

Metadata-Aware Adaptation of a Generative Foundation Model for Conditional CMR Synthesis

1st Marc Rodríguez¹, 2nd Grzegorz Skorupko², 3rd Nay Aung^{3,4}, 4th Steffen E Petersen^{3,4}, 5th Karim Lekadir^{2,5}, 6th Polyxeni Gkontra²

Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain

Barcelona Artificial Intelligence in Medicine Lab (BCN-AIM), Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain

William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University London, Charterhouse Square, London, UK

Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, London, UK
Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, Barcelona, Spain
 polyxeni.gkontra@ub.edu

Abstract—Synthetic image generation is a promising strategy to address data scarcity and the underrepresentation of clinically important phenotypes in medical imaging, yet generating images that faithfully reflect meaningful patient characteristics remains challenging. In this work, we investigate metadata-conditioned cardiac magnetic resonance (CMR) synthesis using a pretrained latent diffusion model, encoding structured clinical metadata and slice position as textual prompts to guide CMR generation. To improve metadata adherence and address the imbalance of clinical attributes, we integrate three strategies: Metadata-Free Classifier-Free Guidance (CFG), Contrastive Batching, and Inverse-Frequency Sampling. The framework was fine-tuned and evaluated on 59,058 short-axis CMR from the UK Biobank using paired image similarity, distributional fidelity, and subgroup-level analyses. The combined approach achieved a Fréchet Inception Distance (FID) of 37.47, improving by 57.04% over the same model fine-tuned without these strategies and by 28.68% over a previous text-conditioned CMR diffusion baseline requiring cardiac geometry as additional input, while relying solely on patient metadata. This distributional gain, driven mainly by Metadata-Free CFG, came with a modest reduction in paired similarity, suggesting that the model prioritizes population-level realism over exact image reproduction. Subgroup analyses demonstrated improved alignment across demographic and acquisition-related metadata, with disease-specific conditioning being the most challenging task. These findings demonstrate the potential of generative foundation models for clinically meaningful CMR synthesis while highlighting the need for more effective metadata-aware conditioning strategies. Our code is available at <https://github.com/rodriguezmarc/conditional-cmr>.

I. INTRODUCTION

Cardiac Magnetic Resonance (CMR) is the reference modality for non-invasive cardiac assessment, playing a pivotal role in the diagnosis, prognosis, and monitoring of cardiovascular disease [1]. Building upon the rich information provided by CMR, recent advances in machine learning have demonstrated significant potential for enhancing automated CMR analysis and supporting clinical decision-making [2], [3], [4]. Nonetheless, privacy restrictions, the high cost of expert annotation, and the inherent scarcity of certain patient populations and disease phenotypes result in limited and imbalanced datasets [5],

[6], [7]. Thus, as medical image analysis has moved from isolated proof-of-concept models towards clinically relevant systems, the primary bottlenecks to developing robust and clinically applicable AI models have shifted from model architecture to data availability, diversity, and quality [8].

Synthetic image generation has emerged as a practical strategy for addressing these limitations, thus supporting data augmentation [9], privacy-preserving experimentation [10] and fairness-aware dataset balancing [11]. Recent diffusion models have substantially improved the realism and diversity of synthetic medical images [12]. However, visual plausibility alone is insufficient; clinically meaningful synthesis also requires generated images to faithfully reflect patient metadata. Recently, the MINIM framework [13] demonstrated that a single text-conditioned generative model can synthesize clinically useful images across multiple modalities and anatomical sites, improving downstream clinical tasks. However, MINIM was not evaluated on CMR, and metadata-conditioned synthesis, as demonstrated for other anatomies such as the brain [14], remains largely unexplored.

