

MOX-Report No. 19/2026

**Scanner-agnostic MRI harmonization via SSIM-guided
disentanglement**

Caldera, L.; Cavinato, L.; Ieva, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

SCANNER-AGNOSTIC MRI HARMONIZATION VIA SSIM-GUIDED DISENTANGLEMENT

Luca Caldera
MOX, Department of Mathematics
Politecnico di Milano
Milan, Italy
luca.caldera@polimi.it

Lara Cavinato
MOX, Department of Mathematics
Politecnico di Milano
Milan, Italy
lara.cavinato@polimi.it

Francesca Ieva
MOX, Department of Mathematics
Politecnico di Milano
Milan, Italy
francesca.ieva@polimi.it

For the Alzheimer’s Disease Neuroimaging Initiative*

ABSTRACT

The variability introduced by differences in MRI scanner models, acquisition protocols, and imaging sites hinders consistent analysis and generalizability across multicenter studies. We present a novel image-based harmonization framework for 3D T1-weighted brain MRI, which disentangles anatomical content from scanner- and site-specific variations. The model incorporates a differentiable loss based on the Structural Similarity Index (SSIM) to preserve biologically meaningful features while reducing inter-site variability. This loss enables separate evaluation of image luminance, contrast, and structural components. Training and validation were performed on multiple publicly available datasets spanning diverse scanners and sites, with testing on both healthy and clinical populations. Harmonization using multiple style targets, including style-agnostic references, produced consistent and high-quality outputs. Visual comparisons, voxel intensity distributions, and SSIM-based metrics demonstrated that harmonized images achieved strong alignment across acquisition settings while maintaining anatomical fidelity. Following harmonization, structural SSIM reached 0.97, luminance SSIM ranged from 0.98 to 0.99, and Wasserstein distances between mean voxel intensity distributions decreased substantially. Downstream tasks showed substantial improvements: mean absolute error for brain age prediction decreased from 5.36 to 3.30 years, and Alzheimer’s disease classification AUC increased from 0.78 to 0.85. Overall, our framework enhances cross-site image consistency, preserves anatomical fidelity, and improves downstream model performance, providing a robust and generalizable solution for large-scale multicenter neuroimaging studies.

Keywords Image harmonization · I2I Translation · Magnetic Resonance Imaging · Disentanglement

1 Introduction

The growing availability of brain MRI datasets presents valuable opportunities for understanding neurological diseases and supporting clinical applications. However, differences in imaging protocols, scanner models, and acquisition settings introduce inconsistencies that compromise the reliability of imaging biomarkers. Scanner type and acquisition site most strongly influence image characteristics such as contrast, brightness, and spatial resolution. Even identical scanner models can yield varying results due to differences in hardware, software, and maintenance, complicating multicenter studies and reducing reproducibility [1, 2]. Therefore, harmonizing MRI data across scanners and sites is essential for consistent, comparable analyses.

*Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Harmonization techniques fall into two categories: feature-based and image-based. Feature-based methods adjust derived metrics—such as cortical thickness, regional volumes, or diffusion metrics—without altering raw images [3, 4, 5]. Although computationally efficient, these methods depend heavily on preprocessing pipelines and risk propagating site-specific biases, which limit their ability to generalize across scanners and capture fine-grained voxel-level variations [6].

To overcome these limitations, image-based methods leverage machine learning to directly correct scanner- and site-induced variability at the voxel level. These approaches include Transformer models, Image-to-Image (I2I) translation, and Style Transfer approaches. Deep learning methods like CALAMITI [7], MURD [8], IGUANE [9], STGAN [10], and DISARM [11, 12] have shown promise in reducing site-related variability and improving harmonization for multicenter MRI studies. Although image-based methods directly address voxel-level variability, they often struggle to balance structural fidelity and visual consistency. Some approaches overemphasize appearance similarity, potentially altering anatomical content, while others preserve structure too rigidly, limiting the extent of harmonization. In addition, many models operate on 2D slices or small patches, which constrains their ability to capture global 3D context.

In this study, we propose a novel 3D MRI harmonization framework that extends the AGUIT domain translation network [13]—originally designed for 2D natural images—to volumetric medical data. This adaptation enables the disentanglement of anatomical structure from scanner- and site-specific variations, allowing the generation of harmonized images that preserve biologically meaningful anatomy while substantially reducing variability across imaging sites. The main contributions of this work are threefold: (1) we design a new harmonization model tailored for medical imaging; (2) we propose a differentiable loss function based on the Structural Similarity Index (SSIM) [14], allowing for end-to-end optimization that enhances perceptual consistency, preserves anatomical fidelity, and promotes appearance consistency across harmonized images; and (3) we conduct extensive validation on multiple publicly available datasets encompassing diverse scanners, sites, and populations. Validation and testing included both healthy individuals and clinical cohorts, such as autism spectrum disorder (ASD) and Alzheimer’s disease (AD) groups. In total, our evaluation comprised 2,733 3D T1-weighted MRI volumes acquired from 88 imaging sites using 19 different scanner models from the three main manufacturers—Siemens, Philips, and GE—providing highly heterogeneous and representative testing conditions that demonstrate the robustness and generalizability of the proposed framework.

The remainder of this paper is organized as follows: Section 2 describes the proposed harmonization framework; Section 3 outlines the datasets and image preprocessing steps; Section 4 details the experimental setup; Section 5 reports and discusses the experimental results; and Section 6 concludes the paper and outlines future directions.

2 Methodology

This section presents the proposed harmonization framework. We first introduce the mathematical formulation of the problem, followed by a detailed description of the model architecture and the training procedure.

2.1 Mathematical Formulation

Let $\mathcal{X}_u, \mathcal{X}_l \in \mathbb{R}^{1 \times H \times W \times D}$ denote the sets of unlabeled and labeled MR images, respectively. Each labeled image in \mathcal{X}_l is associated with a label vector $\mathbf{l} \in \mathcal{L} = \{-1, 1\}^K$, where K is the number of scanners and sites considered. In each label vector, the entry corresponding to the scanner and site where the image was acquired is set to 1, with all other entries set to -1. We assume that each image can be disentangled into two latent spaces:

- the anatomical space \mathcal{B} , which encodes the structural information of the brain, and
- the style space \mathcal{S} , which captures the non-structural features of the MR scan.

