Appendices to our MedAI 2024 Paper*

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

1 How we extract the radiomics features from the phantoms?

The features of the phantom data used in this MedAI 2024 study are the same as in our previously published paper [1]. For more details on feature extraction, please refer to [1].

1.1 Dataset

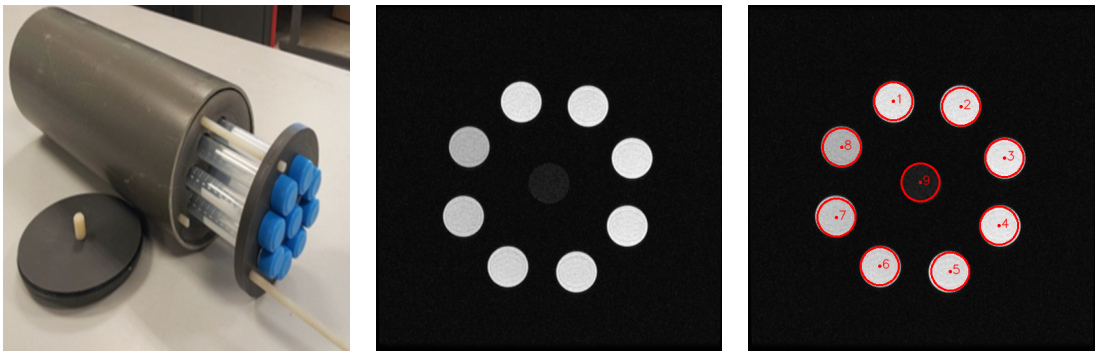
Our dataset contains MRI data of two kinds of phantoms (i.e. homogeneous phantom and heterogeneous phantom) designed by Ammari et al. [2], the details are described as below.

- Homogeneous Phantom

  Homogeneous phantom is designed to mimic cerebrospinal fluid and opacified blood vessels. It is defined with different gadolinium concentrations: eight 30-ml tubes were filled with demineralized water mixed with increasing gadolinium chelate concentration (Gadoteric Acid; Dotarem ®, Guerbet): 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75 and 2 mmol/l and a central demineralized water tube. The nine tubes were numbered as 1 to 9, as shown in Figure 1.

- Heterogeneous Phantom

  Heterogeneous phantom, designed to mimic the brain white matter, are composed of agarose gel coating polystyrene beads. The proton density of agarose has similar characteristics and



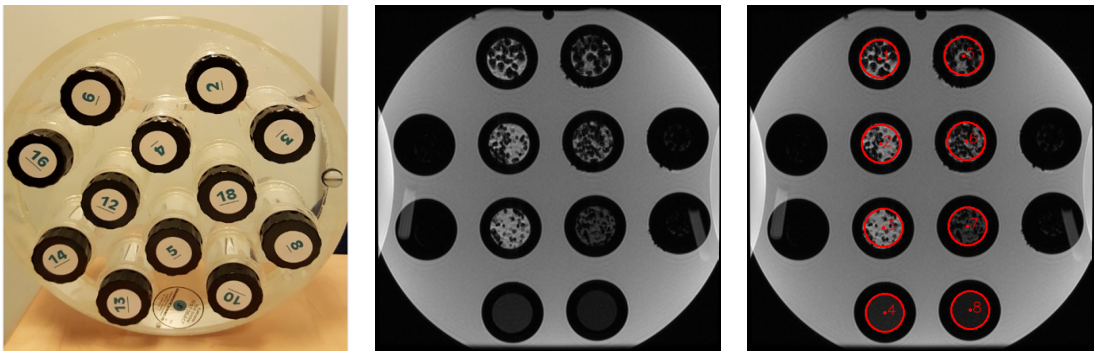
(a) Homogeneous phantom (b) T1 weighted MRI image (c) Number of the tubes

Figure 1: Homogeneous phantom , its T1 weighted MRI image and Number of tubes. Number 1-8 correspond to demineralized water mixed with 2, 1.75, 1.5, 1.25, 1, 0.75, 0.5 and 0.25 mmol/l gadolinium chelate concentration, and number 9 corresponds to pure demineralized water tube.

relaxation times to the biological tissue [3]. Six heterogeneous tubes and two homogeneous tubes are defined in the two central columns, numbered by tube 1 to 8, as shown in Figure 2. Six 30-ml tubes are filled with polystyrene beads of different diameters (1, 2 and 3 mm) in an agarose gel solution (2%), either pure or mixed with 0.25 mmol/l gadolinium chelate, and two 30-ml tubes are filled with agarose gel solution (2%) and two different concentration of gadolinium chelate [4], more details are as listed as follows:

- Tube 1: Agarose gel + 0.25 mmol/l gadolinuim chelate + large polystyrene beads (diameters range from 2 to 3.5 mm)
- Tube 2: Agarose gel + 0.25 mmol/l gadolinuim chelate + medium polystyrene beads (diameters range from 1.5 to 2.5 mm)
- Tube 3: Agarose gel + 0.25 mmol/l gadolinuim chelate + small polystyrene beads (diameters range from 0.9 to 1.5 mm)
- Tube 4: Agarose gel + 0.25 mmol/l gadolinuim chelate
- Tube 5: Agarose gel + large polystyrene beads (diameters range from 2.5 to 3.5 mm)
- Tube 6: Agarose gel + medium polystyrene beads (diameters range from 1.5 to 2.5 mm)
- Tube 7: Agarose gel + small polystyrene beads (diameters range from 0.9 to 1.5 mm)
- Tube 8: Pure agarose gel

The phantoms are scanned by different devices with different acquisition settings to introduce the scanner effects. We can simply summarize it as 10 scanner settings (different in magnetic field strength, field of view and matrix) as listed in Table 1.



(a) Heterogeneous phantom (b) T1 weighted MRI image (c) Number of the tubes

Figure 2: Heterogeneous phantom, its T1 weighted MRI image and Number of tubes.

Setting NO.	Scanner device	Magnetic field strength	FOV (cm)	Matrix (pixels)	Pixel size (mm)
1	Optima MR450w	1.5T	24	256×256	0.469 × 0.469 × 1
2	Optima MR450w	1.5T	24	256×128	0.469 × 0.469 × 1
3	Optima MR450w	1.5T	24	128×128	0.469 × 0.469 × 1
4	Optima MR450w	1.5T	18	256×256	0.352 × 0.352 × 1
5	Optima MR450w	1.5T	12	256×256	0.234 × 0.234 × 1
6	Discovery MR750w	3T	24	256×256	0.469 × 0.469 × 1
7	Discovery MR750w	3T	24	256×128	0.469 × 0.469 × 1
8	Discovery MR750w	3T	24	128×128	0.469 × 0.469 × 1
9	Discovery MR750w	3T	18	256×256	0.352 × 0.352 × 1
10	Discovery MR750w	3T	12	256×256	0.234 × 0.234 × 1

Table 1: Scanner settings used in homogeneous and heterogeneous phantom experiments.

1.2 Region of Interests (ROIs)

The homogenous phantom and heterogeneous phantom are scanned based on each acquisition setting in Table 1, so we have 10 3D MRI images for each phantom. Region of Interests (ROIs) will be extracted from certain slices of these 3D MRI images, that is, 51-th to 120-th slices for homogenous phantom and 21-th to 50-th slices for heterogeneous phantom will be used to extract ROIs.

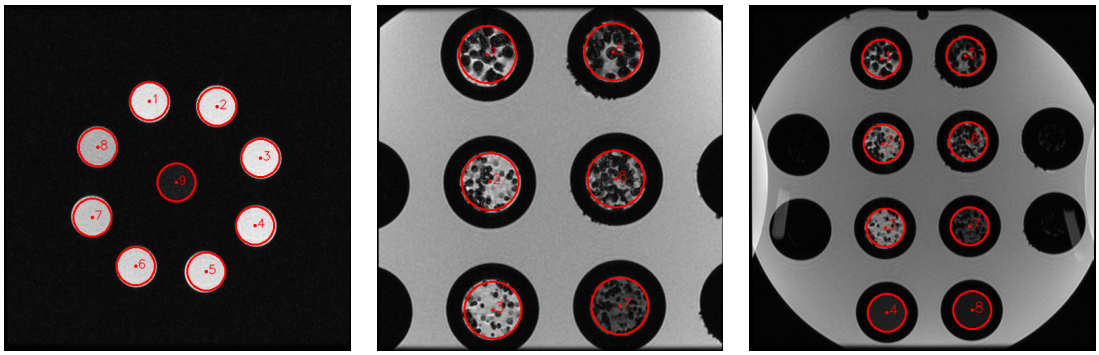
The ROIs are extracted from each slice by three steps. Firstly, detect all the circles in each image slice automatically using the HoughCircles function in OpenCV. Secondly, clip the circles with a fixed radius for scanner settings with the same phantom type and FOV, as listed in Table 2. Thirdly, select the interested circles as ROIs for radiomics analysis, i.e., 9 ROIs (numbered by 1 to 9) for homogeneous phantom, and 6 ROIs (numbered by 1,2,3,5,6,7) for heterogeneous phantom, as shown in Figure 3.

