CovBat removes scanner effects in covariance

Removal of Scanner Effects in Covariance Improves Multivariate Pattern Analysis

Andrew A. Chen [1,]
andrewac@pennmedicine.upenn.edu

Joanne C. Beer[1] **Nicholas J. Tustison**[3] **Philip A. Cook**[4] **Russell T. Shinohara**[1,2,a] **Haochang Shou**[1,2,a] **and the Alzheimer's Disease Neuroimaging Initiative**[5]

[1] *Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania*
[2] *Center for Biomedical Image Computing and Analytics, University of Pennsylvania*
[3] *Department of Radiology and Medical Imaging, University of Virginia, Irvine*
[4] *Department of Radiology, University of Pennsylvania*
[5] *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*
[a] *R.T.S. and H.S. contributed equally to this work.*

Introduction

- A growing number of multi-site neuroimaging studies aggregate data across different locations and equipment
- *Scanner effects* arise from differences in scanner manufacturer, model, magnetic field strength, head coil, and more **(Han et al. 2006)**
- Researchers increasingly use multivariate pattern analysis (MVPA), which leverages the covariance between brain features **(O'Toole et al. 2007)**
- MVPA is able to detect scanner with high accuracy **(Glocker et al. 2019)**
- A state-of-the-art harmonization technique ComBat corrects for site differences in mean and variance, but not covariance **(Fortin et al. 2016, 2018)**
- We introduce CovBat (Correcting Covariance Batch Effects) to remove scanner effects in mean, variance, and covariance

Methods

- Using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we examine if scanner effects exist in the covariance of cortical thickness values acquired on the three scanners with the largest number of subjects (see **ADNI data analysis** for more details)
- We conduct an MVPA experiment for prediction of scanner manufacturer labels using unharmonize, ComBat-harmonized, and CovBat-harmonized data
- We also use MVPA assess whether CovBat maintains associations of cortical thicknesses with age and sex

Results

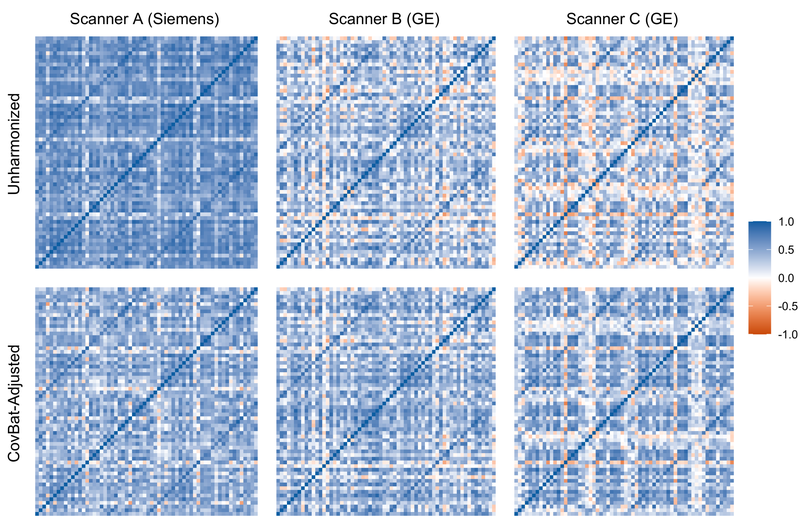

Figure 1: Cortical thickness correlation matrices from three scanners before and after CovBat harmonization. ComBat-adjusted matrices are indistinguishable from unharmonized.

- Striking differences exist between the unharmonized scanner-specific correlation matrices (**Fig. 1**)
- CovBat successfully harmonizes these correlation matrices, but scanner differences still persist (**Fig. 1**)

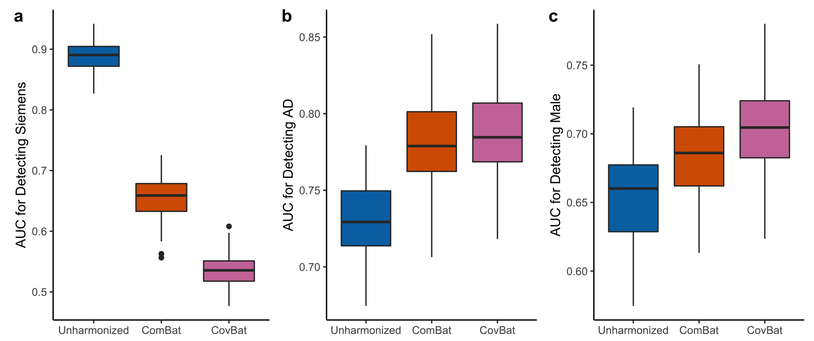

Figure 2: Multivariate pattern analysis experiments for detection of scanner manufacturer, sex, and Alzheimer's disease status using cortical thickness data.

