

OHANA: Optimizing Heterogeneous Multi-Artifact Correction in Neuroimaging Analysis

Alícia Oliveira^{1,2}, Beatriz Cepa^{1,2}, António Sousa^{1,2}, Cláudia Brito^{1,2}

¹HASLab, INESC TEC, Braga, Portugal

²Department of Informatics, University of Minho, Braga, Portugal

{alicia.oliveira, beatriz.cepa, antonio.l.sousa, claudia.v.brito}@inesctec.pt

Abstract—Magnetic Resonance Imaging (MRI) is highly susceptible to acquisition artifacts, which degrade image quality and compromise medical diagnosis. In clinical settings, the absence of robust recovery methods forces technologists to repeat corrupted scans, unnecessarily increasing operational costs and patient discomfort. Although Deep Learning methods have shown potential for correcting MRI artifacts, their development is hindered by the limited availability of artifact-corrupted images. When available, data remain dispersed across hospitals and cannot be centralized due to privacy and regulatory constraints. Compounding this, scanner heterogeneity requires models to be trained on multi-institutional data to achieve efficient correction.

To address these challenges, we propose *OHANA*, an end-to-end solution for synthetic artifact generation and correction in multi-contrast brain MRI. We first introduce a synthetic data generator that simulates multiple artifact types, creating artifact-corrupted training datasets. Using this augmented data, we extend MC2-Net to perform artifact correction across four MRI contrasts. *OHANA* also integrates Federated Learning, allowing multiple institutions to collaboratively train models without sharing data.

Experiments demonstrate that *OHANA* outperforms state-of-the-art artifact-correction approaches, achieving a 17.2% improvement in the Structural Similarity Index Measure. A radiologist assessed the realism of the generated artifacts and the diagnostic fidelity of the corrected images. These results highlight the potential of *OHANA* to improve medical diagnosis.

Index Terms—MRI, medical imaging, artifact correction, synthetic data generation, deep learning, federated learning

I. INTRODUCTION

Magnetic Resonance Imaging (MRI) is widely used in medical diagnosis, providing detailed, non-invasive insights into human anatomy. However, MRI scans are susceptible to acquisition artifacts caused by hardware malfunctions, software glitches, motion, and spatial limitations. These distortions reduce image quality and may compromise diagnostic reliability [1]. Patient motion is the dominant source of artifacts, accounting for approximately 95% of cases [2]. Other artifacts,

including aliasing [3], magnetic susceptibility distortions [4], and noise interference [5], further degrade image quality.

In clinical settings, the absence of robust recovery methods forces technologists to repeat corrupted scans, resulting in extended examination times and increased operational costs. Motion-degraded images require sequence repeats in up to 20% of MRI exams, resulting in estimated annual revenue losses of \$115,000 to \$139,000 per scanner [6].

Deep Learning (DL) has emerged as a promising approach for improving MRI image quality by learning to correct artifacts directly from data [7]. By establishing mappings between corrupted and artifact-free images, DL-based methods can restore degraded scans without repeated acquisitions.

A key obstacle in developing artifact-correction models is the limited availability of corrupted MRI data. Many public datasets undergo strict quality control, removing images that contain artifacts. This is especially problematic for artifact correction, which often requires paired clean scans as Ground Truth (GT) references [7]. Synthetic data generation methods address this gap by simulating realistic artifacts to produce large volumes of corrupted images for training [8].

Another challenge is that medical imaging datasets are dispersed across institutions and cannot be centralized due to privacy and regulatory constraints [9]. As scanner and protocol variability introduce domain shifts, it motivates multi-institutional training for robust artifact correction. As such, Federated Learning (FL) rises as an opportunity for collaboration, i.e., institutions keep data local and share only model updates [10], [11].

To address these challenges, we propose *OHANA*, an open-source¹ end-to-end solution for synthetic artifact generation and correction in multi-contrast brain MRI. *OHANA* integrates two models within an FL framework: a synthetic data generation model that simulates motion, aliasing, magnetic susceptibility, and noise; and an extension of MC2-Net [12] for artifact correction across four MRI contrasts. In addition to T1-weighted (T1w), T2-weighted (T2w), and Fluid Attenuated Inversion Recovery (FLAIR), we include T1-weighted Contrast-Enhanced (T1CE) scans. While four contrasts were used to maximize anatomical coverage, *OHANA* remains applicable to single- and multi-contrast settings. For collaborative setup,

This work was funded jointly by the EuroHPC Joint Undertaking, through the European Union's Connecting Europe Facility and the Horizon 2020 research and innovation programme, as well as Portugal under the project "Operação do Supercomputador Deucalion". This work was funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the support UID/50014/2025 (<https://doi.org/10.54499/UID/50014/2025>) (Alícia Oliveira). This work was supported by FCT IP under Project 2024.09642.CPCA.A1 at Deucalion supercomputer, jointly funded by EuroHPC JU and Portugal and financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, through Ph.D. grant 2025.03394.BDANA (Beatriz Cepa).