1.3 Image Preprocessing and Normalization

For the scanned MRI images, we first perform some image preprocessing and image normalization before extracting the radiomic features. Here two image preprocessing methods are considered. The first one is N4 bias field correction [5], one of the most popular methods to correct low-frequency intensity non-uniformity (also known as bias, inhomogeneity, illumination nonuniformity, or gain field) present in the MRI images. N4 bias field correction was implemented by ANTsPy, with the shrink factor for multi-resolution correction set to 2, the maximum number of iterations set to 100, and other parameters using default values. The second preprocessing method considered is image

Phantom Type	FOV	Radius of ROIs
Homogeneous Phantom	FOV24	14
	FOV18	18
	FOV12	28
Heterogeneous Phantom	FOV24	22
	FOV18	28
	FOV12	42

Table 2: Radius of VOIs corresponding to different phantom type and FOV.



(a) 9 ROIs for homogeneous phantom

(b) 6 ROIs for heterogeneous phantom with FOV=12.

(c) 6 ROIs (1,2,3,5,6,7) for heterogeneous phantom (FOV=18 or 24).

Figure 3: Examples of Region of Interests (ROIs). Note that for heterogeneous phantom with FOV=18 or 24, 8 circles are detected, but to keep consistent with FOV=12, only 6 circles(numbered by 1,2,3,5,6,7) will be used as ROIs.

resampling, which was implemented by ANTsPy using b-spline interpolation strategy, ensuring that all the MRI images to have the same voxel size, namely $1mm \times 1mm \times 1mm$ in our study.

Besides, two normalization methods (Z-Score and Nyúl normalization) [1] are used to normalize these MRI images, see the detailed definition of these methods in [1, 6]. We directly used the ROI masks as normalization masks for Z-Score normalization.

Note that, we have made the features extracted from different preprocessed images publicly available in `https://github.com/Yingping-LI/Harmonization_methods`. However, in the experiments presented in our MedAI 2024 paper, only Nyúl normalization was utilized.

1.4 Extraction of Radiomics Features

Pyradiomics [7], an open-source python package for the extraction of radiomics features from medical imaging, will be used in our research to extract texture features from the region of interests(ROIs). The detail meaning and calculation of these features can be found in the official website of *PyRadiomics* [2], as well as the image biomarker standardisation initiative(IBSI) [8], which defines the standardisation of extracting image biomarkers from acquired imaging for the purpose of high-throughput quantitative image analysis (radiomics).

The 120 radiomic features extracted using *PyRadiomics* can be subdivided into the following classes:

- First Order Statistics: 19 features
- Shape-based (3D): 16 features
- Shape-based (2D): 10 features

[1] `https://github.com/jcreinhold/intensity-normalization`
[2] `https://pyradiomics.readthedocs.io/en/latest/features.html`

- Gray Level Cooccurence Matrix (GLCM): 24 features
- Gray Level Run Length Matrix (GLRLM): 16 features
- Gray Level Size Zone Matrix (GLSZM): 16 features
- Neighbouring Gray Tone Difference Matrix (NGTDM): 5 features
- Gray Level Dependence Matrix (GLDM): 14 features

From all the available features, we will discard these features: 1) We discard all the 2D and 3D shape features. 2) Among the first-order features, we discard standard deviation, because "Stand Deviation= $\sqrt{\text{Variance}}$". 3) GLCM is symmetrical, therefore "Sum Average = 2 × Joint Average", so we only calculate Joint Average in our experiments.

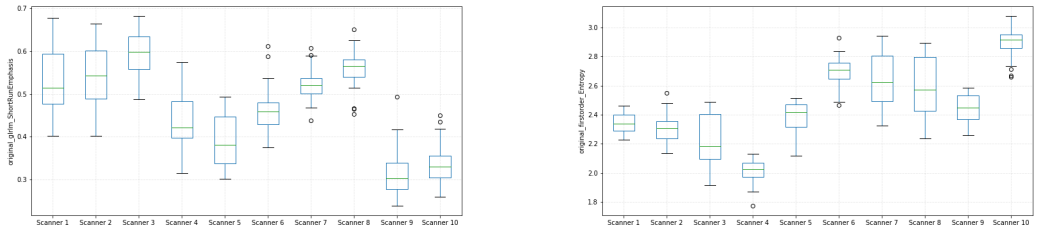
Finally, 120-16-10-1-1=92 radiomic features are extracted for each ROI, as listed in Table 3. Note that When extracting radiomic features, we directly used the fixed bin number strategy and took 32 bins for discretization.

Table 3: The 92 radiomic features used in our study.

Feature Type	**Features**
first-order statistics (18 features)	Energy,Total Energy, Entropy, Minimum, 10th percentile, 90th percentile, Maximum, Mean, Median,Interquartile Range, Range, Mean Absolute Deviation, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Kurtosis, Variance, Uniformity
GLCM (23 features)	Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Informational Measure of Correlation 1, Informational Measure of Correlation 2, Inverse Difference Moment, Maximal Correlation Coefficient, Inverse Difference Moment Normalized, Inverse Difference, Inverse Difference Normalized, Inverse Variance, Maximum Probability, Sum Entropy, Sum of Squares
GLSZM (16 features)	Small Area Emphasis, Large Area Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Size-Zone Non-Uniformity, Size-Zone Non-Uniformity Normalized, Zone Percentage, Gray Level Variance, Zone Variance, Zone Entropy, Low Gray Level Zone Emphasis, High Gray Level Zone Emphasis, Small Area Low Gray Level Emphasis, Small Area High Gray Level Emphasis,Large Area Low Gray Level Emphasis, Large Area High Gray Level Emphasis
GLRLM (16 features)	Short Run Emphasis, Long Run Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Gray Level Variance, Run Variance, Run Entropy, Low Gray Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis,Long Run Low Gray Level Emphasis, Long Run High Gray Level Emphasis
NGTDM (5 features)	Coarseness, Contrast, Busyness, Complexity, Strength
GLDM (14 features)	Small Dependence Emphasis, Large Dependence Emphasis, Gray Level Non-Uniformity, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Gray Level Variance, Dependence Variance, Dependence Entropy, Low Gray Level Emphasis, High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis, Small Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Large Dependence High Gray Level Emphasis

1.5 Existence of Scanner Effects

We have successfully extracted 92 radiomic features for each ROI collected from different scanner settings. Now we will investigate whether there exists scanner effects among these texture features by drawing box plots. Figure 4 shows an example of the box plots for homogeneous phantom and heterogeneous phantom, respectively. Obviously, the same phantom scanned by different scanner settings has different mean and variances for the extracted features, which indicates the existence of scanner effects.



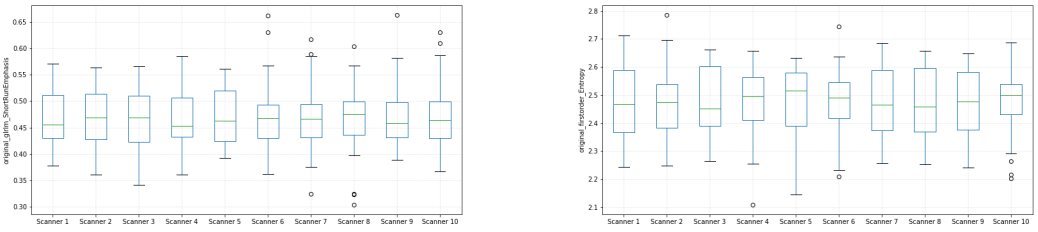
(a) "Short runs emphasis" feature for pattern 1 of homogeneous phantom

(b) "Entropy" feature for pattern 1 of heterogeneous phantom

Figure 4: Box plots to show the existence of scanner effects.

