



# HHS Public Access

Author manuscript

*Med Image Comput Assist Interv.* Author manuscript; available in PMC 2017 July 27.

Published in final edited form as:

*Med Image Comput Assist Interv.* 2016 October ; 9900: 282–290. doi:  
10.1007/978-3-319-46720-7\_33.

## Joint Data Harmonization and Group Cardinality Constrained Classification

Yong Zhang<sup>a,\*</sup>, Sang Hyun Park<sup>b,\*</sup>, and Kilian M. Pohl<sup>a,b</sup>

<sup>a</sup>Department of Psychiatry & Behavioral Sciences, Stanford University, USA

<sup>b</sup>Center of Health Sciences, SRI International, USA

### Abstract

To boost the power of classifiers, studies often increase the size of existing samples through the addition of independently collected data sets. Doing so requires harmonizing the data for demographic and acquisition differences based on a control cohort before performing disease specific classification. The initial harmonization often mitigates group differences negatively impacting classification accuracy. To preserve cohort separation, we propose the first model unifying linear regression for data harmonization with a logistic regression for disease classification. Learning to harmonize data is now an adaptive process taking both disease and control data into account. Solutions within that model are confined by group cardinality to reduce the risk of overfitting (via sparsity), to explicitly account for the impact of disease on the interdependency of regions (by grouping them), and to identify disease specific patterns (by enforcing sparsity via the  $l_0$ -‘norm’). We test those solutions in distinguishing HIV-Associated Neurocognitive Disorder from Mild Cognitive Impairment of two independently collected, neuroimage data sets; each contains controls and samples from one disease. Our classifier is impartial to acquisition difference between the data sets while being more accurate in diseases separation than sequential learning of harmonization and classification parameters, and non-sparsity based logistic regressors.

### 1 Introduction

Popular for improving the power of classifiers is to expand application from a single data set (Fig. 1(a)) to multiple, independently collected sets of the same disease (Fig. 1(b)) [1]. To analyze across multiple sets, neuroimage studies generally first harmonize the data by, for example, regressing out demographic factors from MRI measurements and then train the classifier to distinguish disease from control samples [2]. However, harmonization might mitigate group differences making classification difficult (such as in Fig. 2). To improve classification accuracy, we propose the first approach to jointly learn how to harmonize MR image data and classify disease. Harmonization relies on both controls and disease cohorts to reduce differences in the image based measurements due to acquisition differences between sets while preserving the group separation investigated by the classifier. We evaluate our approach on the challenging task (Fig. 1(c)) of being trained on two data sets

\*Equal contribution by Dr. Zhang and Dr. Park

differing in their acquisition and diseases they sample, and tested on accurately distinguishing the two diseases while being impartial to acquisition differences, *i.e.*, the controls of the two sets.

Training of our approach is defined by an energy function that combines a linear harmonization model with a logistic regression classifier. Minimizing this function is confined to group cardinality constrained solutions, *i.e.*, labeling is based on a small number of groups of image scores (counted via the  $l_0$ -‘norm’). Doing so reduces the risk of overfitting, accounts for inter-dependency between regions (*e.g.*, bilateral impact of diseases), and identifies disease distinguishing patterns (defined by non-zero weights of classifiers) that avoid issues of solutions based on relaxed sparsity constraints [3]. Inspired by [4], our method uses block coordinate descent to find the optimal parameters for harmonizing the two data sets and correctly labeling disease samples while being indifferent to control cohorts (*i.e.*, acquisition differences). During testing, we use those parameters to harmonize the image scores before performing classification ensuring that the data set associated with a subject is not part of the labeling decision.

Using 5-fold cross-validation, we measure the accuracy of our method on distinguishing HIV-Associated Neurocognitive Disorder (HAND) from Mild Cognitive Impairment (MCI) of two independently collected data sets; each set contains controls and samples of one disease only. Distinguishing HAND from MCI is clinically challenging due to similarity in symptoms and missing standard protocols for assessing neuropsychological deficits [5]. Not only is our classifier indifferent to the controls from both groups but is also significantly more accurate in distinguishing the two diseases than the conventional sequential harmonization and classification approach and non-sparsity based logistic regression methods.

