Fed-ComBat: A Generalized Federated Framework for Batch Effect Harmonization in Collaborative Studies

Santiago Silva[a,], Neil Oxtoby[b], Andre Altmann[b], Marco Lorenzi[a]

[a] *Université Côte d'Azur and INRIA, EPIONE team, Sophia Antipolis, France*
[b] *Centre for Medical Image Computing, University College London, London, United Kingdom*

Abstract

In neuroimaging research, the utilization of multi-centric analyses is crucial for obtaining sufficient sample sizes and representative clinical populations. Data harmonization techniques are typically part of the pipeline in multi-centric studies to address systematic biases and ensure the comparability of the data. However, most multi-centric studies require centralized data, which may result in exposing individual patient information. This poses a significant challenge in data governance, leading to the implementation of regulations such as the GDPR and the CCPA, which attempt to address these concerns but also hinder data access for researchers. Federated learning offers a privacy-preserving alternative approach in machine learning, enabling models to be collaboratively trained on decentralized data without the need for data centralization or sharing.

In this paper, we present Fed-ComBat, a federated framework for batch effect harmonization on decentralized data. Fed-ComBat extends existing centralized linear methods, such as ComBat and distributed as d-ComBat,

Email address: santiago-smith.silva-rincon@inria.fr (Santiago Silva)

and nonlinear approaches like ComBat-GAM in accounting for potentially nonlinear and multivariate covariate effects. By doing so, Fed-ComBat enables the preservation of nonlinear covariate effects without requiring centralization of data and without prior knowledge of which variables should be considered nonlinear or their interactions, differentiating it from ComBat-GAM. We assessed Fed-ComBat and existing approaches on simulated data and multiple cohorts comprising healthy controls (CN) and subjects with various disorders such as Parkinson's disease (PD), Alzheimer's disease (AD), and autism spectrum disorder (ASD).

Results indicate that Fed-ComBat outperforms centralized ComBat in the presence of nonlinear effects and is comparable to centralized methods such as ComBat-GAM. Using synthetic data, Fed-ComBat is able to better reconstruct the target unbiased function by 35% (RMSE = 0.5952) with respect to d-ComBat (RMSE = 0.9162) and 12% with respect to our proposal to federate ComBat-GAM, d-ComBat-GAM (RMSE= 0.6751) and exhibits comparable results on MRI-derived phenotypes to centralized methods as ComBat-GAM without the need of prior knowledge on potential nonlinearities.

Keywords: Harmonization, federated learning, medical imaging, multicentric data, batch effects, data privacy, ComBat

---

1. Introduction

With the vast generation of neuroimaging data across multiple institutions, concerns seeking to protect sensitive data have been issued as the General Data Protection Regulation - GDPR and the California Consumer

Privacy Act - CCPA (European Commission, 2016; Bukaty, 2019; Calzada, 2022). Such policies demand changes on the paradigms for multicentric clinical studies which nowadays mostly rely on data centralization.

Additionally, data harmonization is known to be a crucial factor in tackling certain challenges of multicentric data analysis due to systematic errors or *batch effects* commonly present in such studies (Wachinger et al., 2021). Harmonization aims to address biases and variations among data collected from different sources (e.g. scanner brands or institutions), ensuring that the data is comparable and can be combined effectively for analysis. Johnson et al. (2007) proposed ComBat in the field of gene expression analysis, as a way to correct these batch effects while preserving the desired covariate effects (e.g., sex, diagnosis, age). Later, this method was adapted to brain magnetic resonance imaging (MRI) phenotypes (Fortin et al., 2017, 2018). However, the typical use of ComBat in multicentric studies still requires data to be centralized, thus limiting the application to real-life collaborative analysis scenarios. To overcome this problem, Chen et al. (2022) proposed d-ComBat to harmonize data within a distributed setup by extending the optimization of ComBat to allow bias correction without the need to exchange the data across centers. To achieve this, d-ComBat shares full local covariance matrices computed at each site, and the subsequently aggregates these covariance matrices to estimate the effect parameters at a global level. This method was shown to achieve almost an exact solution with respect to its centralized version requiring only a few rounds of communication. However, sharing full covariance matrices represents a potential risk of sensitive data leakage (van Kesteren et al., 2019), and like ComBat, d-ComBat con-

siders that covariates have a linear influence on phenotypes, which may not always be the case as reported by Bethlehem et al. (2022) where age shows a clear nonlinear influence on MRI-derived phenotypes. Afterwards, Pomponio et al. (2020) introduced ComBat-GAM. A method that relies on generalized additive models (GAMs) to address the issue of preserving nonlinear covariate effects. ComBat-GAM uses smoothing functions, such as polynomial or splines, to decompose covariate effects into smooth basis functions, thereby approximating such nonlinear influences. Although ComBat-GAM demonstrates better performance than ComBat, it has limitations such as being designed only for centralized settings. Moreover, ComBat-GAM requires an explicit definition of covariate interactions, typically limited to additive or multiplicative terms (e.g., Sex $\times$ Diagnosis). Despite efforts, a gap still exists in the availability of methods for nonlinear covariate effect preservation without requiring data centralization.

Federated learning (FL) is a popular machine learning paradigm to securely optimize models on decentralized data while preserving institutions' data governance and individuals' sensitive data (Konečnỳ et al., 2016). Rather than requiring data to be centralized, FL allows institutions to collaboratively train a model keeping data on-site, and sharing only model updates, aggregating these updates on a central server, and distributing the updated global model back to each participant. This iterative process continues until convergence is achieved. Federated Averaging (FedAvg) is a popular approach for aggregation in federated learning. It adapts the stochastic gradient descent (SGD) algorithm for the federated setting providing convergence guarantees (McMahan et al., 2017). FL has facilitated access to larger cohorts and it has

successfully been applied in neuroimaging for tumor segmentation and diagnosis (Li et al., 2019a; Mahlool and Abed, 2022), disease studies in functional magnetic resonance imaging (fMRI) (Li et al., 2020b), intracranial hemorrhage (ICH) detection(Cheung et al., 2023) and continues to be increasingly adopted in neuroimaging applications.

Within the context of federated learning (FL) and data harmonization, an opportunity opens for the development of flexible and effective federated harmonization approaches that can be easily integrated into FL pipelines, enabling harmonization of heterogeneous data across sites while accounting for complex covariate effects on heterogeneously distributed data. Such approaches have the potential to fill the current gap and allow for more comprehensive analysis of distributed cohorts using FL. To tackle the mention challenges, our contributions are considered as follows: i) we introduce a formulation of ComBat which generalizes covariate effect modeling allowing the introduction of more complex functions and, ii) we develop Fed-ComBat, a generalized federated ComBat framework, that allows data batch effect harmonization in federated settings.

We benchmarked and compared our proposed methods with existing centralized approaches (ComBat, and ComBat-GAM) and distributed (d-ComBat) (Johnson et al., 2007; Pomponio et al., 2020; Chen et al., 2022). Moreover, for the sake of comparison we also formulated and implemented a distributed version of ComBat-GAM (d-ComBat-GAM), which relies on distributed covariance estimation, and extend the method of Chen et al. (2022) by allowing basis decomposition to interpolate nonlinear covariate effects using smoothing functions.

The different methods were compared for their ability to harmonize batch effects and preserve the quality of covariate effects on simulated data (Section 4). Furthermore, we performed an evaluation on derived phenotypes from MRI-brain images from nine cohorts (A4, ABIDE-I, ABIDE-II, ADNI, AIBL, MIRIAD, OASIS3, PDPB, PPMI) comprising controls and people with different brain disorders: patients with different subtypes of Parkinson's disease (PD), Alzheimer's disease (AD) and Autism spectrum disorder (ASD).

Using synthetic data, we find that Fed-ComBat achieves a 35% improvement in reconstructing the target unbiased function compared to d-ComBat (RMSE = 0.9162), and a 12% improvement compared to our federated extension of ComBat-GAM, d-ComBat-GAM (RMSE = 0.6751). In addition, our application on real data shows that Fed-ComBat on MRI-derived phenotypes is comparable to centralized ComBat-GAM. This implies that the studied cohorts may not have exhibited complex nonlinear effects or interactions that could not be captured by a GAM. However, Fed-ComBat requires no prior assumptions of nonlinearities unlike ComBat-GAM and automatically captures such effects and interactions through the model. Moreover, Fed-ComBat can be considered a better approach towards privacy preserving methods as it only shares model parameters, while d-ComBat and our included extension d-ComBat-GAM share full covariance and cross-covariance matrices (van Kesteren et al., 2019).