1.6 A Simple Example to Remove Scanner Effects

After applying our proposed ComBat-MLE harmonization method to the features for each pattern class separately, most of the scanner effects can be removed from these texture features. Figure 5 shows the box plots of the harmonized features (harmonized by our proposed ComBat-MLE method) corresponding to features in Figure 4. We can easily see that after harmonization by our proposed ComBat-MLE method, the mean and variance of the features are similar for all scanner settings, which verifies the effectiveness of our proposed method to remove scanner effects.



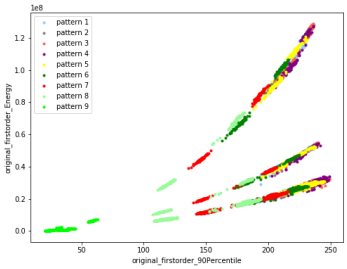
(a) "Short runs emphasis" feature for pattern 1 of homogeneous phantom

(b) "Entropy" feature for pattern 1 of heterogeneous phantom

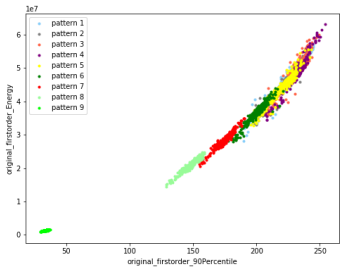
Figure 5: Box plots of the harmonized features by our proposed ComBat-MLE method.

2 New Method for Harmonizing Scanner Effects Across Classes

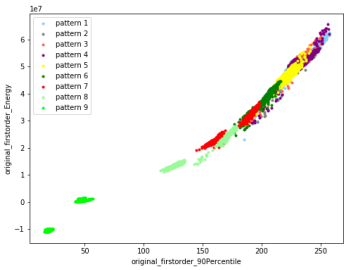
Orlhac *et al.* [9] mentioned one kind of failure cases of ComBat method we also met in our experiments. We explain it in detail in Figure 6. Obviously, before ComBat harmonization, the two example radiomic features suffer from severe scanner effects (Figure 6(a)). If we apply the harmonization methods for different pattern classes separately, then most of the scanner effects can be removed successfully (Figure 6(b)). But if the pattern class is unknown, and we apply harmonization methods one time for the data consisting of all pattern classes, then some scanner effects can't be removed (Figure 6(c)). The reason lies in the unmet assumptions of the data and the ComBat method. ComBat model is built based on the assumption that scanner effects are the same for feature data extracted from the same scanner setting. However, in this example, when applying ComBat harmonization for the data from all different pattern classes together, the feature data from the same scanner setting but from different pattern classes actually has different scanner effects.



(a) Before harmonization



(b) Apply parametric ComBat method for each pattern class separately.



(c) Apply parametric ComBat method for data consisting of all pattern classes.

Figure 6: Compare harmonization results of applying ComBat methods (use parametric ComBat as an example) in different ways. (a) The radiomic features before harmonization suffer from severe scanner effects, appearing as several clusters for each pattern class (i.e. color). (b)If harmonizing features for each pattern class separately, data from same pattern class (i.e. same color) are harmonized to one cluster, indicating that scanner effects are successfully removed. (c)If harmonizing features for all pattern class data one time, then each pattern class (i.e. color) are still composed of two different clusters, meaning that some scanner effects still exist.

In this section, we aim to develop a new harmonization method, with the assumption that, the data are composed of different pattern classes, but the feature data from the same scanner setting has different scanner effects for different pattern classes.

2.1 Notations

Let $\boldsymbol{X}_s \in \mathbb{R}^p$ be a random variable representing the observed feature data 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 $(\boldsymbol{x}_n^{(s)})^\top$.

To represent the pattern class label of $\boldsymbol{X}_s$, we introduce a binary random variable $\boldsymbol{Z}_s \in \mathbb{R}^C$. For each realization $\boldsymbol{z}^{(s)} = (z_1^{(s)}, z_2^{(s)}, ..., z_C^{(s)})^\top$ of variable $\boldsymbol{Z}_s$, it has a 1-of-C representation in which a particular element $z_c^{(s)}$ is equal to 1 and all other elements are equal to 0. Therefore, the values of $z_c^{(s)}$ satisfy $z_c^{(s)} \in \{0, 1\}$ and $\sum_{c=1}^C z_c^{(s)} = 1$. Let the realizations $\{\boldsymbol{z}_1^{(s)}, ..., \boldsymbol{z}_{N_s}^{(s)}\}$ of $\boldsymbol{Z}_s$ be the pattern classes corresponding to $\{\boldsymbol{x}_1^{(s)}, ..., \boldsymbol{x}_{N_s}^{(s)}\}$, then we can use a $N_s \times C$ matrix Z_s with rows $(\boldsymbol{z}_n^{(s)})^\top$ to represent the pattern class corresponding to X_s.

2.2 Model without scanner effects

Suppose random variable $\boldsymbol{X}_s \in \mathbb{R}^p$ representing the observed feature data from scanner setting $s \in \{1, 2, ..., S\}$ doest not suffer from scanner effects, and it consists of data from C pattern classes. The pattern class variable $\boldsymbol{Z}_s \in \mathbb{R}^C$ corresponding to $\boldsymbol{X}_s$ can be specified in terms of mixing coefficients π_{sc}, such that

$$p(z_c^{(s)} = 1) = \pi_{sc}, \tag{1}$$

where for each scanner setting $s \in \{1, 2, ..., S\}$, the parameters $\{\pi_{sc}\}$ must satisfy

$$0 \leq \pi_{sc} \leq 1, \quad \sum_{c=1}^{C} \pi_{sc} = 1, \tag{2}$$

so the distribution of variable $\boldsymbol{Z}_s$ can be written as

$$p(\boldsymbol{Z}_s) = \prod_{c=1}^{C} (\pi_{sc})^{z_c^{(s)}}. \tag{3}$$

Then the conditional distribution of $\boldsymbol{X}_s$ given a particular value of $\boldsymbol{Z}_s$ is

$$p(\boldsymbol{X}_s | z_{sc} = 1) = \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \tag{4}$$

or, equivalently,

$$p(\boldsymbol{X}_s | \boldsymbol{Z}_s) = \prod_{c=1}^{C} \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)^{z_c^{(s)}}, \tag{5}$$

where $\boldsymbol{\mu}_c \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}_c \in \mathbb{R}^{p \times p}$ are the mean and variance matrix of the c-th pattern class. So the joint distribution $p(\boldsymbol{X}_s, \boldsymbol{Z}_s)$ is

$$p(\boldsymbol{X}_s, \boldsymbol{Z}_s) = p(\boldsymbol{Z}_s) p(\boldsymbol{X}_s | \boldsymbol{Z}_s) = \prod_{c=1}^{C} [\pi_{sc}\, \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)]^{z_c^{(s)}}, \tag{6}$$

the marginal distribution of $\boldsymbol{X}_s$ can be obtained by summing joint distribution over all possible states of $\boldsymbol{Z}_s$ as

$$p(\boldsymbol{X}_s) = \sum_{\boldsymbol{Z}_s} p(\boldsymbol{X}_s, \boldsymbol{Z}_s) = \sum_{c=1}^{C} \pi_{sc}\, \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c). \tag{7}$$

So if there is no scanner effects, for each scanner setting $s \in \{1, 2, ..., S\}$, the observed feature data $\boldsymbol{X}_s \in \mathbb{R}^p$ follows a Gaussian mixture distribution:

$$p(\boldsymbol{X}_s) = \sum_{c=1}^{C} \pi_{sc}\, \mathcal{N}(\boldsymbol{X}_s | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \tag{8}$$

where

$$0 \leq \pi_{sc} \leq 1, \quad \sum_{c=1}^{C} \pi_{sc} = 1 \tag{9}$$

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

2.3 Model with scanner effects

Now let us model the case when the observed feature data $\boldsymbol{X}_s \in \mathbb{R}^p$ suffers from scanner effects. Suppose that the scanner effects are different for each pattern class, and they affect the mean and variance of each pattern class, then the observed data $\boldsymbol{X}_s$ suffering scanner effects from scanner setting s can be assumed to follow the following Gaussian mixture distribution:

$$p(\boldsymbol{X}_s) = \sum_{c=1}^{C} \pi_{sc}\, \mathcal{N}(\boldsymbol{X}_s|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c), \tag{10}$$

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

$$0 \le \pi_{sc} \le 1, \quad \sum_{c=1}^{C} \pi_{sc} = 1 \tag{11}$$

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

Note that in model (10), parameters $\boldsymbol{\mu}_c$ and $\boldsymbol{\gamma}_{sc}$ (similarly, $\boldsymbol{\Delta}_{sc}$ and $\boldsymbol{\Sigma}_c$) are not identifiable, so proper constraints should be proposed. As an reasonable option, we propose the following additive and multiplicative constraints:

- additive constraints:

$$\sum_{s=1}^{S} w_s \boldsymbol{\gamma}_{sc} = \mathbf{0},\ c = 1, 2, ..., C. \tag{12}$$

- multiplicative constraints:

$$\prod_{s=1}^{S} (\boldsymbol{\Delta}_{sc})^{w_s} = \mathrm{I},\ c = 1, 2, ..., C. \tag{13}$$

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.