## 2 Jointly Learning of Harmonization and Classification

Let a data set consists of set  $S_A$  of controls and samples with disease  $A$ , and an independently collected set  $S_B$  of controls and samples of disease  $B$ . The four subsets are matched with respect to demographic scores, such as age. Each sample  $s$  of the joint set is represented by a vector of image features  $\mathbf{x}_s$  and a label  $y_s$ , where  $y_s = -1$  if  $s \in S_A$  and  $y_s = +1$  for  $s \in S_B$ . The acquisition differences between  $S_A$  and  $S_B$  are assumed to linearly impact the image features. To extract disease separating patterns from this joint set, we review the training of a sequential model for data harmonization and classification, and then propose to simultaneously parameterize both tasks by minimizing a single energy function.

### 2.1 Sequential Harmonization and Classification

Training the linear regression model for data harmonization results in parameterizing matrix  $U$  so that it minimizes the difference (with respect to the  $l_2$ -norm  $\|\cdot\|_2$ ) between the inferred values  $U \cdot [1 \ y_s]^T$  (1 is the bias term of the linear model) and the raw image scores  $\mathbf{x}_s$  across  $n_C$  controls of the joint data set, *i.e.*,

$$\hat{U} := \arg \min_U h_C(U) \text{ with } h_C(U) := \frac{1}{n_C} \sum_{s=1}^{n_C} \|U \cdot [1 \ y_s]^\top - \mathbf{x}_s\|_2^2. \quad (1)$$

A group sparsity constrained logistic regression classifier is trained on the residuals  $\mathbf{r}_s(\hat{U}) := \hat{U} \cdot [1 \ y_s]^\top - \mathbf{x}_s$ , *i.e.*, harmonized scores, across samples ‘ $s$ ’ of two groups, which are  $n_D$  samples of the two disease cohorts. Note, classification based on the inferred values  $\hat{U} \cdot [1 \ y_s]^\top$  is uninformative as all scores of a region are mapped to two values. Now, let the logistic function be  $\theta(a) := \log(1 + \exp(-a))$ , the weight vector  $\mathbf{w}$  encode the importance of each score of  $\mathbf{r}_s$  and  $v \in \mathbb{R}$  be the label offset, then the penalty (or label) function of the classifier is defined by

$$l_D(\hat{U}, v, \mathbf{w}) := \frac{1}{n_D} \sum_{s=n_C+1}^{n_C+n_D} \theta \left( \mathbf{y}_s \cdot \left( \mathbf{w}^\top \cdot \mathbf{r}_s(\hat{U}) + v \right) \right). \quad (2)$$

The search for point  $(\hat{v}, \hat{\mathbf{w}})$  minimizing  $l_D(\hat{U}, \cdot, \cdot)$  often has to be limited so that  $\hat{\mathbf{w}}$  conforms to disease specific constraints, such as the bilateral impact of HAND or MCI on individual regions. These constraints can be modeled by group cardinality. Specifically, every entry of  $\mathbf{w}$  is assigned to one of  $n_G$  groups  $\mathbf{g}_1(\mathbf{w}), \dots, \mathbf{g}_{n_G}(\mathbf{w})$ . The search space  $\mathbb{S}$  is composed of weights where the number of groups with non-zero weight entries are below a certain threshold  $k \in \mathbb{N}$ , *i.e.*,  $\mathbb{S} := \{\mathbf{w} : \sum_{i=1}^{n_G} \|\mathbf{g}_i(\mathbf{w})\|_2 \|0\|_0 \leq k\}$  with  $\|\cdot\|_0$  being the  $l_0$ -‘norm’. Finally, the training of the classifier is fully described by the following minimization problem

$$(\hat{v}, \hat{\mathbf{w}}) := \arg \min_{v \in \mathbb{R}, \mathbf{w} \in \mathbb{S}} l_D(\hat{U}, v, \mathbf{w}), \quad (3)$$

which can be solved via penalty decomposition [4, 6]. Note, that Eq. (3) turns into a sparsity constrained problem if each group  $\mathbf{g}_i$  is of size 1. Furthermore, setting ‘ $k = n_G$ ’ turns Eq. (3) into a logistic regression problem. Finally, Eq. (3) can distinguish a single disease from controls by simply replacing  $y_s$  in Eq. (2) with a variable encoding assignment to cohorts instead of data sets.

