ComBat-MLE: A novel perspective on developing and solving ComBat harmonization method

Yingping Li, Emilie Chouzenoux, Shuiping Gou, Peng Zeng, Samy Ammari, Nathalie Lassau

ComBat-MLE: A Novel Perspective on Developing and Solving ComBat Harmonization Method

Yingping Li
School of Artificial Intelligence
Xidian University
Xi'an, China
liyingping@xidian.edu.cn

Emilie Chouzenoux
Centre de Vision Numérique & INRIA
Université Paris-Saclay
Paris, France
emilie.chouzenoux@centralesupelec.fr

Shuiping Gou*
School of Artificial Intelligence
Xidian University
Xi'an, China
shpgou@mail.xidian.edu.cn

Peng Zeng
School of Artificial Intelligence
Xidian University
Xi'an, China
zengpeng1107@gmail.com

Samy Ammari
BIOMAPS, UMR1281 Inserm, CEA, CNRS
Département d'Imagerie Médicale, Gustave Roussy
Université Paris-Saclay
Paris, France
samy.ammari@gustaveroussy.fr

Nathalie Lassau
BIOMAPS, UMR1281 Inserm, CEA, CNRS
Département d'Imagerie Médicale, Gustave Roussy
Université Paris-Saclay
Paris, France
nathalie.lassau@gustaveroussy.fr

***Abstract*—This paper provides a novel perspective on the model development of the ComBat method. Two sets of parameter constraints were introduced to help identify parameters during the parameter estimation process. And Maximum Likelihood Estimation (MLE) was employed to solve the re-formulated ComBat using each set of constraints. The simpler and faster solution was selected as the proposed ComBat-MLE harmonization method. A series of experiments were conducted to evaluate the effectiveness of various ComBat variants, both quantitatively and qualitatively. The experiments validated the effectiveness of the proposed ComBat-MLE method in removing non-biological scanner effects while at the same time preserving intrinsic biological information. Although the proposed ComBat-MLE method did not significantly outperform existing ComBat variants, it aids in understanding the ComBat method and offers valuable insights into the development of new harmonization methods.**

***Index Terms*—Harmonization method, ComBat, Radiomics, Scanner effects, Maximum Likelihood Estimation (MLE)**

I. Introduction

In medical studies, medical images are often collected from multiple institutions, with different image acquisition devices or protocols. During this image acquisition process, some unwanted variations are introduced. Fortin [1] termed these unwanted non-biological variations associated with different image acquisition scanners or parameter configurations as "scanner effects". In radiomic studies, scanner effects should be carefully considered, because it may hinder the detection of useful biological information, limit the capabilities of the radiomic models, even lead to unreliable or wrong conclusions. It may also influence the generalizability of the radiomic models when they are applied to new data set collected with different settings. Under this circumstance, harmonization methods have been proposed and can be treated as very effective ways to handle the scanner effects.

The most popular harmonization method in radiomic studies is ComBat [2], a method originally proposed to remove the batch effects present in microarray expression data. In 2017, Fortin et al. [1], [3] firstly adapted ComBat for removing scanner effects in multi-site studies on diffusion tensor imaging and cortical thickness measurements in MRI, reporting that ComBat can remove the unwanted variations associated with sites and preserve the useful biological information at the same time. They also published their codes of ComBat method in github *https://github.com/Jfortin1/ComBatHarmonization*, including R, Matlab and Python versions so far. Subsequently, other researchers adopted their public ComBat codes and further verified the effectiveness of ComBat to remove scanner effects in different kinds of medical image modalities and in different kinds of radiomic models [4]–[9]. For example, Orlhac and her colleagues have validated ComBat as an effective harmonization method in PET radiomics [4], in CT radiomics [5], and in MRI radiomics [6], [7].

In 2022, Orlhac *et al.* [10] systematically explained and illustrated the use of ComBat method, and reported some cases where ComBat may not work. Among these mentioned failure cases, there is one kind of cases we also met in our radiomics experiments. That is, when ComBat is applied separately for each of the pattern classes, the scanner effects can be successfully removed. However, when ComBat is applied for data comprising all pattern classes, it fails to realign the features correctly. The intrinsic reason lies in that, different pattern classes may suffer from different scanner effects. That is to say, the imaging of different tissues, organs, or tumors is affected in a different way by scanner effects introduced through different image protocols, scanners, or reconstruction methods. So determining a single transform by ComBat for data from different pattern classes can not provide satisfactory

* Corresponding author.
This work was supported by the Fundamental Research Funds for the Central Universities (grant number XJSJ24071 and XJSJ24072), and the National Natural Science Foundation of China (grant number 62302355).

data harmonization results.

In this context, we explored developing new harmonization methods based on a different assumption from ComBat: that scanner effects vary across different pattern classes (such as tumors and normal tissues), even if they originate from the same scanner settings. We modeled this case as a Gaussian Mixture Model (GMM) and solved it using the Expectation Maximization (EM) algorithm. Unfortunately, the developed model failed to provide the expected harmonization results due to insufficient parameter constraints needed for parameter identification. The details can be found in the Supplementary Materials at *https://github.com/Yingping-LI/Harmonization_methods*. However, during this process, we found that ComBat can be considered as a special case of the proposed GMM model. This insight offers us a new perspective on developing and solving the ComBat harmonization method, leading to the creation of this paper.

We summarize our main contributions as: (1) we provide a new perspective on developing the ComBat model, and solve it using Maximum Likelihood Estimation (MLE) with two different sets of parameter constraints. This part aids in understanding the ComBat method and provides valuable insights into the development of new harmonization methods. (2) We make our radiomic features extracted from the phantoms publicly available, thus assisting researchers in testing their harmonization methods.

II. Related Work

A. ComBat Method

ComBat method [2] was originally proposed to remove batch effects in microarray expression data. The model assumes the expression value Y_{ijg} for gene g for sample j from batch i to be as follows:

$$Y_{ijg} = \alpha_g + X\beta_g + \gamma_{ig} + \delta_{ig}\varepsilon_{ijg}, \tag{1}$$

where α_g is the mean expression value for gene g, X is a designed covariates of interest matrix and β_g is the vector of regression coefficients corresponding to X, ε_{ijg} is the error term which is assumed to follow a normal distribution $\mathcal{N}(0, \sigma_g^2)$, r_{ig} and δ_{ig} represent the additive and multiplicative batch effects of batch i for gene g, respectively.

Assuming the knowledge of the estimated values $\widehat{\alpha}_g, \widehat{\beta}_g, \widehat{\gamma}_{ig}, \widehat{\delta}_{ig}$ for the parameters $\alpha_g, \beta_g, \gamma_{ig}, \delta_{ig}$, the batch-adjusted data Y^*_{ijg} would be given by

$$Y^*_{ijg} = \widehat{\alpha}_g + X\widehat{\beta}_g + \frac{Y_{ijg} - \widehat{\alpha}_g - X\widehat{\beta}_g - \widehat{\gamma}_{ig}}{\widehat{\delta}_{ig}}. \tag{2}$$

So the problem remains to how to estimate these parameters in model (1). ComBat uses empirical Bayes (EB) method to estimate parameters, with detail steps listed as below.

1) Standardize the data: First, one uses a gene-wise ordinary least-squares approach to estimate the model parameters $\alpha_g, \beta_g, \gamma_{ig}$ as $\widehat{\alpha}_g, \widehat{\beta}_g, \widehat{\gamma}_{ig}$ for $i = 1, ..., m$ and $g = 1, ..., G$, constraining $\sum_i n_i \widehat{\gamma}_{ig} = 0$ for $g = 1, ..., G$ to ensure the identifiability of the parameters, and estimate $\widehat{\sigma}_g^2$ as $\widehat{\sigma}_g^2 = \frac{1}{N}\sum_{ij}(Y_{ijg} - \widehat{\alpha}_g - X\widehat{\beta}_g - \widehat{\gamma}_{ig})^2$ (N is the total number of samples). Then the standardized data Z_{ijg} is

$$Z_{ijg} = \frac{Y_{ijg} - \widehat{\alpha}_g - X\widehat{\beta}_g}{\widehat{\sigma}_g}. \tag{3}$$

The Empirical Bayes (EB) method will be designed to borrow information across genes in order to robustly handle high-dimensional data with small sample size, so standardization is essential here to reduce gene-to-gene variation in the data.

2) Empirical Bayes (EB) method to estimate batch effect parameters: Suppose the standardized data $Z_{ijg} \sim \mathcal{N}(\gamma_{ig}, \delta_{ig}^2)$, the parameters γ_{ig} and δ_{ig}^2 can be simply estimated from the standardized data using method of moments:

$$\begin{cases} \gamma^*_{ig} = \dfrac{1}{n_i}\displaystyle\sum_j Z_{ijg}, \\ \delta^{2*}_{ig} = \dfrac{1}{n_i - 1}\displaystyle\sum_j (Z_{ijg} - \widehat{\gamma}_{ig})^2, \end{cases} \tag{4}$$

where γ^*_{ig} and δ^{2*}_{ig} correspond to the batch i sample mean and variance for gene g, respectively. Then two formats of EB methods can be further applied to more robustly estimate the parameters γ_{ig} and δ^2_{ig}.

