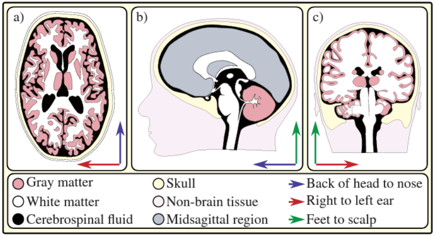
Privacy-Preserving MRI Site-Effect Removal for Brain Age Estimation

What is Human Brain?

The human brain comprises right and left hemispheres, each containing structures similar in size and shape



Source: RICHARD et al.

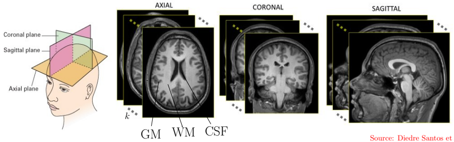
1. Grey matter (GM) - cerebral cortex controlling higher cognitive functions
2. White matter (WM) - tracts/fibers to transmit signals b/w brain regions
3. Cerebrospinal fluid (CSF) - watery fluid transferring nutrients and providing protective cushioning to the brain

Structural MRI (sMRI)

Structural MRI captures the anatomy/structure of the brain

MRI divides the brain into different planes

1. **Axial:** upper and lower parts.
2. **Coronal:** front and back portions.
3. **Sagittal:** left and right halves.



Source: Diedre Santos et

Size $= m \times n \times k$

- Size of a slice/frame $= m \times n$
- Number of slices $= k$

T1-weighted MRI

- GM & CSF appear darker
- WM appears brighter
- Used to outline brain anatomy

T2-weighted MRI

- GM & CSF appear brighter
- WM appears darker
- Used to highlight pathology

Neurological Disorders

1. Neurodegenerative Disorders - Characterized by progressive deterioration of nerve cells and brain tissue
   - **Alzheimer's Disease:** Demonstrates widespread GM atrophy, particularly in memory-related regions like the hippocampus and cortical areas.
   - **Parkinson's Disease:** Shows reduction in GM volume such as degeneration of neurons in the basal ganglia structures.
2. Neurodevelopmental Disorders - Arise from abnormalities in brain development leading to structural differences from early life
   - **Autism Spectrum Disorder (ASD):** Associated with altered connectivity and structural variations in social cognition regions like the amygdala and prefrontal cortex

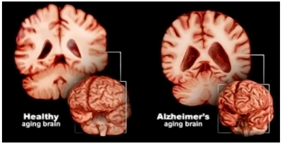
Uses of MRI for Neurological Disorders

Source: Inspiring/Shutterstock

Radiologists use MRI to look at anomalous anatomical structures

For instance, Alzheimer's disease manifests itself on an MRI:

- **GM Changes:** reduced GM volume visualized on MRI scans as regions of decreased intensity
- **WM Changes:** WM hyperintensities (WMHs) on MRI scans indicate damage in WM tracts
- **CSF Changes:** enlarged ventricles, showing an increase in the volume of CSF-filled spaces, are visible on MRI scans
- **Hippocampal Volume Loss:** Reduced hippocampal volume is apparent on MRI scans



Source: TheVisualMD

MRI Data Analysis

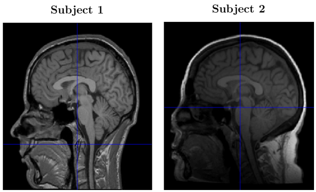
Employing computer vision with DL and ML for designing MRI-based studies includes:

1. Collecting and preparing the MRI training data
2. Either deriving/extracting the relevant information (features) or using the raw MRIs
3. Training and developing an ML/DL model for the specific task

Need to Prepare the MRIs for Analysis

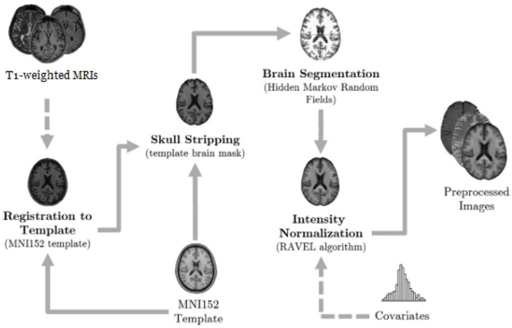
Raw brain MRIs have:

- Undesired information - head, face, and non-brain tissue
- Different head sizes
- Noise - hyperintense or bright areas
- Head motion during a session
- Different MRI scanners, scanner locations, head coils, etc.