## 2.2 Simultaneous Harmonization and Classification

We now determine  $(\hat{U}, \hat{v}, \hat{\mathbf{w}})$  for a single minimization problem composed of the linear (harmonization) and logistic (classification) regression terms, *i.e.*,

$$(\hat{U}, \hat{v}, \hat{\mathbf{w}}) := \arg \min_{U, v, \mathbf{w} \in \mathbb{S}} (1 - \lambda) \cdot l_D(U, v, \mathbf{w}) + \lambda \cdot h_C(U), \quad (4)$$

### Algorithm 1

#### Joint Harmonization and Classification.

- 
- 1: Set  $\rho = 0.1, \sigma = \sqrt{10}, \varepsilon_B = \varepsilon_P = 10^{-3}$  (according to [4]) and  $U' = 0, \mathbf{w}' = 0, \mathbf{q}' = 0, v' = 1$
  - 2: **Repeat**
  - 3:   **Repeat** (Block Coordinate Descent)
  - 4:      $U'' \leftarrow U', \mathbf{w}'' \leftarrow \mathbf{w}', \mathbf{q}'' \leftarrow \mathbf{q}', v'' \leftarrow v'$
  - 5:      $U' \leftarrow \operatorname{argmin}_U (1 - \lambda) l_D(U, v', \mathbf{q}') + \lambda h_C(U)$  (via Gradient Descent)
  - 6:      $(v', \mathbf{q}') \leftarrow \operatorname{argmin}_{v, \mathbf{q}} (1 - \lambda) l_D(U', v, \mathbf{q}) + \rho \|\mathbf{w}' - \mathbf{q}\|_2^2$  (via Gradient Descent)
  - 7:   Update  $\mathbf{w}'$  by
   
    Sort groups  $\mathbf{g}_j(\mathbf{q}')$  so that  $\|\mathbf{g}_{j_1}(\mathbf{q}')\|_2 \geq \|\mathbf{g}_{j_2}(\mathbf{q}')\|_2 \geq \dots \geq \|\mathbf{g}_{j_{n_G}}(\mathbf{q}')\|_2$ 
  
    Define  $\mathbb{I}$  to be the set of indices of  $\mathbf{q}'$  associated with groups  $\mathbf{g}_{j_1}, \dots, \mathbf{g}_{j_k}$   $\mathbf{w}'_i \leftarrow \mathbf{q}'_i$  if  $\mathbf{q}'_i \in \mathbb{I}$  and 0 otherwise.
  - 8:   **Until**  $\max \left\{ \frac{\|U' - U''\|_{\max}}{\max \{\|U'\|_{\max}, 1\}}, \frac{|v' - v''|}{\max \{|v'|, 1\}}, \frac{\|\mathbf{w}' - \mathbf{w}''\|_{\max}}{\max \{\|\mathbf{w}'\|_{\max}, 1\}}, \frac{\|\mathbf{q}' - \mathbf{q}''\|_{\max}}{\max \{\|\mathbf{q}'\|_{\max}, 1\}} \right\} < \varepsilon_B,$
  - 9:    $\rho \leftarrow \sigma \cdot \rho$
  - 10: **Until**  $\|\mathbf{w}' - \mathbf{q}'\|_{\max} < \varepsilon_P$
- 

where  $\lambda \in (0, 1)$ . Note, the model fails to classify when  $\lambda = 1$  ( $\hat{v}$  and  $\hat{\mathbf{w}}$  are undefined) or harmonize when  $\lambda = 0$  (entries of  $\hat{U}$  are undefined). Motivated by [4], we simplify optimization by first parameterizing the classifier with respect to the ‘unconstrained’ vector  $\mathbf{q}$  before determining the corresponding sparse solution  $\hat{\mathbf{w}}$ . The solution to Eq. (4) is estimated by iteratively increasing  $\rho$  of