- Siemens scanners are easily identifiable based on unharmonized cortical thickness measurements and still detected after ComBat is applied (**Fig. 2a**)
- **After CovBat, MVPA's performance for differentiating between scanners is close to chance** (**Fig. 2a**)
- **CovBat aids in recovery of associations with Alzheimer's disease status** (**Fig. 2b**) **and sex** (**Fig. 2c**)

Discussion

- Strong scanner effects in covariance exist and could influence downstream MVPA experiments, which remain after performing the state-of-the-art harmonization
- CovBat effectively removes scanner differences in covariance and improves detection of biological associations via MVPA
- Future work could investigate how predictors influence covariance and identify situations that particularly benefit from CovBat harmonization

Methodological Details

CovBat

We first review ComBat **(Fortin et al. 2016; Johnson, Li, and Rabinovic 2007)** for harmonization of neuroimaging measures. Let $\boldsymbol{y}_{ij},\ i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, n_i$ denote the $p \times 1$ vectors of observed data where p is the number of features and $\boldsymbol{x}_{ij}$ denote the vectors of covariates. ComBat estimates covariate effects and finds empirical Bayes point estimates γ^*_{iv} and δ^*_{iv} by pooling information across features then residualizes with respect to these estimates to obtain ComBat-adjusted data

$$y_{ijv}^{ComBat} = \frac{y_{ijv} - \hat{\alpha}_v - \boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}}_v - \gamma^*_{iv}}{\delta^*_{iv}} + \hat{\alpha}_v + \boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}}_v$$

We now propose the CovBat algorithm, which accounts for the joint distribution of ComBat-adjusted observations as follows:

Step 1. We first perform ComBat and additionally residualize with respect to the intercept and covariates to obtain ComBat-adjusted residuals denoted $\boldsymbol{e}_{ij}^{ComBat}$.

Step 2. The $\boldsymbol{e}_{ij}^{ComBat}$ are assumed to have mean 0; their covariance matrices which we denote by Σ_i, however, may differ across scanners. We thus perform principal components analysis (PCA) on the full data residuals and express the ComBat-adjusted residuals as $e_{ij}^{ComBat} = \sum_{k=1}^{q} \xi_{ijk}\boldsymbol{\phi}_k$ where $\boldsymbol{\phi}_k$ are the principal components and ξ_{ijk} are the principal component scores. This model assumes that the covariance scanner effect is contained within the variance of the principal component scores ξ_{ijk}.

Step 3. Thus, we posit:

$$\xi_{ijk} = \mu_{ik} + \rho_{ik}\epsilon_{ijk}$$

where $\epsilon_{ijk} \sim \mathrm{N}(0, \tau_k^2)$ and $\mu_{ik},\ \rho_{ik}$ are the center and scale parameters corresponding to principal components $k = 1, 2, \ldots K$ where $K \leq q$ is a tuning parameter chosen to capture the majority of the variation in the observations. We can then estimate each of the K pairs of center and scale parameters by finding the values that bring each scanner's mean and variance in scores to the pooled mean and variance. We then remove the scanner effect in the scores via

$$\xi_{ijk}^{CovBat} = (\xi_{ijk} - \hat{\mu}_{ik})/\hat{\rho}_{ik}$$

Step 4. We obtain CovBat-adjusted residuals $\boldsymbol{e}_{ij}^{CovBat}$ by projecting the adjusted scores back into the residual space via $\boldsymbol{e}_{ij}^{CovBat} = \sum_{k=1}^{K} \xi_{ijk}^{CovBat}\boldsymbol{\phi}_k + \sum_{l=K+1}^{q} \xi_{ijl}\boldsymbol{\phi}_l$. We then add the intercepts and covariates effects estimated in Step 1 to obtain CovBat-adjusted observations

$$y_{ijv}^{CovBat} = e_{ijv}^{CovBat} + \hat{\alpha}_v + \boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}}_v$$

CovBat is available as an R package and Python script at https://github.com/andy1764/CovBat_Harmonization.