In this work, building on the MINIM text-conditioning paradigm, we adapt a general-domain latent diffusion model to the CMR domain by encoding structured patient metadata together with slice position as textual prompts. Unlike MINIM, which relies on large-scale multimodal medical pretraining and a self-improving reinforcement loop, we fine-tune from public Stable Diffusion weights, and focus on metadata-aware conditioning and sampling strategies to mitigate class imbalance, enabling the generation of clinically meaningful short-axis CMR images. The proposed framework improved distributional fidelity (FID) over both a fine-tuned baseline and a previous CMR diffusion baseline [15] requiring cardiac geometry, while relying solely on metadata. We further analyze how conditioning reliability varies across metadata attributes with differing levels of class imbalance.

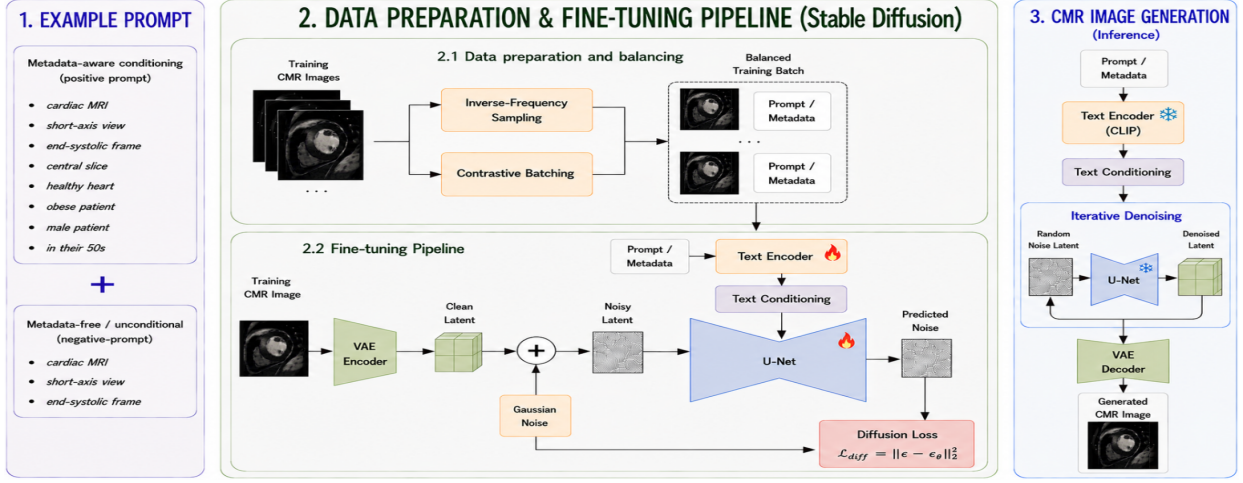


Fig. 1. Overview of the combined conditional CMR image generation pipeline. Target CMR characteristics are encoded as text prompts and fed to a pretrained latent diffusion model, which is then fine-tuned using the proposed techniques to improve metadata adherence and mitigate class imbalance during training.

II. METHODS

The work integrates three complementary metadata-aware strategies into a latent diffusion pipeline for conditional CMR synthesis. Metadata-Free Classifier-Free Guidance strengthens metadata conditioning during inference, Contrastive Batching increases metadata diversity within training batches, and Inverse-Frequency Sampling improves exposure to underrepresented metadata values during data loading. The overall workflow is illustrated in Fig. 1.

A. Metadata-Free Classifier-Free Guidance (MF-CFG)

A common inference-time technique for Stable Diffusion-based architectures to strengthen the conditioning effect is Classifier-Free Guidance (CFG), which combines conditional and unconditional predictions so that the model is optimized according to the difference of both results [16]. The proposed MF-CFG replaces the unconditional branch with a metadata-free version of the input prompt, obtained by removing metadata segments while preserving the acquisition context. The metadata-aware and a metadata-free prompt are then passed as the positive and negative prompts respectively, encouraging the guidance term to emphasize metadata-specific information.