2.4 Solve the model by EM algorithm

The proposed model (10) with constraints (11)(12)(13), which models the case when scanner effects are different for each pattern class, can be solved by Expectation Maximization (EM) algorithm. Moreover, the E (expectation) and M (maximization) step can be summarized as follows:

2.4.1 E step

In the E step, for each scanner setting $s \in \{1, 2, ..., S\}$, we will compute the posterior distribution $p_s(Z_s|X_s, \boldsymbol{\theta}^{\text{old}})$ given the current parameters $\boldsymbol{\theta}^{\text{old}}$, and then use this posterior distribution to find the expected complete-data log likelihood evaluated for some general parameter value $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ represents the set of all the model parameters and is defined as

$$\boldsymbol{\theta} = \{\pi_{sc} \in \mathbb{R},\ \boldsymbol{\mu}_c \in \mathbb{R}^p,\ \boldsymbol{\Sigma}_c \in \mathbb{R}^{p\times p},\ \boldsymbol{\gamma}_{sc} \in \mathbb{R}^p,\ \boldsymbol{\Delta}_{sc} \in \mathbb{R}^{p\times p} \mid s = 1, 2, ..., S;\ c = 1, 2, ..., C\}. \tag{14}$$

Specifically, consider the complete data set (X_s, Z_s) from scanner setting $s \in \{1, 2, ..., S\}$, the complete-data likelihood function takes the form

$$p_s(X_s, Z_s|\boldsymbol{\theta}) = \prod_{n=1}^{N_s} \prod_{c=1}^{C} [\pi_{sc}\, \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)]^{z_{nc}^{(s)}}, \tag{15}$$

and thus the logarithm of complete-data likelihood function takes the form

$$\ln p_s(X_s, Z_s|\boldsymbol{\theta}) = \sum_{n=1}^{N_s}\sum_{c=1}^{C} z_{nc}^{(s)}[\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)], \tag{16}$$

where $z_{nc}^{(s)}$ represents the c^{th} components of $\boldsymbol{z}_n^{(s)}$. The expectation of complete-data log likelihood function, denoted by $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$, can be calculated by

$$\begin{aligned}
\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}_{Z_s|X_s,\boldsymbol{\theta}^{\text{old}}}\big[\ln p_s(X_s, Z_s|\boldsymbol{\theta})\big] \\
&= \sum_{n=1}^{N_s}\sum_{c=1}^{C} \mathbb{E}_{Z_s|X_s,\boldsymbol{\theta}^{\text{old}}}[z_{nc}^{(s)}] \cdot [\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)] \\
&= \sum_{n=1}^{N_s}\sum_{c=1}^{C} \alpha(z_{nc}^{(s)})[\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)],
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
\mathbb{E}_{Z_s|X_s,\boldsymbol{\theta}}[z_{nc}^{(s)}] &= p_s(z_{nc}^{(s)} = 1|\boldsymbol{x}_n^{(s)}, \boldsymbol{\theta}) \times 1 + p_s(z_{nc}^{(s)} = 0|\boldsymbol{x}_n^{(s)}, \boldsymbol{\theta}) \times 0 \\
&= \frac{p_s(\boldsymbol{x}_n^{(s)}|z_{nc}^{(s)} = 1, \boldsymbol{\theta}) p_s(z_{nc}^{(s)} = 1|\boldsymbol{\theta})}{p_s(\boldsymbol{x}_n^{(s)}|\boldsymbol{\theta})} \\
&= \frac{p_s(\boldsymbol{x}_n^{(s)}|z_{nc}^{(s)} = 1, \boldsymbol{\theta}) p_s(z_{nc}^{(s)} = 1|\boldsymbol{\theta})}{\sum_{j=1}^{C} p_s(\boldsymbol{x}_n^{(s)}|z_{nj}^{(s)} = 1, \boldsymbol{\theta}) p_s(z_{nj}^{(s)} = 1|\boldsymbol{\theta})} \\
&= \frac{\pi_{sc}\, \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)}{\sum_{j=1}^{C} \pi_{sj}\, \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_j + \boldsymbol{\gamma}_{sj}, \boldsymbol{\Delta}_{sj}\boldsymbol{\Sigma}_j)} \triangleq \alpha(z_{nc}^{(s)}).
\end{aligned} \tag{18}$$

Note that $\alpha(z_{nc}^{(s)})$ is calculated by the old parameter $\boldsymbol{\theta}^{\text{old}}$, and will be regarded as a constant in M step.

So in the E step, for each scanner setting $s \in \{1, 2, ..., S\}$, we have finished calculating the expectation of complete-data log likelihood function $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ as described in Equation (17).

2.4.2 M step

In the M step, we will maximize the expected complete-data log likelihood function $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$, $s = 1, 2, ..., S$ with respect to the parameters $\boldsymbol{\theta}$. Since we have S expectation function $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ to be optimized, and they share the same parameters $\{\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, c = 1, 2, ..., C\}$, so we cannot maximize only one $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ with respect to $\{\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, c = 1, 2, ..., C\}$, but to optimize these S expectation functions together. In this case, it is natural to define $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ as a weighted sum of the $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}), s = 1, 2, ..., S$ to be final objective function to be optimized as follows:

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \sum_{s=1}^{S} w_s \mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \\
&= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s}\sum_{c=1}^{C} \alpha(z_{nc}^{(s)})[\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)],
\end{aligned} \tag{19}$$

where w_s is the weight of scanner setting s and can be set to

$$w_s = \frac{N_s}{\sum_{s=1}^{S} N_s}, \tag{20}$$

and

$$\mathcal{N}(\boldsymbol{x}_n^{(s)}|\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|^{\frac{1}{2}}} \exp\Big\{-\frac{1}{2}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})^\top (\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big\}. \tag{21}$$

Then we can update parameters $\boldsymbol{\theta}$ as $\boldsymbol{\theta}^{\text{new}}$ by

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}), \tag{22}$$

where $\boldsymbol{\theta}$ is the set of the model parameters as defined by Equation (14).

Following the coordinate descent optimization strategy, each parameter of $\boldsymbol{\theta}$ can be updated iteratively to gradually maximize the objective function $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ by taking the derivative of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ with respect to the parameter, and the method of Lagrange multipliers will be included to deal with the parameter constraints (12) and (13), the details can be found as below.

Update $\boldsymbol{\mu}_c$:
In order to update $\boldsymbol{\mu}_c$, we can calculate the derivatives of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ with respect to $\boldsymbol{\mu}_c$, and set it to zero as follows

$$\begin{aligned}
\frac{\partial \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})}{\partial \boldsymbol{\mu}_c} &= \frac{\partial \Big\{ \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \sum_{c=1}^{C} \alpha(z_{nc}^{(s)}) [\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)] \Big\}}{\partial \boldsymbol{\mu}_c} \\
&= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \frac{\partial \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)}{\partial \boldsymbol{\mu}_c} \\
&= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) (\boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) = 0.
\end{aligned} \tag{23}$$

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

$$\boldsymbol{\mu}_c = \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \boldsymbol{\Delta}_{sc}^{-1} \Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \boldsymbol{\Delta}_{sc}^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_{sc}) \Big]. \tag{24}$$

Update $\boldsymbol{\gamma}_{sc}$:
Now let's update $\boldsymbol{\gamma}_{sc}$. To ensure the identifiability of the parameter $\boldsymbol{\mu}_c$ and $\boldsymbol{\gamma}_{sc}$, we assume

$$\sum_{s=1}^{S} w_s \boldsymbol{\gamma}_{sc} = \mathbf{0}, \;\; c = 1, 2, ..., C. \tag{25}$$

Notice here we use w_s, which means that each scanner setting s have similar amount of data for each pattern class, otherwise, it should use w_{sc}. We will consider this case later.