¹<https://github.com/aliciasoliveira/OHANA>

OHANA uses FedProx [13] to mitigate heterogeneity across institutions.

Thus, *OHANA* is the first solution that simultaneously *i)* provides synthetic generation of multiple MRI artifact types, *ii)* performs artifact correction across four MRI contrasts, and *iii)* enables collaborative training via FL.

II. LITERATURE REVIEW

A. Correction

Motion artifacts. Küstner *et al.* [14] implemented a Multilayer Convolutional Neural Network (CNN) for detecting motion artifacts in MRI images and developed a retrospective correction method [15]. However, this method faced challenges in preserving anatomical features. Johnson and Drangova [16] employed a Conditional Generative Adversarial Network (CGAN) [17] to predict artifact-free images, though some irregularities persisted. Armanious *et al.* [18] used MedGAN [19] to achieve strong correction performance, trading off fine image detail in the process. Lee *et al.* [12] created MC2-Net to address motion artifacts in multi-contrast brain MRI, which exhibited limitations in correcting subtle artifacts.

Aliasing artifacts. Nahas *et al.* [20] proposed a method for aliasing artifact detection using a CNN [21] and a U-Net [22], achieving near 90% accuracy. Lee *et al.* [23] created a DL algorithm for aliasing artifact removal in MRI images, outperforming previous methods that required high computational complexity, meaning they demanded significant processing time and computational resources.

Magnetic Susceptibility artifacts. Duong *et al.* [24] introduced S-Net for correcting susceptibility artifacts using a CNN [21], while TS-Net [25] further enhanced it for 3D Echo Planar Imaging images.

Noise artifacts. Yan *et al.* [26] presented an unsupervised method for MRI denoising using content and noise encoders, but it required a computationally demanding training environment due to its high parameter count.

B. Synthetic Data Generation

Although artifact correction methods have advanced [12], [23], [25], [26], they are often trained and evaluated on synthetic data that does not fully reflect clinical complexity. Most publicly available synthetic data generation models focus on motion or noise artifacts, typically under a mild severity level [27]. Aliasing and magnetic susceptibility artifacts are overlooked, and when synthetically generated, they lack realism [28].

Thus, while previous approaches show limited generalization to real clinical data and limited effectiveness in correcting artifacts, *OHANA* addresses this by introducing a synthetic artifact generation model that mimics artifact types and a correction model that leverages this diversity during training.

III. METHODOLOGY

A. Data Preprocessing

Data preprocessing starts with brain images stored in the Neuroimaging Informatics Technology Initiative (NIfTI) format [29]. These volumes are first converted to 2D `png` format, and then resized to 256×256 pixels to maintain consistency with the model's input requirements.

B. Synthetic Artifact Generation Model

The synthetic artifact generation model addresses the limited availability of MRI datasets containing real corrupted images paired with GT references. By simulating realistic artifacts, it produces corrupted-clean image pairs that can be used to train and evaluate artifact-correction models [30]. The model incorporates four main artifact generation processes (see Fig. 1).

The generation of **motion** is based on a Fourier-domain approach [31], where random rotations and translations are applied in the frequency domain and transformed back to the spatial domain to mimic motion effects. These transformations are controlled by a *corruption scheme*, which determines the severity of the resulting ghosting and blurring distortions.

To simulate **aliasing**, a vertically shifted version of the original image is created, with the *shift amount* determining the extent of the displacement. This shifted copy is then combined with the original image using a *blend factor*, which controls how much the shifted version contributes to the final result [3].

The **magnetic susceptibility** artifact is modeled by deforming the source image with a spatially varying displacement field. The strength of this deformation is controlled by the *susceptibility factor*, which determines the extent of the local distortion. This reproduces the geometric warping caused by magnetic field inhomogeneities [32].