2. Methods

2.1. Generalized ComBat model

Following the original ComBat formulation proposed by Johnson et al. (2007), let us denote a batch (represented by different scanners protocols, machines, or institutions) indexed by $i \in \{1, 2, ..., S\}$ on a particular phenotype (e.g., a brain region) indexed by $g \in \{1, 2, ..., G\}$. Each batch contains n_i number of observations, and the total number of observations is $N = \sum_i^S n_i$. S can denote for simplicity the number of sites in the study, but it can also be extended to the total number of scanners between sites or any other number of batch effects. We can model a specific phenotype g observed in the j-th patient who belongs to the i-th site denoted by y_{ijg} as follows:

$$y_{ijg} = \alpha_g + \phi(\mathbf{x}_{ij}, \boldsymbol{\theta}_g) + \gamma_{ig} + \delta_{ig}\varepsilon_{ijg}, \tag{1}$$

where $\mathbf{x}_{ij}$ denotes the covariate effects expected to be preserved after removing the batch effects (e.g., sex and age), α_g acts as a global fixed intercept (i.e., the mean), while γ_{ig} indicates a random intercept that accounts for the site-specific shift. ε_{ijg} is a noise model that captures the variability of each phenotype $\varepsilon_{ijg} \sim \mathcal{N}(0, \sigma_g^2)$, and δ_{ig} is a multiplicative effect that scales the "unbiased" phenotype variability to fit the one at each site.

This formulation generalizes the original linear model proposed by Johnson et al. (2007) and the nonlinear covariate effect model proposed by Pomponio et al. (2020). While the original model considered $\phi(\mathbf{x}; \boldsymbol{\theta}_g)$ to be a linear function of $\mathbf{x}$, and the nonlinear model considered $\phi(\mathbf{x}; \boldsymbol{\theta}_g)$ to be an univariate spline, this formulation treats $\phi(\mathbf{x}; \boldsymbol{\theta}_g)$ as a general function that can potentially be multivariable and nonlinear and parametrized by $\boldsymbol{\theta}_g$. This

generalization allows for the use of more complex approximating functions such as kernel-based models or neural networks as the multilayer perceptron (MLP), which is known to be an universal function approximator. However, it is essential to use cross-validation techniques to avoid overfitting and ensure the model generalizes well to unseen data.

Some constraints are defined during the fixed effect parameter estimation $(\hat{\alpha}_g, \hat{\boldsymbol{\theta}}_g, \hat{\gamma}_{ig}, \hat{\sigma}_g)$:

$$\arg \max_{\hat{\alpha}_g, \hat{\boldsymbol{\theta}}_g, \hat{\gamma}_{ig}, \hat{\sigma}_g} P(y_{ijg} | \hat{\alpha}_g, \hat{\boldsymbol{\theta}}_g, \hat{\gamma}_{ig}, \hat{\sigma}_g) \tag{2}$$

$$\text{subject to} \quad \mathbb{E}_g[\hat{\gamma}_i] = \sum_i^S \frac{n_i}{N} \hat{\gamma}_{ig} = 0, \, \forall g \in \{1, ..., G\} \tag{3}$$

$$\text{and} \quad \phi(\mathbf{x}, \boldsymbol{\theta}_g)|_{x=0} = 0 \tag{4}$$

A first constraint in Equation (3) is set to allow identifiability of the intercept parameters as explained by Johnson et al. (2007) (explained in practice in Appendix S4). However, in the same regard, the constraint in Equation (4) ensures that no batch effects are captured by the covariate function $\phi(\cdot)$. This second constraint, despite not being mentioned, is also fulfilled by Johnson et al. (2007) and Pomponio et al. (2020); making ComBat and ComBat-GAM a particular case of the proposed formulation in this work.

For a centralized setup, the estimation of all these parameters is performed in three steps: i) maximum likelihood estimation (MLE) for parameters $\hat{\alpha}_g, \hat{\boldsymbol{\theta}}_g, \hat{\gamma}_{ig}$ (see Equation (2)) and of the phenotype variance $\hat{\sigma}_g^2 = \frac{1}{N} \sum_{ij} \left(y_{ijg} - \hat{\alpha}_g - \phi(\mathbf{x}_{ij}; \hat{\boldsymbol{\theta}}_g) - \hat{\gamma}_{ig} \right)^2$, ii) residual standardization mapping the

residuals to satisfy the form $y_{ijg} \rightarrow z_{ijg} \sim \mathcal{N}(\gamma_{ig}, \delta^2_{ig})$ as follows:

$$z_{ijg} = \frac{y_{ijg} - \hat{\alpha}_g - \phi(\mathbf{x}_{ij}; \hat{\boldsymbol{\theta}}_g)}{\hat{\sigma}_g} \tag{5}$$

and iii) estimation of the additive and multiplicative batch effects $\hat{\gamma}^*_{ig}$ and $\hat{\delta}^*_{ig}$ as in Equation (6), using empirical Bayes (EB) with priors on γ_{ig} and δ^2_{ig} to iteratively estimate these parameters as in Equation (7) (Johnson et al., 2007, Sec. 3.2).

$$\gamma_{ig} \sim \mathcal{N}(Y_i, \tau_i^2) \quad \text{and} \quad \delta^2_{ig} \sim \text{Inverse Gamma}(\lambda_i, \vartheta_i) \tag{6}$$

$$\gamma^*_{ig} = \frac{n_i \bar{\tau}_i^2 \hat{\gamma}_{ig} + \delta^{2*}_{ig} \bar{\gamma}_{ig}}{n_i \bar{\tau}_i^2 + \delta^{2*}_{ig}}, \; \delta^*_{ig} = \frac{\bar{\theta}_i + \frac{1}{2} \sum_{j=1}^{n_i} (z_{ijg} - \gamma^*_{ig})^2}{\frac{n_i}{2} \bar{\lambda}_i - 1} \tag{7}$$

Lastly, phenotypes can be harmonized while preserving the covariate effects of interest as follows:

$$y^{\text{ComBat}}_{ijg} = \frac{\hat{\sigma}_g}{\hat{\delta}^*_{ig}} \left(z_{ijg} - \hat{\gamma}^*_{ig} \right) + \hat{\alpha}_g + \phi(\mathbf{x}_{ij}; \hat{\boldsymbol{\theta}}_g) \tag{8}$$

In the following section, we will discuss how this formulation facilitates the incorporation of harmonization within the federated learning framework.

2.2. Federated ComBat: Fed-ComBat

Considering the formulation previously presented in Equation (1), the parameters $\alpha_g, \boldsymbol{\theta}_g$ and γ_{ig} can be optimized by minimizing an objective function F. As data is now siloed, it is only possible to have an evaluation of the cost function at each site F_i, thus defining the federated optimization problem as:

$$\arg \min_{\alpha_g, \boldsymbol{\theta}_g, \gamma_{ig}} F(\alpha_g, \boldsymbol{\theta}_g, \gamma_{ig}), \tag{9}$$

$$\text{where, } F(\alpha_g, \boldsymbol{\theta}_g, \gamma_{ig}) := \sum_{i=1}^{S} \frac{n_i}{N} F_i(\alpha_g, \boldsymbol{\theta}_g, \gamma_{ig}) \tag{10}$$

We rely on *Federated averaging* (FedAVG) to tackle this optimization problem (McMahan et al., 2017) by allowing each site to conduct partial optimization using stochastic gradient descent (SGD) locally on their data, followed by an aggregation step that combines the shared parameters of each site. The convergence of the resulting iterative process has been demonstrated for both IID and non-IID data distribution across clients (Li et al., 2019b).