- Parametric EB method:
  Using parametric empirical priors and suppose
  $$\gamma_{ig} \sim \mathcal{N}(\gamma_i, \tau_i^2) \qquad \text{and} \qquad \delta^2_{ig} \sim \mathcal{IG}(\lambda_i, \theta_i), \tag{5}$$
  where $\mathcal{IG}$ represents the *Inverse Gamma* distribution. Then γ^*_{ig} and δ^{2*}_{ig} are estimated as the expectation of the posterior distributions of γ_{ig} and δ^2_{ig}, respectively.
- Non-parametric EB method:
  The parametric empirical priors may not be appropriate for some data sets, so the author proposes non-parametric empirical priors as a more flexible option. γ^*_{ig}, δ^{2*}_{ig} are estimated as the posterior expectations of γ_{ig}, δ^2_{ig}, respectively. Note that Monte Carlo integration method is used to simulate the integration values when calculating the posterior expectations.

3) Adjust the data for batch effects: With the estimated parameters γ^*_{ig} and δ^{2*}_{ig} by Step 2, now the batch-adjusted data Y^*_{ijg} can be calculated by

$$Y^*_{ijg} = \widehat{\alpha}_g + X\widehat{\beta}_g + \frac{\widehat{\sigma}_g}{\widehat{\delta}^*_{ig}}(Z_{ijg} - \widehat{\gamma}^*_{ig}) \tag{6}$$

in a similar fashion than in the batch-adjust equation (2).

We can summarize the three variants of ComBat as ComBat-MoM, non-parametric ComBat, and parametric ComBat, each corresponding to a different parameter estimation method: the Method of Moments (MoM), non-parametric empirical Bayes, and parametric empirical Bayes, respectively.

B. Other Harmonization Methods

There are also some other harmonization methods proposed in recent years. Beer at al. [11] proposed the Longitudinal ComBat method for harmonizing longitudinal multi-scanner

imaging data. Chen et al. [12] proposed the CovComBat method for removing not only the scanner effects in mean and variance, but also the scanner effects in the covariance between multivariate measurements. Da-Ano et al. [13] proposed three versions of modified ComBat, namely M-ComBat for transforming the features to a chosen reference instead of the overall mean, the B-ComBat using bootstrap and Monte Carlo for improved robustness, and the BM-ComBat adopting the modifications of both M-ComBat and B-ComBat. Radua and his colleagues [14] modified the ComBat method so that it can be used for the missing data, and in addition, the processes of fitting and application of ComBat are separated. Neuroharmony method [15] is another harmonization method which could even harmonize a single image from an unseen scanner, by training a model to learn the correlations between the image quality metrics and the relative volume corrections for each region of the brain standardized by ComBat.

Most of these harmonization methods are built upon the ComBat method and act on radiomic features, but there are also some exceptions. For instance, Chatterjee et al. [16] applied feature standardization separately for each independent set using imbalance adjustments, instead of applying the feature standardization for the overall pooling datasets, as a relatively simple harmonization method. Dewey et al. [17] used deep learning method and proposed a U-Net-based architecture called DeepHarmony for the contrast harmonization across scanner changes in MRI images.

Despite the availability of various harmonization methods, ComBat remains the state-of-the-art and most commonly used approach in radiomics studies.

III. The proposed method

In this section, we will redescribe the ComBat model without considering the covariates, and then use a MLE-based method to estimate the model parameters, with two different sets of parameter constraints considered.

A. Build the Model

Let $\boldsymbol{Y} \in \mathbb{R}^p$ be a random variable representing the clear feature data without scanner effects, and $\boldsymbol{X}_s \in \mathbb{R}^p$ be a random variable representing the observed feature data with scanner effects from scanner setting $s \in \{1, 2, ..., S\}$. Suppose for each scanner setting $s \in \{1, 2, ..., S\}$, we have N_s observations $\{\boldsymbol{x}_1^{(s)}, ..., \boldsymbol{x}_{N_s}^{(s)}\}$ of variable $\boldsymbol{X}_s$, and denote the $N_s \times p$ matrix X_s as the set of all observations from scanner setting s, where the n-th row represents $(\boldsymbol{x}_n^{(s)})^\top$.

Now we will build the model to simulate the scanner effects in the observed feature data. Suppose that clear feature data variable $\boldsymbol{Y} \in \mathbb{R}^p$ follows a multivariate Gaussian distribution

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{7}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are the mean and covariance matrix of the clear data, respectively. Then the variable $\boldsymbol{X}_s \in \mathbb{R}^p, s \in \{1, 2, ..., S\}$, which represents the observed feature data variable with scanner effects from scanner setting s , can be assumed to follow the following multivariate Gaussian distribution

$$\boldsymbol{X}_s \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}), \tag{8}$$

where $\boldsymbol{\gamma}_s \in \mathbb{R}^p$ and $\boldsymbol{\Delta}_s \in \mathbb{R}^{p\times p}$ represent the additive and multiplicative scanner effects, respectively. We suppose the parameters $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{\Delta}_s \in \mathbb{R}^{p\times p}$ to be diagonal matrices.

Let $\boldsymbol{\theta}$ represent all the model parameters, that is,

$$\boldsymbol{\theta} = \{\boldsymbol{\mu} \in \mathbb{R}^p,\ \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p},\ \boldsymbol{\gamma}_s \in \mathbb{R}^p,\ \boldsymbol{\Delta}_s \in \mathbb{R}^{p\times p} \mid s = 1, ...S\}. \tag{9}$$

If we could get an estimation $\boldsymbol{\theta}^* = \{\boldsymbol{\mu}^*,\ \boldsymbol{\Sigma}^*,\ \boldsymbol{\gamma}_s^*,\ \boldsymbol{\Delta}_s^* \mid s = 1, ...S\}$ of $\boldsymbol{\theta}$, then for each observation $\boldsymbol{x}$ of variable $\boldsymbol{X}_s$ which represents the observed feature data from scanner s, the corresponding clear feature data $\boldsymbol{y}$ (an observation of variable $\boldsymbol{Y}$) would be

$$\boldsymbol{y} = \boldsymbol{\mu}^* + (\boldsymbol{\Omega}_s^*)^{-1}(\boldsymbol{x} - \boldsymbol{\mu}^* - \boldsymbol{\gamma}_s^*), \tag{10}$$

where $\boldsymbol{\Omega}_s^* = (\boldsymbol{\Delta}_s^*)^{\frac{1}{2}}$. So the problem becomes to how to estimate the model parameters $\boldsymbol{\theta}$.

Note that the parameters (such as $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}_s$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}_s$) are unidentifiable, so proper constraints should be imposed to help identify the parameters. It is very flexible to choose the parameter constraints as long as they are reasonable. For our model, as listed below, we can either impose constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$, or impose constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Optional constraints 1: impose constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$.

- additive constraint:

$$\sum_{s=1}^{S} w_s\boldsymbol{\gamma}_s = \boldsymbol{0}. \tag{11}$$

- multiplicative constraint:

$$\prod_{s=1}^{S} (\boldsymbol{\Delta}_s)^{w_s} = \mathrm{I}. \tag{12}$$

Optional constraints 2: impose constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

- additive constraint:

$$\boldsymbol{\mu} = \sum_{s=1}^{S} w_s\boldsymbol{\mu}_s, \tag{13}$$

  where $\boldsymbol{\mu}_s = \frac{1}{N_s}\sum_{n=1}^{N_s} \boldsymbol{x}_n^{(s)}, s = 1, 2, ..., S$.
- multiplicative constraint:

$$\boldsymbol{\Sigma} = \sum_{s=1}^{S} w_s\boldsymbol{\Sigma}_s, \tag{14}$$

  where $\boldsymbol{\Sigma}_s = \mathrm{diag}(\boldsymbol{\Gamma}_s)$, where $\boldsymbol{\Gamma}_s = \frac{1}{N_s-1}\sum_{n=1}^{N_s}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s)$ and $\circ$ represents the Hadamard product (also known as element-wise product).

In essence, no matter imposing constraints on $\boldsymbol{\gamma}_s$, $\boldsymbol{\Delta}_s$, or on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, these constraints always aim to adjust the mean and variance of the harmonized data to be the weighted center and weighted variance of the clusters from all scanner

settings, where $w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}$, $s = 1, 2, ..., S$ are the weight coefficients.

We will estimate the model parameters using the two sets of constraints: constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$ in subsection III-B, and constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in subsection III-C.