Noise is simulated by drawing samples from a zero-mean Gaussian distribution [33] and superimposing them onto the original image. The standard deviation of the distribution controls the *noise intensity*, with higher values producing more pronounced graininess [34].

C. Artifact Correction Model

The correction model is based on the MC2-Net architecture [12], which combines image alignment and a motion correction network. In our implementation, we do not use image alignment; instead, we focus exclusively on the correction network. To handle four contrasts, the first and last layers were changed from three-channel to four-channel input/output. Additionally, the number of convolutional filters was doubled throughout the correction backbone (from 16 to 32) to increase feature extraction capacity. The batch size was increased from 1 to 4, which is still relatively small. Group Normalization was introduced after convolutional layers to improve training stability with small batch sizes.

The DL-based model consists of an *Encoder*, a *Transformer*, and a *Decoder*. The *Encoder* uses 2D convolutions to create feature vectors for each contrast, which are then concatenated. These vectors are processed by the *Transformer*, and the *Decoder* reconstructs features using transposed 2D convolutional

layers and Rectified Linear Unit activation. A perceptual loss computed using a pretrained VGG-16 network [35] was used to optimize the artifact correction during training. The model uses the Adaptive Moment Estimation optimizer [36] with a learning rate of 0.001.

D. Architecture

OHANA implements a unified pipeline for MRI artifact correction within an FL environment. Fig. 1 illustrates the architecture in a two-client setup. After the MRI acquisition (❶), the NIfTI volumes are preprocessed (❷) to obtain artifact-free images. These images are then used to generate synthetic artifacts (❸), producing corrupted samples with different artifact levels. The clean and corrupted images are organized by contrast and used as input (❹) to train the artifact correction model (❺) in a collaborative setup (❻). Once the model is trained, it can be applied to new MRI scans to produce artifact-corrected images that are suitable for medical analysis (❼).

Since clinical data are typically dispersed across institutions/clients and cannot be shared, *OHANA* integrates the synthetic generation and correction models within an FL environment. This enables collaborative training without centralizing data, as each client trains locally and shares only model updates.

In each round, the FL server sends the model to the clients, who train on their datasets and return updates for aggregation into a global model. Repeating this process over multiple rounds progressively improves artifact correction across data sources.

E. Implementation

OHANA's synthetic artifact generation model was implemented using Python's Math [37], Nibabel [38], and Pillow [39] libraries. *OHANA*'s correction model is implemented in Python [40], resorting to TensorFlow [41], Keras [42], and NVFlare [43]. NVFlare allows *OHANA* to train models on decentralized data sources without transferring data, ensuring collaborative real-world applications [43]. Furthermore, we implement the FedProx [13] algorithm to optimize the FL process. This algorithm improves the convergence of the DL model, particularly in heterogeneous environments [11].

IV. EXPERIMENTAL SETUP

OHANA's correction model was trained using the Brain Tumor Segmentation Challenge 2021 (BraTS2021) dataset [44]–[46], which contains preprocessed MRI scans from 1248 subjects, each with four contrasts: T1w, T2w, FLAIR, and T1CE. The dataset was partitioned at the *subject level*, meaning that each client has access only to data from unique subjects, i.e., no subject is shared across clients. From these subject-specific partitions, we extracted 2D axial images, resulting in a total of 94,848 artifact-free images (23,712 per contrast). For each contrast, the images were split into 75% for training (17,784), 15% for validation (3,557), and 10% for test (2,371).

To ensure a fair comparison, all models were trained with **2,223** images per contrast. In the collaborative setup, each client was assigned the same number of images per contrast. Consequently, the total amount of training data increases with

the number of clients. In the 8-client setup, the entire training dataset was utilized (17,784 images per contrast).

For evaluating state-of-the-art implementations, each model was trained for 500 epochs. In the collaborative setup, the number of federated rounds was determined based on validation performance, with 100 local epochs per round fixed to ensure stable convergence. The tests were conducted on the GPU partition of the Deucalion supercomputer [47], with each node having 2x AMD EPYC 7742 64-core 2.25 GHz (512 GB RAM) and 4 NVIDIA Ampere A100 80 GB GPUs.

V. RESULTS

A. Quantitative Evaluation

Performance was evaluated using the **Structural Similarity Index Measure (SSIM)** [48], which quantifies the similarity between the corrected images and GT counterparts. We also analyzed the **Normalized Mutual Information (NMI)** [49] performance metric, since it captures statistical dependencies between corrected and GT images independently of intensity scaling, making it robust to the contrast variations inherent in multi-contrast MRI.