B. Contrastive Batching (CB)

Batch training not only defines how many samples are processed at each optimizer step, but also sample diversity. In conditional generation, random sampling from an imbalanced dataset may produce batches dominated by frequent metadata values. Therefore, changing the order of sampled indices alters batch composition and increases exposure to less frequent conditioning values [17], [18]. CB uses this approach to construct batches that contain different values from a selected metadata-attribute whenever sufficient diversity is available, increasing within-batch metadata variation. Metadata values from the selected attribute are sampled without replacement, with one prompt randomly drawn from the corresponding metadata bucket for each selected value. If the batch size exceeds the number of available values, all values are sampled once before being reused cyclically.

C. Inverse-Frequency Sampling (IFS)

Because a sample may be rare for one metadata attribute but common for another in multi-class conditional generation, balancing is performed at metadata level rather than the sample level, as any weighting system would be ambiguous and inexact. Inverse-Frequency Sampling computes weights from the inverse metadata value frequencies to increase the representation of underrepresented values [15]. Let c be a given batch-selected metadata attribute, and $x \in X_c$ be one of its possible values. Furthermore, let N_x be the number of samples from the training split that contain the x value for the c attribute. Its inverse-frequency weight w_x is:

$$w_x = \min\left(\frac{\bar{N}_c}{N_x}, \tau\right) \quad (1)$$

where $\bar{N}_c$ is the mean bucket size across all values of c . The value w_x was capped at $\tau = 2.0$ to prevent extremely rare values from affecting results. During batch construction, metadata values for the selected attribute are sampled proportionally to these weights, prioritizing minority values. Specifically, for a given attribute c , the probability of sampling sample n_i with metadata value x_i^c is:

$$P(n_i | c) = P(x_i^c | c) \cdot P(n_i | x_i^c) = \frac{w_{x_i^c}}{\sum_{x \in X_c} w_x} \cdot \frac{1}{N_{x_i^c}} \quad (2)$$

Since the active metadata attribute varies across batches, balancing is applied across all conditioning metadata during training [15]. Combined with CB, value selection remains contrastive while being weighted by inverse-frequency probabilities, prioritizing minority values.

III. IMPLEMENTATION

A. Dataset and Prompt Construction

The dataset used in this work originates from the UK Biobank [19]. For each subject, the mid-ventricular short-axis slice together with two adjacent slices were extracted from the end-systolic frames of the cine CMR acquisition, along with the corresponding metadata. Samples were then split into

TABLE I

PAIRED AND DISTRIBUTIONAL FIDELITY METRICS TOGETHER WITH THE MAE AND MS-SSIM DIVERSITY RATIOS (DR). STANDARD DEVIATIONS FOR THE DR WERE OBTAINED BY FIRST-ORDER ERROR PROPAGATION. $\downarrow/\uparrow$ INDICATE THAT LOWER/HIGHER VALUES ARE BETTER; FOR DR, VALUES CLOSER TO 1 ($\rightarrow$ 1) INDICATE AGREEMENT WITH THE REFERENCE DISTRIBUTION. METADATA-AWARE STRATEGIES ARE ADDED INCREMENTALLY ON THE FINE-TUNED STABLE DIFFUSION (SD) BASELINE. ABBREVIATIONS: MF-CFG, METADATA-FREE CLASSIFIER-FREE GUIDANCE; IFS, INVERSE-FREQUENCY SAMPLING; CB, CONTRASTIVE BATCHING.

Metric	Fine-tuned SD	MF-CFG	MF-CFG+IFS	MF-CFG+CB	MF-CFG+IFS+CB	Skorupko et al. [15]
MAE $\downarrow$	0.257 ± 0.040	0.254 ± 0.045	0.261 ± 0.041	0.263 ± 0.042	0.260 ± 0.043	0.253 ± 0.040
MS-SSIM $\uparrow$	0.192 ± 0.075	0.177 ± 0.079	0.164 ± 0.072	0.163 ± 0.073	0.171 ± 0.076	0.266 ± 0.075
FID $\downarrow$	87.22	44.84	39.12	37.08	37.47	52.54
FRD $\downarrow$	4.76	4.64	4.27	4.44	4.64	5.65
MAE DR $\rightarrow$ 1	0.931 ± 0.189	1.010 ± 0.127	1.006 ± 0.178	1.050 ± 0.185	1.026 ± 0.191	1.248 ± 0.178
MS-SSIM DR $\rightarrow$ 1	0.940 ± 0.126	1.081 ± 0.197	1.0323 ± 0.120	1.056 ± 0.122	1.063 ± 0.126	1.068 ± 0.111