B. Estimate model parameters with constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$

In this subsection, we will use the MLE method to estimate the model parameters $\boldsymbol{\theta}$ defined by Equation (9), with constraints imposed on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$ by Equation (11) and (12). Consider the data set X_s from scanner setting s , our goal is to maximize the likelihood function $p_s(X_s|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, where

$$p_s(X_s|\boldsymbol{\theta}) = \prod_{n=1}^{N_s} f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}), \tag{15}$$

and the corresponding logarithm likelihood function is

$$\ln p_s(X_s|\boldsymbol{\theta}) = \sum_{n=1}^{N_s} \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}). \tag{16}$$

We have S objective functions $\ln p_s(X_s|\boldsymbol{\theta}), s = 1, 2, ..., S$ to be optimized, and they share the same parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, so it is natural that we design our final objective function to be the weighted sum of these S logarithm likelihood functions, that is,

$$\begin{aligned}\mathcal{L}_0(\boldsymbol{\theta}) &= \sum_{s=1}^{S} w_s \ln p_s(X_s|\boldsymbol{\theta}) \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}),\end{aligned} \tag{17}$$

where

$$\begin{aligned}f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}) &= \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{\frac{1}{2}}} \\ \cdot \exp\Big\{ -\frac{1}{2}&(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)^\top (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)\Big\}\end{aligned} \tag{18}$$

and $w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}$. Our objective is to maximize the logarithm likelihood function $\mathcal{L}_0(\boldsymbol{\theta})$ under the constraints (11) and (12) with respect to the set of model parameters, which is defined as $\boldsymbol{\theta}$ in Equation (9). So our problems becomes

$$\begin{aligned}&\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathcal{L}_0(\boldsymbol{\theta}), \\ &\text{s.t.} \quad \sum_{s=1}^{S} w_s\boldsymbol{\gamma}_s = \mathbf{0}, \\ &\qquad \prod_{s=1}^{S} (\boldsymbol{\Delta}_s)^{w_s} = \mathrm{I}.\end{aligned} \tag{19}$$

Suppose $\boldsymbol{\Delta}_s$, $s = 1, 2, ..., S$ are diagonal matrices and can be represented as $\boldsymbol{\Delta}_s = \mathrm{diag}(\boldsymbol{\Lambda}_s)$, where $\boldsymbol{\Lambda}_s = (\delta_{1s}, ..., \delta_{ps})^\top$, $s = 1, 2, ..., S$. Then the constraint $\prod_{s=1}^{S}(\boldsymbol{\Delta}_s)^{w_s} = \mathrm{I}$ can be re-written as $\sum_{s=1}^{S} w_s \ln \boldsymbol{\Lambda}_s = \mathbf{0}$. By applying the method of Lagrange multipliers, we can define the final Lagrange function to be maximized as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \mathcal{L}_0(\boldsymbol{\theta}) + \sum_{i=1}^{p} \tilde{\lambda}_i\big(\sum_{s=1}^{S} w_s\gamma_{is}\big) + \sum_{i=1}^{p} \hat{\lambda}_i(\sum_{s=1}^{S} w_s \ln \delta_{is}) \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma}) \\ &\quad + \sum_{i=1}^{p} \tilde{\lambda}_i\big(\sum_{s=1}^{S} w_s\gamma_{is}\big) + \sum_{i=1}^{p} \hat{\lambda}_i(\sum_{s=1}^{S} w_s \ln \delta_{is}).\end{aligned} \tag{20}$$

where $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_p)^\top \in \mathbb{R}^p$ and $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p)^\top \in \mathbb{R}^p$ are the Lagrange multipliers corresponding to the two constraints. Now we will update the parameters iteratively to maximize the Lagrange function $\mathcal{L}(\boldsymbol{\theta})$, following a coordinate descent optimization strategy.

Update $\boldsymbol{\mu} \in \mathbb{R}^p$: We start with updating $\boldsymbol{\mu}$ by calculating the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\mu}$, and set it to zero as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\mu}} &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \frac{\partial \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma})}{\partial\boldsymbol{\mu}} \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) = 0.\end{aligned} \tag{21}$$

Multiply $\boldsymbol{\Sigma}$ on both sides and rearrange the equation, we get

$$\boldsymbol{\mu} = \Big[\sum_{s=1}^{S} w_s N_s \boldsymbol{\Delta}_s^{-1}\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \boldsymbol{\Delta}_s^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_s)\Big]. \tag{22}$$

Update $\boldsymbol{\gamma}_s \in \mathbb{R}^p, s = 1, 2, ..., S$: Calculate the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\gamma}_s$, we have

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}_s} &= w_s \sum_{n=1}^{N_s} \frac{\partial \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma})}{\partial\boldsymbol{\gamma}_s} + w_s\tilde{\boldsymbol{\lambda}} \\ &= w_s \sum_{n=1}^{N_s} (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) + w_s\tilde{\boldsymbol{\lambda}},\end{aligned} \tag{23}$$

then set the derivative to zero, we can get

$$\boldsymbol{\gamma}_s = \frac{1}{N_s}(\boldsymbol{\Delta}_s\boldsymbol{\Sigma}) \cdot \Big[(\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}) + \tilde{\boldsymbol{\lambda}}\Big]. \tag{24}$$

Note that the additive constraint $\sum_{s=1}^{S} w_s\boldsymbol{\gamma}_s = \mathbf{0}$ holds, where $w_s = \frac{N_s}{N}$, so by taking equation (24) to the additive constraint and then rearranging the equations, we can get the expression of $\tilde{\boldsymbol{\lambda}}$ as follows:

$$\tilde{\boldsymbol{\lambda}} = -\big[\sum_{s=1}^{S} (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})\big]^{-1} \cdot \big[\sum_{s=1}^{S} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu})\big]. \tag{25}$$

So parameter $\boldsymbol{\gamma}_s$ can be updated by Equation (24), where $\tilde{\boldsymbol{\lambda}}$ is calculated by Equation (25).

Update $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$: Now let us provide the way to update $\boldsymbol{\Delta}_s$ and $\boldsymbol{\Sigma}$. Suppose $\boldsymbol{\Delta}_s$ and $\boldsymbol{\Sigma}$ are diagonal matrices, and can be denoted by $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, ..., \sigma_p)$ and $\boldsymbol{\Delta}_s = \text{diag}(\delta_{1s}, ..., \delta_{ps})$. If we take the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}_s$ directly, we cannot ensure that their update matrices are still diagonal. In order to solve this problem, we introduce $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}_s$ as

$$\boldsymbol{\Gamma} = (\sigma_1, ..., \sigma_p)^\top, \ \boldsymbol{\Lambda}_s = (\delta_{1s}, ..., \delta_{ps})^\top, \tag{26}$$

which obviously satisfy

$$\begin{aligned} \boldsymbol{\Sigma} &= \text{diag}(\boldsymbol{\Gamma}), \ \ \boldsymbol{\Sigma}\boldsymbol{e} = \boldsymbol{\Gamma}, \\ \boldsymbol{\Delta}_s &= \text{diag}(\boldsymbol{\Lambda}_s), \ \ \boldsymbol{\Delta}_s\boldsymbol{e} = \boldsymbol{\Lambda}_s, \end{aligned} \tag{27}$$

where $\boldsymbol{e} = (1, 1, ..., 1)^\top \in \mathbb{R}^p$. We will update $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}_s$ by calculating the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}_s$, respectively, and then update $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}_s$ by Equation (27).

The derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\Gamma}$ takes the form

$$\begin{aligned} &\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\Gamma}} = \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \frac{\partial \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma})}{\partial \boldsymbol{\Gamma}} \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \Bigg\{ \frac{\partial \ln |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial \boldsymbol{\Gamma}} \\ &- \frac{1}{2} \frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)^\top (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big\}}{\partial \boldsymbol{\Gamma}} \Bigg\}, \end{aligned} \tag{28}$$

where

$$\frac{\partial \ln |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial \boldsymbol{\Gamma}} = -\frac{1}{2} |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-1} \cdot \frac{\partial |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|}{\partial \boldsymbol{\Gamma}} = -\frac{1}{2} \boldsymbol{\Sigma}^{-1}\boldsymbol{e}, \tag{29}$$

and

$$\begin{aligned} &\frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)^\top (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big\}}{\partial \boldsymbol{\Gamma}} \\ &= -\boldsymbol{\Sigma}^{-2} \boldsymbol{\Delta}_s^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s), \end{aligned} \tag{30}$$

and $\circ$ represents the Hadamard product. Taking Equation (29) and Equation (30) into Equation (28), and setting the derivative to zero, we have

$$\begin{aligned} &\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\Gamma}} = \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \Big[-\frac{1}{2} \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{e} \\ &+ \frac{1}{2} \boldsymbol{\Sigma}^{-2} \boldsymbol{\Delta}_s^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big] = 0. \end{aligned} \tag{31}$$

Multiplying $\boldsymbol{\Sigma}^2$ on both sides, and rearranging the equation with noticing that $\boldsymbol{\Sigma}\boldsymbol{e} = \boldsymbol{\Gamma}$, we get

$$\begin{aligned} \boldsymbol{\Gamma} &= \Big[\sum_{s=1}^{S} w_s N_s \Big]^{-1} \\ &\cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \boldsymbol{\Delta}_s^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big], \end{aligned} \tag{32}$$

then $\boldsymbol{\Sigma}$ can be updated by $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Gamma})$.