MRI Data Analysis: Preprocessing Pipeline

1. Correct intensity non-uniformities
2. Remove the skull
3. Register to a standard space
4. Intensity normalization
5. Perform smoothing to reduce noise

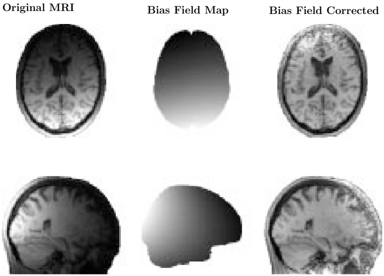


Source: Payares-Garcia et al.

MRI Preprocessing: Intensity Non-uniformity Correction

Uneven brightness across an MRI obscures true anatomical features and hinders accurate analysis

- Bias field refers to a slowly varying (low-frequency) intensity variation present across the entire MRI
- Caused by variations in magnetic field strength, radiofrequency coil sensitivity, and tissue properties
- Correction methods involve spatially smoothing the image to estimate the non-uniformity field, which is then divided to restore uniformity
- Common techniques: **N4 Bias Field Correction** and **Polynomial Fitting**

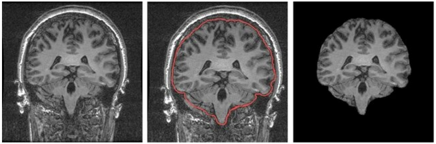


Source: John Sled et al.

MRI Preprocessing: Skull Stripping

Skull stripping or brain extraction removes non-brain tissues from MRI, leaving brain parenchyma (GM, WM) and CSF

- **Intensity thresholding** segments the brain from surrounding tissues
  - Determine an appropriate threshold value that separates the target structure (e.g., brain tissue) from the background or noise
  - Compare each voxel in the MRI image to this threshold and classify it as either brain or skull

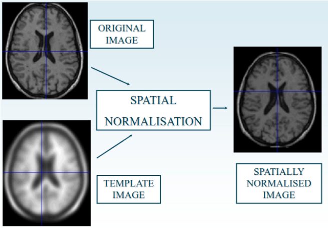


Source: Parsa Hosseini et al.

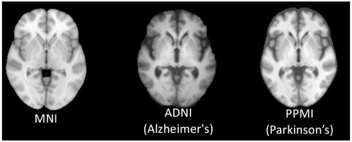
MRI Preprocessing: Registration to a Standard Space

All brain MRIs aligned to a common anatomical or coordinate framework

- Registration to a standard space facilitates group analysis and inter-subject comparison by ensuring spatial correspondence
- Registration algorithms compute spatial transformations (translation, rotation, scaling) and deform individual images to match a template (e.g., MNI)
- Improves spatial accuracy and consistency across MRI volumes and reduces motion-related artifacts



Source: Nicola Hobbs et al.

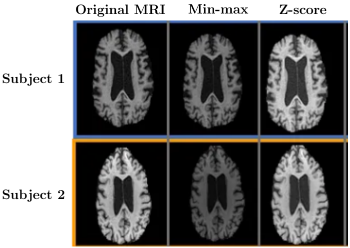


Commonly used templates (MNI-152 is the average of 152 brains)

MRI Preprocessing: Intensity Normalization

Standardizing the intensity values across different MRI volumes or subjects

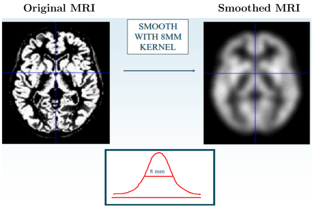
- MRI intensity values can vary due to differences in acquisition parameters, scanner characteristics, and subject-related factors.
- Normalization ensures that MRI data from different sources or individuals are on a common scale
- Methods adjust the intensity values of MRI volumes to match a predefined reference or standard distribution
- Techniques include **linear scaling, histogram matching, and z-score normalization**