Comparison with state-of-the-art models. We compare *OHANA* against state-of-the-art models designed for artifact correction in brain MRI images. Specifically, we trained and evaluated CGAN [16], MC-Net [50], and MC2-Net [12]. Both CGAN and MC-Net were developed to operate on a single MRI contrast, and their implementations focus on T2w images. For consistency and reproducibility, we trained these models solely on T2w images. In contrast, MC2-Net was designed to support multi-contrast correction and was trained on T1w, T2w, and FLAIR images. Results are reported in Table I.

TABLE I: SSIM results of different correction models. The *Training Images* column represents *Number of images per contrast* $\times$ *Number of contrasts*.

Model	Training Images	SSIM			
		T1w	T1CE	T2w	FLAIR
CGAN	2223 $\times$ 1	-	-	0.8192	-
MC-Net	2223 $\times$ 1	-	-	0.8325	-
MC2-Net	2223 $\times$ 3	0.9508	-	0.8424	0.9434
MC2-Net4C	2223 $\times$ 4	0.9573	0.8888	0.8034	0.8222
<i>OHANA</i>	2223$\times$4	0.9785	0.9580	0.9418	0.9526

Single-contrast models (CGAN and MC-Net) achieved SSIM scores of up to 0.8325 for T2w correction. Regarding the multi-contrast model, MC2-Net achieved a high SSIM value for T1w (0.9508). We also implemented a four-contrast baseline, MC2-Net4C, that extends the MC2-Net architecture to handle T1w, T2w, FLAIR, and T1CE contrasts. MC2-Net4C follows the same training setup, hyperparameters, and data preprocessing as the MC2-Net, differing only in the number of contrasts provided as input. This baseline setup allows us to evaluate the impact of including additional contrasts. Despite this extension, MC2-Net4C showed lower performance across all contrasts compared to *OHANA* (see Figure 2), and lower performance

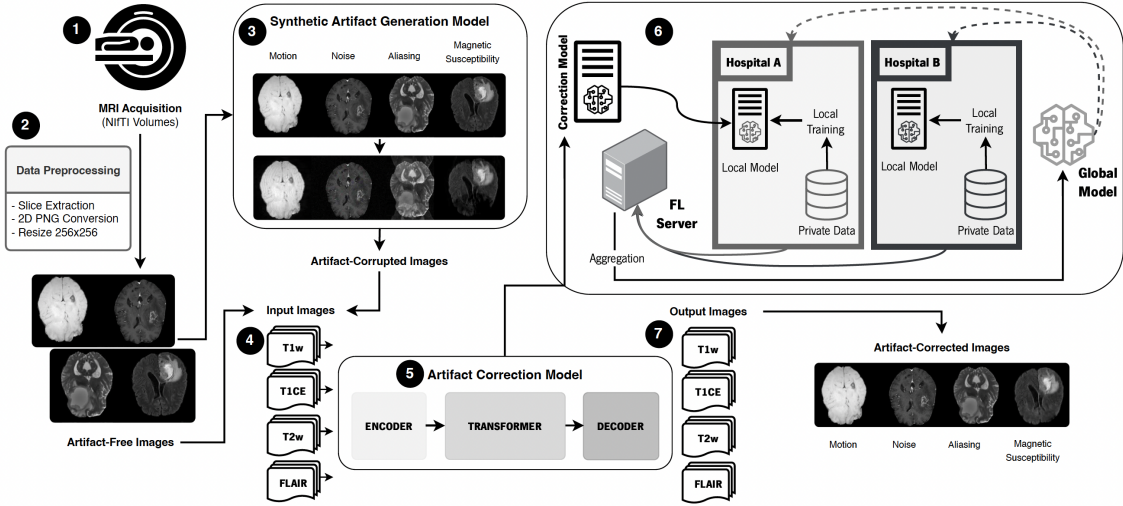


Fig. 1: *OHANA* pipeline with two-client setup, each representing a different hospital (A, B).

on two specific contrasts (T2w and FLAIR) compared to MC2-Net. Our model achieved the highest SSIM scores across all contrasts, with T1w reaching 0.9785.