Formally, an image $\mathbf{x} \in \mathcal{X}$ can be represented by a pair of latent codes, where $\mathbf{c} \in \mathcal{B}$ and $\mathbf{s} \in \mathcal{S}$. The style code \mathbf{s} may itself be decomposed into a noise component $\mathbf{n} \in \mathcal{N}$ and a label component $\mathbf{l} \in \mathcal{L}$, such that $\mathbf{s} = (\mathbf{n}, \mathbf{l})$, where the label component \mathbf{l} captures scanner- and site-specific variations, while the noise component \mathbf{n} accounts for subject-specific, non-anatomical variations that introduce residual inconsistencies in MRI appearance possibly beyond scanner- or site-related effects.

2.2 Model Architecture

The network architecture comprises two encoders and a generator. The brain structure encoder, $E_b : \mathcal{X} \rightarrow \mathcal{B}$, maps an input image $\mathbf{x} \in \mathcal{X}$ to a lower-dimensional space \mathcal{B} , capturing anatomical features. The style encoder, $E_s : \mathcal{X} \rightarrow \mathcal{S}$, extracts non-anatomical features from the same image \mathbf{x} , encoding them into a separate latent space \mathcal{S} . The complete style vector $\mathbf{s} \in \mathcal{S} = (\mathcal{N}, \mathcal{L})$ is constructed by combining the noise component with the label component $\mathbf{l} \in \mathcal{L}$, which

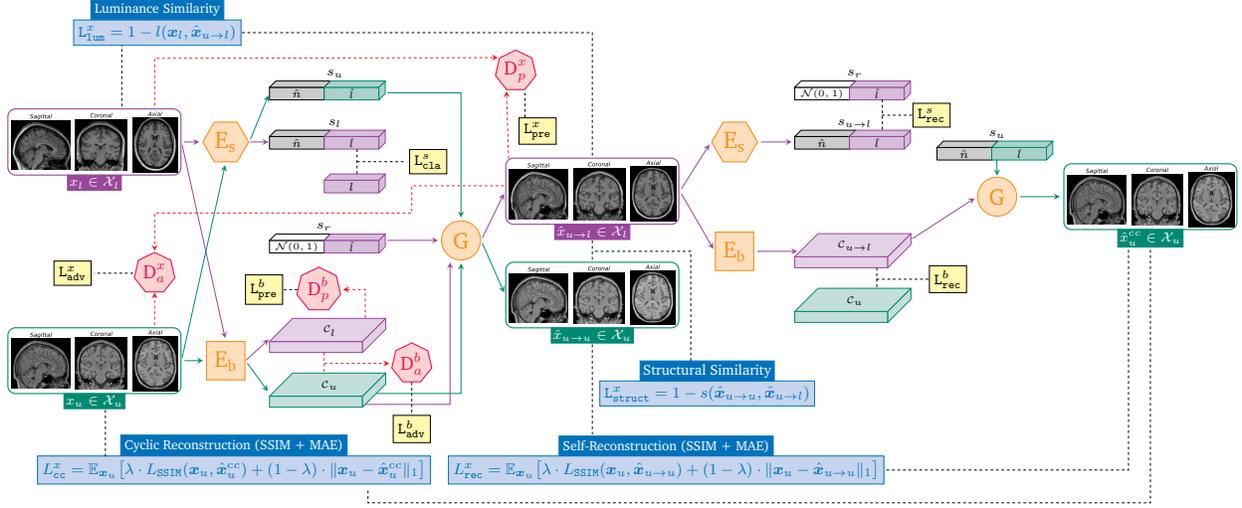


Figure 1: A high-level functional diagram of the model training procedure.

contains information about the scanner and the site of image acquisition. Finally, the generator, $G : (\mathcal{B}, \mathcal{S}) \rightarrow \mathcal{X}$, synthesizes a new image $\hat{x} \in \mathcal{X}$ that retains the anatomical structure from \mathcal{B} while incorporating the scanner-site style attributes from \mathcal{S} . Additionally, adversarial training is supported by incorporating discriminators to enhance realism and scanner-site-specific characteristics alignment.

2.3 Training

The training process, illustrated in Figure 1, is performed on pairs of unlabeled and labeled images, denoted as $x_u \in \mathcal{X}_u$ and $x_l \in \mathcal{X}_l$, respectively. Each image is independently processed by the encoders, E_b and E_s , to extract latent representations: the style codes s_u and s_l , and the anatomical structure codes c_u and c_l . The image x_u is reconstructed by passing its brain structure code c_u and style code s_u to the generator. To translate scanner-site-specific characteristics, the generator is instead given c_u and a style code s_r , where s_r is composed of random noise and the scanner-site label l from x_l . The resulting generated images are denoted as $\hat{x}_{u \rightarrow u}$ and $\hat{x}_{u \rightarrow l}$, respectively. For cyclic reconstruction, the encoders reprocess $\hat{x}_{u \rightarrow l}$ to extract its style and brain structure codes. The brain structure code $c_{u \rightarrow l}$ is then passed into the generator along with the style vector s_u of x_u . We denote the reconstructed image as \hat{x}_u^{cc} . The training objective involves minimizing a combination of loss functions, each guiding specific subnetworks in the architecture, as outlined in Figure 1.

We devise a new Structural Similarity Index Measure (SSIM)-based loss formulation that allows separate evaluation of the image luminance $l(x, y)$, contrast $c(x, y)$, and structure $s(x, y)$ components:

$$L_{SSIM}(x, y) = 1 - \text{SSIM}(x, y) = 1 - l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma$$

with

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

where μ_x, μ_y denote image means; σ_x, σ_y , their standard deviations; and σ_{xy} , their covariance. Constants C_1, C_2, C_3 stabilize computation, while exponents α, β, γ weight the SSIM components, allowing tailored loss functions that emphasize luminance, contrast, or structure.