Update $\boldsymbol{\Delta}_s \in \mathbb{R}^{p\times p}, s = 1, 2, ..., S$: Similarly, calculate the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\Lambda}_s \in \mathbb{R}^p$, we have

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\Lambda}_s} &= w_s \sum_{n=1}^{N_s} \frac{\partial \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu} + \boldsymbol{\gamma}_s, \boldsymbol{\Delta}_s\boldsymbol{\Sigma})}{\partial \boldsymbol{\Lambda}_s} + w_s \boldsymbol{\Delta}_s^{-1} \hat{\boldsymbol{\lambda}} \\ &= w_s \sum_{n=1}^{N_s} \Bigg\{ \frac{\partial \ln |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial \boldsymbol{\Lambda}_s} \\ &- \frac{1}{2} \frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)^\top (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big\}}{\partial \boldsymbol{\Lambda}_s} \Bigg\} \\ &+ w_s \boldsymbol{\Delta}_s^{-1} \hat{\boldsymbol{\lambda}}, \end{aligned} \tag{33}$$

where

$$\frac{\partial \ln |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial \boldsymbol{\Lambda}_s} = -\frac{1}{2} |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|^{-1} \cdot \frac{\partial |\boldsymbol{\Delta}_s\boldsymbol{\Sigma}|}{\partial \boldsymbol{\Lambda}_s} = -\frac{1}{2} \boldsymbol{\Delta}_s^{-1}\boldsymbol{e} \tag{34}$$

and

$$\begin{aligned} &\frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)^\top (\boldsymbol{\Delta}_s\boldsymbol{\Sigma})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big\}}{\partial \boldsymbol{\Lambda}_s} \\ &= -\boldsymbol{\Delta}_s^{-2} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s). \end{aligned} \tag{35}$$

Taking Equation (34) and (35) into Equation (33), we can get the derivatives of $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\Lambda}_s$ as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\Lambda}_s} &= w_s \boldsymbol{\Delta}_s^{-1} \hat{\boldsymbol{\lambda}} + w_s \sum_{n=1}^{N_s} \Big[-\frac{1}{2} \boldsymbol{\Delta}_s^{-1}\boldsymbol{e} \\ &+ \frac{1}{2} \boldsymbol{\Delta}_s^{-2} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big]. \end{aligned} \tag{36}$$

Setting this derivative to zero, multiplying $\boldsymbol{\Delta}_s^2$ on both sides and then rearranging the equation, we can get

$$\begin{aligned} \boldsymbol{\Lambda}_s &= \boldsymbol{\Delta}_s \cdot \boldsymbol{e} = \frac{1}{N_s \cdot \boldsymbol{e} - 2\hat{\boldsymbol{\lambda}}} \\ &\circ \Big[\boldsymbol{\Sigma}^{-1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \Big]. \end{aligned} \tag{37}$$

Since $\prod_{s=1}^{S} (\boldsymbol{\Delta}_s)^{w_s} = \text{I}$, and $\hat{\boldsymbol{\lambda}}$ can't be solved in a closed form, so we will use the Newton method to solve $\hat{\boldsymbol{\lambda}}$ numerically. For each item $\hat{\lambda}_i$ of $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p)^\top \in \mathbb{R}^p$, define

$$\begin{aligned} A_{is} &= (N_s - 2\hat{\lambda}_i)^{w_s}, \\ B_{is} &= \Big[\sigma_i^{-1} \sum_{n=1}^{N_s} (x_{in}^{(s)} - \mu_i - \gamma_{is})^2 \Big]^{w_s}, \end{aligned} \tag{38}$$

then we need to find $\hat{\lambda}_i$ s.t. $\prod_{s=1}^{S} (\delta_{is})^{w_s} = \prod_{s=1}^{S} \frac{B_{is}}{A_{is}} = 1$. Define

$$h(\hat{\lambda}_i) = \prod_{s=1}^{S} B_{is} - \prod_{s=1}^{S} A_{is}, \tag{39}$$

then the derivative of $h(\hat{\lambda}_i)$ with respect to $\hat{\lambda}_i$ is

$$h^{'}(\hat{\lambda}_i) = 2\big(\prod_{s=1}^{S} A_{is} \big) \big(\sum_{s=1}^{S} \frac{w_s}{N_s - 2\hat{\lambda}_i} \big). \tag{40}$$

Then $\hat{\lambda}_i,\ i = 1, 2, ..., p$ can be numerically calculated by Newton method as described in Algorithm 1 as below.

Algorithm 1 Newton method to solve λ s.t. $h(\lambda) = 0$.

Input: a positive number ϵ which is small enough, and the initialization value of λ denoted by $\lambda^{(1)}$.

Output: numerical solution λ^* s.t. $h(\lambda^*) \approx 0$, namely, $\|h(\lambda^*)\| < \epsilon$.

for t=1,2,... **do**
  $\lambda^{(t+1)} = \lambda^{(t)} - \frac{h(\lambda^{(t)})}{h'(\lambda^{(t)})}$
  if $\|h(\lambda^{(t+1)})\| \leq \epsilon$ **then**
    $\lambda^* = \lambda^{(t+1)}$
    break
  end if
end for
return λ^*

In conclusion, we can update $\boldsymbol{\Delta}_s = \text{diag}(\boldsymbol{\Lambda}_s)$, where $\boldsymbol{\Lambda}_s$ can be calculated by Equation (37), and the Lagrange multiplier parameters $\hat{\lambda}_i,\ i = 1, 2, ..., p$ in Equation (37) can be numerically calculated by Algorithm 1.

So far, we have finished the derivation of the formulas to update the parameters in model (8) with constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$ following a coordinate descent optimization strategy. The algorithm can be summarized as below in Algorithm 2.

Algorithm 2 Estimate the parameters of model (8) with constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$.

Input: For each setting class $s \in \{1, 2, ..., S\}$, we have N_s observed data $\{\boldsymbol{x}_1^{(s)}, ..., \boldsymbol{x}_{N_s}^{(s)}\}$. Let ϵ be a positive number that is small enough.

Output: Estimated value $\boldsymbol{\theta}^*$ of the set of model parameters $\boldsymbol{\theta}$ defined by Equation (9) to maximize the logarithm likelihood function (17) under the following constraints:

$$\sum_{s=1}^{S} w_s \boldsymbol{\gamma}_s = \mathbf{0}, \quad \prod_{s=1}^{S} (\boldsymbol{\Delta}_s)^{w_s} = \mathrm{I}, \text{ where } w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}.$$

Initialization. Initialize the parameters $\boldsymbol{\Delta}_s^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)}$ as below, and initialize the logarithm likelihood function $\mathcal{L}_0(\boldsymbol{\theta})^{(0)} = -\infty$.

$$\begin{aligned}
&\boldsymbol{\Delta}_s^{(0)} = \mathrm{I},\ s = 1, 2, ..., S.\\
&\boldsymbol{\mu}^{(0)} = \sum_{s=1}^{S} w_s \boldsymbol{\mu}_s, \text{ where } \boldsymbol{\mu}_s = \frac{1}{N_s} \sum_{n=1}^{N_s} \boldsymbol{x}_n^{(s)}.\\
&\boldsymbol{\Sigma}^{(0)} = \sum_{s=1}^{S} w_s \boldsymbol{\Sigma}_s, \qquad (41)\\
&\text{where } \boldsymbol{\Sigma}_s = \text{diag}(\boldsymbol{\Gamma}_s)\\
&\text{and } \boldsymbol{\Gamma}_s = \frac{1}{N_s - 1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s).
\end{aligned}$$

for k=1,2,... **do**
  update the parameters $\boldsymbol{\gamma}_s^{(k)}, \boldsymbol{\Delta}_s^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}$ by $\boldsymbol{\gamma}_s^{(k-1)}, \boldsymbol{\Delta}_s^{(k-1)}, \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Sigma}^{(k-1)}$ as follows:

- update $\boldsymbol{\gamma}_s^{(k)} \in \mathbb{R}^p,\ s = 1, 2, ..., S$:

$$\begin{aligned}
\boldsymbol{\gamma}_s^{(k)} &= \frac{1}{N_s} (\boldsymbol{\Delta}_s^{(k-1)} \boldsymbol{\Sigma}^{(k-1)})\\
&\cdot \Big[\sum_{n=1}^{N_s} (\boldsymbol{\Delta}_s^{(k-1)} \boldsymbol{\Sigma}^{(k-1)})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k-1)}) + \tilde{\boldsymbol{\lambda}} \Big],
\end{aligned} \qquad (42)$$

where

$$\tilde{\boldsymbol{\lambda}} = -\Big[\sum_{s=1}^{S} \boldsymbol{\Delta}_s^{(k-1)} \boldsymbol{\Sigma}^{(k-1)} \Big]^{-1} \cdot \Big[\sum_{s=1}^{S} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k-1)}) \Big]. \qquad (43)$$

- update $\boldsymbol{\Delta}_s^{(k)} \in \mathbb{R}^{p \times p},\ s = 1, 2, ..., S$ by $\boldsymbol{\Delta}_s^{(k)} = \text{diag}(\boldsymbol{\Lambda}_s^{(k)})$, where

$$\begin{aligned}
\boldsymbol{\Lambda}_s^{(k)} &= \frac{1}{N_s \cdot \boldsymbol{e} - 2\hat{\boldsymbol{\lambda}}} \circ \Big[(\boldsymbol{\Sigma}^{(k-1)})^{-1}\\
&\cdot \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k-1)} - \boldsymbol{\gamma}_s^{(k)}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k-1)} - \boldsymbol{\gamma}_s^{(k)}) \Big],
\end{aligned} \qquad (44)$$

and each element of $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p)^\top$ can be solved numerically by Algorithm 1.