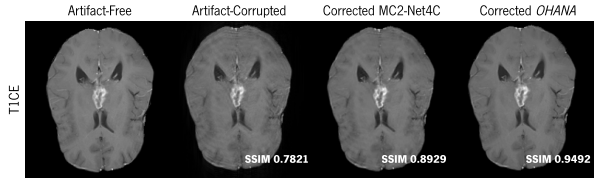


Fig. 2: Analysis of motion correction: MC2-Net4C vs *OHANA* in T1CE contrast.

Regarding NMI, the single-contrast models, CGAN and MC-Net, achieved lower values, ranging from 0.7012 to 0.7294 for T2w. MC2-Net achieved NMI scores of 0.7391-0.7668, while MC2-Net4C showed a slight drop, with NMI ranging from 0.7237 to 0.7523. In contrast, *OHANA* achieved the highest NMI across all contrasts, ranging from 0.7812 to 0.7932.

Collaborative Setup. As the number of clients increased (up to 8, each using 4 GPUs), the total number of training images also grew, reflecting the aggregated contributions of each client. Since each client holds a unique subset of images, with no overlapping patient data between clients, this setup promotes better generalization across diverse data distributions. The results are presented in Table II.

SSIM scores demonstrate a continuous improvement as the number of clients increases, as expected, since the training dataset is expanded and greater variability is introduced across individuals. The 8-client configuration achieved the highest SSIM scores, reaching 0.9858 for the T1w contrast. Figure 3 presents examples of image correction in an 8-client collaborative setup for various artifacts: motion in T1w, noise in T1CE, aliasing in T2w, and magnetic susceptibility in FLAIR. As shown, the corrupted images present substantially lower SSIM values compared to their corrected counterparts.

TABLE II: SSIM results of *OHANA* in collaborative setup for different numbers of clients. The *Training Images* column represents *Number of images per contrast × Number of contrasts × Number of clients*.

Clients	Rounds	Training Images	SSIM			
			T1w	T1CE	T2w	FLAIR
2	3	2223×4×2	0.9821	0.9611	0.9576	0.9616
4	5	2223×4×4	0.9843	0.9627	0.9585	0.9657
8	8	2223×4×8	0.9858	0.9697	0.9666	0.9673

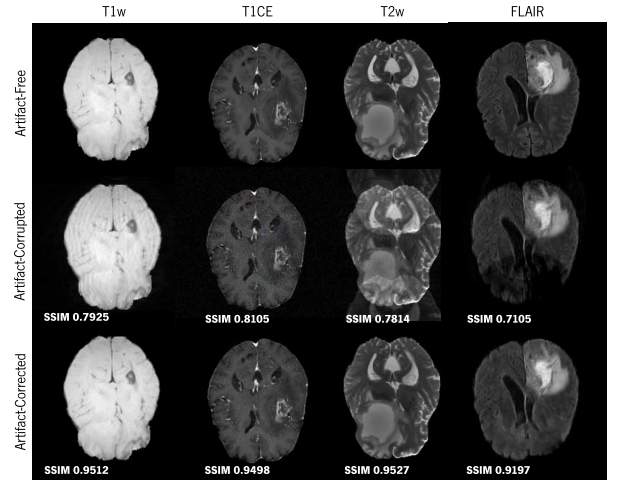


Fig. 3: Image correction of *OHANA* in an 8-client collaborative setup for: motion (T1w), noise (T1CE), aliasing (T2w) and magnetic susceptibility (FLAIR).

NMI scores followed a similar direction to SSIM, increasing with the number of clients. In the 2-client setup, NMI values ranged from 0.7921 to 0.7993, improving to 0.7976-0.8001 with 4 clients. The best results were obtained with 8 clients, achieving values between 0.8081 and 0.8216, indicating

enhanced capture of mutual information across all contrasts.

B. Qualitative Evaluation

To evaluate visual perception, a radiologist with experience interpreting brain MRI images assessed the synthetic, artifact-free, and artifact-corrected images. The evaluation was conducted for the synthetic artifact generation and correction models. Images were rated based on two criteria (see Table III): *i) the realism of the synthetic artifacts*, using an ordinal scale from 0 (“weak artifact generation”) to 3 (“strong artifact generation”); and *ii) the effectiveness of the corrections*, using an ordinal scale from 0 (“weak artifact correction”) to 3 (“strong artifact correction”).

TABLE III: Qualitative evaluation criteria and scoring scales used by the radiologist.