Multivariate pattern analysis experiments

We use a Monte Carlo split-sample experiment where we i) randomly split the subjects into 50% training set and 50% validation set, ii) train a random forests algorithm to detect either scanner manufacturer or a binary clinical covariate, and iii) assess predictive performance on the validation set via AUC. We train separate models for unharmonized, ComBat-harmonized, and CovBat-harmonized data where both harmonization methods are performed including age, sex, and diagnosis status as covariates. We perform steps (i)-(iii) 100 times for each dataset.

ADNI data analysis

All data for this paper are obtained from ADNI and processed using the ANTs longitudinal single-subject template pipeline **(Tustison et al. 2019)**. Code for implementing this pipeline is available on GitHub https://github.com/ntustison/CrossLong.

Subjects are considered to be acquired on the same scanner if they share the same location of scan, scanner manufacturer, scanner model, head coil, and magnetic field strength. Our definition yields 142 distinct scanners of which 78 had less than three subjects and were removed from analyses. The final sample consists of 505 subjects across 64 scanners, with 213 subjects imaged on scanners manufactured by Siemens, 70 by Philips, and 222 by General Electric. The sample has a mean age of 75.3 (SD 6.70) and is comprised of 278 (55%) males, 115 (22.8%) Alzheimer's disease (AD) patients, 239 (47.3%) late mild cognitive impairment (LMCI), and 151 (29.9%) cognitively normal (CN) individuals.

The subsample of the three largest sites consisted of 23 subjects from Scanner A, 20 subjects from Scanner B, and 20 subjects from Scanner C. Scanner A was a Siemens Symphony 1.5T scanner and the other scanners are General Electric Signa Excite 1.5T scanners.

References

Fortin, Jean-Philippe, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, et al. 2018. "Harmonization of Cortical Thickness Measurements Across Scanners and Sites." *NeuroImage* 167 (February): 104–20. https://doi.org/10.1016/j.neuroimage.2017.11.024.

Fortin, Jean-Philippe, Elizabeth M. Sweeney, John Muschelli, Ciprian M. Crainiceanu, and Russell T. Shinohara. 2016. "Removing Inter-Subject Technical Variability in Magnetic Resonance Imaging Studies." *NeuroImage* 132 (May): 198–212. https://doi.org/10.1016/j.neuroimage.2016.02.036.

Glocker, Ben, Robert Robinson, Daniel C. Castro, Qi Dou, and Ender Konukoglu. 2019. "Machine Learning with Multi-Site Imaging Data: An Empirical Study on the Impact of Scanner Effects." *arXiv:1910.04597 [Cs, Eess, Q-Bio]*, October. http://arxiv.org/abs/1910.04597.

Han, Xiao, Jorge Jovicich, David Salat, Andre van der Kouwe, Brian Quinn, Silvester Czanner, Evelina Busa, et al. 2006. "Reliability of MRI-Derived Measurements of Human Cerebral Cortical Thickness: The Effects of Field Strength, Scanner Upgrade and Manufacturer." *NeuroImage* 32 (1): 180–94. https://doi.org/10.1016/j.neuroimage.2006.02.051.

Johnson, W. Evan, Cheng Li, and Ariel Rabinovic. 2007. "Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods." *Biostatistics* 8 (1): 118–27. https://doi.org/10.1093/biostatistics/kxj037.

O'Toole, Alice J., Fang Jiang, Hervé Abdi, Nils Pénard, Joseph P. Dunlop, and Marc A. Parent. 2007. "Theoretical, Statistical, and Practical Perspectives on Pattern-Based Classification Approaches to the Analysis of Functional Neuroimaging Data." *Journal of Cognitive Neuroscience* 19 (11): 1735–52. https://doi.org/10.1162/jocn.2007.19.11.1735.

Tustison, Nicholas J., Andrew J. Holbrook, Brian B. Avants, Jared M. Roberts, Philip A. Cook, Zachariah M. Reagh, Jeffrey T. Duda, et al. 2019. "Longitudinal Mapping of Cortical Thickness Measurements: An Alzheimer's Disease Neuroimaging Initiative-Based Evaluation Study." *Journal of Alzheimer's Disease* 71 (1): 165–83. https://doi.org/10.3233/JAD-190283.