Biostatistics (2022) **00**, 00, *pp.* 1–16
https://doi.org/10.1093/biostatistics/kxac031

Penalized decomposition using residuals (PeDecURe) for feature extraction in the presence of nuisance variables

SARAH M. WEINSTEIN*

Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, Penn Statistics in Imaging and Visualization Center, 108/109B, Blockley Hall, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104, USA

smweinst@pennmedicine.upenn.edu

CHRISTOS DAVATZIKOS, JIMIT DOSHI

Department of Radiology, Perelman School of Medicine, Center for Biomedical Image Computing and Analytics, 3700 Hamilton Walk, Richards Building 7th Floor, University of Pennsylvania, Philadelphia, PA 19104, USA

KRISTIN A. LINN†, RUSSELL T. SHINOHARA†

Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, Penn Statistics in Imaging and Visualization Center, 2nd Floor, Blockley Hall, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104, USA and Department of Radiology, Perelman School of Medicine, Center for Biomedical Image Computing and Analytics, 3700 Hamilton Walk, Richards Building 7th Floor, University of Pennsylvania, Philadelphia, PA 19104, USA

FOR THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE‡

Summary

Neuroimaging data are an increasingly important part of etiological studies of neurological and psychiatric disorders. However, mitigating the influence of nuisance variables, including confounders, remains a challenge in image analysis. In studies of Alzheimer's disease, for example, an imbalance in disease rates by age and sex may make it difficult to distinguish between structural patterns in the brain (as measured by neuroimaging scans) attributable to disease progression and those characteristic of typical human aging or sex differences. Concerningly, when not properly accounted for, nuisance variables pose threats to the generalizability and interpretability of findings from these studies. Motivated by this critical issue, in this work, we examine the impact of nuisance variables on feature extraction methods and propose Penalized Decomposition Using Residuals (PeDecURe), a new method for obtaining nuisance variable-adjusted features.

*To whom correspondence should be addressed.
†These authors contributed equally to this work.
‡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.

PeDecURe estimates primary directions of variation which maximize covariance between partially residualized imaging features and a variable of interest (e.g., Alzheimer's diagnosis) while simultaneously mitigating the influence of nuisance variation through a penalty on the covariance between partially residualized imaging features and those variables. Using features derived using PeDecURe's first direction of variation, we train a highly accurate and generalizable predictive model, as evidenced by its robustness in testing samples with different underlying nuisance variable distributions. We compare PeDecURe to commonly used decomposition methods (principal component analysis (PCA) and partial least squares) as well as a confounder-adjusted variation of PCA. We find that features derived from PeDecURe offer greater accuracy and generalizability and lower correlations with nuisance variables compared with the other methods. While PeDecURe is primarily motivated by challenges that arise in the analysis of neuroimaging data, it is broadly applicable to data sets with highly correlated features, where novel methods to handle nuisance variables are warranted.

Keywords: Feature extraction; Image analysis; Multivariate data.

1. INTRODUCTION

The incorporation of neuroimaging data in research and clinical practice has brought new insights into how structural and functional patterns in the brain are associated with behavioral, psychiatric, and neurological phenotypes. For instance, structural magnetic resonance imaging (MRI) has provided a more nuanced understanding of how changes in the brain are associated with presymptomatic markers and longitudinal progression of Alzheimer's disease (AD) (Frisoni *and others*, 2010; Ferreira and Busatto, 2011) and has proven useful for diagnosing and evaluating treatment effectiveness for AD (Islam and Zhang, 2018; Struyfs *and others*, 2020; Mueller *and others*, 2006).

While methodological developments over the last few decades have improved the overall quality of neuroimaging studies, including about AD, questions about the generalizability (including reproducibility and replicability) of research in this field have gained traction (Poldrack *and others*, 2017). Recent work has brought attention to the importance of large sample sizes (Marek *and others*, 2022), representative sampling (LeWinn *and others*, 2017), and harmonization of data across multiple sites or scanners (Fortin *and others*, 2017) to improve the generalizability of neuroimaging research.

In the present work, we consider another important factor in generalizability: how confounders and other nuisance variables are handled. In many neuroimaging applications, confounding by demographic variables such as age and sex is ubiquitous. For instance, it can be challenging to distinguish structural changes in the brain observed in patient populations with AD from those that are characteristic of aging in healthy individuals (Double *and others*, 1996; Hua *and others*, 2010; Fjell *and others*, 2014). Additionally, AD is more prevalent among females than males, but sex differences in the brain can also be unrelated to Alzheimer's (Singh *and others*, 2006; Fjell *and others*, 2009; Eskildsen *and others*, 2013). Without properly accounting for sources of variability that are not of direct interest but are still correlated with both imaging features and phenotypes of interest, generalizability and interpretability of findings from neuroimaging studies may be severely compromised.

One common approach to handling nuisance variables in neuroimaging studies is "mass-univariate" residualization. This typically involves first fitting a separate model (e.g., linear regression) across study participants at each image location (e.g., voxel or vertex), where the image measurement is the dependent variable and confounders such as age and sex are covariates. Estimated effects of confounders are then subtracted from participant- and location-specific image measurements, and subsequent analyses (e.g., association studies or training predictive models) are conducted using these residuals. While mass-univariate methods underlie a number of important methods in neuroimaging analysis, including voxel-based morphometry (Ashburner and Friston, 2000; Davatzikos *and others*, 2001) and statistical

parametric mapping (Friston *and others*, 1991, 1994), it is also widely understood that these methods do not account for the inherent spatial structure of the brain and the correlatedness of measurements derived from neuroimaging scans. Multivariate methods, which do take this structure into account, are therefore increasingly preferred (Habeck *and others*, 2008; Habeck and Stern, 2010; Westman *and others*, 2011).

Broadly defined, multivariate methods consider joint relationships across features in models studying associations between imaging features and phenotypes (e.g., behavior or a diagnosis). In other words, unlike mass-univariate methods, multivariate methods leverage rather than ignore relationships across features to study joint image-phenotype associations rather than marginal associations. In neuroimaging research, multivariate methods sometimes involve maintaining the full set of imaging features in an association study or prediction model, while in other cases, feature extraction is done to reduce both the dimensionality and redundancy of highly correlated features (Mwangi *and others*, 2014). To our knowledge, only a limited number of methods using a multivariate framework have been proposed to handle confounders and other nuisance variables in neuroimaging. Linn *and others* (2016) proposed the inverse probability weighted support vector machine (IPW-SVM), which involves weighting the slack variables of the SVM objective function with weights similar to those widely used in the causal inference literature to control for confounding in observational studies (Hernán and Robins, 2006; Cole and Hernán, 2008). Predictions (e.g., AD classification) from the IPW-SVM model are considered confounder-adjusted in the sense that they are based on patterns (i.e., sets of features) in the brain that are directly related to disease progression, rather than patterns that are common among those with the disease simply because of overlap in confounders like older age or sex. Rao *and others* (2017) extended this idea, using "instance weighting" to train confounder-adjusted machine learning models for predicting continuous outcomes (e.g., clinical risk scores). More recently, More *and others* (2020) considered a variety of regression and normalization-based approaches to mitigate nuisance variables. In association studies involving brain networks, multivariate distance matrix regression can adjust for nuisance variables by including them as model covariates (Shehzad *and others*, 2014; Simpson and Laurienti, 2015).