The Fed-ComBat framework requires the following steps for harmonization: i) a federated standardization step to avoid scaling issues in gradient descent as proposed by Silva et al. (2019), ii) a federated estimation of $\alpha_g, \boldsymbol{\theta}_g$ and γ_{ig} as presented in Equation (10), and iii) a local estimation of the random effects following using EB as in Equations (6) and (7). A description of the steps followed in federated harmonization using Fed-ComBat is described in Algorithm 1 including standardization as preprocessing, local updates relying on SGD, global updates across sites of shared parameters using FedAVG and random effect estimation using Empirical Bayes.

2.3. Related works

2.3.1. d-ComBat

Chen et al. (2022) proposed a distributed version of ComBat that allows correction for batch effects in a distributed manner. This is achieved by taking advantage of the multivariate linear regression problem solved in ComBat, requiring to compute the product between the inverse of a covariance matrix and the data-targets cross-covariance. Both can be computed by aggregating the individual matrices provided by each site. This allows us to reformulate Equation (1) as done by (Chen et al., 2022):

Algorithm 1: Fed-ComBat

Data: Non-harmonized phenotypes (y_{ijg}) and covariates ($\mathbf{x}_{ij}$).

Result: Harmonized phenotypes with siloed data.

$\mathbf{x} \leftarrow$ FederatedStandardization($\mathbf{x}$);

Estimation of fixed effects and random intercept:

initialize of parameter space $\boldsymbol{\Omega} := \{\boldsymbol{\theta}_g, \alpha_g, \gamma_{ig}\}_{g=1}^{G}$;

while *not converged* **do**

  foreach *site i* **do**

    $\boldsymbol{\Omega}_i^t \leftarrow \boldsymbol{\Omega}^{(t+1)}$; `// Initialization.`

    `// Partial local optimization using SGD.`

    foreach *local gradient step t* **do**

      $\boldsymbol{\Omega}_i^{(t+1)} = \boldsymbol{\Omega}_i^{(t)} - \eta \nabla_{\boldsymbol{\Omega_i}} F(\boldsymbol{\Omega_i}^{(t)})$;

  `// Aggregate and update every parameter using FedAVG.`

  $\boldsymbol{\Omega}^{(t+1)} \leftarrow \sum_i \frac{n_i}{N} \boldsymbol{\Omega}_i^t$

forall *site i* **do**

  `// Standardize using estimated parameters`

  $z_{ijg} \leftarrow \frac{y_{ijg} - \hat{\alpha}_g - \phi(\mathbf{x}_{ij}; \hat{\boldsymbol{\theta}}_g)}{\hat{\sigma}_g}$

  `// Estimate` γ_{ig}^* `and` δ_{ig}^* `using EB`

  $\gamma_{ig}^*, \delta_{ig}^* \leftarrow$ EmpiricalBayes($\mathbf{x}$) ; `// Equations (6) and (7)`

  `// Correct data`

  return $y_{ijg}^{\text{ComBat}} \leftarrow \frac{\hat{\sigma}_g}{\hat{\delta}_{ig}^*}\left(z_{ijg} - \hat{\gamma}_{ig}^*\right) + \hat{\alpha}_g + \phi(\mathbf{x}_{ij}; \hat{\boldsymbol{\theta}}_g)$

$$\underbrace{\begin{bmatrix} y_{11g} \\ \vdots \\ y_{ijg} \\ \vdots \\ y_{Sn_sg} \end{bmatrix}}_{\mathbf{y}_g} = \underbrace{\left[\begin{array}{ccc|ccc|c} x_{111} & \dots & x_{11C} & 1 & \cdots & 0 & 1 \\ \vdots & x_{ijc} & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{Sn_S1} & \dots & x_{Sn_sC} & 0 & \dots & 1 & 1 \end{array}\right]}_{\mathbf{X}} \underbrace{\left[\begin{array}{c} \hat{\theta}_{g1} \\ \vdots \\ \hat{\theta}_{gc} \\ \hline \hat{\gamma}_{1g} \\ \vdots \\ \hline \hat{\alpha}_g \end{array}\right]^{\intercal}}_{\hat{\mathbf{\Theta}}_g} + \boldsymbol{\delta}_g \boldsymbol{\varepsilon}_g \tag{11}$$

Starting from the panel formulation where $\mathbf{y_g}$ is the stacked vector of all subjects across all sites i, and $\mathbf{X}$ is a design matrix containing the covariate effects to be preserved indexed by c in subject j from site i (x_{ijc}), an indicator matrix encoding the site, and a column of ones to capture α_g, we can estimate the augmented parameter matrix $\hat{\mathbf{\Theta}}_g$ using maximum likelihood. The estimator can be decomposed as a sum of the covariance and the cross-covariance matrices at each site, as shown in Equation 12. This formulation has also been used to distribute latent variable models, such as PCA and PLS, among multiple sites (Silva et al., 2019; Lorenzi et al., 2018). The authors have also suggested sharing a partial-eigen decomposition to mitigate the risk of data leakage, as shown in Equation 13.

$$\hat{\mathbf{\Theta}}_g = \left(\mathbf{X}^{\intercal}\mathbf{X}\right)^{-1}\mathbf{X}^{\intercal}\mathbf{y}_g = \left(\sum_{i=1}^{S}\mathbf{X}_i^{\intercal}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{S}\mathbf{X}_i^{\intercal}\mathbf{y}_{ig}\right) \tag{12}$$

$$= \left(\sum_{i=1}^{S}\mathbf{Q}_i\mathbf{\Lambda}_i\mathbf{Q}_i^{-1}\right)\left(\sum_{i=1}^{S}\mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^{\intercal}\right) \tag{13}$$

2.3.2. d-ComBat-GAM

Seeking a federated nonlinear approach, we here propose d-ComBat-GAM as an adaptation for d-ComBat for non-linear covariate effect preservation

via generalized additive models (GAM) as proposed in its centralized version by Pomponio et al. (2020). This method, as d-ComBat, allows a fast estimation in a closed form by sharing either full or partial decomposition of the covariance and cross covariance matrices but transforming covariates in a design matrix using smooth functions such as polynomials or splines.

We can consider d-ComBat-GAM as a particular case of Fed-ComBat where $\phi(\cdot)$ becomes a linear combination parametrized by $\boldsymbol{\theta}$. Let $b(\cdot)$ be an arbitrary basis representation, the covariate function approximate is then defined as follows:

$$\phi(\mathbf{x}_{ij}; \boldsymbol{\theta}_g) = \sum_{c=1}^{C} \sum_{k=1}^{K_c} \theta_{gk} b_{gk}(x_{ijc}) \quad \text{where} \quad \boldsymbol{\theta}_g = [\theta_{g1}, \ldots, \theta_{gK_c}] \tag{14}$$

Note that Equation (14) expresses a linear combination of parameters, hence the same procedure as in Section 2.3.1 can be followed to estimate the fixed effects. Although polynomial decomposition is straightforward as it is the equivalent of transforming and appending the variable to a design matrix, splines require a set of defined control points based on the covariate range. We rely on federated standardization to map the nonlinear covariate to a normalized distribution where the range is known and thus the control points of the splines can be predefined.

3. Materials

We evaluated Fed-ComBat on synthetic data accounting for different sources of batch- and covariate-wise heterogeneity.

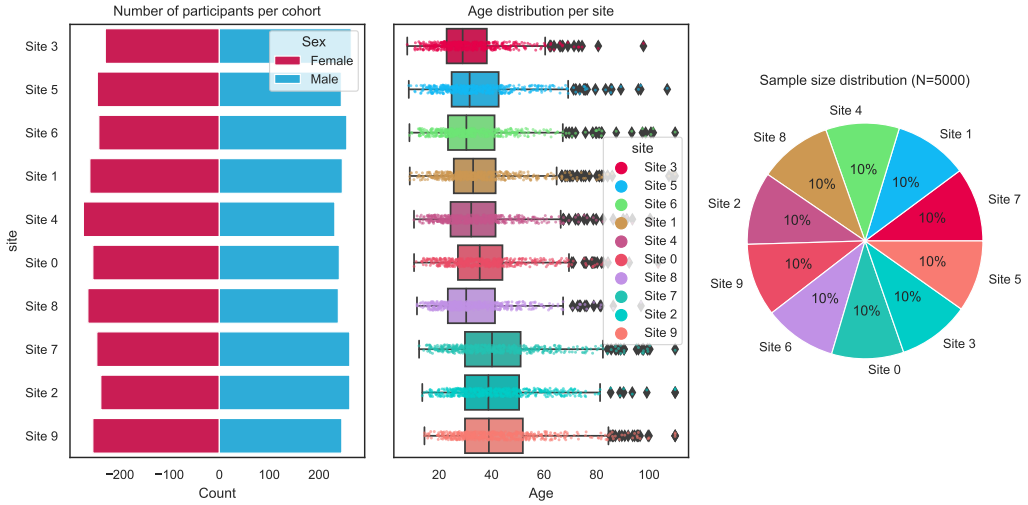
We also benchmarked this approach on a collection of nine cohorts corresponding to different studies in neurodegenerative disorders including partic-

ipants diagnosed with or at risk of developing Parkinson's and Alzheimer's disease, and Autism spectrum disorder as well as controls.