Let $\boldsymbol{\lambda}_c = (\lambda_{1c}, \lambda_{2c}, ..., \lambda_{pc})^T \in \mathbb{R}^p$, and define the Lagrange function as follows:

$$\begin{aligned}
\mathcal{L}_1 &= \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + \sum_{i=1}^{p} \lambda_{ic} \big(\sum_{s=1}^{S} w_s \gamma_{isc} \big) \\
&= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \sum_{c=1}^{C} \alpha(z_{nc}^{(s)}) [\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)] + \sum_{i=1}^{p} \lambda_{ic} \big(\sum_{s=1}^{S} w_s \gamma_{isc} \big).
\end{aligned} \tag{26}$$

The derivatives of $\mathcal{L}_1$ with respect to $\boldsymbol{\gamma}_{sc}$ is

$$\begin{aligned}
\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\gamma}_{sc}} &= w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \frac{\partial \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)}{\partial \boldsymbol{\gamma}_{sc}} + w_s \boldsymbol{\lambda}_c \\
&= w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) (\boldsymbol{\Delta}_{sc} \boldsymbol{\Sigma}_c)^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) + w_s \boldsymbol{\lambda}_c = 0.
\end{aligned} \tag{27}$$

Re-arranging the above equation, we can get

$$\boldsymbol{\gamma}_{sc} = \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}\Big]^{-1} \cdot \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c) + \boldsymbol{\lambda}_c\Big]. \tag{28}$$

Since $\sum_{s=1}^{S} w_s \boldsymbol{\gamma}_{sc} = \mathbf{0},\ c = 1, 2, ..., C$, so we can get $\boldsymbol{\lambda}_c$ as follows:

$$\begin{aligned}\boldsymbol{\lambda}_c = &-\Bigg[\sum_{s=1}^{S} w_s \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}\Big]^{-1}\Bigg]^{-1} \\ &\cdot \Bigg\{\sum_{s=1}^{S} w_s \Bigg[\Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}\Big]^{-1} \cdot \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c)\Big]\Bigg]\Bigg\}.\end{aligned} \tag{29}$$

So $\boldsymbol{\gamma}_{sc}$ can be updated by Equation (28), where $\boldsymbol{\lambda}_c$ is calculated by Equation (29).

Update $\boldsymbol{\Sigma}_c$:
Now let's analyze the way to update $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Delta}_{sc}$. We assume them to be diagonal matrices, namely, suppose $\boldsymbol{\Sigma}_c = \mathrm{diag}(\sigma_{1c}, ..., \sigma_{pc})$, $\boldsymbol{\Delta}_{sc} = \mathrm{diag}(\delta_{1sc}, ..., \delta_{psc})$. If we take the derivatives of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})$ with respect to $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Delta}_{sc}$ directly, we can't assure that the updated matrices are still diagonal. In order to solve this problem, we introduce $\boldsymbol{\Gamma}_c$ and $\boldsymbol{\Lambda}_{sc}$ as follows

$$\boldsymbol{\Gamma}_c = (\sigma_{1c}, ..., \sigma_{pc})^T, \quad \boldsymbol{\Lambda}_{sc} = (\delta_{1sc}, ..., \delta_{psc})^T, \tag{30}$$

which obviously satisfy

$$\begin{aligned}\boldsymbol{\Sigma}_c = \mathrm{diag}(\boldsymbol{\Gamma}_c), \ \ \boldsymbol{\Sigma}_c \cdot \boldsymbol{e} = \boldsymbol{\Gamma}_c, \\ \boldsymbol{\Delta}_{sc} = \mathrm{diag}(\boldsymbol{\Lambda}_{sc}), \ \ \boldsymbol{\Delta}_{sc} \cdot \boldsymbol{e} = \boldsymbol{\Lambda}_{sc},\end{aligned} \tag{31}$$

where $\boldsymbol{e} = (1, 1, ..., 1)^T \in \mathbb{R}^p$. So we will update $\boldsymbol{\Gamma}_c$ and $\boldsymbol{\Lambda}_{sc}$ by calculating the derivatives of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})$ with respect to $\boldsymbol{\Gamma}_c$ and $\boldsymbol{\Lambda}_{sc}$, respectively, and then update $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Delta}_{sc}$ by

$$\boldsymbol{\Sigma}_c = \mathrm{diag}(\boldsymbol{\Gamma}_c), \quad \boldsymbol{\Delta}_{sc} = \mathrm{diag}(\boldsymbol{\Lambda}_{sc}). \tag{32}$$

The derivatives of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})$ with respect to $\boldsymbol{\Gamma}_c$ takes the form

$$\begin{aligned}\frac{\partial \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})}{\partial \boldsymbol{\Gamma}_c} &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \frac{\partial \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)}{\partial \boldsymbol{\Gamma}_c} \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \Bigg\{ -\frac{1}{2}\frac{\partial \ln |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Gamma}_c} - \frac{1}{2}\frac{\partial \Big\{(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})^T (\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big\}}{\partial \boldsymbol{\Gamma}_c} \Bigg\}.\end{aligned} \tag{33}$$

On the one hand,

$$-\frac{1}{2}\frac{\partial \ln |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Gamma}_c} = -\frac{1}{2}|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|^{-1} \cdot \frac{\partial |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Gamma}_c}, \tag{34}$$

since we have supposed $\boldsymbol{\Sigma}_c = \mathrm{diag}(\sigma_{1c}, ..., \sigma_{pc})$ and $\boldsymbol{\Delta}_{sc} = \mathrm{diag}(\delta_{1sc}, ..., \delta_{psc})$, so

$$|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c| = \prod_{i=1}^{p} \delta_{isc}\sigma_{ic}. \tag{35}$$

Then

$$\frac{\partial|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial\boldsymbol{\Gamma}_c} = \frac{\partial(\prod_{i=1}^{p}\delta_{isc}\sigma_{ic})}{\partial\boldsymbol{\Gamma}_c} = \begin{pmatrix} \frac{\partial(\prod_{i=1}^{p}\delta_{isc}\sigma_{ic})}{\partial\sigma_{1c}} \\ \frac{\partial(\prod_{i=1}^{p}\delta_{isc}\sigma_{ic})}{\partial\sigma_{2c}} \\ \vdots \\ \frac{\partial(\prod_{i=1}^{p}\delta_{isc}\sigma_{ic})}{\partial\sigma_{pc}} \end{pmatrix} = (\prod_{i=1}^{p}\delta_{isc}\sigma_{ic}) \begin{pmatrix} \frac{1}{\sigma_{1c}} \\ \frac{1}{\sigma_{2c}} \\ \vdots \\ \frac{1}{\sigma_{pc}} \end{pmatrix} = |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c| \cdot \boldsymbol{\Sigma}_c^{-1} \cdot \boldsymbol{e}, \tag{36}$$

so we can get

$$-\frac{1}{2}\frac{\partial \ln|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial\boldsymbol{\Gamma}_c} = -\frac{1}{2}|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|^{-1} \cdot \frac{\partial|\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial\boldsymbol{\Gamma}_c} = -\frac{1}{2}\boldsymbol{\Sigma}_c^{-1} \cdot \boldsymbol{e}\ . \tag{37}$$

On the other hand,

$$\begin{aligned} &\frac{\partial\Big\{(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})^T(\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big\}}{\partial\boldsymbol{\Gamma}_c} \\ &= \begin{pmatrix} \frac{\partial(x_{1n}^{(s)} - \mu_{1c} - \gamma_{1sc})^2(\delta_{1sc}\sigma_{1c})^{-1}}{\partial\sigma_{1c}} \\ \frac{\partial(x_{2n}^{(s)} - \mu_{2c} - \gamma_{2sc})^2(\delta_{1sc}\sigma_{2c})^{-1}}{\partial\sigma_{2c}} \\ \vdots \\ \frac{\partial(x_{pn}^{(s)} - \mu_{pc} - \gamma_{psc})^2(\delta_{psc}\sigma_{pc})^{-1}}{\partial\sigma_{pc}} \end{pmatrix} = -\begin{pmatrix} (x_{1n}^{(s)} - \mu_{1c} - \gamma_{1sc})^2 \cdot \delta_{1sc}^{-1} \cdot \sigma_{1c}^{-2} \\ (x_{2n}^{(s)} - \mu_{2c} - \gamma_{2sc})^2 \cdot \delta_{2sc}^{-1} \cdot \sigma_{2c}^{-2} \\ \vdots \\ (x_{pn}^{(s)} - \mu_{pc} - \gamma_{psc})^2 \cdot \delta_{psc}^{-1} \cdot \sigma_{pc}^{-2} \end{pmatrix} \\ &= -\boldsymbol{\Sigma}_c^{-2}\boldsymbol{\Delta}_{sc}^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}). \end{aligned} \tag{38}$$

where $\circ$ represents the Hadamard product (also known as element-wise product). Taking Equation (37) and Equation (38) into Equation (33), we have

$$\frac{\partial\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})}{\partial\boldsymbol{\Gamma}_c} = \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})\Big[-\frac{1}{2}\boldsymbol{\Sigma}_c^{-1} \cdot \boldsymbol{e} + \frac{1}{2}\boldsymbol{\Sigma}_c^{-2}\boldsymbol{\Delta}_{sc}^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big]. \tag{39}$$