Notably, the existing approaches to adjust for nuisance variables in the multivariate framework preserve the full set of imaging features. However, feature extraction methods, such as independent components analysis, canonical correlation analysis, partial least squares (PLS), and principal component analysis (PCA) are increasingly popular in neuroimaging (Mwangi *and others*, 2014), and tools for mitigating nuisance variables in this context are needed. Outside the neuroimaging literature, there is some existing work in this area. Lin *and others* (2016) propose adding a penalty term to the PCA objective function to achieve simultaneous dimension reduction and correction for batch effects (a type of nuisance variable). Aliverti *and others* (2021) propose a similar penalized decomposition approach, aiming to extract features that are uncorrelated with variables that may introduce unfair biases into predictive models. Tabak *and others* (2020) use a domain adaptation framework for dimension reduction and nuisance variable adjustment in cellular imaging. There is also some work using adversarial learning for the removal of confounding variables (Adeli *and others*, 2019; Zhao *and others*, 2020), highlighting the importance of mitigating nuisance variables to achieve generalizability. While these methods vary in their applications, they are similar in that they map a set of correlated features to a space that is free from the influence of nuisance variables.

In the present work, we examine the challenge of feature extraction in neuroimaging when the goal is to both adjust for nuisance variables and also to study associations with a phenotype of interest. To address these goals, we propose a new method: Penalized Decomposition Using Residuals (PeDecURe). PeDecURe extracts a small number of features, which can be easily incorporated into any downstream analysis, including classical regression or machine learning models. While PeDecURe is primarily motivated by methodological issues in the context of neuroimaging, it undertakes an issue that is pervasive across many areas of applied research and can be easily applied to data in other settings.

2. METHODS

2.1. *Notation and previous methods*

Let X be an $n \times p$ matrix of p image-derived features, such as volumes of different regions of interest (ROIs) in the brain, for n study subjects. Let A be an $n \times q$ matrix of nuisance variables, which (on the basis of subject area knowledge) must be adjusted for before analyzing associations between X and Y, an outcome of interest (say, AD diagnosis). Without loss of generality, we assume each column of X, A, and Y has been centered to have mean zero. We also assume X, A, and Y have some degree of interdependence, which we illustrate as a Venn diagram in Figure 1(a). While A may include confounders as defined in a traditional causal inference framework (Greenland *and others*, 1999), in this article, we use the term "nuisance variables" to refer more generally to sources of variation in the data that are not of direct interest but are correlated with variables that are of interest.

Using this notation, the objective functions of PCA and PLS are expressed in Figure 1(b). Briefly, each method estimates v_j ($j = 1, ..., p^*$ for $p^* \leq p$), which are directions of variation or primary components

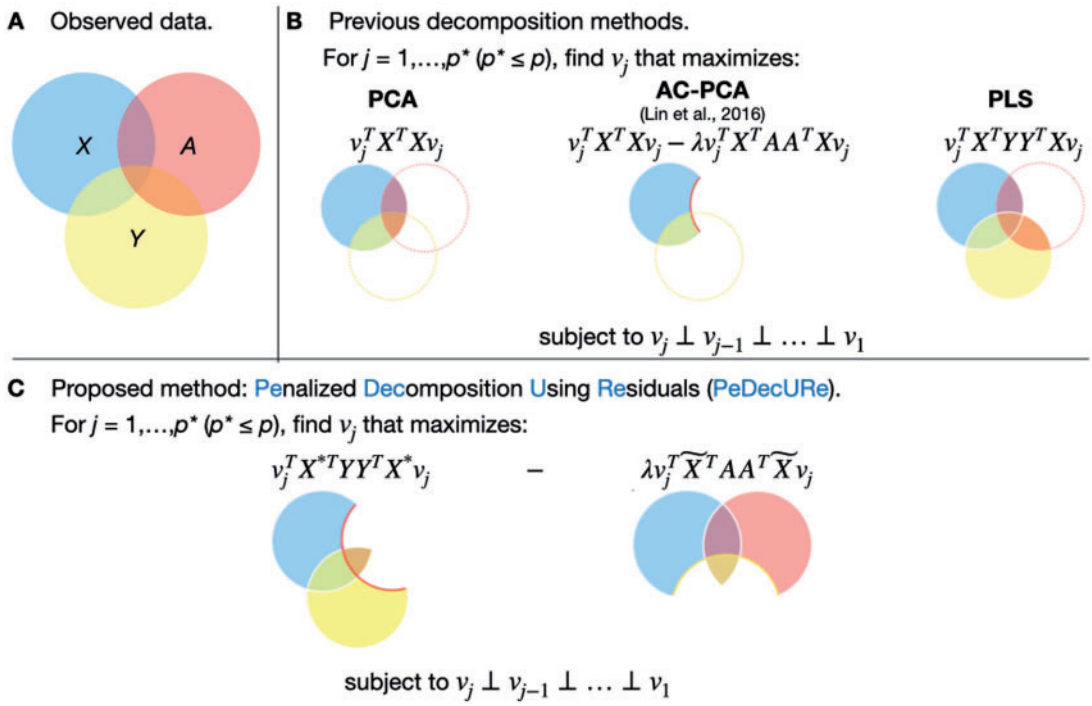

Fig. 1. Visualization of decomposition methods considered in this article. (A) How we broadly assume interdependence between the observed features (X), outcome of interest (Y), and nuisance variables (A). (B) Objective functions for three existing methods and illustrates the portions of the data whose covariance will be explained by v_j. In particular, for $j = 1, \ldots, p^*$ ($p^* \leq p$), PCA finds v_j which maximize covariance of X, AC-PCA solves a similar maximization, but in a subspace that is (nearly) orthogonal to A (as controlled by a tuning parameter λ), and PLS finds v_j which maximize the covariance between X and Y (equivalent to the eigenvectors of the covariance of $X^T Y$). Panel C illustrates our proposed method, where we find v_j that simultaneously maximizes associations between X and Y—but not the portions of X that are also associated with A (hence using the residual X^* instead of the original X))—and minimizes covariance between X and A—without penalizing portions of X that are also associated with Y (hence the residual $\widetilde{X}$).

(PCs) that maximize the covariance of X (PCA) or covariance between X and Y (PLS), subject to orthogonality to previously estimated PCs (as well as other potential constraints). Figure 1(b) also illustrates that the portion of observed data whose covariance is maximized in both PCA and PLS include regions of overlap with A. To address this problem in the context of PCA, Lin *and others* (2016) propose AC-PCA, which maximizes covariance of X while penalizing covariance between X and A (to an extent determined by a tuning parameter λ). As we illustrate in Figure 1(b), AC-PCA may result in the loss of information about Y, which is of interest.