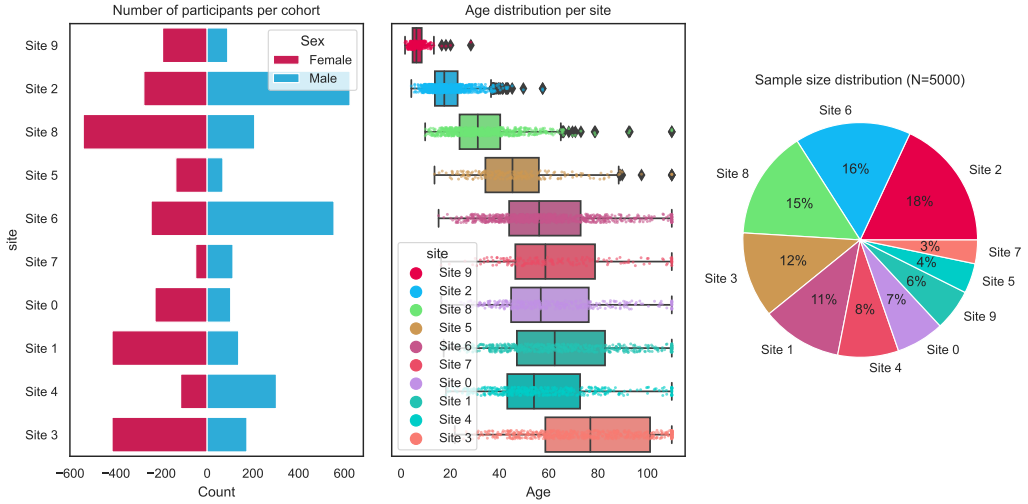
3.1. Synthetic data

We created simulated datasets under three main scenarios: i) when cohorts have similar distributions for covariates and sample size effects (IID), ii) when covariate distributions slightly overlap but follow different patterns, and when there are differences in sample sizes (mid-non-IID), and iii) when there is no overlap between covariate distributions between cohorts and there are evident differences in sample sizes (non-IID). We simulated ComBat parameters following the methodology outlined in Reynolds et al. (2022), using the graphical model depicted in Figure 2. The primary difference between our approach and that of Reynolds et al. (2022) was the model used to relate the covariates and phenotypes ϕ. We used a nonlinear function in this step, which was employed as the unbiased target function to be learned after the harmonization process for the synthetically biased phenotypes.

Sex was generated by sampling from a Bernoulli($p_i^{(\text{sex})}$) distribution, which modulated sex proportion distribution across sites. Age was drawn from a $\mathcal{N}(\mu_i^{(\text{age})}, {\sigma_i^{(\text{age})}}^2)$ distribution, where $\mu_i^{(\text{age})}$ and $\sigma_i^{(\text{age})}$ were sampled from a Uniform(a, b) distribution, with heterogeneity modulated by the width of the distribution $(b - a)$. Sample size heterogeneity was simulated using a concentration model $p_i \sim \text{Dir}(\boldsymbol{\alpha})$, as previously suggested for measuring the effects of heterogeneity in Federated Learning by Hsu et al. (2019).



(a)



(b)

Figure 1: Population distribution of synthetic data across cohorts in two different scenarios. The left panel shows the sex distribution, the middle panel shows the age distribution, and the right panel shows the sample size distribution. Panel 1a shows a homogeneous and IID population, while panel 1b shows a non-IID distribution of covariates, which is more commonly observed and used in this work.

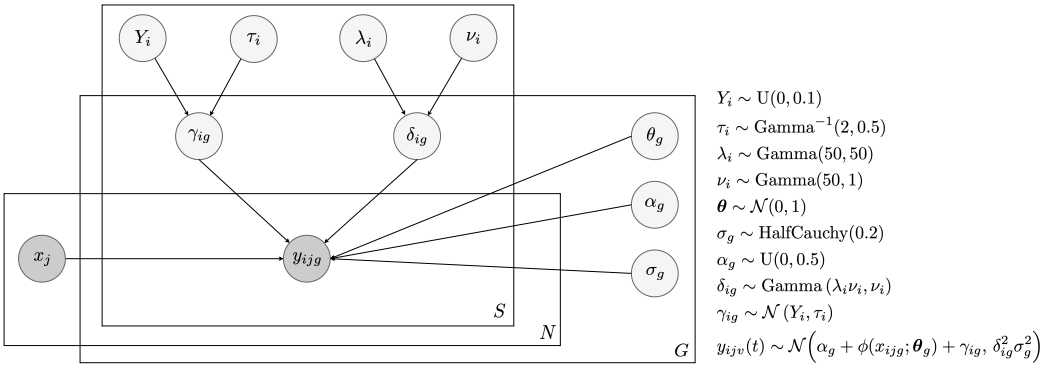


Figure 2: Graphical model used to generate synthetic data. The shaded circles indicate observed measurements, including covariates and imaging feature values, while unshaded circles represent latent parameters.

3.2. Brain MRI-Data

We evaluated our method on a cross-sectional cohort of 7265 participants from nine public studies, utilizing MRI-derived phenotypes from baseline structural T1-weighted magnetic resonance imaging (MRI).

Participants were categorized by diagnosis as follows: 3992 cognitively normal (CN), of which 731 of them were considered at risk of cognitive impairment from the A4 study (Sperling et al., 2014); 1065 were identified as having mild cognitive impairment (MCI) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al., 2013) and the Australian Imaging, Biomarker and Lifestyle Flagship Study of Aging (AIBL) (Ellis et al., 2009); 998 had been diagnosed with Autism Spectrum Disorder (ASD) from ABIDE-I and ABIDE II (Di Martino et al., 2014); and 538 were patients diagnosed with Alzheimer's disease (AD) from ADNI and Open Access Series of Imaging Studies (OASIS) dataset (Marcus et al., 2010) and, 686 participants diagnosed in multiple phases of Parkinson's disease from the Parkinson's Dis-

ease Biomarkers Program (PDPB) (Rosenthal et al., 2016) and the Parkinson's Progression Markers Initiative (PPMI) (Marek et al., 2018). Further demographic information can be found in Table 1.

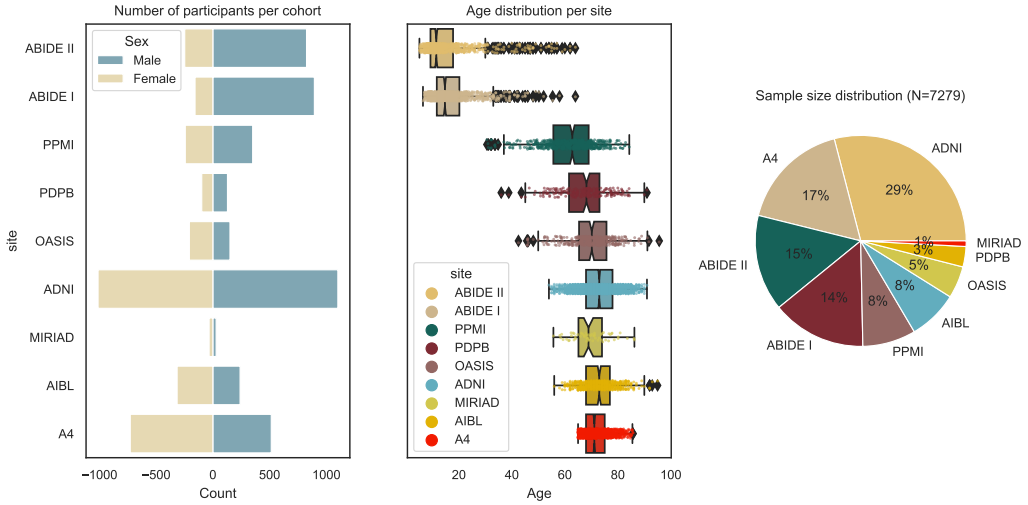


Figure 3: Population pyramid representing the demographics of the real data used in this study. The left panel shows the distribution of sex, the middle panel shows the distribution of age, and the right panel shows the sample size distribution. Cohorts were sorted by ascending median age. The demographics for the population are described in Table 1. The figure shows a clear non-IID distribution, similar to the distribution observed in Figure 1b.