Set this derivative to zero, and multiply $2\boldsymbol{\Sigma}_c^2$ on both sides, we have

$$\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})\Big[-\boldsymbol{\Sigma}_c \cdot \boldsymbol{e} + \boldsymbol{\Delta}_{sc}^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big] = 0. \tag{40}$$

Notice $\boldsymbol{\Sigma}_c \cdot \boldsymbol{e} = \boldsymbol{\Gamma}_c$, so rearranging the above equation, we have

$$\boldsymbol{\Gamma}_c = \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})\boldsymbol{\Delta}_{sc}^{-1}(\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})\Big]. \tag{41}$$

So we get the equation 41 to update $\boldsymbol{\Gamma}_c$, then $\boldsymbol{\Sigma}_c$ can be easily updated by $\boldsymbol{\Sigma}_c = \text{diag}(\boldsymbol{\Gamma}_c)$.

Update $\boldsymbol{\Delta}_{sc}$:
We will use the similar way to update $\boldsymbol{\Delta}_{sc} = \text{diag}(\boldsymbol{\Lambda}_{sc})$. To ensure the identifiability of the parameter $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Delta}_{sc}$, we assume

$$\prod_{s=1}^{S} \boldsymbol{\Delta}_{sc} = \mathbb{I},\ c = 1, 2, ..., C. \tag{42}$$

or

$$\prod_{s=1}^{S} \boldsymbol{\Lambda}_{sc} = \boldsymbol{e},\ c = 1, 2, ..., C. \tag{43}$$

where $\mathbb{I} \in \mathbb{R}^{p\times p}$ is the identity matrix and $\boldsymbol{e} = (1, 1, ..., 1)^T \in \mathbb{R}^p$.

Let $\hat{\boldsymbol{\lambda}}_c = (\hat{\lambda}_{1c}, \hat{\lambda}_{2c}, ..., \hat{\lambda}_{pc})^T \in \mathbb{R}^p$, and define the Lagrange function as

$$\begin{aligned}\mathcal{L}_2 &= \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + \sum_{i=1}^{p} \hat{\lambda}_{ic}(\prod_{s=1}^{S} \delta_{isc} - 1) \\ &= \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \sum_{c=1}^{C} \alpha(z_{nc}^{(s)})[\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)] + \sum_{i=1}^{p} \hat{\lambda}_{ic}(\prod_{s=1}^{S} \delta_{isc} - 1).\end{aligned} \tag{44}$$

Calculate the derivatives of $\mathcal{L}_2$ with respect to $\boldsymbol{\Lambda}_{sc}$, we have

$$\begin{aligned}\frac{\partial \mathcal{L}_2}{\partial \boldsymbol{\Lambda}_{sc}} &= w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \frac{\partial \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)}{\partial \boldsymbol{\Lambda}_{sc}} + \boldsymbol{\Delta}_{sc}^{-1} \hat{\boldsymbol{\lambda}}_c \\ &= w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \Bigg\{ -\frac{1}{2} \frac{\partial \ln |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Lambda}_{sc}} - \frac{1}{2} \frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})^T (\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \Big\}}{\partial \boldsymbol{\Lambda}_{sc}} \Bigg\} \\ &\quad + \boldsymbol{\Delta}_{sc}^{-1} \hat{\boldsymbol{\lambda}}_c,\end{aligned} \tag{45}$$

where

$$-\frac{1}{2} \frac{\partial \ln |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Lambda}_{sc}} = -\frac{1}{2} |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|^{-1} \cdot \frac{\partial |\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c|}{\partial \boldsymbol{\Lambda}_{sc}} = -\frac{1}{2} \boldsymbol{\Delta}_{sc}^{-1} \cdot \boldsymbol{e} \tag{46}$$

and

$$\begin{aligned}&\frac{\partial \Big\{ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc})^T (\boldsymbol{\Delta}_{sc}\boldsymbol{\Sigma}_c)^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \Big\}}{\partial \boldsymbol{\Lambda}_{sc}} \\ =& - \boldsymbol{\Delta}_{sc}^{-2} \boldsymbol{\Sigma}_c^{-1} \cdot (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}).\end{aligned} \tag{47}$$

Taking Equation (46) and (47) into Equation (45), we can get the derivatives of $\mathcal{L}_2$ with respect to $\boldsymbol{\Lambda}_{sc}$ as follows

$$\begin{aligned}\frac{\partial \mathcal{L}_2}{\partial \boldsymbol{\Lambda}_{sc}} = w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \Big[-\frac{1}{2} \boldsymbol{\Delta}_{sc}^{-1} \cdot \boldsymbol{e} + \frac{1}{2} \boldsymbol{\Delta}_{sc}^{-2} \boldsymbol{\Sigma}_c^{-1} \cdot (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \Big] \\ + \boldsymbol{\Delta}_{sc}^{-1} \hat{\boldsymbol{\lambda}}_c\end{aligned} \tag{48}$$

Set this derivative to zero, multiply $\boldsymbol{\Delta}_{sc}^2$ on both sides and then rearranging the equation, we can get

$$\begin{aligned}\boldsymbol{\Lambda}_{sc} = &\frac{1}{\Big[w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \cdot \boldsymbol{e} - 2\hat{\boldsymbol{\lambda}}_c \Big]} \\ &\circ \Big[w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \cdot \boldsymbol{\Sigma}_c^{-1} \cdot (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}) \Big].\end{aligned} \tag{49}$$

Now we aim to use the constraints $\prod_{s=1}^{S} \boldsymbol{\Lambda}_{sc} = \boldsymbol{e},\ c = 1, 2, ..., C$ to solve $\hat{\boldsymbol{\lambda}}_c$. Define

$$A_{isc} = w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}), \quad B_{isc} = w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \cdot \sigma_{ic}^{-1} (x_{in}^{(s)} - \mu_{ic} - \gamma_{isc})^2, \tag{50}$$

for $i = 1, ..., p$, then δ_{isc} can be represented as

$$\delta_{isc} = \frac{B_{isc}}{A_{isc} - 2\hat{\lambda}_{ic}}. \tag{51}$$

Since the constraints $\prod_{s=1}^{S} \mathbf{\Lambda}_{sc} = \boldsymbol{e},\ c = 1, 2, ..., C$ hold, we can easily conclude

$$\prod_{s=1}^{S} \delta_{isc} = \prod_{s=1}^{S} \frac{B_{isc}}{A_{isc} - 2\hat{\lambda}_{ic}} = 1 \tag{52}$$

hold for $i = 1, ..., p,\ \ c = 1, 2, ..., C$. Define

$$h(\hat{\lambda}_{ic}) = \prod_{s=1}^{S} B_{isc} - \prod_{s=1}^{S} (A_{isc} - 2\hat{\lambda}_{ic}), \tag{53}$$

then

$$h^{'}(\hat{\lambda}_{ic}) = 2\Big[\prod_{s=1}^{S} (A_{isc} - 2\hat{\lambda}_{ic})\Big] \cdot \Big[\sum_{s=1}^{S} \frac{1}{A_{isc} - 2\hat{\lambda}_{ic}}\Big]. \tag{54}$$

Then $\hat{\lambda}_{ic}$ can be calculated numerically by Newton method as follows:

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

Input: a positive number ϵ which is small enough; one initialization value $\lambda^{(1)}$.
Output: numerical solution λ^* s.t. $h(\lambda^*) \approx 0$, namely, $\|h(\lambda^*)\| < \epsilon$

```
for t=1,2,... do
    λ^(t+1) = λ^(t) − h(λ^(t)) / h'(λ^(t))
    if ‖h(λ^(t+1))‖ ≤ ε then
        λ* = λ^(t+1)
        break
    end if
end for
return λ*
```

In summary, we can update $\mathbf{\Delta}_{sc} = \text{diag}(\mathbf{\Lambda}_{sc})$ by equation 49, where $\hat{\boldsymbol{\lambda}}_c$ can be solved numerically by the Newton method (Algorithm 1).