2.2. *Proposed method: PeDecURe*

In settings where the outcome of interest Y is observed, a natural extension of Lin *and others* (2016)'s AC-PCA would be to find directions of variation which maximize the covariance between X and Y (as in PLS), subject to the same constraint as in AC-PCA:

$$\underset{v_j}{\text{argmax}} \left(v_j^T X^T Y Y^T X - \lambda v_j X^T A A^T X v_j \right) \quad \text{s.t. } v_j \perp v_{j-1} \perp \cdots \perp v_1. \tag{2.1}$$

However, given interdependence between X, Y, and A, it is conceivable that directions of variation derived from solving (2.1) may still be driven by associations with A or have diluted associations with Y. To avoid this possibility, we propose using residualized features in the proposed objective function.

Similar to mass-univariate residualization, we first fit the following linear model at each image location:

$$X^{(j)} = \mathbf{1}\beta_0^{(j)} + Y\beta_Y^{(j)} + A\beta_A^{(j)} + \epsilon \quad j = 1, \ldots, p, \tag{2.2}$$

where $X^{(j)}$ is an n-dimensional vector of image measurements at the jth location or image unit (e.g., voxel), $\mathbf{1}$ is an n-dimensional vector of 1's, $\beta_0^{(j)}$ is the location-specific intercept parameter, $\beta_Y^{(j)}$ represents the effect of Y on the jth image feature (conditional on A), $\beta_A^{(j)}$ is a $q \times 1$ vector of the effects of each column of A on the jth image feature (conditional on Y), and ϵ is a normally distributed error term. In the present work, we assume the normality of each imaging feature in (2.2), as this has been the assumption in voxelwise models of neuroimaging features (Friston *and others*, 1994; Ashburner and Friston, 2000; Davatzikos *and others*, 2001; Oakes *and others*, 2007). However, depending on the imaging modality and covariates, or in nonimaging extensions of our method, other types of models could be substituted.

Parameter estimates from each location-specific model (2.2) are used to compute the following partial residuals for each feature:

$$X^{(j)*} = X^{(j)} - A\hat{\beta}_A^{(j)}, \tag{2.3}$$

and $X^* = \left[X^{(1)*} \quad X^{(2)*} \quad \ldots \quad X^{(p)*}\right]$ denotes the matrix of these residuals—that is, the result of removing the effects of nuisance variables A from each feature ($X^{(j)}$), conditional on Y. We also define the partial residuals that remain when we remove Y from $X^{(j)}$ conditional on A:

$$\widetilde{X}^{(j)} = X^{(j)} - Y\hat{\beta}_Y^{(j)}, \tag{2.4}$$

and $\widetilde{X} = \left[\widetilde{X}^{(1)} \quad \widetilde{X}^{(2)} \quad \ldots \quad \widetilde{X}^{(p)}\right]$ denotes the matrix of this second set of residuals.

The objective function for our proposed method is as follows. For $j \leq p$, we solve:

$$\underset{v_j}{\text{argmax}} \left(v_j^T X^{*T} Y Y^T X^* v_j - \lambda v_j \widetilde{X}^T A A^T \widetilde{X} v_j \right) \quad \text{s.t.} \quad v_j \perp v_{j-1} \perp \cdots \perp v_1. \tag{2.5}$$

In words, PeDecURe identifies PCs which maximize covariance between X^* and Y, while simultaneously penalizing the covariance between $\widetilde{X}$ and A. As we illustrate in Figure 1(c), by using the residuals X^* and $\widetilde{X}$, we aim to control the influence of A on our estimated PCs while preserving the influence of Y, except for the portions of Y that overlap with A. While the primary analyses in this paper will compare PeDecURe to PCA, PLS, and AC-PCA, we also provide a comparison to penalized PLS without residualization (Equation (2.1)) in Section S1 of the Supplementary material available at *Biostatistics* online based on the simulation settings described below.

Features (or PC scores) extracted using each method, including PeDecURe, are calculated as the matrix-vector product $\text{PC}_j = Xv_j$. We note that the same feature matrix X is used to estimate PC scores across all methods considered in this article; that is, even though PeDecURe's objective function involves residuals (from a model that requires that both A and Y have been observed in the sample used for estimating the directions of variation, v_j), the features estimated using PeDecURe's v_js are based on the original, not residualized, X.

2.2.1. *Selection of tuning parameter* λ. The tuning parameter λ in PeDecURe (2.5) is based on a similar idea as in Lin *and others* (2016)'s AC-PCA, where the goal is to select λ in order to balance in the extent to which we penalize associations with nuisance variables versus maximize covariance involving variables of interest. In PeDecURe, we select a λ which, on average (across nuisance variables and PCs), maximizes the difference between the absolute value of partial correlations between PC scores with Y and those between PC scores with A_1 and A_2. Out of a set of prespecified candidate λs, we consider the following maximization:

$$\lambda_{\text{best}} = \underset{\lambda}{\text{argmax}} \frac{1}{q} \sum_{j=1}^{q} \sum_{k=1}^{p^*} \frac{\alpha_k}{\sum_{k^*=1}^{p^*} \alpha_{k^*}} \left(\mid \text{Cor}(Xv_k^{(\lambda)}, Y \mid A_j) \mid - \mid \text{Cor}(Xv_k^{(\lambda)}, A_j \mid Y) \mid \right), \tag{2.6}$$

where q is the number of nuisance variables, p^* ($\leq p$) is the number of PCs that we estimate, α_k is the eigenvalue corresponding to the kth PC, and $Xv_k^{(\lambda)}$ is the kth PC score from PeDecURe, estimated when the tuning parameter is set to λ. We use the proportion of variation explained by each PC ($\frac{\alpha_k}{\sum_{k^*=1}^{p^*} \alpha_{k^*}}$) to up-weight earlier components in this maximization. In Figure S1.1 of the Supplementary material available at *Biostatistics* online, we illustrate for the two simulation settings described below that PeDecURe appears robust to the specific choice of λ.

3. PARAMETRIC SIMULATION STUDIES

3.1. *Simulation setting*

Using the framework and notation introduced above, we conduct a parametric simulation study to compare PeDecURe to PCA, AC-PCA, and PLS. We simulate 1000 data sets, each with sample size $n = 200$ and $p = 1000$ observed features and $q = 2$ observed nuisance variables per participant. The observed feature matrix X is constructed as a linear combination of W (a matrix of underlying patterns of interest) and A (a matrix of nuisance variables). We simulate the outcome Y as a function of both W and A, and a preferred method would extract features that are correlated with Y and uncorrelated with A. A detailed description of the simulation setup is provided in Section S2 of the Supplementary material available at *Biostatistics* online.