- update $\boldsymbol{\mu}^{(k)} \in \mathbb{R}^p$:

$$\begin{aligned}
\boldsymbol{\mu}^{(k)} &= \Big[\sum_{s=1}^{S} w_s N_s (\boldsymbol{\Delta}_s^{(k)})^{-1} \Big]^{-1}\\
&\cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} (\boldsymbol{\Delta}_s^{(k)})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_s^{(k)}) \Big].
\end{aligned} \qquad (45)$$

- update $\boldsymbol{\Sigma}^{(k)} \in \mathbb{R}^{p \times p}$ by $\boldsymbol{\Sigma}^{(k)} = \text{diag}(\boldsymbol{\Gamma}^{(k)})$, where

$$\begin{aligned}
\boldsymbol{\Gamma}^{(k)} &= \Big[\sum_{s=1}^{S} w_s N_s \Big]^{-1}\\
&\cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} (\boldsymbol{\Delta}_s^{(k)})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k)} - \boldsymbol{\gamma}_s^{(k)})\\
&\qquad \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{(k)} - \boldsymbol{\gamma}_s^{(k)}) \Big].
\end{aligned} \qquad (46)$$

  calculate the logarithm likelihood function

$$\mathcal{L}_0(\boldsymbol{\theta})^{(k)} = \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \ln f(\boldsymbol{x}_n^{(s)};\ \boldsymbol{\mu}^{(k)} + \boldsymbol{\gamma}_s^{(k)}, \boldsymbol{\Delta}_s^{(k)} \boldsymbol{\Sigma}^{(k)}), \qquad (47)$$

  if $\mathcal{L}_0(\boldsymbol{\theta})^{(k)} - \mathcal{L}_0(\boldsymbol{\theta})^{(k-1)} \leq \epsilon$, **then**
    $\boldsymbol{\theta}^* = \{\boldsymbol{\mu}^{(k)},\ \boldsymbol{\Sigma}^{(k)},\ \boldsymbol{\gamma}_s^{(k)},\ \boldsymbol{\Delta}_s^{(k)} \mid s = 1, ...S\}$,
    break
  end if
end for
return $\boldsymbol{\theta}^*$

C. Estimate model parameters with constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

Similar to subsection III-B, in this subsection, we will use the MLE method to estimate the parameters of model (8), but with constraints imposed on parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by Equation (13) and (14). The set of the model parameters is defined as $\boldsymbol{\theta}$ by Equation (9).

To begin with, we consider the case when there is no parameter constraints, and calculate the formulas to maximize the logarithm likelihood function $\mathcal{L}_0(\boldsymbol{\theta})$ expressed by (17) following the coordinate descent optimization strategy. The results are listed as below.

- **Update $\boldsymbol{\mu} \in \mathbb{R}^p$:**

$$\boldsymbol{\mu} = \Big[\sum_{s=1}^{S} w_s N_s \boldsymbol{\Delta}_s^{-1}\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \boldsymbol{\Delta}_s^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_s)\Big]. \tag{48}$$

- **Update $\boldsymbol{\gamma}_s \in \mathbb{R}^p, s = 1, 2, ..., S$:**

$$\boldsymbol{\gamma}_s = \frac{1}{N_s} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}). \tag{49}$$

- **Update $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ by $\boldsymbol{\Sigma} = \mathbf{diag}(\boldsymbol{\Gamma})$:**

$$\begin{aligned}\boldsymbol{\Gamma} = &\Big[\sum_{s=1}^{S} w_s N_s\Big]^{-1} \\ &\cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \boldsymbol{\Delta}_s^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)\Big].\end{aligned} \tag{50}$$

- **Update $\boldsymbol{\Delta}_s \in \mathbb{R}^{p\times p}, s = 1, 2, ..., S$ by $\boldsymbol{\Delta}_s = \mathbf{diag}(\boldsymbol{\Lambda}_s)$:**

$$\boldsymbol{\Lambda}_s = \frac{1}{N_s} \cdot \Big[\boldsymbol{\Sigma}^{-1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu} - \boldsymbol{\gamma}_s)\Big]. \tag{51}$$

In general, in the same spirit than a coordinate descent optimization method, the parameters will be updated iteratively by the above equations to maximize the logarithm likelihood function $\mathcal{L}_0(\boldsymbol{\theta})$ expressed by Equation (17). But in our case, we can prove that the logrithm function $\mathcal{L}_0(\boldsymbol{\theta})$ will stop change and reach its local maxima after one iteration. The detail proof is given by first initializing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as $\boldsymbol{\mu}^{\text{old}}$ and $\boldsymbol{\Sigma}^{\text{old}}$, then $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$ can be updated as updated as $\boldsymbol{\gamma}_s^{\text{new}}$ and $\boldsymbol{\Delta}_s^{\text{new}}$ as follows:

$$\boldsymbol{\gamma}_s^{\text{new}} = \frac{1}{N_s} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{\text{old}}), \qquad \boldsymbol{\Delta}_s^{\text{new}} = \text{diag}(\boldsymbol{\Lambda}_s^{\text{new}}), \tag{52}$$

where

$$\begin{aligned}\boldsymbol{\Lambda}_s^{\text{new}} = &\frac{1}{N_s} \cdot \Big[(\boldsymbol{\Sigma}^{\text{old}})^{-1} \\ &\cdot \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{\text{old}} - \boldsymbol{\gamma}_s^{\text{new}}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{\text{old}} - \boldsymbol{\gamma}_s^{\text{new}})\Big].\end{aligned} \tag{53}$$

Re-update $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ as $\boldsymbol{\mu}^{\text{new}}$ and $\boldsymbol{\Gamma}^{\text{new}}$ as follows:

$$\begin{aligned}\boldsymbol{\mu}^{\text{new}} &= \Big[\sum_{s=1}^{S} w_s N_s (\boldsymbol{\Delta}_s^{\text{new}})^{-1}\Big]^{-1} \\ &\quad \cdot \Big[\sum_{s=1}^{S} w_s (\boldsymbol{\Delta}_s^{\text{new}})^{-1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_s^{\text{new}})\Big] \\ &= \Big[\sum_{s=1}^{S} w_s N_s (\boldsymbol{\Delta}_s^{\text{new}})^{-1}\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s (\boldsymbol{\Delta}_s^{\text{new}})^{-1} \cdot N_s \cdot \boldsymbol{\mu}^{\text{old}}\Big] \\ &= \boldsymbol{\mu}^{\text{old}},\end{aligned} \tag{54}$$

and

$$\begin{aligned}\boldsymbol{\Gamma}^{\text{new}} &= \Big[\sum_{s=1}^{S} w_s N_s\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s (\boldsymbol{\Delta}_s^{\text{new}})^{-1} \\ &\quad \cdot \sum_{n=1}^{N_s} (\boldsymbol{x}_n - \boldsymbol{\mu}^{\text{new}} - \boldsymbol{\gamma}_s^{\text{new}}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^{\text{new}} - \boldsymbol{\gamma}_s^{\text{new}})\Big] \\ &= \Big[\sum_{s=1}^{S} w_s N_s\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s N_s \cdot \boldsymbol{\Gamma}^{\text{old}}\Big] \\ &= \boldsymbol{\Gamma}^{\text{old}},\end{aligned} \tag{55}$$

then we can conclude that $\boldsymbol{\mu}^{\text{new}} = \boldsymbol{\mu}^{\text{old}}$ and $\boldsymbol{\Sigma}^{\text{new}} = \boldsymbol{\Sigma}^{\text{old}}$. So the logarithm likelihood function $\mathcal{L}_0(\boldsymbol{\theta})$ will stop change and reach its local maxima after one iteration. In this case, we don't need to use the Lagrange maximize $\mathcal{L}_0(\boldsymbol{\theta})$ with constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and the constraints can be imposed directly by initializing the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as additive and multiplicative constraints. To conclude, the algorithm to estimate the parameters of model model (8) with constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be summarized as below.

Algorithm 3 Estimate the parameters of model (8) with constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Input: For each setting class $s \in \{1, 2, ..., S\}$, we have N_s observed data $\{\boldsymbol{x}_1^{(s)}, ..., \boldsymbol{x}_{N_s}^{(s)}\}$.

Output: Estimated value $\boldsymbol{\theta}^*$ of the set of model parameters $\boldsymbol{\theta}$ defined by Equation (9) to maximize the logarithm likelihood function (17) under the constraints (13) and (14).

1: The set of the estimated model parameters denoted by $\boldsymbol{\theta}^* = \{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \boldsymbol{\gamma}_s^*, \boldsymbol{\Delta}_s^* \mid s = 1, ...S\}$ would be calculated directly by the following formulas.