3.2.1. Image processing

Cortical thickness and subcortical volume were extracted from MRI scans using FreeSurfer v7.1.1 (documented and freely available for download online at: `http://surfer.nmr.mgh.harvard.edu/`) (Dale et al., 1999; Fischl et al., 2004). Steps for phenotypical measures extraction included skull stripping (Ashburner and Friston, 2005), Non-uniform intensity Normalization (N3) (Sled et al., 1998), segmentation using the Desikan–Killiany atlas (Desikan

et al., 2006) and extraction of the MRI-derived phenotypes.

Harmonization was carried out preserving age, sex, diagnosis and intracranial volume (ICV / eTIV), and each study was considered a site whose systematic biases are expected to be corrected (Reynolds et al., 2022).

Table 1: Subject demographics of real data used in this work.

		N		Age in years ± SD [age range]	
site	**Group \ Sex**	**Female**	**Male**	**Female**	**Male**
A4	**CN**	728	515	71.3 ± 4.5 [65.0 - 85.7]	72.9 ± 5.1 [65.0 - 85.7]
ABIDE I	**Autism**	59	434	16.3 ± 8.1 [8.1 - 45.0]	17.4 ± 8.6 [7.0 - 64.0]
	CN	98	446	15.4 ± 6.6 [7.8 - 46.0]	17.4 ± 7.6 [6.5 - 48.0]
ABIDE II	**Autism**	74	416	13.1 ± 8.1 [5.2 - 54.0]	15.0 ± 9.3 [5.1 - 62.0]
	CN	175	397	13.5 ± 7.4 [5.9 - 46.6]	15.6 ± 10.0 [5.9 - 64.0]
ADNI	**AD**	156	190	73.6 ± 7.8 [55.0 - 91.0]	75.0 ± 7.6 [55.0 - 90.0]
	CN	444	345	71.8 ± 6.2 [55.0 - 90.0]	74.0 ± 6.1 [60.0 - 90.0]
	MCI	406	555	71.5 ± 7.8 [55.0 - 88.0]	73.6 ± 7.2 [54.0 - 90.0]
AIBL	**AD**	34	25	75.6 ± 8.1 [58.0 - 93.0]	72.7 ± 6.9 [60.0 - 83.0]
	CN	237	169	71.9 ± 6.0 [60.0 - 89.0]	73.0 ± 6.3 [60.0 - 90.0]
	MCI	44	46	75.8 ± 8.4 [56.0 - 95.0]	74.1 ± 5.6 [64.0 - 87.0]
MIRIAD	**AD**	22	17	69.5 ± 6.3 [58.1 - 80.2]	69.4 ± 7.5 [55.7 - 86.1]
	CN	11	12	66.3 ± 4.9 [58.8 - 73.8]	73.6 ± 7.2 [63.5 - 86.3]
OASIS	**AD**	45	41	75.6 ± 6.8 [62.9 - 95.6]	75.2 ± 7.3 [60.5 - 91.7]
	CN	162	109	67.6 ± 8.6 [45.7 - 88.8]	69.1 ± 8.3 [42.5 - 86.2]
PDPB	**PD**	97	129	67.8 ± 9.1 [50.7 - 90.0]	67.2 ± 9.7 [36.0 - 91.0]
PPMI	**CN**	51	84	59.1 ± 11.6 [31.0 - 81.9]	60.8 ± 11.7 [30.6 - 82.8]
	GenCohort PD	29	27	65.2 ± 8.0 [51.7 - 81.2]	63.9 ± 9.3 [32.2 - 78.6]
	GenCohort Unaff	52	35	61.1 ± 8.2 [33.7 - 84.3]	61.3 ± 7.6 [46.8 - 75.0]
	PD	94	164	61.2 ± 10.2 [33.5 - 81.7]	62.2 ± 9.4 [34.8 - 82.9]
	Prodromal	2	14	70.1 ± 4.3 [67.1 - 73.2]	69.0 ± 6.7 [61.7 - 82.5]
	SWEDD	15	26	61.7 ± 8.9 [46.8 - 77.6]	61.9 ± 9.9 [39.2 - 77.0]

4. Results

Two versions of Fed-ComBat were used for comparison: a first one defining a linear model ($\phi(\mathbf{x};\boldsymbol{\theta}) = \mathbf{x}^\intercal\boldsymbol{\theta}$) to show equivalence with respect to ComBat and d-ComBat, and a second one defining ϕ as a multi-layer perceptron model (MLP) for nonlinear effect preservation. These two versions of Fed-ComBat were tested jointly with our extended proposal d-ComBat-GAM against their two centralized versions: ComBat[1] (a.k.a. NeuroComBat), ComBat-GAM[2] and the only reported distributed version d-ComBat-GAM[3].

4.1. Synthetic data

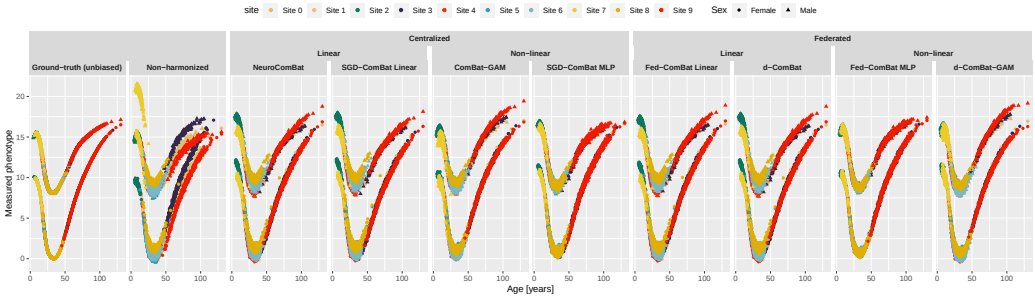
Reconstruction of an unbiased phenotype function depending on age and sex after harmonizing using different methods is shown in Figure 4a. As observed, centralized ComBat, d-ComBat, and Fed-ComBat in its linear version follow similar trends. It can also be observed that for nonlinear covariate effects linear correction can even induce biases due to its limitation to capture complex relationships between covariates and phenotypes (see Non-harmonized vs. NeuroCombat in Figure 4a).

While both ComBat-GAM and our distributed adaptation d-ComBat-GAM demonstrate improved correction for nonlinear covariate effects compared to ComBat (as shown in Figure 4a), it is worth noting that d-ComBat-GAM requires a priori definition of covariate interactions. In cases where these interactions are not defined, the performance of d-ComBat-GAM is

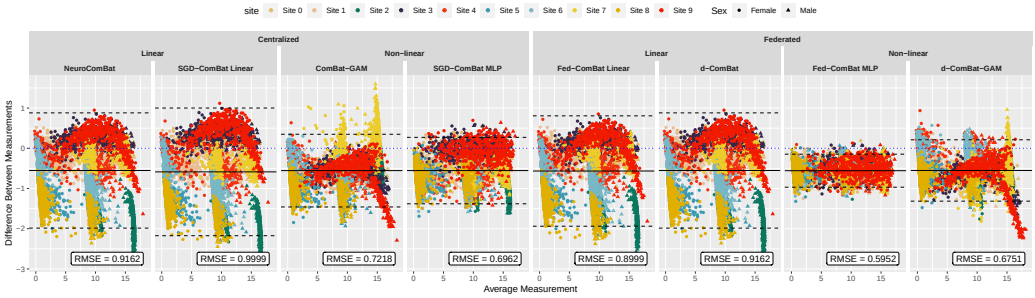
[1] `https://github.com/Jfortin1/neuroCombat`
[2] `https://github.com/rpomponio/neuroHarmonize`
[3] `https://github.com/andy1764/Distributed-ComBat`

reduced to that of ComBat. A better reconstruction of the target unbiased trajectories is achieved with Fed-ComBat using a two-layer perceptron neural network with 100 hidden units and with no hyperparameter tuning. Hypertuning was out of the scope of this work. Root mean square errors (RMSE) evaluating the reconstruction quality is presented in Table 2 for all centralized and federated methods.