We apply each method to 70% of each simulated sample and reserve the remaining 30% for testing. Partial Pearson correlations between the top three PC scores from each method and Y (conditional on A_1 and A_2), A_1 (conditional on Y and A_2), and A_2 (conditional on Y and A_1) are calculated. We consider a method to perform well if the distribution of the absolute values of partial correlations between the first

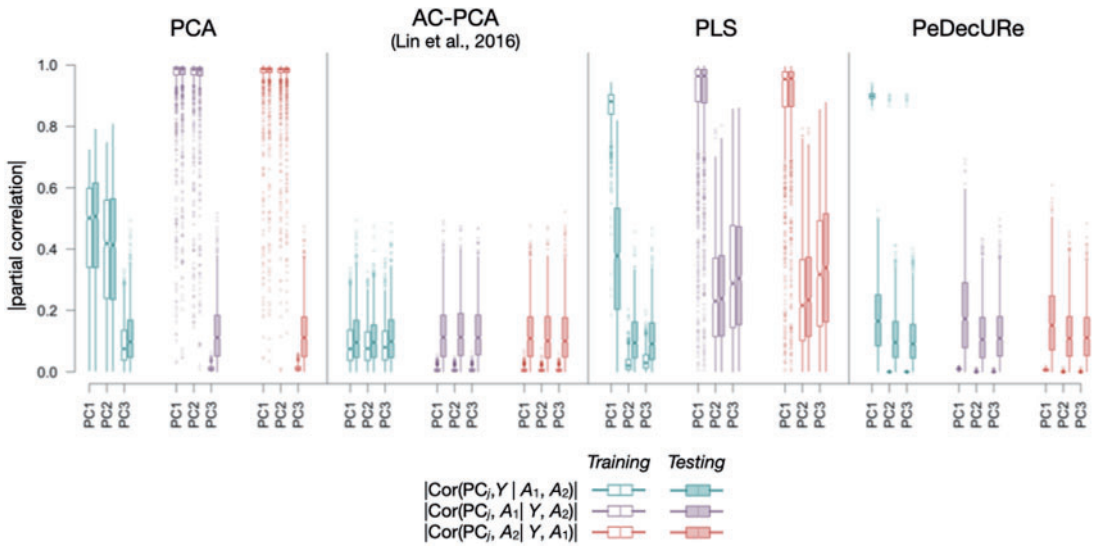

Fig. 2. Parametric simulation results in 1000 random training and testing samples. Groups of boxplots under each method for primary component (PC) scores 1-3 indicate associations with the outcome variable (Y (first group, in green)) or one of the two nuisance variables (A_1 (second group, in purple), A_2 (third group, in orange)), conditional on the other variables. In the training sample (white boxes) PeDecURe is the only method where we observe a high distribution of absolute values of partial correlations between the first PC score and Y and low distributions of partial correlations between all three PCs and each nuisance variables. In testing samples (shaded boxes), the distribution of absolute values of partial correlations between PC1 and Y is not substantially higher than it is between PC1 and the nuisance variables.

PC score and Y is higher than the distribution of absolute partial correlations between all PC scores with A_1 and A_2.

3.2. *Simulation study results*

Results from our parametric simulation study are plotted in Figure 2. Scores on PCA's first and second PCs and PLS's first PC have high magnitudes of partial correlations with A_1 (purple boxplots) and scores on PCA's first and second PCs and PLS's first and third PCs have elevated magnitudes of partial correlations A_2 (orange). The absolute values of partial correlations between PC scores derived using AC-PCA and Y (green), A_1, and A_2 are all low, which corroborates the intuition behind a method that combines strengths of PLS (which preserves associations with Y) and AC-PCA.

PeDecURe's distributions of absolute values of partial correlations between PC scores and A_1 and A_2 are low, while distributions of partial correlations between PeDecURe's PC1 are high in training and remain slightly elevated in testing, although they are not substantially higher than the distributions involving the two nuisance variables. Figure S1.2 of the Supplementary material available at *Biostatistics* online provides a comparison between PeDecURe and the penalized PLS approach without residualization (2.1).

4. Application to structural MRI data

4.1. *Experiments using data from the ADNI*

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private

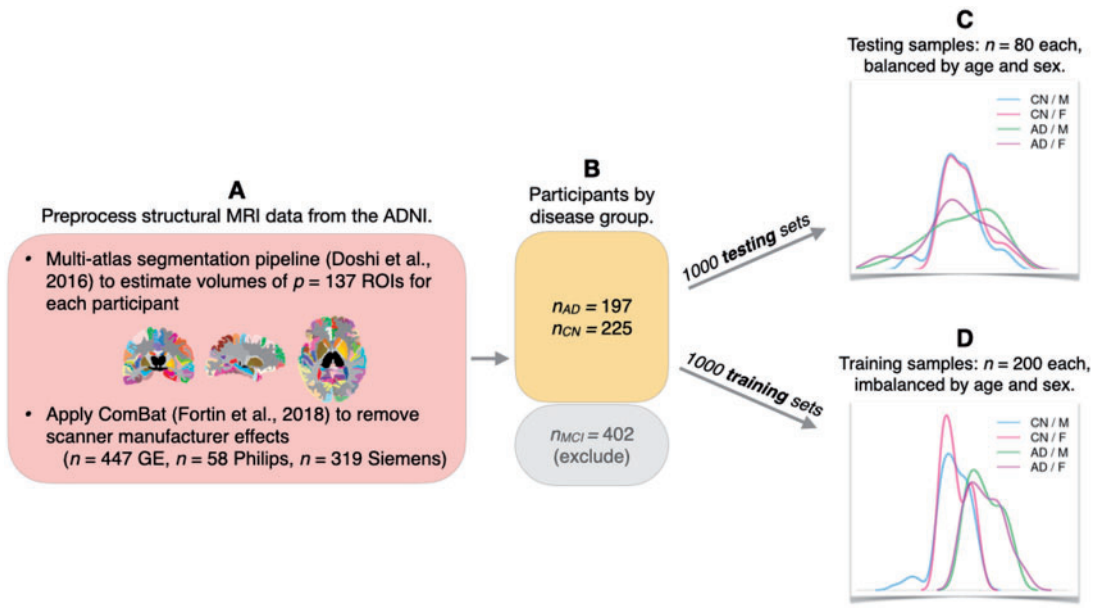

Fig. 3. Preprocessing of data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and sampling of training and testing samples for plasmode simulation study. Note in Panels C and D that each testing sample is drawn before the corresponding training sample, as the test sample preserves the distribution of the original ADNI study, while each training sample is selected through a biased sampling procedure to ensure that the age distribution for individuals with AD is higher than for those in the CN group that and females are more likely to have AD than males (see Section 4.1). AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; M/F, males/females; MCI, mild cognitive impairment; ROIs, regions of interest.

partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early AD.

In this study, we apply feature extraction methods to ROI volumes (X) derived from baseline structural MRI scans in the ADNI study (The ADNI team, 2021). Extracted features are then used to evaluate associations with and predict a variable of interest (AD diagnosis, Y), while controlling for age (A_1) and sex (A_2). Both age and sex confound the relationship between brain structure and AD diagnosis, and controlling these sources of variation is a necessary intermediate step to allow for the generalizability of association studies or predictive models of AD diagnosis.

Preprocessing steps are summarized in Figure 3. Briefly, we applied Doshi *and others* (2016)'s multi-atlas segmentation pipeline to estimate volumes for $p = 137$ ROIs for $n = 840$ ADNI participants. We also applied ComBat (Fortin *and others*, 2017) to correct for batch effects associated with different scanner manufacturers (adjusting for diagnosis, sex, and age). We then restricted our sample to $n = 422$ participants who, based on baseline cognitive assessments, were classified as "cognitively normal" (CN, $n = 225$) or met the criteria for an AD diagnosis ($n = 197$).