We thus further combine the SSIM with Mean absolute error (MAE) in the Cycle Consistency Loss L_{cc}^x to ensure consistency in cyclic image translation, and in the Reconstruction Loss L_{rec}^x to ensure faithful reconstruction of the input image, with λ being a weighting factor:

$$L_{cc}^x = \mathbb{E}_{x_u} [\lambda \cdot L_{SSIM}(x_u, \hat{x}_u^{cc}) + (1 - \lambda) \cdot \|x_u - \hat{x}_u^{cc}\|_1]$$

$$L_{rec}^x = \mathbb{E}_{x_u} [\lambda \cdot L_{SSIM}(x_u, \hat{x}_{u \rightarrow u}) + (1 - \lambda) \cdot \|x_u - \hat{x}_{u \rightarrow u}\|_1]$$

To enforce the preservation of anatomical integrity and luminance during translation, we also introduce a Structural Consistency Loss L_{struct}^x and a Luminance Consistency Loss L_{lum}^x as:

$$L_{\text{struct}}^x = \mathbb{E}_{\mathbf{x}_u, \mathbf{x}_l} [1 - s(\hat{\mathbf{x}}_{u \rightarrow u}, \hat{\mathbf{x}}_{u \rightarrow l})]$$

$$L_{\text{lum}}^x = \mathbb{E}_{\mathbf{x}_u, \mathbf{x}_l} [1 - l(\mathbf{x}_l, \hat{\mathbf{x}}_{u \rightarrow l})]$$

Additional loss components are adapted from the AGUIT framework [13], including: Brain Structure Adversarial Loss, Style Continuity Loss, Content-Style Separation Loss, Image Adversarial Loss, Image Classification Loss, and Feature Consistency Loss. The overall objective for E_s, E_b, G and for $D_a^b, D_a^x, D_p^b, D_p^x$ become respectively:

$$\begin{aligned} L_{G, E_b, E_s} &= \lambda_{\text{lum}}^x L_{\text{lum}}^x + \lambda_{\text{struct}}^x L_{\text{struct}}^x + \lambda_{\text{cla}}^s L_{\text{cla}}^s + \lambda_{\text{adv}}^b L_{\text{adv}}^b - \lambda_{\text{pre}}^b L_{\text{pre}}^b + \lambda_{\text{rec}}^x L_{\text{rec}}^x \\ &\quad + \lambda_{\text{adv}}^x L_{\text{adv}}^x + \lambda_{\text{pre}}^{x, G} L_{\text{pre}}^{x, G} + \lambda_{\text{cc}} L_{\text{cc}} + \lambda_{\text{lat}} L_{\text{lat}} \\ L_D &= -\lambda_{\text{adv}}^c L_{\text{adv}}^c + \lambda_{\text{pre}}^b L_{\text{pre}}^b - \lambda_{\text{adv}}^x L_{\text{adv}}^x + \lambda_{\text{pre}}^{x, D} L_{\text{pre}}^{x, D} \end{aligned}$$

2.4 Inference

At inference, a new image $x_t \in \mathcal{X}$ is encoded by E_s and E_b to obtain its style s_t and anatomical structure c_t . The attribute component of s_t can be modified—for instance, by setting values corresponding to a target scanner or site to +1—to synthesize an image that retains the anatomy of x_t while exhibiting the desired scanner-site characteristics (Figure 2a). Alternatively, the noise component can be sampled from $z \sim \mathcal{N}(0, 1)$, with all attributes set to -1 , enabling the generation of images with consistent scanner-site effects independent of any training data context (Figure 2b).

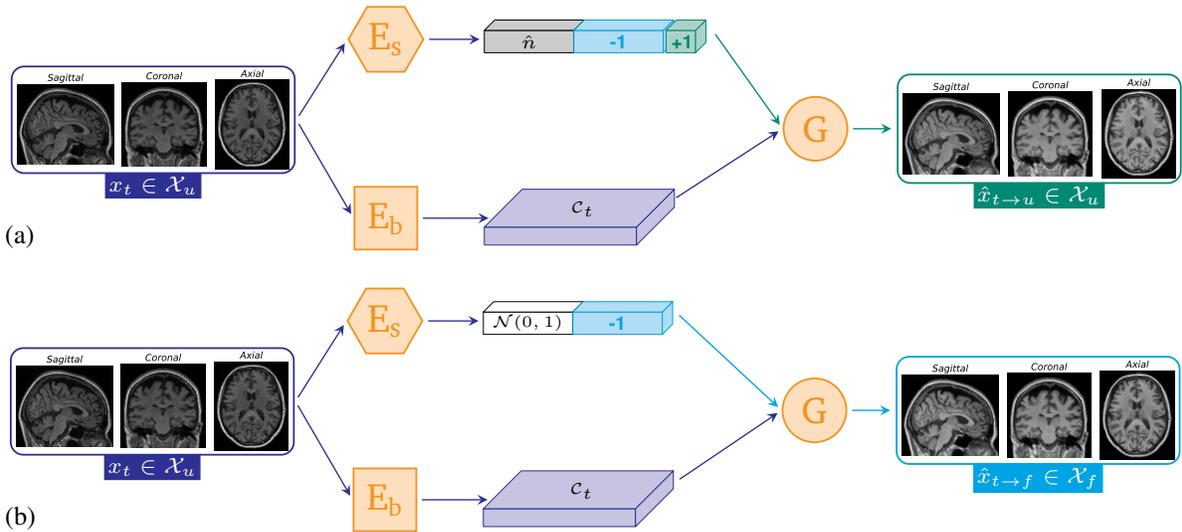


Figure 2: Visualizations of image generation during inference. (a) Generation with target scanner-site attributes. (b) Generation using style-agnostic.

3 Data Acquisition and Preprocessing

We conducted experiments using T1-weighted MRI data from five large-scale, publicly available datasets, comprising both healthy controls and diverse clinical populations across multiple acquisition sites and scanner platforms. Table 1 summarizes all datasets, including scanner models, manufacturers, acquisition sites, sources, and image counts.