- Calculate $\boldsymbol{\mu}^* \in \mathbb{R}^p$:

$$\boldsymbol{\mu}^* = \sum_{s=1}^{S} w_s \boldsymbol{\mu}_s, \tag{56}$$

where

$$\boldsymbol{\mu}_s = \frac{1}{N_s} \sum_{n=1}^{N_s} \boldsymbol{x}_n^{(s)}, \; w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}, \quad s = 1, 2, ..., S. \tag{57}$$

- Calculate $\boldsymbol{\Sigma}^* \in \mathbb{R}^{p\times p}$:

$$\boldsymbol{\Sigma}^* = \sum_{s=1}^{S} w_s \boldsymbol{\Sigma}_s, \tag{58}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_s &= \text{diag}(\boldsymbol{\Gamma}_s), \\ \boldsymbol{\Gamma}_s &= \frac{1}{N_s - 1} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_s). \end{aligned} \tag{59}$$

- Calculate $\boldsymbol{\gamma}_s^* \in \mathbb{R}^p,\ s = 1, 2, ..., S$:

$$\boldsymbol{\gamma}_s^* = \frac{1}{N_s} \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^*). \tag{60}$$

- Calculate $\boldsymbol{\Delta}_s^* \in \mathbb{R}^{p\times p},\ s = 1, 2, ..., S$:

$$\boldsymbol{\Delta}_s^* = \text{diag}(\boldsymbol{\Lambda}_s^*) \tag{61}$$

where

$$\begin{aligned} \boldsymbol{\Lambda}_s^* = \frac{1}{N_s} \cdot \Big[&(\boldsymbol{\Sigma}^*)^{-1} \\ &\cdot \sum_{n=1}^{N_s} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^* - \boldsymbol{\gamma}_s^*) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}^* - \boldsymbol{\gamma}_s^*) \Big]. \end{aligned} \tag{62}$$

2: **return** $\boldsymbol{\theta}^* = \{\boldsymbol{\mu}^*,\ \boldsymbol{\Sigma}^*,\ \boldsymbol{\gamma}_s^*,\ \boldsymbol{\Delta}_s^* \mid s = 1, ...S\}$.

D. Summary of the Proposed Method

As described above, we used two ways to estimate the parameters of model (8), namely, 1) with constraints on $\boldsymbol{\gamma}_s$, $\boldsymbol{\Delta}_s$, and 2) with constraints on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$. In essence, these two ways of imposing parameter constraints are similar, and they all aim to adjust the mean and variance of the harmonized data to be the weighted center and weighted variance of the clusters from all scanner settings. But, obviously, imposing constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is much simpler and faster than imposing constraints on $\boldsymbol{\gamma}_s$ and $\boldsymbol{\Delta}_s$. So we choose to impose constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and use Algorithm 3 as our finally proposed method. We name it ComBat-MLE, as it follows the ComBat method's principle of modeling scanner effects additively and multiplicatively, albeit with a different model development method.

IV. Experiments

A. Dataset

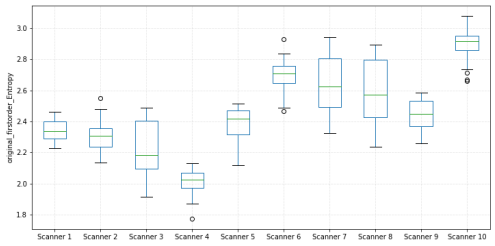
Ammari et al. [18] designed a homogeneous phantom to mimic cerebrospinal fluid and opacified blood vessels, and a heterogeneous phantom to mimic brain white matter. These phantoms underwent MRI scanning under ten distinct scanner configurations, employing two MRI scanners with varying FOVs and matrices, to acquire their T1 sequences [18], [19]. Nyúl normalization [20], [21] was used to normalize these MRI images [19]. Subsequently, we extracted nine circular Region of Interests (ROIs) from 70 continuous slices of the homogeneous phantom, and six circular ROIs from 30 continuous slices of the heterogeneous phantom. 92 radiomic features were then extracted from each 2D ROI in the MRI images using *PyRadiomics*.

Notably, these ROIs correspond to the same phantom tubes, ensuring that any observed differences in radiomic features between scanner settings are solely due to variations in image acquisition parameters, such as magnetic field strengths, FOVs, and matrices. These features are used to assess the presence and elimination of scanner effects. We made these extracted features publicly available to support researchers in testing their harmonization methods. More details can be found at *https://github.com/Yingping-LI/Harmonization_methods*.

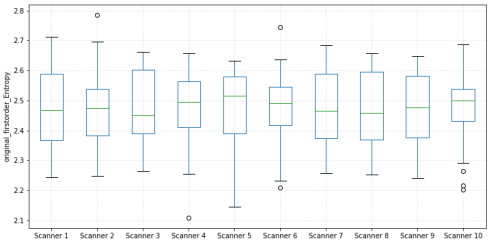
B. Existence of scanner effects and effectiveness of our proposed ComBat-MLE

We successfully extracted 92 radiomic features from each ROI collected under different scanner settings. To investigate the presence of scanner effects, we visualized these features using box plots. Fig. 1(a) displays the box plots of an example feature from the heterogeneous phantom. Obviously, the same phantom scanned by different scanner settings shows distinct mean and variance differences for the extracted features, indicating the existence of scanner effects.

After applying our ComBat-MLE method to each pattern class, most scanner effects were effectively removed. Fig. 1(b) shows the box plots of the harmonized features corresponding to those in Fig. 1(a). After harmonization, the mean and variance of the features are similar across all scanner settings, confirming the effectiveness of our proposed ComBat-MLE method in removing scanner effects.



(a) Before harmonization



(b) After harmonization by our proposed ComBat-MLE

Fig. 1. Box plots to demonstrate the effectiveness of our ComBat-MLE method to remove scanner effects.

TABLE I
STATISTICS ON RATIO OF FEATURES WITH DIFFERENT DISTRIBUTIONS.

Harmonization method	Ratio of features with different distributions	
	homogeneous phantom	heterogeneous phantom
Before harmonization	78.73%	59.49%
Non-parametric ComBat	13.84%	6.88%
Parametric ComBat	0.73%	0.36%
ComBat-MoM ComBat	0.25%	0.15%
Our proposed ComBat-MLE	0.25%	0.15%

TABLE II
CLASSIFICATION ACCURACY BY LDA WITH FEATURE DATA HARMONIZED BY DIFFERENT METHODS.

Harmonization methods	Classification accuracy	
	homogeneous phantom	heterogeneous phantom
Before harmonization	70.78%	91.83%
Non-parametric ComBat	**91.78%**	98.33%
Parametric ComBat	87.14%	97.56%
ComBat-MoM	88.70%	**99.83%**
Our proposed ComBat-MLE	88.38%	99.72%

C. Quantitative Comparison Results

1) Statistics on features with different distributions: To quantitatively assess the effectiveness of these harmonization methods to remove scanner effects, statistic tests were applied to see whether the features extracted from the same phantom but scanned with different scanner settings have the same distribution. In detail, a Friedman test [22] was used for the comparison between three feature distributions (for example, features from matrix 256×256, 256×128 and 128×128 pixels), whereas the Wilcoxon test [23] was used for the comparison between two distributions (for example, features from 1.5T and 3T MRI). Then Bonferroni correction [24]–[26] was then used for multiple testing corrections. In our study, a P value less than 0.05 indicates significantly different feature distributions, while a P value greater than 0.05 suggests the absence of scanner effects.

TABLE I displays the statistical results on the ratio of features which have different distributions across different scanner settings among the 92 features. A smaller ratio indicates fewer scanner effects. For homogeneous phantom, 78.73% features have different distributions between different scanner settings, and after harmonization by our ComBat-MLE, only 0.25% features are left to have different distributions, demonstrating that most of the scanner effects have been removed. For heterogeneous phantom, only 0.15% features have different distributions compared to 59.49% before harmonization. The results by other harmonization methods are also listed, and we can see that our proposed ComBat-MLE method works similar with ComBat-MoM, and works better than non-parametric ComBat and parametric ComBat.

2) Classification of different pattern classes: For harmonization methods, we not only care about their ability to remove the non-biological scanner effects, but also care about whether they retain the intrinsic biological information, thus to help the downstream analysis. So we use a supervised classification task as an example to further explain the power of these harmonization methods. Linear discriminant analysis (LDA) [27], [28], a popular supervised dimension reduction method that maximizes the projection coordinates to predict the data classes, was used. The classification results of different pattern classes (i.e. phantom tubes) are listed in TABLE II. Evidently, the classification accuracy improves significantly for both the homogeneous and heterogeneous phantom after applying these harmonization methods.

D. Qualitative Comparison Results

1) Toy Examples: A toy example is given to verify the effectiveness of our proposed ComBat-MLE method. In Fig. 2, different colors correspond to data from different classes, and different clusters with the same color correspond to data from the same class but from different scanner settings. As shown in Fig. 2, our proposed ComBat-MLE can successfully remove the variations caused by different scanner settings, and at the same time preserve the intrinsic data information which helps to distinguish different classes. For this toy example, our proposed method works similar to the ComBat-MoM method, and performs better than parametric and non-parametric ComBat.