Bland-Altman plots contrasting the difference between the different harmonization methods and the ground truth are presented in Figure 4b highlighting that there is less error variability (σ_ε) in estimating the unbiased function after harmonization for nonlinear methods and, particularly for Fed-ComBat in terms of the Root mean square error (RMSE = 0.5952) compared to the second federated best d-ComBat-GAM (RMSE = 0.6751).



(a) Reconstruction of an unbiased groundtruth function.



(b) Bland-Altmann plots and root mean square errors (RMSE) contrasting the different centralized and federated harmonization methods against the groundtruth.

Figure 4: Qualitative comparison across harmonization methods evaluating the quality of harmonization for a simulated phenotype with a nonlinear relationship with age and following a different trajectory per group (males and females). Results are shown for the extreme non-IID covariate distribution is depicted in Figure 1b. This can be seen as certain centers contain extremely young or old cohorts (tails). (a) Expected result after harmonization: groundtruth (unbiased), worst case scenario: biased data (non-harmonized), and harmonized phenotypes using the different ComBat methods considered and proposed in this work. As shown, ComBat could even insert biases when data is too heterogeneous. Fed-ComBat preserves better the real trajectories than the linear approaches and ComBat-GAM without the need of defining where the nonlinearities may be. Fed-ComBat using a multi-layer perceptron (MLP) shows the best reconstruction with an improvement of 35% in RMSE with respect to d-ComBat and 12% with respect to d-ComBat-GAM.

Setup	Approach	Method	$\lvert\bar{\varepsilon}\rvert$ (MAE)	σ_ε	RMSE
Federated	Non-linear	Fed-ComBat MLP	0.55710	**0.20960**	**0.59520**
Federated	Non-linear	d-ComBat-GAM	**0.55160**	0.38930	0.67510
Centralized	Non-linear	SGD-ComBat MLP	0.55420	0.42150	0.69620
Centralized	Non-linear	ComBat-GAM	0.55530	0.46120	0.72180
Federated	Linear	Fed-ComBat Linear	0.56540	0.70020	0.89990
Centralized	Linear	NeuroComBat	0.55230	0.73110	0.91620
Federated	Linear	d-ComBat	0.55230	0.73110	0.91620
Centralized	Linear	SGD-ComBat Linear	0.58640	0.81000	0.99990

Table 2: Quantitative results for Figure 4a in terms of mean absolute error between methods and the target hidden unbiased trajectories ($\lvert\bar{\varepsilon}\rvert$), its standard deviation (σ_ε) and root mean square error (RMSE). Highlighted values correspond to the best metric across different methods.

4.2. Brain MRI-data

Seeking evidence of nonlinear covariate effects of age on brain phenotypes, we compared the goodness of fit of two models: a linear model and a GAM. The criterion used to evaluate the presence of nonlinearities was the difference in the Akaike Information Criterion (AIC) between the GAM and the linear model. When the difference is negative, it indicates that the GAM is a better fit, and the magnitude of the difference indicates by how much. Figure 5b shows the AIC metric across the regions of the brain using the Desikan-Killiany parcellation and cortical thickness as the endogenous variable and the ASEG atlas for subcortical volumes. The data was controlled for sex, diagnosis, ICV, and group, and age was considered an exogenous variable. The data was also previously harmonized using ComBat. Figure 5a shows the top three regions that are better explained by a nonlinear model of age

in terms of AIC, as well as the residuals after the regression illustrating remaining effects or trends. More detailed results on the AICs differences is presented in Table S3.

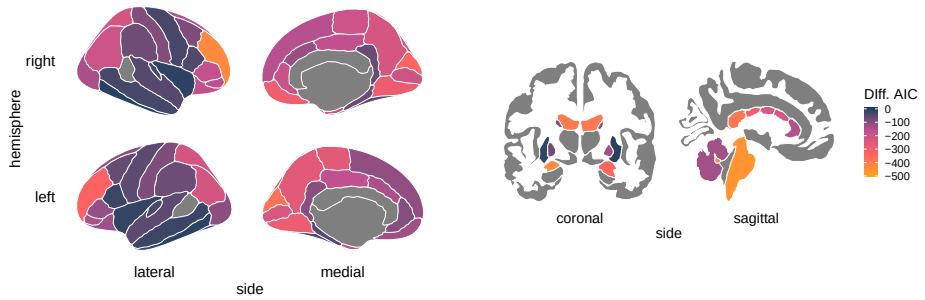
Brain trajectories segregated using diagnosis as Normal and Abnormal for ventral diencephalon, right hippocampus and hemisphere rostral middle frontal gyrus before and after harmonization are shown in Figure 7. The harmonized phenotypes were obtained using as linear centralized method: ComBat, nonlinear centralized: d-ComBat, linear federated: d-ComBat and our adaptation Fed-ComBat using a linear model, and nonlinear federated: our extension d-ComBat-GAM and Fed-ComBat using an MLP with two layers and 100 hidden units.

The residual harmonized term $\frac{\hat{\sigma}_g}{\hat{\delta}^*_{ig}}\left(z_{ijg} - \hat{\gamma}^*_{ig}\right)$ should follow a normal standard noise distribution by definition and should not have center residual effects. Such residuals for the top three regions of interest in Figure 5b are illustrated per cohort in Figure 6 for the decentralized harmonization methods considered and proposed in this work.

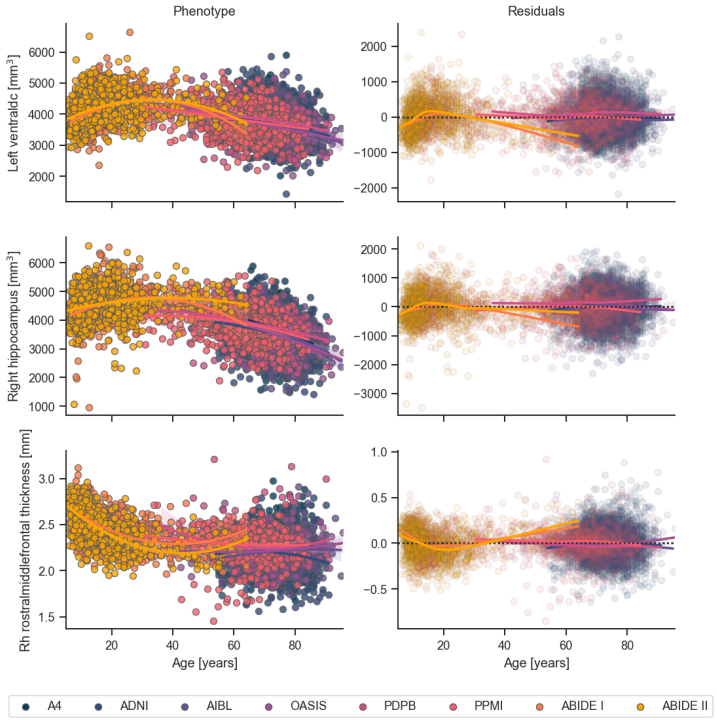
5. Limitations and Future Directions

While Fed-ComBat has demonstrated promising results for data harmonization in a federated learning setting, there are potential limitations and challenges that should be considered, as well as opportunities for future research directions.

Model architecture: The choice of model architecture, such as the two-layer perceptron function employed in the current implementation of Fed-ComBat, can influence the performance of the harmonization process.



(a) This figure displays the regions of the brain where a generalized additive model (GAM) provides a better fit than a linear model, as measured by the difference in Akaike Information Criterion (AIC) between the two models. The parameters for the GAM are set as in Pomponio et al. (2020), with age as a smoothing term, B-splines with 10 degrees of freedom or control points uniformly distributed between the minimum and maximum values, and a maximum polynomial of 3rd degree. A negative difference indicates a better fit with the GAM, while a positive difference indicates a better fit with the linear model.