4.2. *Plasmode simulation setting*

Since models that ignore or inadequately control for nuisance variables pose threats to generalizability, our simulation study involving the ADNI data uses a flexible procedure to sample training and testing

samples, such that the distribution of age and sex differs between these samples. Thus, our assessment of method performance in the testing sample should provide insight into each method's generalizability. We first set aside each testing sample, including $n = 80$ participants randomly subsetted from the 422 ADNI participants who met our inclusion criteria (Figure 3(a) and (b)). Thus, the typical testing sample preserves the same age and sex distributions across the disease group that is present in the original ADNI data (Figure 3(c)). We then sample $n = 200$ individuals from the remaining 342 participants using a biased sampling procedure to induce imbalance in both age and sex in each testing sample, which would reflect confounding patterns by age and sex, an exaggeration of we would expect in a random sample of adults (Figure 3(d)).

We first evaluate each method's performance based on the distribution of partial correlations in training and testing samples between each PC score and Y, A_1, and A_2, similar to our evaluation of parametric simulation results. Second, we consider the prediction of the disease group in each test sample. Since the first direction of variation captures the largest percentage of variation (of some covariance matrix, which differs across methods), we consider the prediction of Y, A_1, and A_2 based on the PC1 score from each method using simple logistic (for Y and A_2) or linear (for A_1) regression models. In practice, we would consider a method to perform well if there is high predictive accuracy for AD status (as measured by area under the curve [AUC]) and poor prediction of both age (as measured by root mean squared errors [RMSE]) and sex (as measured by 1−AUC). We also compare distributions of these prediction metrics to those obtained from cross-validated models trained using the $n = 422$ ADNI participants meeting the inclusion criteria described above. Unlike the samples used to train the models used for prediction in our plasmode simulation setting, the sample of 422 ADNI participants has a balanced distribution of age and sex across disease groups.

Finally, since the PC loadings (e.g., v_1) are each vectors with dimension $p = 137$, the number of ROIs, it is possible to interpret these loadings anatomically in the brain. For example, the estimated value in the jth position of v_1 may be interpreted as a weight indicating the relative importance of that ROI. To identify the most important combination of ROIs according to each feature extraction method, we conduct an exploratory analysis where we identify the ROIs with weights exceeding the 95th percentile of weights (all the values in v_1 from a given training sample and method). After identifying the top 5% of ROIs for each simulation, we identify which ROIs meet these top 5% criteria for at least 50% of all simulations. Finally, we visualize and compare the top ROIs from the first direction of variation across methods. To facilitate interpretation in this exploratory analysis, we also identify the top 5% of ROIs in PC1 estimated by each method in the ADNI sample of $n = 422$. Since the ADNI sample of $n = 422$ participants has a balanced distribution of age and sex across disease groups, the top 5% of ROIs estimated in this sample may reflect which subset of ROIs would be most important for explaining variation in the data in a setting that is, in theory, free from the nuisance variation that we induced into each training sample in our plasmode simulation setting.

4.3. *Results from ADNI plasmode simulations*

Distributions of the absolute values of partial correlations are plotted in Figure 4(a) and suggest that in both training and testing samples, PeDecURe performs as expected. Specifically, associations between the PC1 score and diagnosis (Y, green) are the highest for PeDecURe compared to PCA, AC-PCA, and PLS in both the training and testing samples. While there is a higher variability in the partial correlation distributions in the test samples, the same trend is present, which supports PeDecURe's out-of-sample generalizability given the different sampling procedure used for training and testing (Figure 3(c) and (d)). Furthermore, PeDecURe's scores on all of the first three PC scores have low distributions of absolute partial correlations with both nuisance variables (age (A_1, purple) and sex (A_2, orange)), while PCA, AC-PCA, and PLS have noticeably higher distributions for at least one PC.

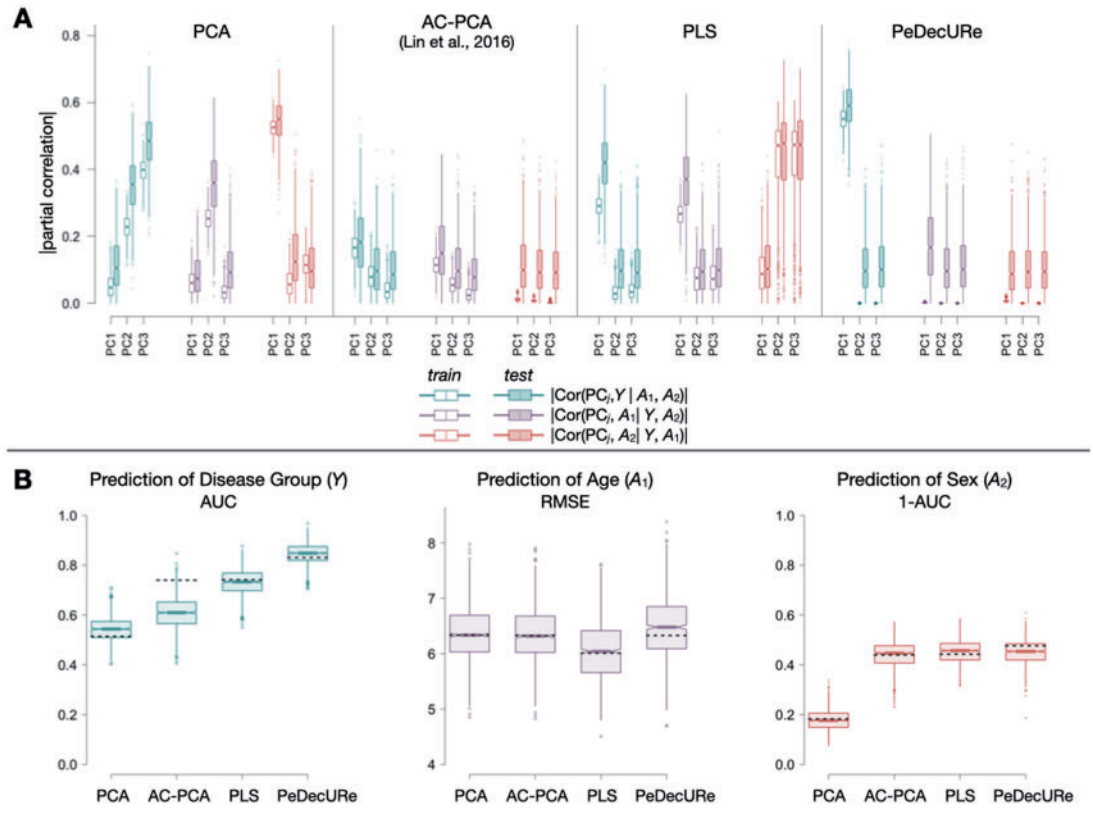

Fig. 4. Results from plasmode simulation studies using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). (A) Groups of boxplots under each method are labeled with primary component (PC) scores 1-3 and show distributions of partial correlations with Alzheimer's disease status (Y (first group of boxes, in green), conditional on age (A_1) and sex (A_2)), age (second group of boxes, in purple), conditional on disease and sex, and sex (third group of boxes, in orange), conditional on age and disease. Distributions from 1000 training samples of $n = 200$ participants each are plotted next to those from 1000 testing samples of $n = 80$ participants each. Note that the distribution of age and sex across disease groups differs in the training and testing samples, allowing us to evaluate the generalizability of each method. PeDecURe is the only method with high distributions of partial correlations between the PC1 score and Y and low distributions of partial correlations between all three PC scores and each nuisance variable, in both training and testing samples. (B) Prediction of disease group (Y, left) age (A_1, center), and sex (A_2, right) in 1000 testing samples using only the first PC score from each method. Predictive performance is evaluated using area under the curve (AUC) for disease group, root mean squared error (RMSE) for age, and one minus AUC (1−AUC) for sex, so that higher values on each y-axis correspond to more desirable performance, given that we would prefer a method with high predictive accuracy for disease and low predictive accuracy for age and sex. Horizontal dotted lines shown with each distribution show prediction based on 5-fold cross-validation in models trained using a subsample of $n = 422$ from ADNI, which is balanced across disease group with respect to age and sex (while the actual training data was imbalanced with respect to these nuisance variables).