MRI Preprocessing: Noise Reduction

Noise refers to random fluctuations in MRI signals that can obscure underlying anatomical structures and affect image quality

- Noise can arise from various sources, including electronic circuitry, thermal motion of molecules, and external interference
- Noise reduction techniques aim to enhance signal-to-noise ratio (SNR) while preserving image detail
- Common methods include **spatial filtering** (e.g., Gaussian smoothing), **temporal averaging**, and advanced denoising algorithms (e.g., wavelet-based methods)



Source: Nicola Hobbs et al.

- Excessive noise reduction may lead to loss of fine detail & blurring of boundaries

MRI Data Analysis: Brain Features & Extraction Methods

1. Voxel-wise
   - Voxel intensity values representing different brain tissues and CSF
   - Voxel-based morphometry (VBM)
2. Region-wise
   - Geometric measurements derived from MRIs such as volumes, thickness, and surface area of different regions
   - Region-based morphometry (RBM)
3. Surface-wise
   - Geometry of the cerebral cortex or cortical surface (GM only), such as thickness and sulcal depth
   - Surface-based morphometry (SBM)

Feature extraction: Voxel-based morphometry (VBM)

MRI voxel intensities are used in two broad ways:

Reduced voxel-wise features

- Size of an MRI 3D-volume $= m \times n \times k$
  - Typically, $182 \times 218 \times 182$ voxel intensity values per MRI with $1mm^3$ voxel size
  - Approx 2 million features per MRI (after masking - removal of the background)
- Number of features $\gg$ number of samples - Curse of dimensionality
  - Perform dimensionality reduction
- Train ML models on the reduced voxel features

Minimally preprocessed raw MRIs

- Provided directly as images to DL models such as 3D CNN and ResNet
- Learn the spatial patterns and correspondence in the images

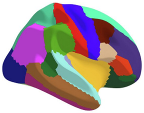
Feature extraction: Region-based morphometry (RBM)

RBM computes the geometric measurements of the regions of interest (ROIs) in an MRI with a reference atlas

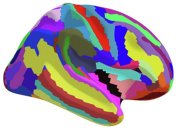
Commonly used reference atlases (parcellation)

- **Desikan atlas:** (68 ROIs)
  - A gyral-based atlas: a gyrus includes the part visible on the pial view
- **Destrieux atlas:** (148 ROIs)
  - Divides brain the cortex into gyral and sulcal regions
- **Neuromorphometrics atlas:** (284 ROIs)
  - Parcelates the whole brain (GM, WM, and CSF)

Desikan-Killiany atlas (Desikan et al., 2006)
68 ROIs, structural parcellation



Destrieux atlas (Destrieux et al., 2010)
148 ROIs, structural parcellation



Source: FreeSurfer

Region-based morphometry (RBM) Pipeline

- Compare the input MRI to the parcellation
- Divide the input MRI into parcels, aka regions of interest (ROI)
- Compute the geometric measurements for each parcel such as
  - Surface area (mm^2), cortical thickness (mm), and curvature (mm^{-1}) for each ROI

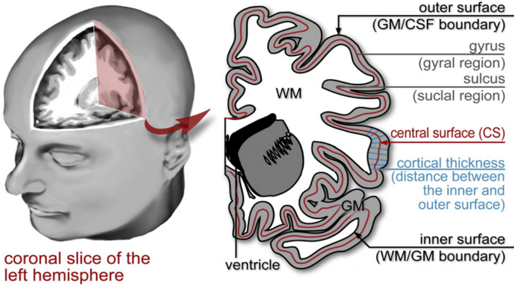
Freesurfer and SPM (CAT12 toolbox) are mainly used for RBM

Feature extraction: Surface-based morphometry (SBM)

SBM constructs the geometry of the brain tissue boundaries or the cortical (central) surface