$$(U_\rho, v_\rho, \mathbf{w}_\rho, \mathbf{q}_\rho) := \operatorname{argmin}_{U, v, \mathbf{w} \in \mathbb{S}, \mathbf{q}} (1 - \lambda) \cdot l_D(U, v, \mathbf{q}) + \lambda \cdot h_C(\mathbf{U}) + \rho \cdot \|\mathbf{w} - \mathbf{q}\|_2^2 \quad (5)$$

until the maximum of the absolute difference between elements of  $\mathbf{w}_\rho$  and  $\mathbf{q}_\rho$  is below a threshold, *i.e.*,  $\|\mathbf{w}_\rho - \mathbf{q}_\rho\|_{\max} < \varepsilon_P$ .  $(U_\rho, v_\rho, \mathbf{w}_\rho, \mathbf{q}_\rho)$  are determined by block coordinate descent (BCD). As outlined in Alg. 1, let  $(U', v', \mathbf{w}', \mathbf{q}')$  be estimates of  $(U_\rho, v_\rho, \mathbf{w}_\rho, \mathbf{q}_\rho)$ , then  $U'$  is updated by solving Eq. (5) with fixed  $(v', \mathbf{w}', \mathbf{q}')$ :

$$U' := \operatorname{argmin}_U (1 - \lambda) \cdot l_D(U, v', \mathbf{q}') + \lambda \cdot h_C(\mathbf{U}). \quad (6)$$

As this minimization is over a convex and smooth function, Eq. (6) is solved via gradient descent. Note, that determining  $U'$  is equivalent to increasing the separation between the two disease groups by minimizing  $l_D(\cdot, v', \mathbf{q}')$  while reducing the difference between the two control groups by minimizing  $h_C(\cdot)$ .

Next, BCD updates  $\mathbf{v}'$  and  $\mathbf{q}'$  by keeping  $(U', \mathbf{w}')$  fixed in Eq. (5), *i.e.*,

$$(\mathbf{v}', \mathbf{q}') := \arg \min_{\mathbf{v}, \mathbf{q}} (1 - \lambda) \cdot l_D(U', \mathbf{v}, \mathbf{q}) + \varrho \cdot \|\mathbf{w}' - \mathbf{q}\|_2^2. \quad (7)$$

This minimization problem is defined by a smooth and convex function. Its solution is thus also estimated via gradient descent. Finally,  $\mathbf{w}'$  is updated by solving Eq. (5) with fixed  $(U', \mathbf{v}', \mathbf{q}')$ , *i.e.*,

$$\mathbf{w}' := \arg \min_{\mathbf{w} \in \mathcal{S}} \|\mathbf{w} - \mathbf{q}'\|_2^2. \quad (8)$$

As shown in [4, 6], the closed form solution of Eq. (8) first computes  $\|\mathbf{g}_i(\mathbf{q}')\|_2$  for each group  $i$  and then sets  $\mathbf{w}'$  to the entries of  $\mathbf{q}'$ , who are assigned to the  $k$  groups with the highest norms. The remaining entries of  $\mathbf{w}'$  are set to 0. The procedures (6)~(8) are repeated until the relative changes of  $(U', \mathbf{v}', \mathbf{w}', \mathbf{q}')$  between iterations are smaller than a threshold  $\epsilon_B$ .  $(U_{\varrho}, \mathbf{v}_{\varrho}, \mathbf{w}_{\varrho}, \mathbf{q}_{\varrho})$  is updated with the converged  $(U', \mathbf{v}', \mathbf{w}', \mathbf{q}')$ ,  $\varrho$  is increased and another BCD loop is initiated until  $\mathbf{w}_{\varrho}$  and  $\mathbf{q}_{\varrho}$  converge towards each other (see Alg. 1 for details).