(b) Results of regression using a GAM on the top 3 regions with the most discrepant AIC values on the left and residual plots illustrating goodness of fit on the right. Top: Ventral diencephalon, middle: Right hippocampus, bottom: Right hemisphere rostral middle frontal gyrus.

Figure 5: Evidence of nonlinear effects present in used cohorts. Subjects were harmonized by cohort using ComBat (Fortin et al., 2018).

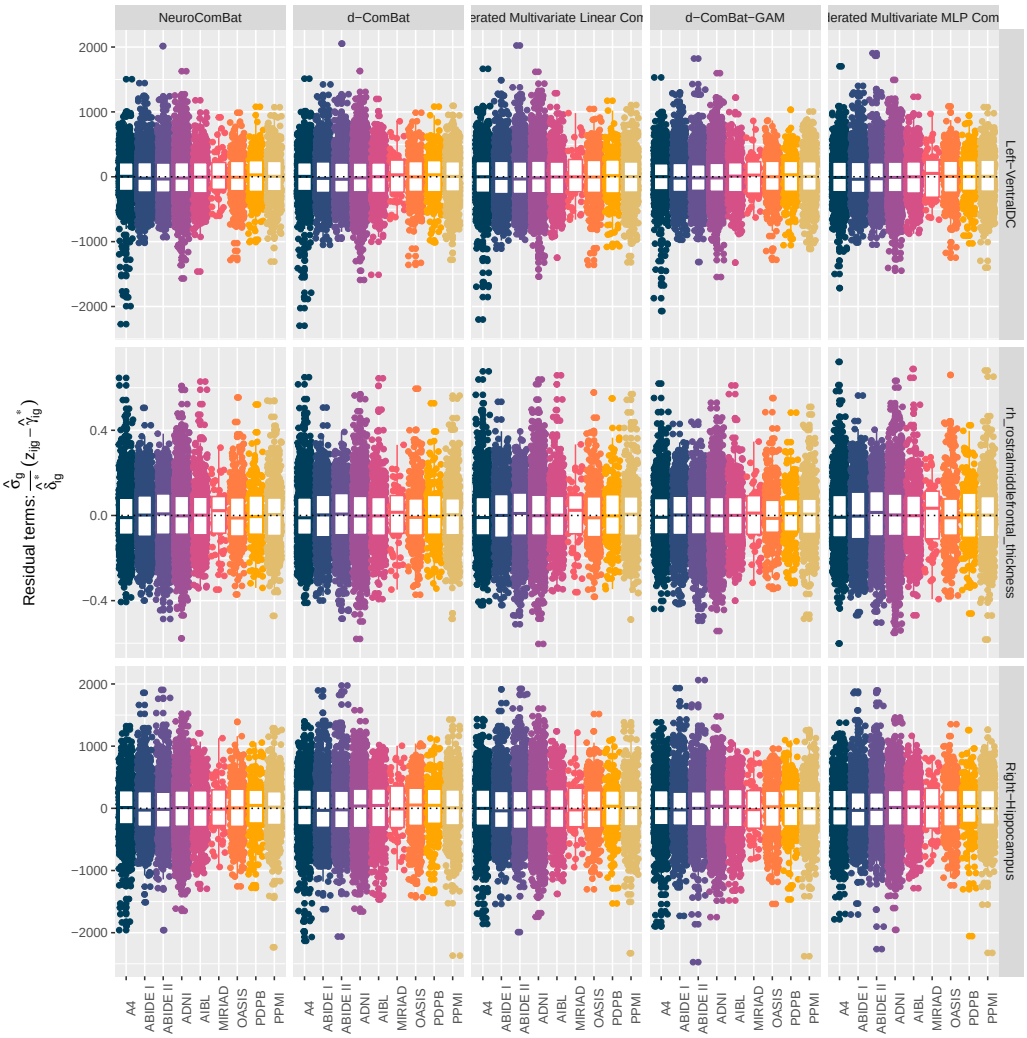


Figure 6: Residual terms after harmonization for top three regions discuses in Figure 5a for federated methods. Additional comparisons using all the methods are presented in Appendix S3.

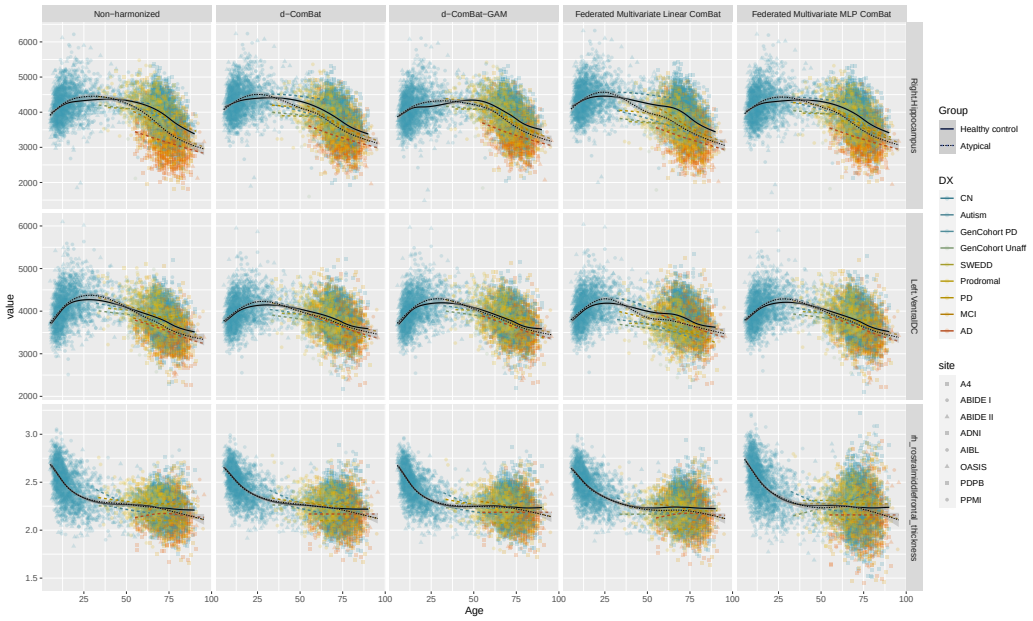


Figure 7: Harmonized phenotypes mentioned in Figure 5a in each row. Non-harmonized, d-ComBat and all the proposed methods in this work in each column. Trajectories are drown after accounting for sex, and ICV after harmonization preserving age, sex, diagnosis and ICV.

In order to better capture complex nonlinearities in various harmonization problems, it may be useful to explore alternative architectures and conduct hyperparameter tuning if necessary, while respecting the constraints required for integration within the Fed-ComBat framework (see Equation (4)).

Heterogeneity: The performance of Fed-ComBat may be affected by the level of heterogeneity across participating institutions. In cases where there is extreme disparity in data distributions, demographic compositions, or data collection protocols, the method might struggle to correct biases effectively. Therefore, the suitability of Fed-ComBat or any ComBat method in such cases should be carefully evaluated. Alternative directions could include exploring optimization methods such as FedProx(Li et al., 2020a).

Computational resources and connectivity: The current implementation of Fed-ComBat relies on the assumption that the participating institutions have enough computational resources and connectivity to perform federated learning. In practice, resource constraints or network issues might limit the applicability of this method, particularly in low-resource settings or in cases where institutions have unreliable or slow network connections.

Application to imaging data: The current implementation of Fed-ComBat focuses on derived phenotypes just like the first version of ComBat (Johnson et al., 2007), but future work could explore the application of this method directly to imaging data. By making $\phi(\mathbf{x}_{ij}, \boldsymbol{\theta}_g)$ a convolutional neural network (CNN), Fed-ComBat could be extended to handle image data directly, which would open up new opportunities for harmonization in multi-centric imaging studies in federated setups.

By addressing these limitations and challenges, and exploring potential

future directions, Fed-ComBat could be further improved and extended to a wider range of applications, ultimately enhancing its utility in federated learning and data harmonization for healthcare research.