Update π_{sc}:
Finally, for each scanner setting $s \in \{1, 2, ..., S\}$, we will analyze the way to update the mixing coefficients π_{sc}. we will maximize $\mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ with respect to π_{sc} under the constraints $\sum_{c=1}^{C} \pi_{sc} = 1$ by taking a Lagrange multiplier and maximizing the following quantity

$$\begin{aligned} \mathcal{L}_s &= \mathcal{Q}_s(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + \lambda(\sum_{c=1}^{C} \pi_{sc} - 1) \\ &= \sum_{n=1}^{N_s} \sum_{c=1}^{C} \alpha(z_{nc}^{(s)})[\ln \pi_{sc} + \ln \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}, \mathbf{\Delta}_{sc}\mathbf{\Sigma}_c)] + \lambda(\sum_{c=1}^{C} \pi_{sc} - 1), \end{aligned} \tag{55}$$

the derivative of $\mathcal{L}_s$ with respect to π_{sc} takes the form

$$\frac{\partial \mathcal{L}_s}{\partial \pi_{sc}} = \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}) \frac{1}{\pi_{sc}} + \lambda. \tag{56}$$

Set $\frac{\partial \mathcal{L}_s}{\partial \pi_{sc}} = 0$, then we can get π_{sc} as

$$\pi_{sc} = -\frac{1}{\lambda} \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}). \tag{57}$$

Notice that

$$\sum_{c=1}^{C} \pi_{sc} = 1, \tag{58}$$

taking Equation (57) to Equation (58), we can get the value of λ as

$$\lambda = -\sum_{n=1}^{N_s} \sum_{c=1}^{C} \alpha(z_{nc}^{(s)}) = -N_s, \tag{59}$$

taking Equation (59) into Equation (57), we get π_{sc} as follows

$$\pi_{sc} = \frac{1}{N_s} \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)}). \tag{60}$$

2.4.3 Summary of the algorithm

We summarize the algorithm for estimating the set of the model parameters $\boldsymbol{\theta}$ in Equation (14) for our proposed model (10) with constraints (11)(12)(13) as follows.

Algorithm 2 Estimate the set of the parameters $\boldsymbol{\theta}$ of the proposed model (10) with constraints (11)(12)(13).

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 $\boldsymbol{\theta}$ as defined in Equation (14).

1: Choose the results of Gaussian mixture model as an initialization of $\boldsymbol{\theta}$, denoted as $\boldsymbol{\theta}^{(0)}$ as below. And initialize the logarithm likelihood function $\mathcal{L}(\boldsymbol{\theta}^{(0)}) = -\infty$.

$$\begin{aligned} \boldsymbol{\theta}^{(0)} = \{&\pi_{sc}^{(0)} \in \mathbb{R},\ \boldsymbol{\mu}_c^{(0)} \in \mathbb{R}^p,\ \boldsymbol{\Sigma}_c^{(0)} \in \mathbb{R}^{p \times p}, \\ &\boldsymbol{\gamma}_{sc}^{(0)} \in \mathbb{R}^p,\ \boldsymbol{\Delta}_{sc}^{(0)} \in \mathbb{R}^{p \times p} \mid s = 1, 2, ..., S;\ c = 1, 2, ..., C\}. \end{aligned} \tag{61}$$

2: **for** k=1,2,... **do**

3: **E Step**: Evaluate the responsibilities using the parameters $\boldsymbol{\theta}^{(\text{k-1})}$.

$$\alpha(z_{nc}^{(s)})^{(\text{k})} = \frac{\pi_{sc}\, \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c^{(\text{k-1})} + \boldsymbol{\gamma}_{sc}^{(\text{k-1})}, \boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})}{\sum_{j=1}^{C} \pi_{sj}\, \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_j^{(\text{k-1})} + \boldsymbol{\gamma}_{sj}^{(\text{k-1})}, \boldsymbol{\Delta}_{sj}^{(\text{k-1})} \boldsymbol{\Sigma}_j^{(\text{k-1})})}) \tag{62}$$

4: **M Step**: Re-estimate the parameters $\boldsymbol{\theta}$ as $\boldsymbol{\theta}^{(\text{k})}$ using the responsibilities $\alpha_s(z_{nc})^{(\text{k})}$ and the parameters $\boldsymbol{\theta}^{(\text{k-1})}$.

- **Update $\boldsymbol{\gamma}_{sc}^{(\text{k})}$:**

$$\boldsymbol{\gamma}_{sc}^{(\text{k})} = \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\text{k})} (\boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})^{-1} \Big]^{-1} \cdot \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\text{k})} (\boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\text{k-1})}) + \boldsymbol{\lambda}_c^{(\text{k})} \Big] \tag{63}$$

where

$$\begin{aligned} \boldsymbol{\lambda}_c^{(\text{k})} = &-\Bigg[\sum_{s=1}^{S} w_s \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\text{k})} (\boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})^{-1} \Big]^{-1} \Bigg]^{-1} \\ &\cdot \Bigg\{ \sum_{s=1}^{S} w_s \Bigg[\Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\text{k})} (\boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})^{-1} \Big]^{-1} \cdot \Big[\sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\text{k})} (\boldsymbol{\Delta}_{sc}^{(\text{k-1})} \boldsymbol{\Sigma}_c^{(\text{k-1})})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\text{k-1})}) \Big] \Bigg] \Bigg\} \end{aligned} \tag{64}$$

- **Update $\boldsymbol{\Delta}_{sc}^{(\mathrm{k})} = \mathbf{diag}(\boldsymbol{\Lambda}_{sc})^{(\mathrm{k})}$:**

$$\begin{aligned}\boldsymbol{\Lambda}_{sc}^{(\mathrm{k})} &= \frac{1}{\left[w_s \sum_{n=1}^{N_s} \alpha_(z_{nc}^{(s)})^{(\mathrm{k})} \cdot \boldsymbol{e} - 2\hat{\boldsymbol{\lambda}}_c^{(\mathrm{k})}\right]} \\ &\circ \Big[w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})} \cdot (\boldsymbol{\Sigma}_c^{(\mathrm{k\text{-}1})})^{-1} \cdot (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\mathrm{k\text{-}1})} - \boldsymbol{\gamma}_{sc}^{(\mathrm{k})}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\mathrm{k\text{-}1})} - \boldsymbol{\gamma}_{sc}^{(\mathrm{k})})\Big]\end{aligned} \tag{65}$$

  and $\hat{\boldsymbol{\lambda}}_c^{(\mathrm{k})}$ can be solved numerically by Newton method.

- **Update $\boldsymbol{\mu}_c^{(\mathrm{k})}$:**

$$\boldsymbol{\mu}_c^{(\mathrm{k})} = \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})} (\boldsymbol{\Delta}_{sc}^{(\mathrm{k})})^{-1}\Big]^{-1} \cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})} (\boldsymbol{\Delta}_{sc}^{(\mathrm{k})})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\gamma}_{sc}^{(\mathrm{k})})\Big] \tag{66}$$

- **Update $\boldsymbol{\Sigma}_c^{(\mathrm{k})} = \mathbf{diag}(\boldsymbol{\Gamma}_c^{(\mathrm{k})})$:**

$$\begin{aligned}\boldsymbol{\Gamma}_c^{(\mathrm{k})} &= \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})}\Big]^{-1} \\ &\cdot \Big[\sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})} (\boldsymbol{\Delta}_{sc}^{(\mathrm{k})})^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\mathrm{k})} - \boldsymbol{\gamma}_{sc}^{(\mathrm{k})}) \circ (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c^{(\mathrm{k})} - \boldsymbol{\gamma}_{sc}^{(\mathrm{k})})\Big]\end{aligned} \tag{67}$$

- **Update $\pi_{sc}^{(\mathrm{k})}$:** $\pi_{sc}^{(\mathrm{k})} = \frac{1}{N_s} \sum_{n=1}^{N_s} \alpha(z_{nc}^{(s)})^{(\mathrm{k})}$