As shown in Figure 4(b), PeDecURe also outperforms the three other methods for disease group prediction. Not surprisingly, as PLS is the only other method that prioritizes maximizing covariance with the outcome, its PC1 score is also more highly predictive of diagnosis than both PCA and AC-PCA. It is conceivable that the higher AUC distribution we obtain from a model involving PeDecURe's PC1 score compared with a model involving PLS's PC1 score may be due to the lack of nuisance variable adjustment in estimating PLS's directions of variation. In addition, although PeDecURe has a similar

RMSE distribution for age prediction as PCA and AC-PCA, and PeDecURe also has a similar 1−AUC distribution for sex prediction as AC-PCA and PLS, our method's favorability over the other methods is apparent in that its prediction of Y exceeds that of other methods and its prediction of confounding variables is no worse than others.

The results in Figure 4(b) also suggest that PeDecURe's out-of-sample performance is similar to its in-sample performance, as the 5-fold cross-validated AUC for predicting disease group based on a sample with a balanced distribution of age and sex across disease groups is close to the distribution of AUCs in a similarly balanced test sample (Figure 3(c), even though each method was trained in a sample with substantial imbalance by sex and age (Figure 3(d)).

Similar to the parametric simulation setting, we also compare PeDecURe to penalized PLS without residualization (2.1), with results provided in Section S1 of the Supplementary material available at *Biostatistics* online. In the ADNI simulation setting, PeDecURe's PC1 has a higher distribution of absolute partial correlations with Y (Figure S1.3 (a) of the Supplementary material available at *Biostatistics* online) and more highly predictive of Y (Figure S1.3(b) of the Supplementary material available at *Biostatistics* online) compared to penalized PLS without residualization.

4.4. *Exploratory pattern interpretation*

Finally, we consider patterns extracted from each method by comparing the top ROIs from the 1000 plasmode simulation training samples to the top ROIs identified when each method was applied to the $n = 422$ ADNI participant subsample whose age and sex distributions were balanced across disease groups.

It is interesting to note that the top 5% of ROIs identified by both methods that involve maximizing the covariance of X (PCA and AC-PCA) are the same in the imbalanced simulation and balanced analysis settings. Specifically, the same set of top ROIs we identify using PCA in the plasmode simulation setting and in the balanced ADNI sample of $n = 422$ (Figure 5, left and right panels of the first row are the same). The same is true for AC-PCA (Figure 5, left and right panels of the second row are the same). We also note that the same is not true for PLS and PeDecURe (i.e., left and right panels differ in the last two rows of Figure 5), suggesting that the selection of ROIs for feature extraction methods which prioritize shared variation in Y may be more sensitive to changes in the sampling distribution, even though PeDecURe's prediction of and correlations with Y were consistent across training and testing samples (Figure 4). This will be interesting to explore in greater detail in future work.

Since our simulations using the ADNI data suggested PeDecURe outperformed all other methods (Figure 4), it seems plausible that PeDecURe may be best at identifying the specific combination of ROIs that balance attenuated associations with nuisance variables but are still jointly associated with disease group. Nevertheless, across all methods, the top ROIs identified in both the plasmode simulations and analysis of the full ADNI subsample of $n = 422$ are consistent with those found in previous structural neuroimaging-based studies of AD (Magnin *and others*, 2009; Habeck *and others*, 2008; Westman *and others*, 2011; Jiang *and others*, 2014). In future work, it will be helpful to develop hypothesis testing methods to conduct a more rigorous assessment of these exploratory findings.

5. Discussion

In this article, we proposed PeDecURe, a new feature extraction method with built-in nuisance variable adjustment. By combining the strengths of previous methods, PeDecURe extracts features that are not only uncorrelated with nuisance variables but that also preserve strong associations with a variable of interest. Through simulation studies and real data analyses, we showed that features derived using PeDecURe can be used to conduct association studies and train generalizable predictive models.

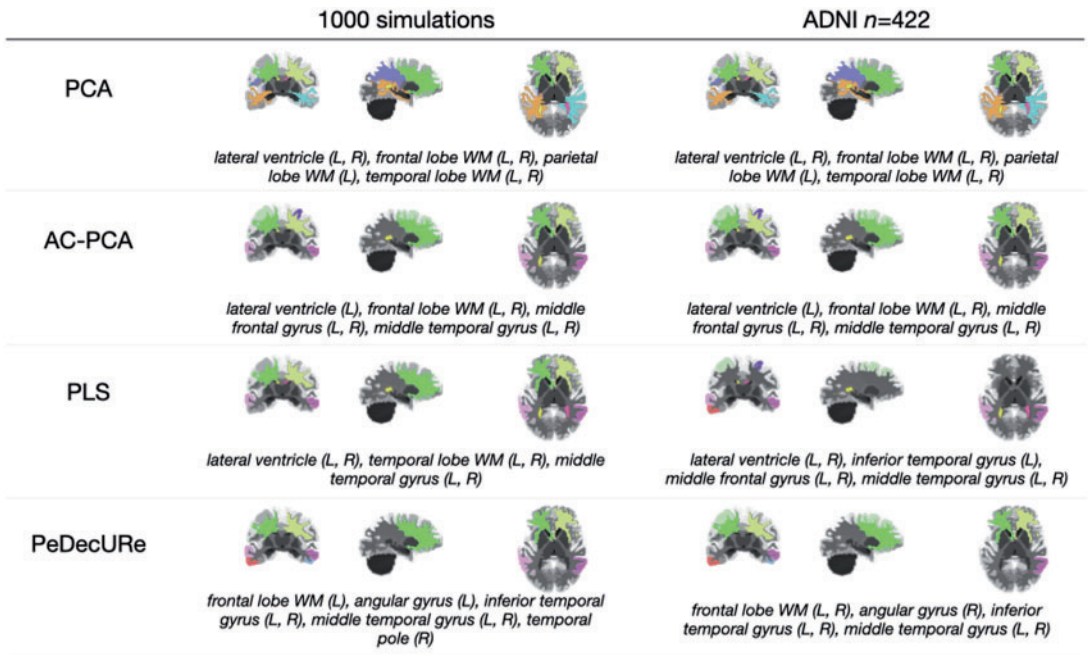

Fig. 5. Top regions of interest (ROIs) according to the first primary component (PC) loadings from each method. ROIs above 95th percentile of PC1 for at least 50% of ADNI plasmode simulations are shown on the left, where the training data are imbalanced with respect to age and sex across disease groups. On the right, we show the top 5% of ROIs after applying each method to a balanced ADNI sample of n=422 (after applying exclusion criteria outlined in Figure 3). L/R, left/right; WM, white matter.