6. Discussion and Conclusion

In conclusion, we have shown that it is possible to achieve acceptable results compared to the linear case without the need to share full covariance matrices or assume nonlinearities. The proposed Fed-ComBat method offers a novel, generalized approach for data harmonization, inspired by ComBat, and is suitable for scenarios where data sharing is restricted and nonlinear covariate effects might be present. By relying on federated learning, Fed-ComBat can capture these nonlinearities and interactions through its fully connected architecture, allowing for comparable results with centralized approaches and d-ComBat while keeping data decentralized and potentially avoiding data leakage.

In contrast to ComBat-GAM, Fed-ComBat does not require predetermined definitions of variables to be decomposed or possible interactions across variables (i.e., smoothing factor). This is because complex approximation functions like multi-layer perceptrons can effectively capture nonlinearities and interactions without explicit definitions.

As a future direction, it would be beneficial to explore the performance of Fed-ComBat on larger cohorts, as in Bethlehem et al. (2022), where nonlinear effects are better exacerbated. This would provide further evidence of the method's effectiveness in handling complex harmonization problems in neuroimaging research.

7. Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847579 (Marie Skłodowska-Curie Actions), and by the INRIA Sophia Antipolis-Méditerranée "NEF" computation cluster. This study was partially supported by the Early Detection of Alzheimer's Disease Subtypes (E-DADS) project, an EU Joint Programme - Neurodegenerative Disease Research (JPND) project (see www.jpnd.eu). The project is supported under the aegis of JPND through the following funding organizations: United Kingdom, Medical Research Council (MR/T046422/1); Netherlands, ZonMW (733051106); France, Agence Nationale de la Recherche (ANR-19-JPW2-000); Italy, Italian Ministry of Health (MoH); Australia, National Health & Medical Research Council (1191535); Hungary, National Research, Development and Innovation Office (2019-2.1.7-ERA-NET-2020-00008).

8. Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`https://adni.loni.usc.edu/`). 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 `https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`. Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth

Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (https://adni.loni.usc.edu/aibl-australian-imaging-biomarkers-and-lifestyle-study-of-ageing-18-month-data-now-released/). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/). Data used in this article were provided by the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (https://adni.loni.usc.edu/aibl-australian-imaging-biomarkers-and-lifestyle-study-of-ageing-18-month-data-now-released/). Data used in the preparation of this article were obtained from the Autism Brain Imaging Data Exchange (ABIDE) I database. Data used in this article were obtained from the Autism Brain Imaging Data Exchange (ABIDE) II database (http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html). The AIBL researchers contributed data but did not participate in the analysis or writing of this report. AIBL researchers are listed at https://www.aibl.csiro.au. Data used in the preparation of this article were obtained from the MIRIAD database. Data used in the preparation of this article were obtained from the Open Access Series of Imaging Studies (OASIS) database. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/). Data used in the preparation of this article were obtained from the Pre-symptomatic Evaluation of Experimental or Novel Treatments for Alzheimer's Disease (A4) study.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bethlehem, R.A., Seidlitz, J., White, S.R., Vogel, J.W., Anderson, K.M., Adamson, C., Adler, S., Alexopoulos, G.S., Anagnostou, E., Areces-Gonzalez, A., et al., 2022. Brain charts for the human lifespan. Nature 604, 525–533.

Bukaty, P., 2019. The California Consumer Privacy Act (CCPA): An implementation guide. IT Governance Publishing. URL: `http://www.jstor.org/stable/j.ctvjghvnn`.

Calzada, I., 2022. Citizens' data privacy in china: The state of the art of the personal information protection law (pipl). Smart Cities 5, 1129–1150.

Chen, A.A., Luo, C., Chen, Y., Shinohara, R.T., Shou, H., Initiative, A.D.N., et al., 2022. Privacy-preserving harmonization via distributed combat. Neuroimage 248, 118822.

Cheung, A.T., Nasir-Moin, M., Kwon, Y.J.F., Guan, J., Liu, C., Jiang, L., Raimondo, C., Chotai, S., Chambless, L., Ahmad, H.S., et al., 2023. Methods and impact for using federated learning to collaborate on clinical research. Neurosurgery 92, 431–438.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9, 179–194.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage 31, 968–980.

Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19, 659–667.

Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., et al., 2009. The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. International psychogeriatrics 21, 672–687.

European Commission, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). URL: `https://eur-lex.europa.eu/eli/reg/2016/679/oj`.

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., et al.,

2004. Automatically parcellating the human cerebral cortex. Cerebral cortex 14, 11–22.

Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., et al., 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 167, 104–120.

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., et al., 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161, 149–170.

Hsu, T.M.H., Qi, H., Brown, M., 2019. Measuring the effects of non-identical data distribution for federated visual classification. arXiv preprint arXiv:1909.06335 .

Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics 8, 118–127.

van Kesteren, E.J., Sun, C., Oberski, D.L., Dumontier, M., Ippel, L., 2019. Privacy-preserving generalized linear models using distributed block coordinate descent. arXiv preprint arXiv:1911.03183 .

Konečnỳ, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D., 2016. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 .

Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V., 2020a. Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems 2, 429–450.

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M.J., et al., 2019a. Privacy-preserving federated brain tumour segmentation, in: International workshop on machine learning in medical imaging, Springer. pp. 133–141.

Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P., Duncan, J.S., 2020b. Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results. Medical Image Analysis 65, 101765.

Li, X., Huang, K., Yang, W., Wang, S., Zhang, Z., 2019b. On the convergence of fedavg on non-iid data. arXiv preprint arXiv:1907.02189 .

Lorenzi, M., Altmann, A., Gutman, B., Wray, S., Arber, C., Hibar, D.P., Jahanshad, N., Schott, J.M., Alexander, D.C., Thompson, P.M., et al., 2018. Susceptibility of brain atrophy to trib3 in alzheimer's disease, evidence from functional prioritization in imaging genetics. Proceedings of the National Academy of Sciences 115, 3162–3167.

Mahlool, D.H., Abed, M.H., 2022. Distributed brain tumor diagnosis using a federated learning environment. Bulletin of Electrical Engineering and Informatics 11, 3313 – 3321. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85139133360&doi=10.11591%2feei.v11i6.4131&partnerID=40&md5=1ad3123e9db42a98e13e51bd26fdcbc1,

doi:10.11591/eei.v11i6.4131. cited by: 0; All Open Access, Gold Open Access, Green Open Access.

Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2010. Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. Journal of cognitive neuroscience 22, 2677–2684.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C.S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C.M., Trojanowski, J.Q., et al., 2018. The parkinson's progression markers initiative (ppmi)–establishing a pd biomarker cohort. Annals of clinical and translational neurology 5, 1460–1477.

McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data, in: Artificial intelligence and statistics, PMLR. pp. 1273–1282.

Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I.M., Satterthwaite, T.D., Fan, Y., et al., 2020. Harmonization of large mri datasets for the analysis of brain imaging patterns throughout the lifespan. NeuroImage 208, 116450.

Reynolds, M., Chaudhary, T., Torbati, M.E., Tudorascu, D.L., Batmanghelich, K., 2022. Combat harmonization: Empirical bayes versus fully bayes approaches. bioRxiv .

Rosenthal, L.S., Drake, D., Alcalay, R.N., Babcock, D., Bowman, F.D., Chen-Plotkin, A., Dawson, T.M., Dewey Jr, R.B., German, D.C., Huang,

X., et al., 2016. The ninds parkinson's disease biomarkers program. Movement Disorders 31, 915–923.

Silva, S., Gutman, B.A., Romero, E., Thompson, P.M., Altmann, A., Lorenzi, M., 2019. Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data, in: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), IEEE. pp. 270–274.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE transactions on medical imaging 17, 87–97.

Sperling, R.A., Rentz, D.M., Johnson, K.A., Karlawish, J., Donohue, M., Salmon, D.P., Aisen, P., 2014. The a4 study: stopping ad before symptoms begin? Science translational medicine 6, 228fs13–228fs13.

Wachinger, C., Rieckmann, A., Pölsterl, S., Initiative, A.D.N., et al., 2021. Detect and correct bias in multi-site neuroimaging datasets. Medical Image Analysis 67, 101879.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., et al., 2013. The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer's & Dementia 9, e111–e194.