Score	Synthetic Artifact Realism	Correction Effectiveness
0	Weak artifact realism	Ineffective correction
1	Limited realism	Partial correction
2	Realistic artifacts	Mostly effective correction
3	Highly realistic artifacts	Highly effective correction

To compare MC2-Net4C and *OHANA* corrections, a radiologist assessed 120 T1CE images balanced across artifact types and severity, and representative examples are presented in Figure 2. Examination of the T1CE-corrected images reveals that the correction produced by *OHANA* was more effective, receiving the highest score (3) for correction realism due to its close resemblance to the artifact-free reference. MC2-Net4C was rated lower (2) due to visible residual motion traces.

To evaluate the performance of the synthetic artifacts generation and *OHANA* models, a radiologist assessed 720 images balanced across artifact types and severity, with examples illustrated in Figure 3. Among the synthetic artifacts generated, motion and noise were perceived as the most realistic, receiving the highest scores (3), reflecting their resemblance to clinical artifacts. These distortions reproduced the spatial inconsistencies and intensity variations observed in corrupted MR images. The aliasing and magnetic susceptibility artifacts were also considered visually plausible, although slightly less rigorous, earning moderate to high scores (2).

Analyzing the *OHANA* correction, motion and aliasing artifacts were perceived as the most successfully corrected, each receiving the highest correction realism score (3) due to the restoration of clinically relevant structural details and their close resemblance to the artifact-free references. Noise correction also achieved a high score (3), with minimal residual graininess. The correction of magnetic susceptibility artifacts, while still effective, was rated slightly lower (2) due to subtle distortions around tissue contrast regions.

VI. DISCUSSION

The centralized results demonstrate that *OHANA* generalizes well across multiple MRI contrasts and delivers superior artifact correction compared to state-of-the-art models, as shown in Figure 2. *OHANA* achieves consistent improvements across all

contrasts, with T2w improving from 0.8424 to 0.9418 (+11.8%) compared to MC2-Net and from 0.8034 to 0.9418 (+17.2%) compared to MC2-Net4C. In terms of NMI, *OHANA* attains the highest scores across all contrasts, increasing from 0.7668 to 0.7932 (+3.44%) relative to MC2-Net and from 0.7523 to 0.7932 (+5.44%) relative to MC2-Net4C.

In a comparison between centralized and collaborative versions of *OHANA*, the collaborative version consistently achieves higher performance. Notably, with eight clients, T2w shows the largest gains: SSIM increases from 0.9418 to 0.9666 (+2.63%) and NMI from 0.7993 to 0.8216 (+2.79%).

In the collaborative setup, as the number of clients scales from two to eight, the SSIM increases from 0.9576 to 0.9666 (+0.94%) for the T2w contrast. This collaborative version benefits from broader data distributions, enabling more effective model refinement and improved generalization. In Figure 3, the corrected images show a clear restoration of anatomical structures and a substantial increase in SSIM compared to the corrupted inputs.

Overall, the qualitative assessment confirms the quantitative results, indicating that *OHANA* produces visually consistent corrections across different artifact types and MRI contrasts.

VII. CONCLUSION

In summary, *OHANA* provides a collaborative framework for artifact generation and correction in brain MRI. First, the synthetic artifact generator produces corrupted images that closely resemble artifacts, enabling effective training of artifact-correction models. Second, the correction model consistently outperforms state-of-the-art methods across all MRI contrasts in single- and multi-client settings. Finally, *OHANA* maintains performance as the number of clients and training samples increases, highlighting its potential for real-world deployments.

ACKNOWLEDGMENT

Data used in this publication were obtained as part of the RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge project through Synapse ID (syn25829067).

REFERENCES

- [1] T. Budrys, V. Veikutis, S. Lukosevicius, R. Gleizniene, E. Monastyreckiene, and I. Kulakiene, “Artifacts in magnetic resonance imaging: how it can really affect diagnostic image quality and confuse clinical diagnosis?” *Journal of Vibroengineering*, vol. 20, no. 2, pp. 1202–1213, 2018.
- [2] A. Sreekumari, D. Shanbhag, D. Yeo, T. Foo, J. Pilitsis, J. Polzin, U. Patil, A. Coblenz, A. Kapadia, J. Khinda *et al.*, “A deep learning-based approach to reduce rescan and recall rates in clinical MRI examinations,” *American Journal of Neuroradiology*, vol. 40, no. 2, pp. 217–223, 2019.
- [3] E. Pusey, C. Yoon, M. L. Anselmo, and R. B. Lufkin, “Aliasing artifacts in MR imaging,” *Computerized medical imaging and graphics*, vol. 12, no. 4, pp. 219–224, 1988.
- [4] S. Balac and G. Caloz, “Magnetic susceptibility artifacts in magnetic resonance imaging: calculation of the magnetic field disturbances,” *IEEE transactions on magnetics*, vol. 32, no. 3, pp. 1645–1648, 1996.
- [5] P. J. Basser and S. Pajevic, “Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 44, no. 1, pp. 41–50, 2000.

