12.008>
- van Oostveen, W. M., & de Lange, E. C. M. (2021). Imaging techniques in Alzheimer's disease: A review of applications in early diagnosis and longitudinal monitoring. *International Journal of Molecular Sciences*, 22(4), 2110. <https://doi.org/10.3390/ijms22042110>
- Xu, L., Wu, X., Li, R., Chen, K., Long, Z., Zhang, J., Guo, X., & Yao, L.; Alzheimer's Disease Neuroimaging Initiative. (2016). Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. *Journal of Alzheimer's Disease: JAD*, 51(4), 1045–1056. <https://doi.org/10.3233/JAD-151010>
- Zhang, S., Smailagic, N., Hyde, C., Noel-Storr, A. H., Takwoingi, Y., McShane, R., & Feng, J.; Cochrane Dementia and Cognitive Improvement Group. (2014). 11 C-PIB-PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, 7, CD010386. <https://doi.org/10.1002/14651858.CD010386.pub2>

- Zhang, T., & Shi, M. (2020). Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease. *Journal of Neuroscience Methods*, 341, 108795. <https://doi.org/10.1016/j.jneumeth.2020.108795>
- Zhou, T., Liu, M., Thung, K. H., & Shen, D. (2019). Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE Transactions on Medical Imaging*, 38(10), 2411–2422. <https://doi.org/10.1109/TMI.2019.2913158>
- Zhou, T., Thung, K. H., Liu, M., Shi, F., Zhang, C., & Shen, D. (2020). Multi-modal latent space inducing ensemble SVM classifier for early dementia diagnosis with neuroimaging data. *Medical Image Analysis*, 60, 101630. <https://doi.org/10.1016/j.media.2019.101630>
- Zhou, T., Thung, K. H., Zhu, X., & Shen, D. (2019). Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping*, 40(3), 1001–1016. <https://doi.org/10.1002/hbm.24428>
- Zhu, Q., Yuan, N., Huang, J., Hao, X., & Zhang, D. (2019). Multi-modal AD classification via self-paced latent correlation analysis. *Neurocomputing*, 355, 143–154. <https://doi.org/10.1016/j.neucom.2019.04.066>

Appendix

Proof of Proposition 1: Let $\gamma = (\{\alpha_{i \in D_1}^1\}, \dots, \{\alpha_{i \in D_{C-1}}^{C-1}\})^T$, ν_i^k , μ_k for i and k , be the Lagrange multipliers, and λ^k , $k = 1, \dots, C-1$, are tuning parameters. The Lagrangian for the primal ADPacer optimization in (3)–(5) is

$$\begin{aligned} L = & \frac{1}{2} w^T w - \sum_{k=1}^{C-1} \sum_{i \in D_k^l} \alpha_i^k \left((w^T \phi(x_i) + b_k) - 1 + \zeta_i^k \right) \\ & - \sum_{k=1}^{C-1} \sum_{i \in D_k^r} \alpha_i^k \left(- (w^T \phi(x_i) + b_k) - 1 + \zeta_i^k \right) \\ & + \sum_{k=1}^{C-1} \lambda_k \left(\sum_{i \in D_k^l} \zeta_i^k + \sum_{j \in D_k^r} \zeta_j^k \right) - \sum_{k=1}^{C-1} \sum_{i \in D_k^l} \nu_i^k \zeta_i^k \\ & - \sum_{k=1}^{C-1} \sum_{j \in D_k^r} \nu_j^k \zeta_j^k - \sum_{k=1}^{C-2} \mu_k (b_{k+1} - b_k). \end{aligned}$$

By introducing a new notation τ satisfying $\tau_i^k = \zeta_i^k$ for $i \in D_k^l$ and $\tau_i^k = \zeta_i^k$ for $i \in D_k^r$, $k = 1, \dots, C-1$, L can be simplified as

$$\begin{aligned} L = & \frac{1}{2} w^T w - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k \left(y_i^k (w^T \phi(x_i) + b_k) - 1 + \tau_i^k \right) \\ & + \sum_{k=1}^{C-1} \lambda_k \sum_{i \in D_k} \tau_i^k - \sum_{k=1}^{C-1} \sum_{i \in D_k} \nu_i^k \tau_i^k - \sum_{k=1}^{C-2} \mu_k (b_{k+1} - b_k), \end{aligned} \quad (12)$$

where $y_i^k = 1$ or -1 indicates if $x_i \in D_k^l$ or $x_i \in D_k^r$, respectively.