Fig. 2 showcases the differences between sequential and joint harmonization and classification. Two synthetic data sets consist of a control and disease cohort, where the raw scores for each cohort are 20 random samples from a Gaussian distribution with the covariance being the identity matrix multiplied by 0.01. The mean of the Gaussian for Disease A of Set I (blue) is (1.3,2) resulting in samples that are somewhat separated from those of Disease B of Set II (mean=(1.5,2), red). The difference in data acquisition between the two sets is simulated by an even larger separation of the means between the two control groups (Set I: mean=(0.9,1), green; Set B: mean=(1.2,1), black). The *Sequential* method (see Sec. 2.1 without sparsity) harmonizes the scores so that the classifier assigns the controls to one set, *i.e.*, the separating plane (black line) is impartial to acquisition differences. This plane fails to perfectly separate the two disease cohorts as the cohorts are ‘pushed’ together with the mean of Disease B being now to the right of the mean of Disease A. Higher accuracy in disease classification is achieved by our joint model (omitting sparsity) with  $\lambda = 0.8$ . Comparing this plot to the results with  $\lambda = 0.5$  shows that as  $\lambda$  decreases the emphasis on separating the two disease increases as intended by Eq. (6). The classifier is still impartial to acquisition differences and perfectly labels the samples of the two disease cohort. In summary, the joint model enables data harmonization that preserves group differences, which was not the case for the sequential approach.

### 3 Distinguishing HAND from MCI

We tested our method on a joint set of two independently collected data sets: the ‘UHES’ set [7] contained cross-sectional MRIs of 15 HAND cases and 21 controls while the ‘ADNI’ set contained MRIs of 20 MCIs and 18 controls. The 4 groups were matched with respect to sex and age. Each MRI was acquired on a 3T Siemens scanner (with the two sets having

different acquisition protocols) and was processed by [8] resulting in 102 regional volume scores. We assigned those scores to ( $n_G = 52$ ) groups to account for the bilateral impact of MCI and HAND on both hemispheres. We now review our experimental setup and results highlighting that jointly learning harmonization and classification results in findings that are indifferent to the two control cohorts and more accurate in distinguishing the two disease groups compared to alternative learning models.

To measure the accuracy of our method, we used 5 fold-cross validation with each fold containing roughly 20% of each cohort. On the training sets, we used parameter exploration to determine the optimal group cardinality  $k \in \{1, 2, \dots, 10\}$  and weighing  $\lambda \in \{0.1, 0.2, \dots, 0.9\}$ . Across all folds and settings, our algorithm converged within 5 iterations while each BCD optimization converged within 500 iterations. For each setting, we then computed the disease accuracy of the classifier by separately computing the accuracy for each cohort (MCI and HAND) to be assigned to the right set (UHES or ADNI) and then averaging the scores to consider the imbalance of cohort sizes. We determined the control accuracy repeating those computations with respect to the two control groups. Since higher control accuracy infers worse harmonization, an *unbiased* control accuracy, *i.e.*, around 50%, coupled with a high disease accuracy was viewed as preferable. Note, an unbiased control accuracy only qualifies the harmonization in that the remaining acquisition differences do not impact the separating plane (or weights  $\hat{w}$ ) of the classifier. For each training set, we then chose the setting  $(\lambda, k)$  with corresponding weights  $\hat{w}$  and harmonization parameters  $\hat{U}$  that produced the largest difference between the two accuracy scores. This criteria selected a unique setting for 2 folds, 2 settings for 2 folds, and 3 settings for 1 fold.

On the test set, we computed the harmonized scores (residuals) of all samples for each selected setting. We then hid the subjects' data set associations from the ensemble of selected classifiers by applying the residuals to classifiers parameterized by the corresponding weight settings. Based on those results, we computed control and disease accuracy as well as the corresponding  $p$ -values via the Fisher exact test [9]. We viewed  $p < 0.05$  as significantly more accurate than randomly assigning samples to the two sets. An ensemble of classifiers was viewed as informative, if separating HAND from MCI resulted in a significant  $p$ -value and a non-significant one for controls.

We repeated the above experiment for the sequential harmonization and classification approach of Section 2.1, called **SeqGroup**, to show the advantages of our joint learning approach, called **JointGroup**. To generalize our findings about joint learning of harmonization and classification parameters to non-sparsity constrained models, we also tested the approach omitting the group sparsity constraint, *i.e.*,  $k = 52$ . **JointNoGroup** refers to the results of the corresponding joint method and **SeqNoGroup** to the sequential results.