- [6] J. B. Andre, B. W. Bresnahan, M. Mossa-Basha, M. N. Hoff, C. P. Smith, Y. Anzai, and W. A. Cohen, "Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations," *Journal of the American College of Radiology*, vol. 12, no. 7, pp. 689–695, 2015.
- [7] V. Spieker, H. Eichhorn, K. Hammernik, D. Rueckert, C. Preibisch, D. C. Karampinos, and J. A. Schnabel, "Deep Learning for Retrospective Motion Correction in MRI: A Comprehensive Review," *arXiv:2305.06739*, 2023.
- [8] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Medical image analysis*, vol. 43, pp. 66–84, 2018.
- [9] H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," *Pattern Recognition*, p. 110424, 2024.
- [10] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3347–3366, 2021.
- [11] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A hybrid approach to privacy-preserving federated learning," in *12th ACM workshop on artificial intelligence and security*, 2019, pp. 1–11.
- [12] J. Lee, B. Kim, and H. Park, "MC2-net: motion correction network for multi-contrast brain MRI," *Magnetic Resonance in Medicine*, vol. 86, no. 2, pp. 1077–1092, 2021.
- [13] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.
- [14] T. Küstner, A. Liebgott, L. Mauch, P. Martirosian, F. Bamberg, K. Nikolaou, B. Yang, F. Schick, and S. Gatidis, "Automated reference-free detection of motion artifacts in magnetic resonance images," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 31, pp. 243–256, 2018.
- [15] T. Küstner, K. Armanious, J. Yang, B. Yang, F. Schick, and S. Gatidis, "Retrospective correction of motion-affected MR images using deep learning frameworks," *Magnetic resonance in medicine*, vol. 82, no. 4, pp. 1527–1540, 2019.
- [16] P. M. Johnson and M. Drangova, "Conditional generative adversarial network for 3D rigid-body motion correction in MRI," *Magnetic resonance in medicine*, vol. 82, no. 3, pp. 901–910, 2019.
- [17] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv:1411.1784*, 2014.
- [18] K. Armanious, S. Gatidis, K. Nikolaou, B. Yang, and T. Küstner, "Retrospective correction of rigid and non-rigid mr motion artifacts using gans," in *ISBI*. IEEE, 2019, pp. 1550–1554.
- [19] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "MedGAN: Medical image translation using GANs," *Computerized medical imaging and graphics*, vol. 79, p. 101684, 2020.
- [20] H. Nahas, J. S. Au, T. Ishii, B. Y. Yiu, A. J. Chee, and C. Alfred, "A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging," *IEEE UFEC*, vol. 67, no. 12, pp. 2615–2628, 2020.
- [21] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv:1511.08458*, 2015.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.
- [23] D. Lee, J. Yoo, and J. C. Ye, "Deep artifact learning for compressed sensing and parallel MRI," *arXiv:1703.01120*, 2017.
- [24] S. T. Duong, S. L. Phung, A. Bouzerdoun, and M. M. Schira, "An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images," *Magnetic Resonance Imaging*, vol. 71, pp. 1–10, 2020.
- [25] S. T. Duong, S. L. Phung, A. Bouzerdoun, S. P. Ang, and M. M. Schira, "Correcting susceptibility artifacts of MRI sensors in brain scanning: a 3D anatomy-guided deep learning approach," *Sensors*, vol. 21, no. 7, p. 2314, 2021.
- [26] X. Yan, M. Xiao, W. Wang, Y. Li, and F. Zhang, "A self-guided deep learning technique for mri image noise reduction," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 01, pp. 109–117, 2024.
- [27] J. P. Cohen, T. Cao, J. D. Viviano, C.-W. Huang, M. Fralick, M. Ghassemi, M. Mamdani, R. Greiner, and Y. Bengio, "Problems in the deployment of machine-learned models in health care," *Cmaj*, vol. 193, no. 35, pp. E1391–E1394, 2021.