Healthy Control Cohorts. We included data from four datasets containing healthy participants: the Alzheimer’s Disease Neuroimaging Initiative (ADNI3) [15], the Parkinson’s Progression Markers Initiative (PPMI) [16], the IXI Brain Development Dataset (IXI) [17], and the Southwest University Adult Lifespan Dataset (SALD) [18]. Together, these datasets include MRI scans acquired from 9 different scanners across 28 sites.

Table 1: Overview of the training and test datasets used in this study. The table lists scanner models, manufacturers, the number of acquisition sites, dataset sources, and image counts. Training and test sets for healthy controls (*HC*) are shown, along with the test sets containing both ASD, AD, and HC participants. The "Images (n)" column indicates the number of images used.

	Scanner Model	Manufacturer	Sites (n)	Dataset	Images (n)
Training (<i>HC</i>)	Prisma Fit	SIEMENS	3	ADNI 3	69
	Prisma	SIEMENS	5	ADNI 3	86
	TrioTim	SIEMENS	2	SALD, PPMI	190
	Gyrosan Intera	Philips	1	IXI	133
	Intera	Philips	1	IXI	177
Test (<i>HC</i>)	Unknown	GE	1	IXI	74
	Achieva dStream	Philips	3	ADNI 3, PPMI	31
	Achieva	Philips	5	ADNI 3, PPMI	27
	Skyra	SIEMENS	4	ADNI3	38
	Gyrosan Intera	Philips	1	IXI	189
	TrioTim	SIEMENS	2	SALD, PPMI	345
Test (<i>ASD + HC</i>)	Signa	GE	2	ABIDE I	185
	MR750	GE	1	ABIDE I	36
	Verio	SIEMENS	2	ABIDE I	84
	Allegra	SIEMENS	3	ABIDE I	277
	Achieva	Philips	2	ABIDE I	104
	Intera	Philips	2	ABIDE I	94
	TrioTim	SIEMENS	5	ABIDE I	332
Test (<i>AD + HC</i>)	Prisma	SIEMENS	6	ADNI 3-4	39
	Prisma Fit	SIEMENS	9	ADNI 3-4	59
	Biograph mMR	SIEMENS	1	ADNI 3	3
	Skyra	SIEMENS	6	ADNI 3-4	54
	Achieva	Philips	3	ADNI 3-4	16
	Achieva dStream	Philips	2	ADNI 3	3
	Ingenia	Philips	4	ADNI 3-4	17
	MR 7700	Philips	2	ADNI 4	22
	DISCOVERY MR750	GE	6	ADNI 3-4	30
	Signa Premier	GE	2	ADNI 3	9
	Verio	SIEMENS	1	ADNI 3	1
	Vida	SIEMENS	1	ADNI 4	9

Clinical Cohorts. We additionally incorporated T1-weighted MR images from two large multi-site datasets containing both clinical and healthy participants, providing a diverse and challenging evaluation setting. The first dataset, the Autism Brain Imaging Data Exchange I (ABIDE I) [19], includes structural MRI data from 1,112 individuals—comprising 539 participants with autism spectrum disorder (ASD) and 573 healthy controls. ABIDE I spans 17 international sites with heterogeneous acquisition protocols, scanner manufacturers, and demographic distributions, thus offering a rigorous benchmark for evaluating the robustness and generalizability of harmonization methods across diverse scanning environments. The second dataset, the Alzheimer’s Disease Neuroimaging Initiative (ADNI; Phases 3–4) [15], includes participants diagnosed with Alzheimer’s disease (AD) as well as healthy controls. From these phases, we included a subset of 94 participants with AD and 168 healthy controls. ADNI 3–4 provides structural MRI data from 43 imaging sites, acquired using Siemens, Philips, and GE scanners, serving as a complementary benchmark for evaluating harmonization performance in neurodegenerative populations.

Together, these datasets enable a comprehensive evaluation of the proposed model across a wide range of acquisition conditions, scanner types, and population characteristics, effectively encompassing most real-world scanning environments encountered in both clinical and research settings.

Image Preprocessing Pipeline. Image preprocessing was conducted using the FSL library [20, 21]. Initially, the images were standardized to a common orientation. Next, bias-field correction was applied to address magnetic field variations in the MRI scanner. The images were then registered to the Standard MNI152-T1-1mm space. The final preprocessed images had a shape of (1, 182, 218, 182).

4 Experimental setup

The model was trained on a dataset of 655 T1-weighted MR images from healthy controls, acquired across 5 scanners and 12 sites. For evaluation, we used an independent test set comprising 704 healthy control images collected using 6 scanners across 16 sites. Additional test sets included both clinical and healthy participants from the ABIDE I and ADNI 3–4 datasets. A detailed summary of the training and test datasets is presented in Table 1.

To evaluate our model, we harmonized all test images to match the domains of two distinct scanner-site pairs—referred to as A and B (Gyrosan Intera – Guy’s Hospital, and TrioTim – Southwest University, China)—as well as to a style-agnostic reference. The effectiveness of the harmonization was evaluated using three complementary criteria: (i) preservation of anatomical structures, (ii) consistency of image appearance across sites, and (iii) performance on downstream predictive tasks. In addition, axial, coronal, and sagittal slices were visually inspected to qualitatively assess the realism and inter-site consistency of the harmonized images.

Preservation of Anatomical Structure. We used the structural component of the SSIM to assess the preservation of anatomical structures following harmonization. Given the absence of ground truth data for contrast and luminance, we focus specifically on SSIM between the original and harmonized images. This metric quantifies the extent to which the underlying anatomical features are maintained after harmonization. Specifically, we computed SSIM values across image pairs representing diverse anatomical structures, as well as across the entire test set, for all three harmonization targets (scanner-site pairs A and B, and the style-agnostic reference), considering each harmonized–original image pair.

Consistency of image appearance. We assess appearance consistency by comparing mean voxel intensity distributions across scanner-site pairs before and after harmonization. To quantify alignment, we compute the Wasserstein distance between all pairs of scanner-site distributions, both pre- and post-harmonization, for all three harmonization targets. We report the mean and standard deviation in both cases. Additionally, to further assess consistency, we calculate the SSIM, focusing on the luminance component, between all pairs of original test images and, separately, all pairs of harmonized images. Again, we report the mean and standard deviation before and after harmonization. These analyses were conducted using the test dataset composed exclusively of healthy controls (Table 1).