Create a triangular mesh around the cortical surface with n nodes

- Commonly used methods to create meshes are **marching cubes algorithm**, **level set method**, or **the deformable surfaces method**



Source: CAT12 Manual

Compute cortical thickness (mm), sulcal depth, curvature (mm^{-1}) at each node

- Methods such as **Euclidean distance mapping** or **geodesic distance calculations** estimate the cortical thickness

MRI Data Analysis: MRI Data Collection/Integration

ML/DL models are data-hungry - robust models require big datasets

Larger samples of MRIs could represent the population and the pathalogy

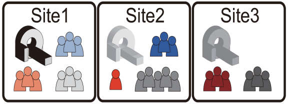
1. MRI acquisition is expensive and time-consuming
   - Hence, MRI studies combine data from different sources and sites, such as ADNI and BraTS
   - Raises data sharing and privacy concerns due to human subjects private data
2. Integrating different MRI sources introduces “unwanted variability” known as batch, scanner or site effects
   - This variability is present even after executing the current MRI preprocessing pipelines

Resultantly, downstream tasks (such as brain tumor detection and brain age estimation) are not “accurate” and not “generalizable”

MRI Data Integration: Site, Scanner & Batch Effects

Site effects: variability across different MRI collection sites or locations

- Different imaging protocols, hardware, software, personnel, or environment
- Differences in the preprocessing methods, techniques, and subjects



Source: Ayumu Yamashita et al.

Scanner effects: differences in MRIs due to variations in MRI scanners

- Different scanner manufacturers, models, magnetic field strengths, head coils, voxel sizes, acquisition parameters, and image parameters and reconstruction algorithm

Source: Ayumu Yamashita et al.

Batch effects: variability due to differences in data acquisition or processing

These effects (site, scanner & batch) are collectively referred to as "Site Effects"

Site Effects: Impacts on the MRI

Site-related signatures present in the MRI include:

1. Intensity inhomogeneity (contrast), where the same tissue appears with varying signal intensities across MRIs
   - The arbitrary nature of MRI intensity scale
2. Variations in voxel sizes and MRI resolutions between scanners
   - Impact the registration to the template and ROI classification
3. Shading artifacts and biases produced in the MRI
   - Head placement in the scanner results in the unanticipated activation of the receiver coils
4. Overestimation of brain volume and cortical measurements
   - Higher cortical thickness value on GE scanners than on Siemens due to low field strength and low resolution

MRI Site Effects: Impacts on MRI Data Analysis

MRIs with site-related signatures significantly impact the downstream analysis

The famous “name that dataset (site)” experiment

- Predict whether an MRI belongs to the site S given its MRI-derived features
- Approx. 90% accuracy with site effects and reduced accuracy after site-effect removal (almost random)

Multi-site MRI datasets produce less accurate results for downstream tasks

- Brain age estimation and Alzheimer's disease prediction accuracy increase after removing site-effects from MRI-derived metrics

The results are less generalizable and not representative of the population because of the non-biological information

Hence, these MRI site effects need to estimated and removed for a better analysis

MRI Data Collection/Integration: Privacy Issues

Gathering large MRI training datasets for large-scale robust analyses is challenging because:

- Cost-prohibitiveness of MRI data collection
- Varying institutional data-sharing policies
- Constrained data-usage agreements such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA)
- Human subjects data-privacy concerns

Need to find/estimate and remove MRI site-related signatures by preserving the data privacy

Privacy-preserving MRI site-effect removal or Privacy-preserving multi-site MRI debiasing problem