- [28] C. WangNicholas, C. NollDouglas, M. KimMichelle *et al.*, "Simulated MRI artifacts: Testing machine learning failure modes," *BME frontiers*, 2022.
- [29] S. Samuel, M. Moore, B. Patterson, H. Sheridan, and C. Sorensen, "Neuroimaging Data Primer: A Resource for Curating Digital Imaging and Communications in Medicine (DICOM) and Neuroimaging Informatics Technology Initiative (NIFTI) Files," 2020.
- [30] K. Somasundaram and P. Kalavathi, "Analysis of imaging artifacts in MR brain images," *Oriental J Comput Sci Technol*, vol. 5, no. 1, pp. 135–141, 2012.
- [31] B. A. Duffy, L. Zhao, F. Sepehrband, J. Min, D. J. Wang, Y. Shi, A. W. Toga, H. Kim, A. D. N. Initiative *et al.*, "Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions," *NeuroImage*, vol. 230, p. 117756, 2021.
- [32] L. Czervionke, D. Daniels, F. Wehrli, L. Mark, L. Hendrix, J. Strandt, A. Williams, and V. Haughton, "Magnetic susceptibility artifacts in gradient-recalled echo MR imaging," *American journal of neuroradiology*, vol. 9, no. 6, pp. 1149–1155, 1988.
- [33] C. E. Puentes, M. M. López, J. E. Pinzón, and J. M. Angulo, "The Gaussian distribution revisited," *Advances in applied probability*, vol. 28, no. 2, pp. 500–524, 1996.
- [34] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.
- [35] A. Bagaskara and M. Suryanegara, "Evaluation of VGG-16 and VGG-19 deep learning architecture for classifying dementia people," in *IC2IE*. IEEE, 2021, pp. 1–4.
- [36] S. Bock and M. Weiß, "A proof of local convergence for the Adam optimizer," in *IJCNN*. IEEE, 2019, pp. 1–8.
- [37] Python Software Foundation, "Documentation: The Python Standard Library - Numeric and Mathematical Modules." [Online]. Available: <https://docs.python.org/3/library/math.html>
- [38] NiBabel Developers, "NiBabel: Access a cacophony of neuro-imaging file formats." [Online]. Available: <https://nipy.org/nibabel/>
- [39] Pillow Developers, "Pillow (PIL Fork) Documentation." [Online]. Available: <https://pillow.readthedocs.io/en/stable/>
- [40] Python Software Foundation, "Python Language Reference, version 3.12." [Online]. Available: <https://www.python.org/>
- [41] B. Pang, E. Nijkamp, and Y. N. Wu, "Deep learning with tensorflow: A review," *Journal of Educational and Behavioral Statistics*, vol. 45, no. 2, pp. 227–248, 2020.
- [42] N. Ketkar, "Introduction to keras," in *Deep learning with python: a hands-on introduction*, 2017, pp. 97–111.
- [43] H. R. Roth, Y. Cheng, Y. Wen, I. Yang, Z. Xu, Y.-T. Hsieh, K. Kersten, A. Harouni, C. Zhao, K. Lu *et al.*, "Nvidia flare: Federated learning from simulation to real-world," *arXiv:2210.13291*, 2022.
- [44] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati *et al.*, "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv:2107.02314*, 2021.
- [45] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [46] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.
- [47] "Deucalion: a new EuroHPC supercomputer has been inaugurated." [Online]. Available: https://www.eurohpc-ju.europa.eu/deucalion-new-eurohpc-supercomputer-has-been-inaugurated-2023-09-06_en
- [48] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli, and L. Vanneschi, "Structural similarity index (SSIM) revisited: A data-driven approach," *Expert Systems with Applications*, vol. 189, p. 116087, 2022.
- [49] A. Mahmoudi and D. Jemielniak, "Proof of biased behavior of Normalized Mutual Information," *Scientific Reports*, vol. 14, no. 1, p. 9021, 2024.
- [50] L. Zhang, X. Wang, M. Rawson, R. Balan, E. H. Herskovits, E. Melhem, L. Chang, Z. Wang, and T. Ernst, "Motion correction in MRI using deep learning and a novel hybrid loss function," *arXiv:2210.14156*, 2022.