Table 1 shows the accuracy scores and  $p$ -values of all implementations listed according to the difference between disease and control accuracy. The method with the smallest difference, SeqNoGroup, was the only approach recording a significant control accuracy score (68.4%,  $p = 0.024$ ). The corresponding separating plane was thus not impartial to acquisition difference so that the relatively high disease accuracy (82.5%) was insignificant. The next approach, SegGroup, obtained non-significant accuracy scores for controls (64.6%)

indicating that the group sparsity constraint improved generalizing the results from the relatively small training set to testing. The difference between control (52.8%) and disease accuracy (83.3%) almost doubled (30.5%) for the joint approach JointNoGroup. As reflected in Fig. 2, learning data harmonization separated from classification does not generalize as well as the joint approach, who harmonizes data so that differences between disease groups are preserved. Confirming all previous findings, the overall best approach (*i.e.*, control and disease accuracy differ most) is our proposed JointGroup approach achieving a disease accuracy of 90%.

The group sparsity constraint aided in separating diseases and identified patterns of regions (*i.e.*, non-zero weights  $\hat{w}$ ) impacted by either MCI or HAND (or HIV). Each column of Fig. 3 shows the largest, unique pattern associated with a training set. For those training sets that selected multiple patterns (*i.e.*,  $\hat{w}$  settings), patterns with less regions were always included in the largest pattern. The precentral gyrus, cerebellum VIII, and lateral ventricle were parts of all patterns. HIV is known to impact the cerebellum [10] and accelerated enlargement of ventricles is linked to both HIV [11] and MCI [12]. These findings indicate that the extracted patterns are informative with respect to MCI and HAND (and HIV), which requires an in-depth morphemic analysis for confirmation.

## 4 Conclusion

We proposed an approach that simultaneously learned the parameters for data harmonization and disease classification. The search for the optimal separating plane was confined by group cardinality to reduce the risk of overfitting, to explicitly account for the impact of disease on the inter-dependency of regions, and to identify disease specific patterns. On separating HAND from MCI samples of two disease specific data sets, our joint approach achieved better classification accuracy than the non-sparsity based model and sequential approaches.

## Acknowledgments

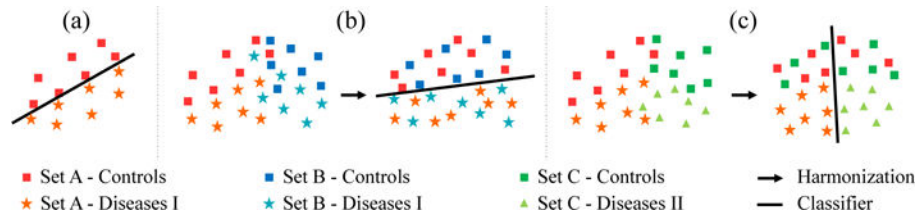
This research was supported in part by the NIH grants U01 AA017347, AA010723, K05-AA017168, K23-AG032872, and P30 AI027767. We thank Dr. Valcour for giving us access to the UHES data set. With respect to the ADNI data, collection and sharing for this project was funded by the NIH Grant U01 AG024904 and DOD Grant W81XWH-12-2-0012. Please see [https://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_DSP\\_Policy.pdf](https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf) for further details.

## References

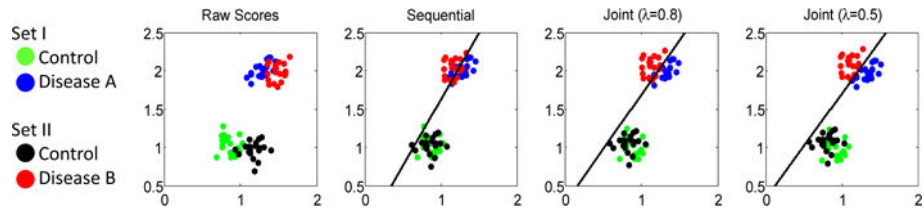
1. Jovicich J, et al. Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. *NI*. 2014; 101:390–403.
2. Moradi E, et al. Predicting symptom severity in autism spectrum disorder based on cortical thickness measures in agglomerative data. *bioRxiv*. 2016
3. Sabuncu M. A universal and efficient method to compute maps from image-based prediction models. *MICCAI*. 2014
4. Zhang Y, et al. Computing group cardinality constraint solutions for logistic regression problems. *Medical Image Analysis*. 2016 In Press.
5. Sanmarti M, et al. HIV-associated neurocognitive disorders. *JMP*. 2014; 2(2)
6. Lu Z, Zhang Y. Sparse approximation via penalty decomposition methods. *SIAM J Optim*. 2013; 23(4):2448–2478.