Then the optimal solution of the primal problem in (3)–(5) is equivalent to the solution of the following optimization:

$$\max_{\gamma, \nu, \mu} \min_{w, b, \tau} L. \quad (13)$$

The KKT conditions for the primal problem require the following to hold:

$$\begin{aligned} \nabla_w L = w - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k \phi(x_i) &= 0, \\ \nabla_{b_k} L = - \sum_{i \in D_k} \alpha_i^k y_i^k - \mu_{k-1} + \mu_k &= 0, \quad k = 1, \dots, C-1, \\ \text{with } \mu_0 = 0, \quad \mu_{C-1} &= 0. \end{aligned} \quad (14)$$

$$\nabla_{\tau_i^k} L = -\alpha_i^k + \lambda_k - \nu_i^k = 0, \quad i \in D_k, \quad k = 1, \dots, C-1. \quad (15)$$

Then we have

$$w = \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k \phi(x_i), \quad (16)$$

$$\nu_i^k = -\alpha_i^k + \lambda_k, \quad i \in D_k, \quad k = 1, \dots, C-1. \quad (17)$$

Inserting (16) and (17) into the optimization in (13), after simplification we can get

$$\begin{aligned} \max_{\gamma, \mu} L = & \frac{1}{2} w^T w - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k \left(y_i^k (w^T \phi(x_i) + b_k) - 1 + \tau_i^k \right) \\ & + \sum_{k=1}^{C-1} \lambda_k \sum_{i \in D_k} \tau_i^k - \sum_{k=1}^{C-1} \sum_{i \in D_k} (-\alpha_i^k + \lambda_k) \tau_i^k \\ & - \sum_{k=1}^{C-1} \mu_k (b_k - b_{k-1}) \\ = & \frac{1}{2} w^T w - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k \left(y_i^k (w^T \phi(x_i) + b_k) - 1 \right) \\ & - \sum_{k=1}^{C-1} \mu_k (b_k - b_{k-1}) \\ = & \frac{1}{2} w^T w - \sum_{k=1}^{C-1} \sum_{i \in D_k} w^T \alpha_i^k y_i^k \phi(x_i) - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k b_k \\ & + \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k - \sum_{k=1}^{C-1} \mu_k (b_k - b_{k-1}). \end{aligned} \quad (18)$$

Furthermore, according to (14), we have

$$\begin{aligned} & - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k b_k - \sum_{k=1}^{C-1} \mu_k (b_k - b_{k-1}) \\ = & - \sum_{k=1}^{C-1} b_k (\mu_{k+1} - \mu_k) - \sum_{k=1}^{C-1} \mu_k (b_k - b_{k-1}) = 0 \end{aligned}$$

Then inserting (16) into the optimization in (18), we can have

$$\begin{aligned} \max_{\gamma} L = & - \frac{1}{2} \left(\sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k \phi(x_i) \right)^T \left(\sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k \phi(x_i) \right) \\ & + \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k = - \frac{1}{2} \gamma^T YKY\gamma + \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k \end{aligned}$$

Additionally, the conditions in (14) give rise to the constraints of

$$\mu_k = \sum_{k=1}^L \sum_{i \in D_k} \alpha_i^k y_i^k \geq 0, \quad L = 1, \dots, C-2,$$

$$\mu_{C-1} = \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k = 0.$$

The conditions in (17) give rise to the constraints of

$$0 \leq \alpha_i^k \leq \lambda_k, \quad i \in D_k, \quad k = 1, \dots, C-1$$

Finally, the dual problem becomes

$$\min_{\gamma} \frac{1}{2} \gamma^T YKY\gamma - \sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k,$$

subject to

$$\sum_{k=1}^{C-1} \sum_{i \in D_k} \alpha_i^k y_i^k = 0,$$

$$\sum_{k=1}^L \sum_{i \in D_k} \alpha_i^k y_i^k \geq 0, \quad L = 1, \dots, C-2,$$

$$0 \leq \alpha_i^k \leq \lambda_k, \quad i \in D_k, \quad k = 1, \dots, C-1.$$