Downstream Analysis. To evaluate the impact of harmonization on downstream neuroimaging applications, we performed two predictive tasks using convolutional neural networks (CNNs): (i) *age prediction* and (ii) *disease classification*. In both tasks, we employed 3D convolutional architectures based on the ResNet family. Each model was trained separately on non-harmonized and harmonized data to assess the effect of harmonization on predictive performance. Model evaluation was conducted using 5-fold cross-validation, reporting the mean and standard deviation of the respective metrics.

For *age prediction*, the CNN was trained to estimate chronological age from structural MRI scans using the ABIDE I dataset. Performance was assessed via mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2).

For *disease classification*, a binary 3D ResNet model was trained to distinguish Alzheimer’s disease (AD) patients from healthy controls using the ADNI 3–4 dataset. Evaluation metrics included accuracy, balanced accuracy, and area under the receiver operating characteristic curve (AUC).

5 Results

In this section, we present the results of the experiments described in Section 4.

Figure 3 provides a visual comparison of the harmonization outcomes across the three target domains. The original images exhibit substantial inter-domain variability, whereas the harmonized outputs demonstrate visually consistent

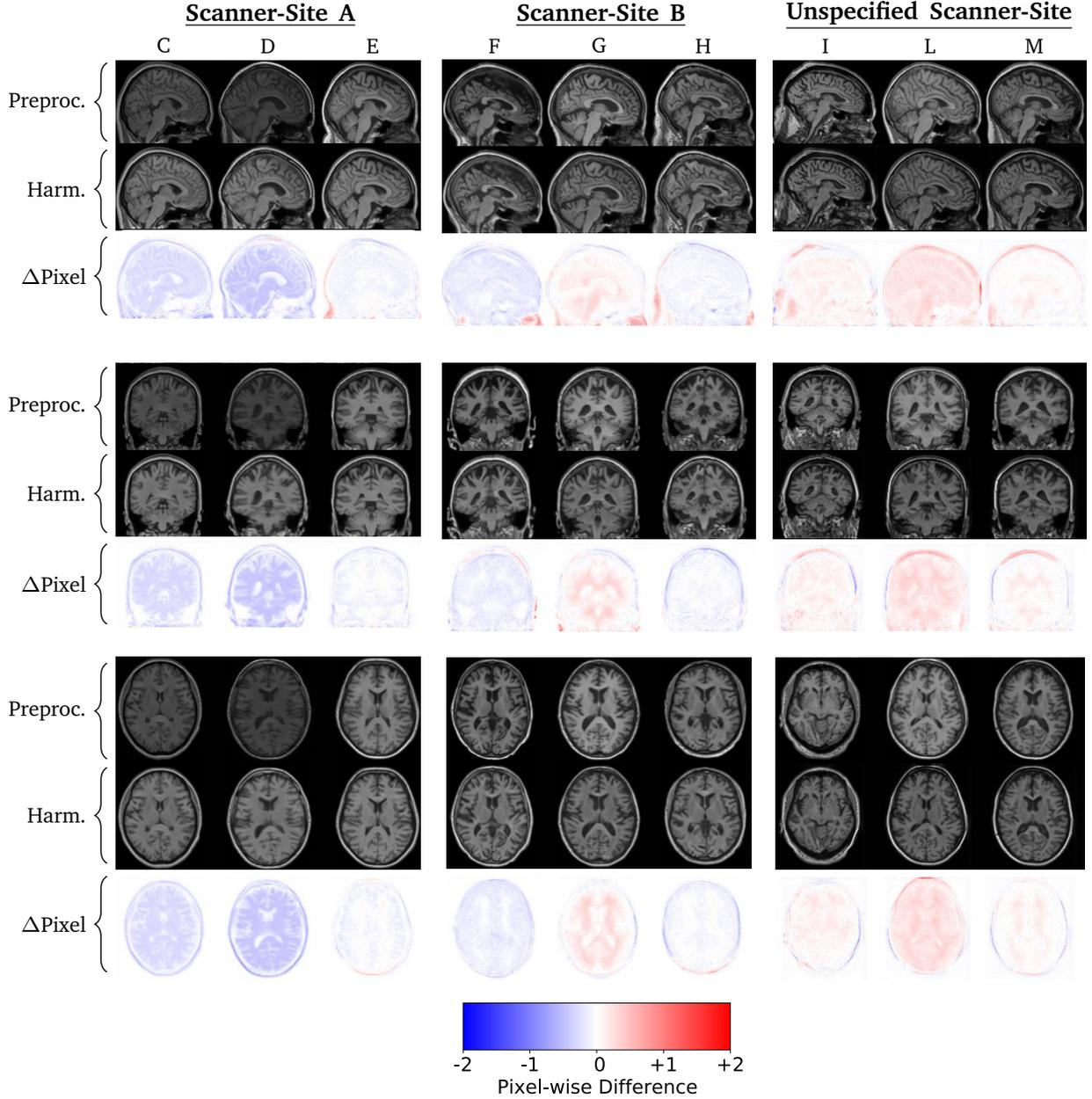


Figure 3: Axial, coronal, and sagittal slices from 9 MR images, both from the original scans (C-M) and their harmonized versions according to scanner-site pair A (left column), scanner-site pair B (middle column), and a style-agnostic target (right column). Heatmaps indicate pixel-wise differences between the original and harmonized images.

appearances across all style targets. This consistency is particularly evident in the heatmaps, which illustrate pixel-wise differences. The anatomical structures remain visually well-preserved after harmonization, indicating that the proposed method effectively reduces style-related variability while maintaining underlying anatomical integrity.

Nevertheless, visual assessment alone is insufficient to draw definitive conclusions regarding the quality of harmonization or the preservation of anatomical details. Therefore, we complement the qualitative evaluation with quantitative analyses to rigorously assess both the harmonization performance and the fidelity of anatomical preservation.