5: Calculate the logarithm likelihood function

$$\mathcal{L}(\boldsymbol{\theta}^{(k)}) = \sum_{s=1}^{S} w_s \sum_{n=1}^{N_s} \ln\Big\{\sum_{c=1}^{C} \pi_{sc}^{(\mathrm{k})}\, \mathcal{N}(\boldsymbol{x}_n^{(s)} | \boldsymbol{\mu}_c^{(\mathrm{k})} + \boldsymbol{\gamma}_{sc}^{(\mathrm{k})}, \boldsymbol{\Delta}_{sc}^{(\mathrm{k})} \boldsymbol{\Sigma}_c^{(\mathrm{k})})\Big\}, \tag{68}$$

6: **if** $\mathcal{L}(\boldsymbol{\theta}^{(k)}) - \mathcal{L}(\boldsymbol{\theta}^{(k-1)}) \leq \epsilon$, **then**
7: $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(k)}$
8: **break**
9: **end if**
10: **end for**
11: **return** $\boldsymbol{\theta}^*$

2.5 Remove scanner effects with the estimated parameters

With the estimated parameters of model (10), then given an observed data $\boldsymbol{x}_n^{(s)}$ from scanner setting $s \in \{1, 2, ..., S\}$, the pattern class label of $\boldsymbol{x}_n^{(s)}$ can be estimated as

$$c = \arg\max_{c \in \{1,2,...,C\}} \alpha(z_{nc}^{(s)}), \tag{69}$$

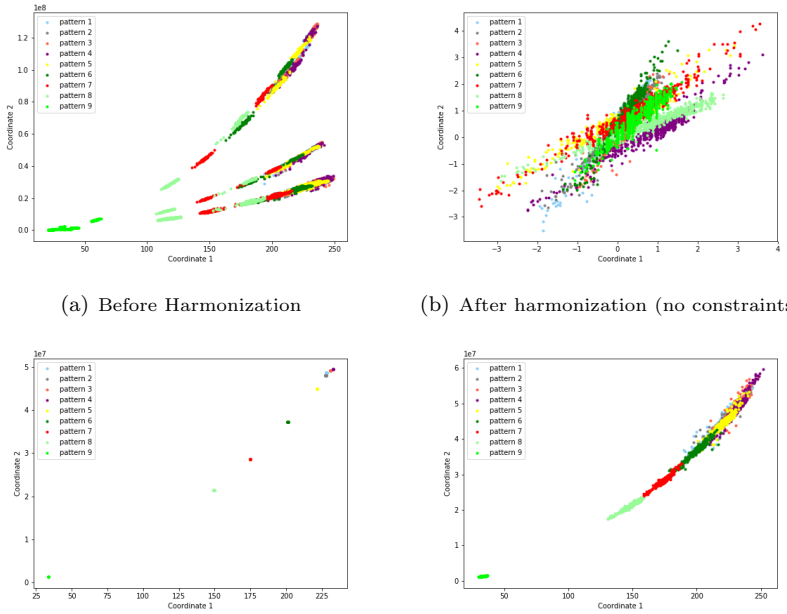
where $\alpha(z_{nc}^{(s)})$ represents the posterior probability of $\boldsymbol{x}_n^{(s)}$ belonging to pattern class c, or the responsibility that pattern class c takes for explaining the observation $\boldsymbol{x}_n^{(s)}$. Then the scanner effects adjusting data $\boldsymbol{y}_n^{(s)}$ corresponding to $\boldsymbol{x}_n^{(s)}$ can be expressed as

$$\boldsymbol{y}_n^{(s)} = \boldsymbol{\mu}_c + \boldsymbol{\Omega}_{sc}^{-1} (\boldsymbol{x}_n^{(s)} - \boldsymbol{\mu}_c - \boldsymbol{\gamma}_{sc}), \tag{70}$$

where $\boldsymbol{\Omega}_{sc} = (\boldsymbol{\Delta}_{\boldsymbol{sc}})^{\frac{1}{2}}$.

2.6 Experiment results and discussion

To better understand the importance of the additive constraints (12) and multiplicative constraints (13), we apply our proposed model (10) for each pattern class separately with the number of pattern class C equal to 1. As shown in Figure (7), we can see that the additive and multiplicative constraints



(a) Before Harmonization

(b) After harmonization (no constraints)

(c) After harmonization (additive constraints only)

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

Figure 7: Effects of the additive and multiplicative constraints. Our proposed model is applied separately for each pattern class with the number of pattern class set to 1, and the model parameters are initialized randomly.

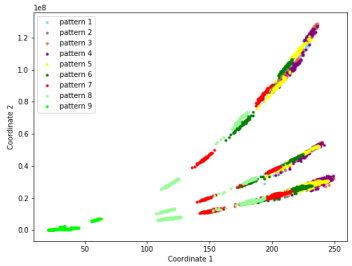
are essential because they help to make identifiable of the parameters and make the adjusted data to have reasonable mean and variance.

Now having realized the importance of the constraint conditions, we now show our experiment results in Figure 8. Note that the true pattern class labels are unknown and are predicted by our proposed model. As shown in Figure 8(b), the predicted pattern class labels are completely wrong, which leads to unacceptable harmonization results as shown in Figure 8(c). In other words, our proposed model fails to provide satisfying harmonization results.

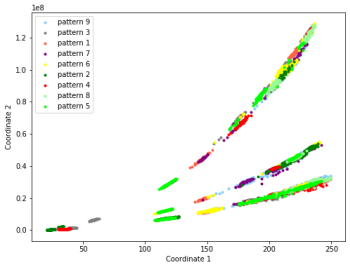
By further analyzing the model in detail, we come to the conclusion that our model fails because it lacks enough constraints to help identify the parameters. More specifically, in model (10), we can see that for a given scanner setting s, the mean of the c-th Gaussian distribution is $\boldsymbol{\mu}_c + \boldsymbol{\gamma}_{sc}$, 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 $\boldsymbol{\Delta}_{sc}$.

We have made it clear that our model fails to provide satisfying harmonization results because it lacks of enough constraints on parameters. Then it is natural to ask whether proper constraints could be imposed to help solve the problem? Unfortunately, the answer is No. No proper constraints could be proposed since we have no proper prior experience of the mean(or variance) and scanner effects for each pattern class. So more information should be provided if we want to have good harmonization results with the assumption that scanner effects are different for each pattern class.

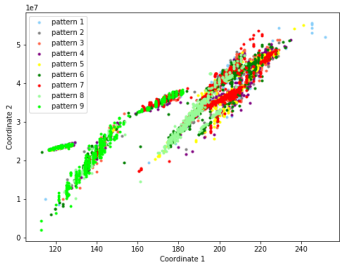
Now our problem can be regarded as solving an unsupervised classification problem and a harmonization problem simultaneously, it is indeed a tricky problem and impossible to solve without enough constraints to help identify the parameters. So we think to solve the problem by involving more information. For example, consider the problem as a learning problem (consisting of a training process and a prediction process), and suppose the pattern class labels are known during the training, then the problem becomes a supervised classification problem and at the same time a harmonization problem. We will consider this case as a future work.



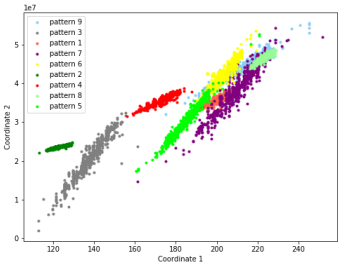
(a) Before Harmonization (colored by true pattern class labels)



(b) Before Harmonization (colored by predicted pattern class labels)



(c) After harmonization (colored by true pattern class labels)



(d) After harmonization (colored by predicted pattern class labels)

Figure 8: Harmonization results by our proposed model. The number of pattern class is set to 9 (the real pattern number), and the model parameters are initialized by the results of Gaussian mixture model.

References

[1] 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.

[2] 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.

[3] F. Orlhac, C. Nioche, M. Soussan, and I. Buvat, "Understanding changes in tumor texture indices in pet: a comparison between visual assessment and index values in simulated and patient data," *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 387–392, 2017.

[4] H. Benoit-Cattin, *Texture analysis for magnetic resonance imaging.* Texture Analysis Magn Resona, 2006.

[5] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: improved n3 bias correction," *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[6] J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, "Evaluating the impact of intensity normalization on mr image synthesis," in *Medical Imaging 2019: Image Processing*, vol. 10949, p. 109493H, International Society for Optics and Photonics, 2019.

[7] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[8] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," *arXiv preprint arXiv:1612.07003*, 2016.

[9] 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.