Our proposed method differs from other approaches to handling nuisance variables in neuroimaging, including methods which train predictive models based on the full set of imaging features (Linn *and others*, 2016; Rao *and others*, 2017) or transform the original set of features to be adjusted for confounders without any dimension reduction (More *and others*, 2020). Our present work also differs from previous work in confounder-adjusted feature extraction outside neuroimaging, including Lin *and others* (2016)'s AC-PCA, which adds a penalty term to the PCA objective function (similar to the penalty term in our proposed method). In particular, results from our simulations suggest that an unintended consequence of AC-PCA is the loss of information relevant about an outcome of interest. While AC-PCA has a number of important strengths, features derived using this method may not be suitable for outcome prediction.

We also found that it was not sufficient to implement a penalized PLS using the AC-PCA penalty term without residualizing features (2.1). Notably, in simulations using data from the ADNI, we showed that compared with features derived using penalized PLS without residualization, PeDecURe's features have higher magnitudes of partial correlations with the disease group (Figure S1.3(a) of the Supplementary material available at *Biostatistics* online) and higher AUCs for predicting disease (Figure S1.3(b) of the Supplementary material available at *Biostatistics* online).

Our method has several limitations that warrant further discussion. Like PLS and unlike PCA and AC-PCA, PeDecURe can only be applied in samples where the outcome of interest Y is observed, and in this sense, it is a supervised method. However, like PLS, the directions of variation that we estimate with PeDecURe can be flexibly applied in external testing samples where Y is unobserved, since there is no need to residualize the testing data, as we calculate PC scores using the original features (X). The results from our ADNI simulation study suggested that PeDecURe's performance is robust in testing samples with

a different sampling distribution from the training sample. As concerns regarding the generalizability of neuroimaging studies have transpired in recent years (Marek *and others*, 2022; LeWinn *and others*, 2017), a feature extraction method that both adjusts for nuisance variables and performs well out-of-sample is appealing.

In addition, while imaging data are often high-dimensional, in this paper we considered settings with at most 1000 features per participant. As we show in Section S3 of the Supplementary material available at *Biostatistics* online, PeDecURe's run-time is reasonable in moderate-dimensional settings with $p > n$, despite its computational complexity. However, our method has not yet been extended to high-dimensional settings ($p >> n$), where standard eigendecomposition-based methods may not be feasible due to computational cost. A sparse PLS-based adaptation of PeDecURe is one possible approach, which may improve the viability of our method in high-dimensional settings. Another limitation of PeDecURe is that by maximizing and penalizing covariance (as in PLS) as opposed to correlation (as in canonical correlation analysis (CCA)), our method is not scale-invariant. A CCA-based adaptation of PeDecURe would be both a useful and natural extension of PeDecURe, which may be more relevant in the context of unstructured imaging modalities (Krishnan *and others*, 2011; Zhuang *and others*, 2020).

Another important extension of this work will be to consider alternative models for estimating the residuals (X^* and $\widetilde{X}$). While researchers routinely assume normality in models of structural neuroimaging measurements (Friston *and others*, 1994; Ashburner and Friston, 2000; Davatzikos *and others*, 2001; Oakes *and others*, 2007), and we make this assumption when we fit models to estimate residuals for PeDecURe, there are settings where this assumption may not be suitable (e.g., functional MRI), and more flexible models that account for temporal or nonlinear aspects of the data may be preferred. It would also be interesting to consider multivariate approaches to residualization, as our current approach involves fitting a separate model at each image location.

Lastly, it will be important to evaluate the stability of PeDecURe (Helmer *and others*, 2021; Marek *and others*, 2022), including quantifying the possible trade-off between improvements in generalizability offered by PeDecURe through nuisance variable adjustment and the possible loss of generalizability in small-sample or higher-dimensional settings. These limitations present abundant opportunities for future research.

6. Software

An R package with code for implementing PeDecURe is publicly available at https://github.com/smweinst/PeDecURe.

Supplementary material

Supplementary material is available at online http://biostatistics.oxfordjournals.org.

Acknowledgments

Conflict of Interest: Russell T. Shinohara receives consulting income from Octave Bioscience and compensation for reviewership duties from the American Medical Association. Kristin A. Linn receives consulting income from Correlation One and compensation for reviewership duties from the American Medical Association.

Funding

This work is supported by the following grants from the National Institute of Mental Health: R01MH112847 (RTS), R01MH123550 (RTS), and R01MH112070 (CD), the following grants from the National Institute of Neurological Disorders and Stroke: R01NS112274 (RTS) and R01NS060910 (RTS), and the following grants from the National Institute on Aging: U01AG068057 (CD) and RF1AG054409

(CD). This work is also supported by the National Science Foundation Graduate Research Fellowship Program (SMW). 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

ADELI, E., ZHAO, Q., PFEFFERBAUM, A., SULLIVAN, E. V., LI, F.-F., NIEBLES, J. C. AND POHL, K. M. (2019). Bias-resilient neural network. https://openreview.net/forum?id=Bke8764twr

ALIVERTI, E., LUM, K., JOHNDROW, J. E. AND DUNSON, D. B. (2021). Removing the influence of group variables in high-dimensional predictive modelling. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **184**, 791–811.

ASHBURNER, J. AND FRISTON, K. J. (2000). Voxel-based morphometrythe methods. *Neuroimage* **11**, 805–821.

COLE, S. R. AND HERNÁN, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 656–664.

DAVATZIKOS, C., GENC, A., XU, D. AND RESNICK, S. M. (2001). Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* **14**, 1361–1369.

DOSHI, J., ERUS, G., OU, Y., RESNICK, S. M, GUR, R. C., GUR, R. E., SATTERTHWAITE, T. D., FURTH, S., DAVATZIKOS, C., ALZHEIMER'S NEUROIMAGING INITIATIVE *and others*. (2016). MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *Neuroimage* **127**, 186–195.

DOUBLE, K. L., HALLIDAY, G. M., KRILL, J. J., HARASTY, J. A., CULLEN, K., BROOKS, W. S., CREASEY, H. AND BROE, G. A. (1996). Topography of brain atrophy during normal aging and Alzheimer's disease. *Neurobiology of Aging* **17**, 513–521.

ESKILDSEN, S. F., COUPÉ, P., GARCÍA-LORENZO, D., FONOV, V., PRUESSNER, J. C., COLLINS, D. L., ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE *and others*. (2013). Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* **65**, 511–521.

FERREIRA, L. K. AND BUSATTO, G. F. (2011). Neuroimaging in alzheimer's disease: current role in clinical practice and potential future applications. *Clinics* **66**, 19–24.