Preservation of Anatomical Structure. Across 100 randomly selected image pairs representing diverse anatomical structures, the average SSIM was 0.68 ± 0.05 . When all test images and all three harmonization targets were considered,

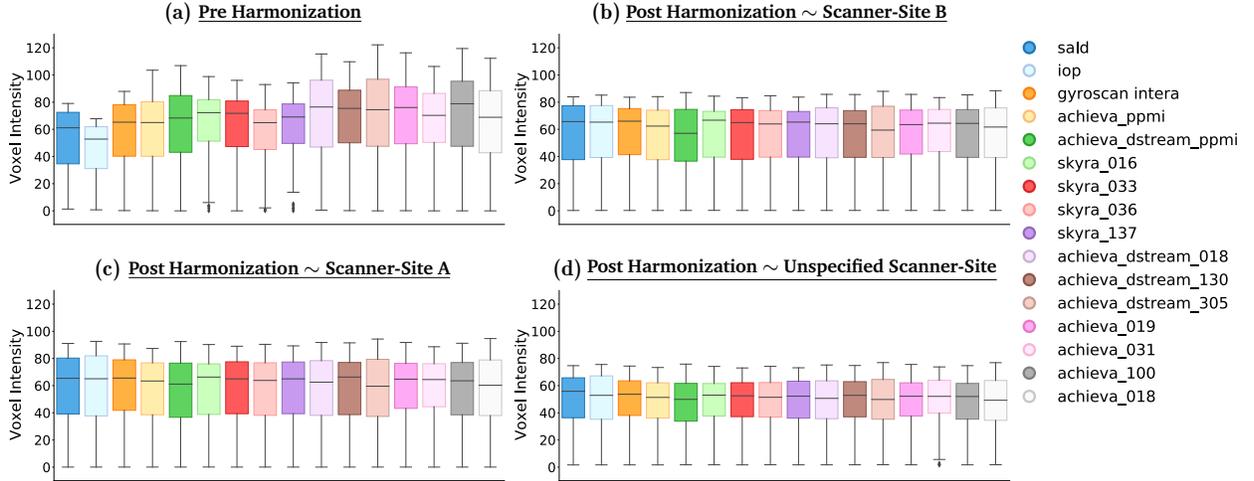


Figure 4: Comparison of voxel intensity mean distributions for each scanner-site pair in the test images, before (a) and after harmonization using the scanner-site pair B (b), the scanner-site pair A (c), and the unspecified scanner-site as the target scanner-site attributes.

the average SSIM increased to 0.971 ± 0.012 . This marked improvement demonstrates that the proposed model achieves strong anatomical preservation, reflected in the significantly higher SSIM scores between original and harmonized images of the same subject.

Consistency of image appearance. We compared the voxel intensity mean distributions for each of the 16 scanner-site pairs in the test images, both before and after harmonization with respect to the three target attributes (Figure 4). As shown in the plots, the proposed model achieves strong alignment of voxel intensity distributions across all target types, consistent with visual inspection.

This observation is quantitatively supported by the results summarized in Table 2. The Wasserstein distance between scanner-site distributions is markedly reduced after harmonization, from 8.454 ± 5.347 pre-harmonization to 1.989 ± 0.703 , 1.601 ± 0.492 , and 1.677 ± 0.559 for targets A, B, and the style-agnostic reference, respectively. This substantial reduction indicates that intensity distributions across scanners and sites are brought into close agreement, effectively mitigating scanner- and site-related variability. Corresponding heatmaps in Figure 5 visualize the pairwise Wasserstein distances.

Similarly, the luminance similarity increases from 0.952 ± 0.037 pre-harmonization to 0.981 ± 0.018 , 0.986 ± 0.016 , and 0.989 ± 0.014 after harmonization to the three targets. These results demonstrate that the harmonized images not only exhibit well-aligned voxel intensity distributions but also maintain high voxel-level luminance consistency across sites.

Table 2: Evaluation of luminance similarity between images and Wasserstein distance between the mean voxel intensity distribution. We report the values before and after harmonization to all three targets considered.

Metric	Pre-Harmonization	A (Scanner-Site)	B (Scanner-Site)	Unspecified (Scanner-Site)
Wasserstein distance	8.454 ± 5.347	1.989 ± 0.703	1.601 ± 0.492	1.677 ± 0.559
Luminance Similarity	0.952 ± 0.037	0.981 ± 0.018	0.986 ± 0.016	0.989 ± 0.014

Downstream Analysis. For age prediction, Table 3 shows that the mean absolute error (MAE) decreases from 5.36 ± 1.89 years pre-harmonization to 3.30 ± 0.48 , 3.39 ± 0.36 , and 3.31 ± 0.78 years for the three harmonization targets (scanner-site A, scanner-site B, and unspecified target, respectively). Similarly, the coefficient of determination (R^2) increases from 0.091 ± 0.571 to 0.643 ± 0.051 , 0.545 ± 0.094 , and 0.639 ± 0.140 , while the root mean squared error (RMSE) similarly decreases across all targets. These improvements indicate that harmonization effectively minimizes scanner- and site-related variability, allowing the CNN to extract biologically relevant age-related information without interference from acquisition-related biases.