7. Nir TM, et al. Mapping White Matter Integrity in Elderly People with HIV. *Human Brain Mapping*. 2014; 35(3):975–992. [PubMed: 23362139]
8. Pfefferbaum A, et al. Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85 years) measured with atlas-based parcellation of MRI. *NI*. 2013; 65:176–193.
9. Fisher R. The logic of inductive inference. *J Roy Stat Soc*. 1935; 1(98):38–54.
10. Chang L, et al. Impact of apolipoprotein E  $\epsilon$ 4 and HIV on cognition and brain atrophy: antagonistic pleiotropy and premature brain aging. *NI*. 2011; 4(58)
11. Thompson PM, et al. 3D mapping of ventricular and corpus callosum abnormalities in HIV/AIDS. *NI*. 2006; 31(1):12–23.
12. Nestor SM, et al. Ventricular enlargement as a possible measure of alzheimer’s disease progression validated using the alzheimer’s disease neuroimaging initiative database. *Brain*. 2008; 131(9): 2443–2454. [PubMed: 18669512]



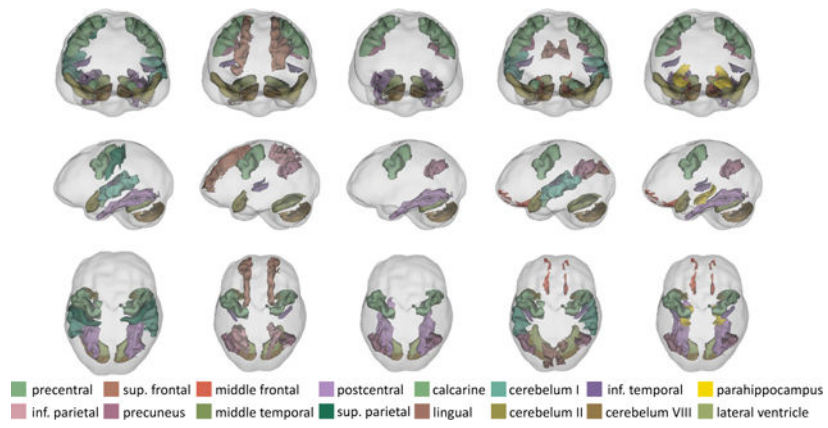


**Fig. 1.** Examples for classifying data: (a) Classification based on a single set, (b) separating healthy and disease based on multiple sets requiring harmonization, and (c) part two disease groups based on harmonizing controls of disease specific data sets.



**Fig. 2.**

Impact of harmonization on classification (black line) of two synthetic sets. Compared to the raw scores, the *Sequential* approach mitigated differences between the two disease groups by ‘pushing’ them together. Our joint model with  $\lambda = 0.5$  is the only approach that is indifferent to controls and correctly labels all disease cases.



**Fig. 3.** Each column shows the largest, unique pattern extracted by JointGroup on one of the 5 training sets. Identified regions are impacted by HAND, HIV, or MCI.

**Table 1**

Implementations listed according to the difference between disease (MCI vs. HAND) and control (UHES controls vs. ADNI controls) accuracy. In bold are the findings that are statistically significant different from chance, *i.e.*,  $p < 0.05$ . Our method, JointGroup, achieves the highest disease accuracy and largest spread in scores.

|              | Classification accuracy |         |            | <i>p</i> -value with ground truth |              |
|--------------|-------------------------|---------|------------|-----------------------------------|--------------|
|              | Disease                 | Control | Difference | Disease                           | Control      |
| SeqNoGroup   | 82.5                    | 68.4    | 14.1       | < <b>0.001</b>                    | <b>0.024</b> |
| SeqGroup     | 80.0                    | 64.6    | 15.4       | < <b>0.001</b>                    | 0.073        |
| JointNoGroup | 83.3                    | 52.8    | 30.5       | < <b>0.001</b>                    | 0.753        |
| JointGroup   | 90.0                    | 52.8    | 37.2       | < <b>0.001</b>                    | 0.753        |