MRI Site-effect Removal: Existing Broad Approaches

The process is referred to as Harmonization, site-effect removal, debiasing

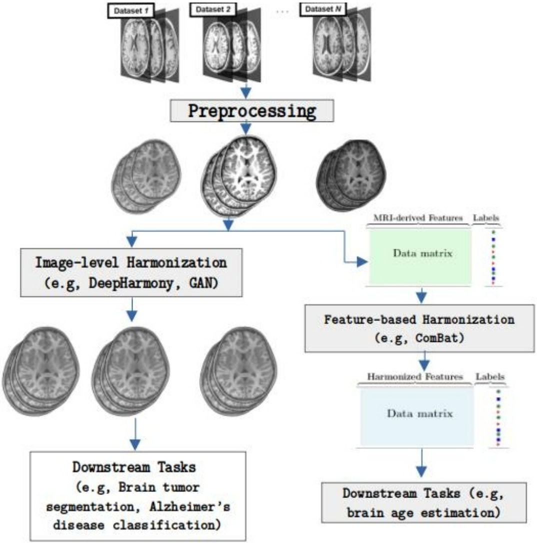
1. Standardized Acquisition
   - Inclusion & exclusion criteria, imaging protocols, quality control measures, and use of phantoms
2. MRI Preprocessing - explicit harmonization
   - Image-level contrast harmonization and intensity non-uniformity correction - doesn't explicitly account for batch
3. Statistical Harmonization - explicitly accounting for batch
   1. Image-level contrast harmonization (explicit), e.g., DeepHarmony
   2. Performed after feature extraction (implicit), e.g., ComBat
4. Robust Downstream Analysis
   - Meta and mega analysis, Hierarchial Bayesian Regression, e.g., ENIGMA

MRI Site-effect Removal: Statistical Harmonization

Harmonization homogenizes data set

Two broad approaches:

1. **Feature-Level:** Statistical harmonization of MRI-derived features (e.g., region-wise)
2. **Image-Level:** Transforming preprocessed MRIs (aka MRI-to-MRI or image-to-image harmonization)



Feature-Level MRI Harmonization

This process takes as input the multi-site MRI-derived feature vectors and outputs the standardized site-invariant features

E.g., ComBat is a popular feature-level MRI harmonization method:

- Models data as a combination of biological effects of interest, batch (site) effects, and noise
- Estimates mean and variance of batch effects for each feature for all batches
- Then adjusts the data by subtracting the estimated batch effect from each feature, and rescaling the data to have the same variance as the original data

Limitations:

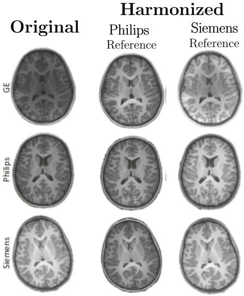
- Extracted features are study-specific (e.g., ageing or pathology-related)
- Assumptions about data distribution or underlying biological processes may not hold true in all cases, leading to biased results
- Images are lost, cannot be labelled after harmonization

Image-level Multi-site MRI Harmonization

This process aims to adjust the intensity values of individual MRI images to reduce or eliminate the multi-site effect

Applying a transformation function to the intensity values of each image to align them with a reference image or a target distribution

- The transformation function or representation can be learned using ML/DL algorithms such as Autoencoders and GANs
- Given source and target MRIs, the encoder of Autoencoder is trained to learn a shared latent space, while the decoder is trained to map the latent space back to the target data space



Source: Siyuan Liu et al.

Image-level MRI Harmonization: Approaches & Advantages

The goal is to learn an MRI-debiasing or transforming function f that takes a minimally-preprocessed MRI $X \in \mathbb{R}^{m \times n}$ and transforms it to a site-invariant feature space $X' \in \mathbb{R}^{m \times n}$

Two broad approaches based on the available multi-site MRI datasets:

1. Paired data: MRIs from one site/batch are chosen as reference/target (typically better quality) while the remaining sites are harmonized with respect to the target site (e.g., traveling subjects datasets)
2. Unpaired data: a standard reference MRI (just like MNI152 for registration) is chosen and the MRIs from all other sites are harmonized by adjusting their style/appearance/contrast to the reference MRI

Advantages over Feature-Level Harmonization:

- Retains variability across all aspects of MRI data, including spatial patterns, intensity distributions, and anatomical structures
- Harmonized MRI can be used for different downstream tasks (e.g., classification, regression, or segmentation)

Image-level MRI Harmonization: Evaluation Metrics

Evaluated by measuring the distance between images of different batches/sites

When paired data is available:

1. Distance quantified as voxel-level difference between harmonized image and true image from reference batch using MAE/MSE
2. Peak signal to noise ratio (PSNR) measures image quality by taking ratio of the maximum image value and the RMSE

In case of unpaired data:

1. Structural similarity index measure (SSIM), as the name implies, measures the degree to which structures are preserved post-transformation
   - SSIM is applied in unpaired data under the assumption that key structures are largely the same between subjects
2. Fréchet Inception Distance (FID), a common evaluation metric for GANs, measures distance between ground truth and generated image distributions as opposed to images themselves

Image-level MRI Harmonization: Research Gap

Existing MRI-to-MRI harmonization methods pool multi-site MRI and learn the site-invariant feature representation

- Data may not be available for pooling/sharing
- Pooling data creates privacy concerns
- Can't get more data to train the image-to-image MRI debiasing function
- Accuracy and reliability of the debiasing function is compromised
- Downstream tasks or analysis are not generalizable

Need to learn an MRI-to-MRI debiasing function with privacy preserving

MRI Data Privacy Preservation: Federated Learning (FL)

FL approach allows to train models on distributed data without pooling

Training local models on local servers and exchanging parameters (e.g., the weights and biases of a deep neural network) iteratively with the main site

Local sites/servers

- Gather/hold private MRI data
- Train local models
- Send weights to the main site

Main site/server

- Does not hold MRI data
- Aggregates the local model weights
- Returns updated weights to local sites

Model can be initialized either at the main or local servers

- In distributed learning, data is centrally stored (e.g., in a data center) - main goal is just to train faster
- In FL, data is naturally distributed and generated locally

MRI Data Privacy Preservation: FL Challenges

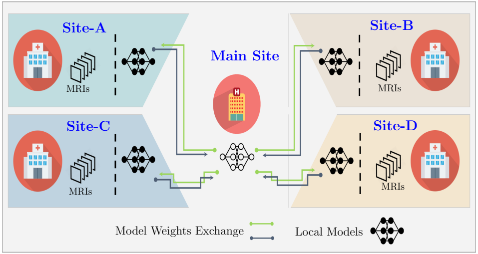
FL-based MRI analyses, though preserving privacy, pose many challenges

1. Non-IID data: Local datasets are heterogeneous, having different sizes and statistical distributions (age, gender, ethnicity, etc.)
   - Leads to biased model aggregation, where certain sites' data may dominate the final model, impacting its representativeness and generalization
2. Require frequent communication between the local sites and remote server
   - Leads to increased network bandwidth and latency requirements
3. Some data points may introduce "noise" in the training process (inputs, parameters, or outputs)
   - Degrade the accuracy of the model predictions (essentially privacy vs accuracy)

Heterogeneous FL-based Multi-site MRI Harmonization

Given brain MRIs from multiple heterogeneous local sites, train an MRI-to-MRI debiasing function or model without pooling the data

- Define the architecture of the local and global models
- Run the standard MRI preprocessing pipelines on the local sites
- Learn the weights/parameters of the MRI-harmonizing function



Source: Dianwen Ng et al.

1. Need to list FL approaches/applications on MRI (e.g. a slide)
2. The above FL assume homogeneous MRI (no site effect) ?
3. FL for Feature-Level MRI Harmonization
4. No FL for Image-level MRI Harmonization

These will lead to better specified Specific Aims

FL-based Image-level Multi-site MRI Harmonization

The specific aims of the proposed approach are:

1. To develop a novel MRI-to-MRI harmonizing function/model in an FL setting
2. To benchmark the performance of the proposed approach on a multi-site MRI dataset (such as OpenBHB and ADNI) and compare it with existing centralized MRI-to-MRI debiasing methods such as DeepHarmony and CycleGAN
3. To compare the performance of the proposed whole-image MRI debiasing model in two downstream tasks involving healthy and diseased MRIs
   - Healthy MRIs will be harmonized for brain age estimation and compared with the state-of-the-art centralized counterparts, while MRIs of the AD patients will be debiased to classify different AD stages and compared with their counterparts