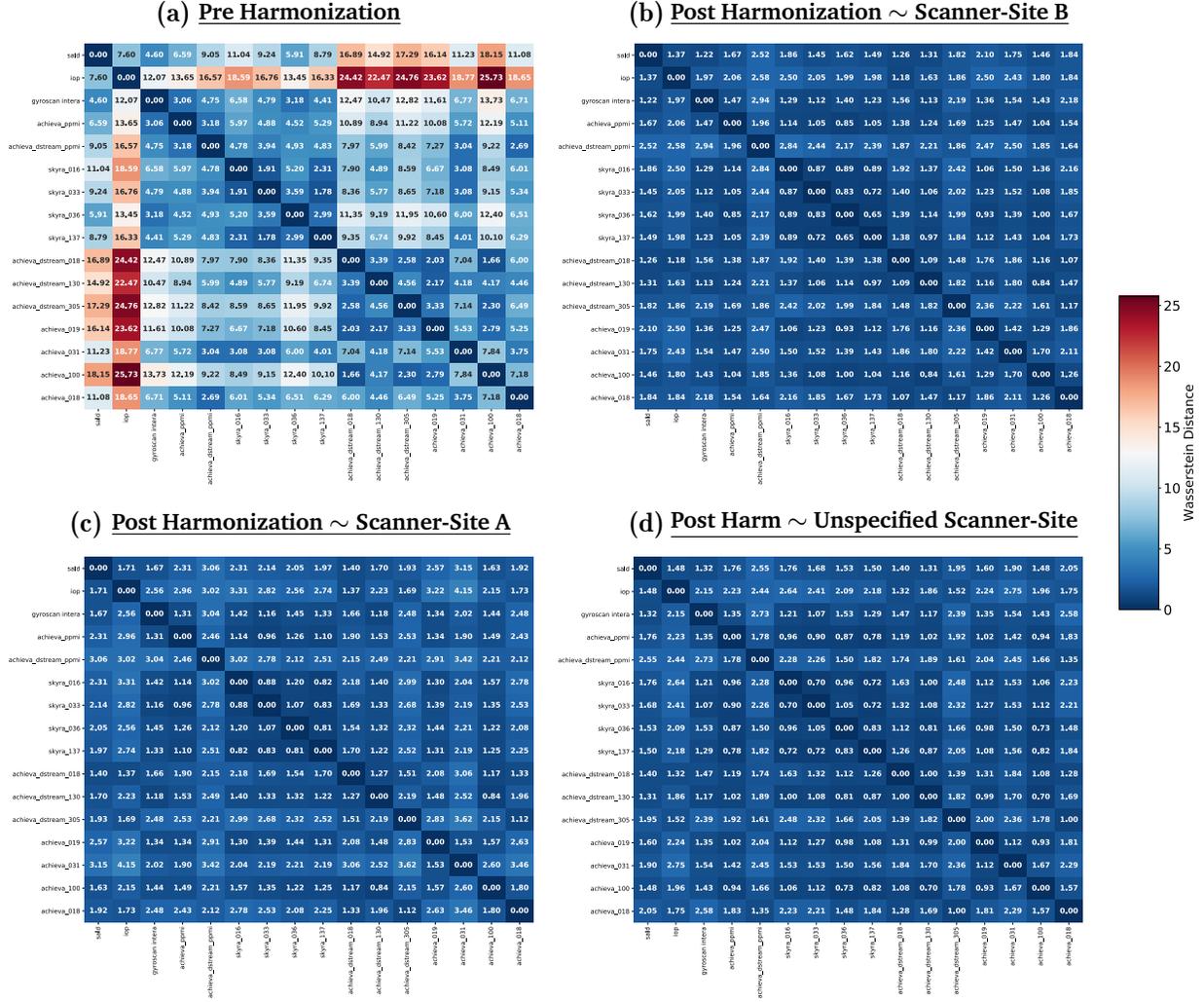


Figure 5: Heatmaps showing the pairwise Wasserstein distances between mean voxel intensity distributions for all scanner-site pairs in the healthy control test dataset. Panel (a) shows distances before harmonization, while panels (b), (c), and (d) show distances after harmonization to scanner-site pair B, scanner-site pair A, and the style-agnostic target, respectively.

For *Alzheimer’s disease (AD) classification*, harmonization leads to clear improvements (Table 4). The area under the ROC curve (AUC) increases from 0.778 ± 0.062 pre-harmonization to 0.848 ± 0.082 , 0.822 ± 0.077 , and 0.836 ± 0.072 for scanner-site A, scanner-site B, and the unspecified target, respectively. Other metrics, including accuracy and balanced accuracy, show similar improvements, reflecting enhanced overall model performance across all harmonization targets.

The consistency of these improvements across both tasks and all harmonization targets demonstrates the robustness of the approach. Overall, these downstream analyses indicate that harmonization not only improves inter-site consistency and preserves structural information but also enhances the predictive utility for both regression and classification. This confirms that the method effectively reduces site-specific biases while retaining biologically relevant information critical for neuroimaging applications.

6 Conclusion

We have presented a novel 3D T1-weighted MRI harmonization framework that disentangles anatomical content from scanner- and site-specific variations. By incorporating an SSIM-based loss alongside traditional reconstruction and

Table 3: Performance of the age prediction model before and after data harmonization. The table reports mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) with standard deviations across folds.

Metric	Pre-Harmonization	A (Scanner-Site)	B (Scanner-Site)	Unspecified (Scanner-Site)
<i>MAE</i>	5.36 ± 1.89	3.30 ± 0.48	3.39 ± 0.36	3.31 ± 0.78
<i>RMSE</i>	7.20 ± 2.38	4.79 ± 0.46	5.41 ± 0.80	4.74 ± 1.05
R^2	0.091 ± 0.571	0.643 ± 0.051	0.545 ± 0.094	0.639 ± 0.140

Table 4: Classification performance (AD vs Healthy Controls) on MRI data before and after harmonization. The table reports the mean and standard deviation across folds for accuracy, balanced accuracy, and AUC.

Metric	Pre-Harmonization	A (Scanner-Site)	B (Scanner-Site)	Unspecified (Scanner-Site)
<i>Accuracy</i>	0.702 ± 0.062	0.759 ± 0.095	0.736 ± 0.052	0.744 ± 0.065
<i>Balanced Accuracy</i>	0.647 ± 0.062	0.727 ± 0.091	0.667 ± 0.067	0.704 ± 0.059
<i>AUC</i>	0.778 ± 0.062	0.848 ± 0.082	0.822 ± 0.077	0.836 ± 0.072

adversarial objectives, our method achieves high structural fidelity while effectively reducing inter-site variability. Extensive evaluation across multiple large-scale datasets—including both healthy controls and clinical populations (ASD and AD)—demonstrated that harmonized images exhibit consistent voxel intensity distributions, improved luminance similarity, and preserved anatomical structures.