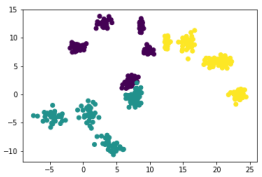
2) Visualizing two example features in a 2D plane: Fig. 3 visually illustrates the harmonization results by plotting two example radiomic features from the homogeneous phantom in a 2D plane. As depicted, the four harmonization methods exhibit similar results. That is, after harmonization, each texture pattern (represented by different colors) is no longer composed of several separate and different clusters, while at the same time better distinguishable between texture patterns. These findings underscore the efficacy of these harmonization methods in removing non-biological scanner effects while preserving the intrinsic biological information.

3) Visualizing LDA-transformed features in a 2D plane: LDA was applied to transform the features into a two-dimensional subspace while maximizing the separation between classes. Fig. 4 visualizes the LDA-transformed data to help compare the separability of pattern classes for homogeneous phantom. As shown, after applying these harmonization methods, the different pattern classes (represented by different colors) become more clearly distinguishable.

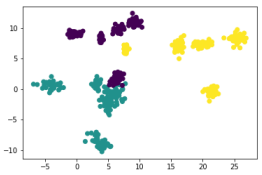
V. DISCUSSION

A. Summary

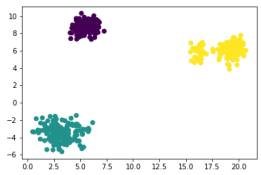
In this paper, we re-describe the state-of-the-art data harmonization method ComBat in another perspective and then discuss two different sets of parameter constraints to help identify the parameters. That is, one with constraints on the additive ($\boldsymbol{\gamma}_s$) and multiplicative ($\boldsymbol{\Delta}_s$) scanner effects, and another with constraints on the mean ($\boldsymbol{\mu}$) and covariance ($\boldsymbol{\Sigma}$) of the clear data without scanner effects. In essence, both of these two sets of parameter constraints aim to adjust the mean



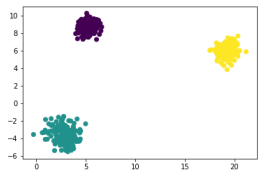
(a) Before harmonization



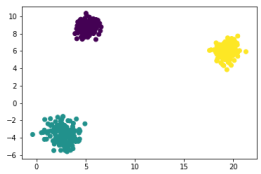
(b) Harmonized by non-parametric ComBat



(c) Harmonized by parametric ComBat

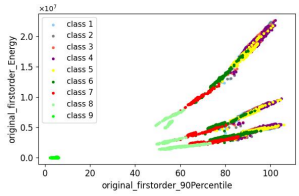


(d) Harmonized by ComBat-MoM

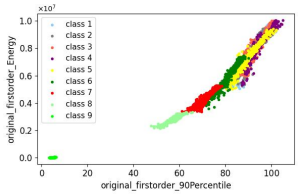


(e) Harmonized by our proposed ComBat-MLE

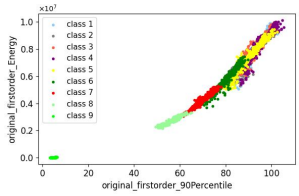
Fig. 2. Toy examples to test different harmonization methods. The data, which is generated by scikit-learn data generator 'make_blobs' in *Python*, consists of data from 5 scanner settings. For each scanner setting, 100 samples composed of 5 feature attributes and 3 clusters are randomly generated. Then additive (randomly chosen between 0 and 10) and multiplicative (randomly chosen between 0.05 and 1.95) scanner effects are generated for each scanner setting and then attached to all the generated data in this scanner setting. The 2D plots show the first two features. Different colors correspond to data from different classes, and different clusters with the same color correspond to data from the same class but from different scanner settings.



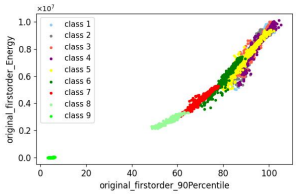
(a) Before harmonization



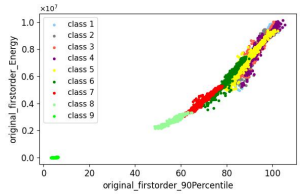
(b) Harmonized by non-parametric ComBat



(c) Harmonized by parametric ComBat



(d) Harmonized by ComBat-MoM



(e) Harmonized by our proposed ComBat-MLE

Fig. 3. Visualizing two example features in a 2D plane before and after harmonized by different harmonization methods.

and variance of the harmonized data to the weighted center and variance of the clusters from all scanner settings. Then we solved the model using MLE with each set of constraints. The latter constraints proved simpler and faster, leading us to propose our ComBat-MLE harmonization method.

Then, experiments were conducted to compare the efficacy of different ComBat variants, including the proposed ComBat-MLE. In some cases, ComBat-MLE and ComBat-MoM outperformed the widely used non-parametric and parametric ComBat methods. This may be because both ComBat-MoM and ComBat-MLE assume that features undergo independent scanner effects. In contrast, the parametric and non-parametric ComBat methods make a stricter assumption to borrow information across features, which is helpful for high-dimensional data with a small sample size. However, this stricter assumption—that after standardization, the scanner effects of all features follow a certain distribution—may contradict the real data and lead to worse harmonization results.

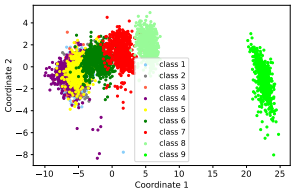
B. *Future work*

These ComBat variants, including our proposed ComBat-MLE, assume that scanner effects are consistent for features extracted from the same scanner setting, regardless of the pattern class. However, this assumption does not always hold true. In such cases, it becomes impossible to completely remove the scanner effects. Orlhac *et al.* also mention this type of failure case in their paper [10].

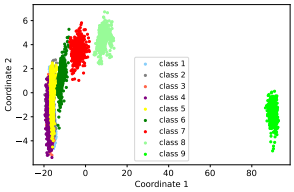
Future work could concentrate on constructing a model under a new assumption – that scanner effects vary across different pattern classes (like tumors and normal tissues), even when scanned with the same scanner setting. While our proposed ComBat-MLE may not exhibit a significant improvement over existing ComBat variants, it offers valuable insights for harmonization model development. Based on ComBat-MLE and this new assumption, we can reformulate the model as a Gaussian Mixture Model (GMM). Namely, the observed data $\boldsymbol{X}_s \in \mathbb{R}^p$ suffering scanner effects from scanner setting s can be assumed to follow the following Gaussian mixture distribution:

$$p(\boldsymbol{X}_s) = \sum_{c=1}^{C} \pi_{sc}\, \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c), \qquad (63)$$

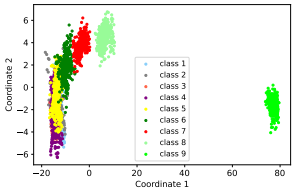
where $\boldsymbol{\mu}_c \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}_c \in \mathbb{R}^{p\times p}$ correspond to the mean and variance matrix of the c-th pattern class. The terms $\boldsymbol{\gamma}_{sc} \in \mathbb{R}^p$



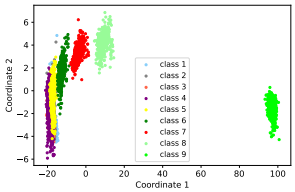
(a) Before harmonization



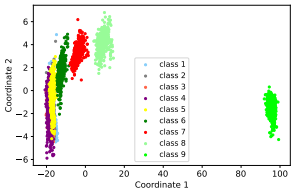
(b) Harmonized by ComBat-NonParametric



(c) Harmonized by ComBat-Parametric



(d) Harmonized by ComBat-MoM



(e) Harmonized by our proposed ComBat-MLE

Fig. 4. Visualizing LDA-transformed features in a 2D plane before and after harmonized by different harmonization methods.

and $\boldsymbol{\Delta}_{sc} \in \mathbb{R}^{p\times p}$ represent the additive and multiplicative scanner effects, respectively. Additionally,

$$0 \leq \pi_{sc} \leq 1, \quad \sum_{c=1}^{C} \pi_{sc} = 1 \tag{64}$$

holds for s=1,2,...,S.

The model (63) can be solved by the Expectation Maximization (EM) algorithm [29], [30]. Note that, parameter constraints are essential to help identify between the parameters. For example, for model (63), we propose the following additive and multiplicative constraints:

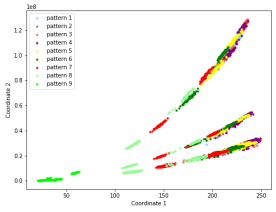
- additive constraints:

$$\sum_{s=1}^{S} w_s \boldsymbol{\gamma}_{sc} = \mathbf{0}, \ c = 1, 2, ..., C. \tag{65}$$

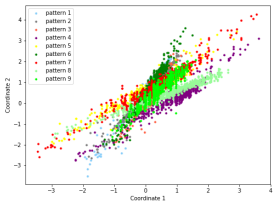
- multiplicative constraints:

$$\prod_{s=1}^{S} (\boldsymbol{\Delta}_{sc})^{w_s} = \mathrm{I}, \ c = 1, 2, ..., C. \tag{66}$$

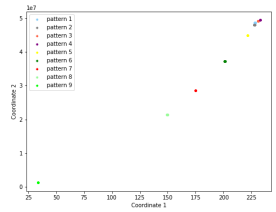
With these additive and multiplicative constraints, after removing scanner effects, the mean (variance) of pattern class c will be the weighted center (variance) of predicted pattern classes of c in all scanner settings, where $w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}$, $s = 1, 2, ..., S$ are the weight coefficients.