FJELL, A. M., WESTLYE, L. T., AMLIEN, I., ESPESETH, T., REINVANG, I., RAZ, N., AGARTZ, I., SALAT, D. H., GREVE, D. N., FISCHL, B. *and others*. (2009). Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. *Journal of Neuroscience* **29**, 8774–8783.

Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Dale, A. M., Walhovd, K. B. and Alzheimer Disease Neuroimaging Initiative. (2014). Accelerating cortical thinning: unique to dementia or universal in aging? *Cerebral Cortex* **24**, 919–934.

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E. *and others*. (2017). Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **161**, 149–170.

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* **6**, 67–77.

Friston, K. J., Frith, C. D., Liddle, P. F. and Frackowiak, R. S. J. (1991). Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism* **11**, 690–699.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D. and Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* **2**, 189–210.

Greenland, S., Pearl, J. and Robins, J. M. (1999). Confounding and collapsibility in causal inference. *Statistical Science* **14**, 29–46.

Habeck, C., Foster, N. L., Perneczky, R., Kurz, A., Alexopoulos, P., Koeppe, R. A., Drzezga, A. and Stern, Y. (2008). Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. *Neuroimage* **40**, 1503–1515.

Habeck, C. and Stern, Y. (2010). Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease. *Cell Biochemistry and Biophysics* **58**, 53–67.

Helmer, M., Warrington, S., Mohammadi-Nejad, A.-R., Ji, J. L., Howell, A., Rosand, B., Anticevic, A., Sotiropoulos, S. N. and Murray, J. D. (2021). On stability of canonical correlation analysis and partial least squares with application to brain-behavior associations. *BioRxiv*, 2020–2008. https://www.biorxiv.org/content/10.1101/2020.08.25.265546v2

Hernán, M. A. and Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health* **60**, 578–586.

Hua, X., Hibar, D. P., Lee, S., Toga, A. W., Jack Jr, C. R., Weiner, M. W., Thompson, P. M., Alzheimer's Disease Neuroimaging Initiative *and others*. (2010). Sex and age differences in atrophic rates: an ADNI study with n = 1368 MRI scans. *Neurobiology of Aging* **31**, 1463–1480.

Islam, J. and Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics* **5**, 1–14.

Jiang, J., Sachdev, P., Lipnicki, D. M., Zhang, H., Liu, T., Zhu, W., Suo, C., Zhuang, L., Crawford, J., Reppermund, S. *and others*. (2014). A longitudinal study of brain atrophy over two years in community-dwelling older individuals. *Neuroimage* **86**, 203–211.

Krishnan, A., Williams, L. J., McIntosh, A. R. and Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475.

LeWinn, K. Z., Sheridan, M. A., Keyes, K. M., Hamilton, A. and McLaughlin, K. A. (2017). Sample composition alters associations between age and brain structure. *Nature Communications* **8**, 1–14.

Lin, Z., Yang, C., Zhu, Y., Duchi, J., Fu, Y., Wang, Y., Jiang, B., Zamanighomi, M., Xu, X., Li, M. *and others*. (2016). Simultaneous dimension reduction and adjustment for confounding variation. *Proceedings of the National Academy of Sciences United States of America* **113**, 14662–14667.

Linn, K. A., Gaonkar, B., Doshi, J., Davatzikos, C. and Shinohara, R. T. (2016). Addressing confounding in predictive models with an application to neuroimaging. *The International Journal of Biostatistics* **12**, 31–44.

MAGNIN, B., MESROB, L., KINKINGNÉHUN, S., PÉLÉGRINI-ISSAC, M., COLLIOT, O., SARAZIN, M., DUBOIS, B., LEHÉRICY, S. AND BENALI, H. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* **51**, 73–83.

MAREK, S., TERVO-CLEMMENS, B., CALABRO, F. J., MONTEZ, D. F., KAY, B. P., HATOUM, A. S., DONOHUE, M. R., FORAN, W., MILLER, R. L., HENDRICKSON, T. J. *and others*. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660.

MORE, S., EICKHOFF, S. B., CASPERS, J. AND PATIL, K. R. (2020). Confound removal and normalization in practice: a neuroimaging based sex prediction case study. *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track* **12461**, 3.

MUELLER, S. G., SCHUFF, N. AND WEINER, M. W. (2006). Evaluation of treatment effects in Alzheimer's and other neurodegenerative diseases by MRI and MRS. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo* **19**, 655–668.

MWANGI, B., TIAN, T. S. AND SOARES, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* **12**, 229–244.

OAKES, T. R., FOX, A. S., JOHNSTONE, T., CHUNG, M. K., KALIN, N. AND DAVIDSON, R. J. (2007). Integrating VBM into the general linear model with voxelwise anatomical covariates. *Neuroimage* **34**, 500–508.

POLDRACK, R. A., BAKER, C. I., DURNEZ, J., GORGOLEWSKI, K. J., MATTHEWS, P. M., MUNAFÒ, M. R., NICHOLS, T. E., POLINE, J.-B., VUL, E. AND YARKONI, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience* **18**, 115–126.

RAO, A., MONTEIRO, J. M., MOURAO-MIRANDA, J., ALZHEIMER'S DISEASE INITIATIVE *and others*. (2017). Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage* **150**, 23–49.

SHEHZAD, Z., KELLY, C., REISS, P. T., CRADDOCK, R. C., EMERSON, J. W., MCMAHON, K., COPLAND, D. A., CASTELLANOS, F. X. AND MILHAM, M. P. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *Neuroimage* **93**, 74–94.

SIMPSON, S. L. AND LAURIENTI, P. J. (2015). A two-part mixed-effects modeling framework for analyzing whole-brain network data. *NeuroImage* **113**, 310–319.

SINGH, V., CHERTKOW, H., LERCH, J. P., EVANS, A. C., DORR, A. E. AND KABANI, N. J. (2006). Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain* **129**, 2885–2893.

STRUYFS, H., SIMA, D. M., WITTENS, M., RIBBENS, A., DE BARROS, N. P., VÂN PHAN, T., MEYER, M. I. F., CLAES, L., NIEMANTSVERDRIET, E., ENGELBORGHS, S. *and others*. (2020). Automated MRI volumetry as a diagnostic tool for Alzheimer's disease: validation of icobrain dm. *NeuroImage: Clinical* **26**, 102243.

TABAK, G, FAN, M., YANG, S., HOYER, S. AND DAVIS, G. (2020). Correcting nuisance variation using Wasserstein distance. *PeerJ* **8**, e8594.

THE ADNI TEAM. (2021). *ADNIMERGE: Alzheimer's Disease Neuroimaging Initiative*. R package version 0.0.1. https://adni.bitbucket.io/

WESTMAN, E., SIMMONS, A., ZHANG, Y., MUEHLBOECK, J.-S., TUNNARD, C., LIU, Y., COLLINS, L., EVANS, A., MECOCCI, P., VELLAS, B. *and others*. (2011). Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage* **54**, 1178–1187.

ZHAO, Q., ADELI, E. AND POHL, K. M. (2020). Training confounder-free deep learning models for medical applications. *Nature Communications* **11**, 1–9.

ZHUANG, X., YANG, Z. AND CORDES, D. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping* **41**, 3807–3833.

[*Received January 25, 2022; revised July 12, 2022; accepted for publication July 16, 2022*]