Importantly, harmonization led to substantial improvements in downstream neuroimaging tasks: brain age prediction errors decreased markedly, and Alzheimer’s disease classification performance improved across all harmonization targets. These results confirm that our framework not only standardizes MRI appearance across scanners and sites but also enhances the extraction of biologically meaningful features relevant for clinical and research applications.

Overall, our approach provides a robust, generalizable solution for large-scale, multicenter neuroimaging studies, facilitating reproducible analyses and cross-site comparability. Future work will explore the extension of this framework to other imaging modalities and anatomical regions, as well as the integration of additional covariates—such as age, sex, and pathological markers—into the style representation.

Acknowledgements

L. Caldera is funded by Health Big Data project sponsored by the Italian Ministry of Health (CCR-2018-23669122). L. Cavinato is funded by the National Plan for NRRP Complementary Investments “Advanced Technologies for Human-centred Medicine” (PNC0000003). The present research is part of the activities of “Dipartimento di Eccellenza 2023-2027”.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

- [1] Hidemasa Takao, Naoto Hayashi, and Kuni Ohtomo. Effect of scanner in longitudinal studies of brain volume changes. *Journal of Magnetic Resonance Imaging*, 34(2):438–444, 2011.
- [2] Russell T Shinohara, Jiwon Oh, Govind Nair, et al. Volumetric analysis from a harmonized multisite brain mri study of a single subject with multiple sclerosis. *American Journal of Neuroradiology*, 38(8):1501–1509, 2017.
- [3] Mahbaneh Eshaghzadeh Torbati, Davneet S Minhas, Ghasan Ahmad, et al. A multi-scanner neuroimaging data harmonization using ravel and combat. *Neuroimage*, 245:118703, 2021.
- [4] Joaquim Radua, Eduard Vieta, Russell Shinohara, et al. Increased power by harmonizing structural mri site differences with the combat batch adjustment method in enigma. *Neuroimage*, 218:116956, 2020.
- [5] Daniel Moyer, Greg Ver Steeg, Chantal MW Tax, and Paul M Thompson. Scanner invariant representations for diffusion mri harmonization. *Magnetic resonance in medicine*, 84(4):2174–2189, 2020.
- [6] Soolmaz Abbasi, Haoyu Lan, Jeiran Choupan, Nasim Sheikh-Bahaei, Gaurav Pandey, and Bino Varghese. Deep learning for the harmonization of structural mri scans: a survey. *BioMedical Engineering OnLine*, 23(1):90, 2024.
- [7] Lianrui Zuo, Blake E Dewey, Yihao Liu, et al. Unsupervised mr harmonization by learning disentangled representations using information bottleneck theory. *NeuroImage*, 243:118569, 2021.
- [8] Siyuan Liu and Pew-Thian Yap. Learning multi-site harmonization of magnetic resonance images without traveling human phantoms. *Communications Engineering*, 3(1):6, 2024.
- [9] Vincent Roca, Grégory Kuchcinski, Jean-Pierre Pruvo, et al. Iguane: A 3d generalizable cyclegan for multicenter harmonization of brain mr images. *Medical Image Analysis*, 99:103388, 2025.
- [10] Yunje Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197, 2020.
- [11] Luca Caldera, Lara Cavinato, Andrea Cappozzo, et al. Disarm: Disentangled scanner-free image generation via unsupervised image2image translation. In *International Workshop on Machine Learning in Clinical Neuroimaging*, pages 102–112. Springer, 2025.
- [12] Luca Caldera, Lara Cavinato, Alessio Cirone, et al. Disarm++: Beyond scanner-free harmonization. *arXiv preprint arXiv:2505.03715*, 2025.
- [13] Xinyang Li, Jie Hu, Shengchuan Zhang, et al. Attribute guided unpaired image-to-image translation with semi-supervised learning. *arXiv preprint arXiv:1904.12428*, 2019.
- [14] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

- [15] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
- [16] Kenneth Marek, Danna Jennings, Shirley Lasch, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.
- [17] Brain Development Project. IXI – Information eXtraction from Images Dataset. <https://brain-development.org/ixi-dataset/>, 2019.
- [18] Dongtao Wei, Kaixiang Zhuang, Lei Ai, et al. Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset. *Scientific data*, 5(1):1–10, 2018.
- [19] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.
- [20] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.
- [21] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, et al. Fsl. *Neuroimage*, 62(2):782–790, 2012.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 17/2026** Caldera, L.; Bottacini, G.; Cavinato, L.
MAGIC-Flow: multiscale adaptive conditional flows for generation and interpretable classification
- 18/2026** Caldera, L.; Cavinato, L.; Ieva, F.
Scanner-agnostic MRI harmonization via SSIM-guided disentanglement
- 15/2026** Zecchi, A. A.; Sanavio, C.; Cappelli, L.; Perotto, S.; Roggero, A.; Succi, S.
Block encoding of sparse matrices with a periodic diagonal structure
- 14/2026** Agasisti, T.; Cannistrà, M.; Paganoni, A.M.
Nudging communication for students at risk: experimental evidence from an Italian university
- 13/2026** Dimola, N.; Coclite, A.; Zunino, P.
Neural Preconditioning via Krylov Subspace Geometry
- 12/2026** Corbetta A; Logan K.M.; Ferro M; Zuccolo L; Perola M.; Ganna A.; Di Angelantonio E.;Ieva F.
Longitudinal patterns of statin adherence and factors associated with decline in over one million individuals in Finland and Italy
- 11/2026** Cicalese, G.; Ciaramella, G.; Mazzieri, I.; Gander, M. J.
Optimized Schwarz Waveform Relaxation for the Damped Wave Equation
- 10/2026** Dimola, N.; Franco, N. R.; Zunino, P.
Numerical Solution of Mixed-Dimensional PDEs Using a Neural Preconditioner
- 09/2026** Manzoni, V.; Ieva, F.;Larranaga, A.C.; Vetrano, D.L.; Gregorio, C.
Hidden multistate models to study multimorbidity trajectories
- 08/2026** Micheletti, S.
Newmark time marching as a preconditioned iteration for large SPD linear systems