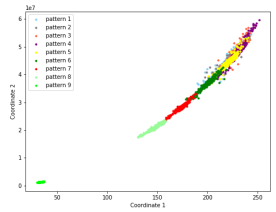
(a) Before Harmonization



(b) After harmonization (without constraints)



(c) After harmonization (with additive constraints only)



(d) After harmonization (with additive and multiplicative constraints)

Fig. 5. Effects of the additive and multiplicative parameter constraints.

We highlight the significance of the parameter constraints. Taking model (63) as an example, when considering only one pattern class, the model simplifies to our proposed ComBat-MLE model, with constraints (65) and (66) adequate for its resolution. By applying model (63) with $C = 1$ to each pattern class separately, as shown in Fig. 5, we can conclude that the additive and multiplicative constraints are essential, as they enhance parameter identifiability and ensure the adjusted data exhibit reasonable means and variances.

However, these constraints are not sufficient for model (63) with pattern class $C > 1$. More specifically, in model (63), we can see that for a given scanner setting s, the mean of the c-th Gaussian distribution is $\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}$, but $\boldsymbol{\mu}_c$ and $\boldsymbol{\gamma}_{sc}$ can not be identified. In other words, each point from scanner setting s can be classified to an arbitrary pattern class just by defining a different additive scanner effects $\boldsymbol{\gamma}_{sc}$. Similar problem exists for identify the variance $\boldsymbol{\Sigma}_c$ of pattern class c and the multiplicative scanner effects $\boldsymbol{\Delta}_{sc}$. As a future work, more proper constraints on parameters can be investigated.

VI. Conclusion

This paper provides a concise re-description of the ComBat method and discusses two sets of parameter constraints to help identify the parameters in the parameter estimation process. Then we solved the re-described ComBat method using Maximum Likelihood Estimation (MLE) with each set of constraints. The simpler and faster solution was chosen as our proposed ComBat-MLE harmonization method. Our experiments validated the power of the proposed ComBat-MLE harmonization method in removing non-biological scanner effects while keeping the intrinsic biological information. Although our method did not significantly outperform existing ComBat variants, it enhances the understanding of the ComBat method and serves as a foundation for developing

new harmonization methods, offering valuable insights into harmonization model development.

REFERENCES

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

[2] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.

[3] J.-P. Fortin, D. Parker, B. Tunç, T. Watanabe, M. A. Elliott, K. Ruparel, D. R. Roalf, T. D. Satterthwaite, R. C. Gur, R. E. Gur, *et al.*, "Harmonization of multi-site diffusion tensor imaging data," *Neuroimage*, vol. 161, pp. 149–170, 2017.

[4] F. Orlhac, S. Boughdad, C. Philippe, H. Stalla-Bourdillon, C. Nioche, L. Champion, M. Soussan, F. Frouin, V. Frouin, and I. Buvat, "A postreconstruction harmonization method for multicenter radiomic studies in pet," *Journal of Nuclear Medicine*, vol. 59, no. 8, pp. 1321–1328, 2018.

[5] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, and I. Buvat, "Validation of a method to compensate multicenter effects affecting ct radiomics," *Radiology*, vol. 291, no. 1, pp. 53–59, 2019.

[6] M.-J. Saint Martin, F. Orlhac, P. Akl, F. Khalid, C. Nioche, I. Buvat, C. Malhaire, and F. Frouin, "A radiomics pipeline dedicated to breast mri: validation on a multi-scanner phantom study," *Magnetic Resonance Materials in Physics, Biology and Medicine*, pp. 1–12, 2020.

[7] F. Orlhac, A. Lecler, J. Savatovski, J. Goya-Outi, C. Nioche, F. Charbonneau, N. Ayache, F. Frouin, L. Duron, and I. Buvat, "How can we combat multicenter variability in mr radiomics? validation of a correction procedure," *European Radiology*, pp. 1–9, 2020.

[8] R. Mahon, M. Ghita, G. Hugo, and E. Weiss, "Combat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets," *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015010, 2020.

[9] C. T. Arendt, D. Leithner, M. E. Mayerhoefer, P. Gibbs, C. Czerny, C. Arnoldner, I. Burck, M. Leinung, Y. Tanyildizi, L. Lenga, *et al.*, "Radiomics of high-resolution computed tomography for the differentiation between cholesteatoma and middle ear inflammation: effects of post-reconstruction methods in a dual-center study," *European Radiology*, vol. 31, no. 6, pp. 4071–4078, 2021.

[10] F. Orlhac, J. J. Eertink, A.-S. Cottereau, J. M. Zijlstra, C. Thieblemont, M. Meignan, R. Boellaard, and I. Buvat, "A guide to combat harmonization of imaging biomarkers in multicenter studies," *Journal of Nuclear Medicine*, vol. 63, no. 2, pp. 172–179, 2022.

[11] J. C. Beer, N. J. Tustison, P. A. Cook, C. Davatzikos, Y. I. Sheline, R. T. Shinohara, K. A. Linn, A. D. N. Initiative, *et al.*, "Longitudinal combat: A method for harmonizing longitudinal multi-scanner imaging data," *Neuroimage*, vol. 220, p. 117129, 2020.

[12] A. A. Chen, J. C. Beer, N. J. Tustison, P. A. Cook, R. T. Shinohara, and H. Shou, "Removal of scanner effects in covariance improves multivariate pattern analysis in neuroimaging data," *bioRxiv*, p. 858415, 2020.

[13] R. Da-Ano, I. Masson, F. Lucia, M. Doré, P. Robin, J. Alfieri, C. Rousseau, A. Mervoyer, C. Reinhold, J. Castelli, *et al.*, "Performance comparison of modified combat for harmonization of radiomic features for multicenter studies," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[14] J. Radua, E. Vieta, R. Shinohara, P. Kochunov, Y. Quidé, M. J. Green, C. S. Weickert, T. Weickert, J. Bruggemann, T. Kircher, *et al.*, "Increased power by harmonizing structural mri site differences with the combat batch adjustment method in enigma," *NeuroImage*, vol. 218, p. 116956, 2020.

[15] R. Garcia-Dias, C. Scarpazza, L. Baecker, S. Vieira, W. H. Pinaya, A. Corvin, A. Redolfi, B. Nelson, B. Crespo-Facorro, C. McDonald, *et al.*, "Neuroharmony: A new tool for harmonizing volumetric mri data from unseen scanners," *NeuroImage*, vol. 220, 2020.

[16] A. Chatterjee, M. Vallières, A. Dohan, I. R. Levesque, Y. Ueno, S. Saif, C. Reinhold, and J. Seuntjens, "Creating robust predictive radiomic models for data from independent institutions using normalization," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 210–215, 2019.

[17] B. E. Dewey, C. Zhao, J. C. Reinhold, A. Carass, K. C. Fitzgerald, E. S. Sotirchos, S. Saidha, J. Oh, D. L. Pham, P. A. Calabresi, *et al.*, "Deepharmony: a deep learning approach to contrast harmonization across scanner changes," *Magnetic resonance imaging*, vol. 64, pp. 160–170, 2019.

[18] S. Ammari, S. Pitre-Champagnat, L. Dercle, E. Chouzenoux, S. Moalla, S. Reuze, H. Talbot, T. Mokoyoko, J. Hadchiti, S. Diffetocq, *et al.*, "Influence of magnetic field strength on magnetic resonance imaging radiomics features in brain imaging, an in vitro and in vivo study," *Frontiers in Oncology*, 2021.

[19] Y. Li, S. Ammari, C. Balleyguier, N. Lassau, and E. Chouzenoux, "Impact of preprocessing and harmonization methods on the removal of scanner effects in brain mri radiomic features," *Cancers*, vol. 13, no. 12, p. 3000, 2021.

[20] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of mri scale standardization," *IEEE transactions on medical imaging*, vol. 19, no. 2, pp. 143–150, 2000.

[21] M. Shah, Y. Xiao, N. Subbanna, S. Francis, D. L. Arnold, D. L. Collins, and T. Arbel, "Evaluating intensity normalization on mris of human brain with multiple sclerosis," *Medical image analysis*, vol. 15, no. 2, pp. 267–282, 2011.

[22] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[23] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in statistics*, pp. 196–202, Springer, 1992.

[24] J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference. Part I.* University of California Press, 2020.

[25] O. J. Dunn, "Multiple comparisons among means," *Journal of the American statistical association*, vol. 56, no. 293, pp. 52–64, 1961.

[26] R. A. Armstrong, "When to use the b onferroni correction," *Ophthalmic and Physiological Optics*, vol. 34, no. 5, pp. 502–508, 2014.

[27] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and information Processing*, vol. 18, no. 1998, pp. 1–8, 1998.

[28] P. Xanthopoulos, P. M. Pardalos, T. B. Trafalis, P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," *Robust data mining*, pp. 27–33, 2013.

[29] C. M. Bishop, "Pattern recognition and machine learning," *Springer google schola*, vol. 2, pp. 1122–1128, 2006.

[30] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.