training and testing sets at the patient level using a 90 : 10 ratio, resulting in 53,298 training images and 5,760 testing.

CMR generation is formulated as a text-conditioned task that uses prompts to describe image characteristics and patient metadata [13]. Each sample is transformed into a standardized record consisting of the processed image path and a textual prompt beginning with a fixed acquisition context (imaging modality, view, and cardiac phase) and followed by patient metadata: slice position (mid-ventricular ± 1), pathology (healthy, heart failure, myocardial infarction, ischemic heart disease, atrial fibrillation), BMI group (underweight: < 18.5 , normal: $18.5 - 24.9$, overweight: $25.0 - 29.9$, obese: ≥ 30.0 kg/m²), sex (male, female), and age group (40s, 50s, 60s, 70s, 80s). An example prompt is: *Cardiac MRI, short-axis view, end-systolic frame, mid-ventricular + 1, healthy heart, obese patient, male patient, patient in their 50s.*

The dataset exhibits substantial class imbalance across several metadata attributes. This is most pronounced for pathology, where healthy subjects dominate, and for BMI and age, where underweight and extreme age groups are underrepresented. In contrast, sex is relatively balanced, while slice position is inherently balanced. These imbalances motivate the metadata-aware strategies proposed in this work to improve the representation of minority metadata values.

B. Model and Training Workflow

In this work, we built on the released MINIM text-conditioning approach and codebase [13]. As the pretrained MINIM weights are not publicly available, we initialized the model from publicly available Stable Diffusion weights (CompVis/stable-diffusion-v1-4, Hugging Face) [20] and fine-tuned it combining the proposed metadata-aware strategies. Each prompt is injected as conditioning and paired with its reference image. The implementation follows a Stable Diffusion-like latent diffusion architecture: a frozen VAE maps sample images to latent space, the U-Net is fine-tuned to denoise CMR latents, and a text encoder provides prompt embeddings to the denoising process [20]. During inference, synthetic images are generated for the testing split, paired with the corresponding reference images, and evaluated using quantitative metrics.

C. Evaluation Protocol

The evaluation considers three complementary aspects: paired-image fidelity, distributional fidelity, and subgroup per-

formance. Paired fidelity compares each generated image with its reference image using Mean Absolute Error (MAE), which quantifies pixel-level differences, and Multi-Scale Structural Similarity (MS-SSIM) [21], used to investigate anatomical similarities. Distributional fidelity compares generated and real testing sets using the Fréchet Inception Distance (FID) [22] and the Fréchet Radiomics Distance (FRD) [23], while subgroup analyses evaluate model performance across the different metadata categories. In addition, distributional matrices compare generated and real image distributions across metadata values. The diagonal entries compare generated and real samples with the same metadata value, while off-diagonal values reveal potential overlap or poor separation between metadata categories. Finally, diversity ratios are computed for the pair fidelity metrics, indicating variation in comparison to the reference testing split.

IV. RESULTS AND DISCUSSION

A. Overall Performance

Table I summarizes the global results. The model incorporating the combined metadata-aware strategies achieved an FID of 37.47, improving by 57.04% over the same latent diffusion model fine-tuned without the proposed strategies and by 28.68% compared with the text-conditioned CMR diffusion model of Skorupko et al. [15], which requires cardiac geometry as additional conditioning input. Ablation results indicate that MF-CFG drives most of the distributional improvement (FID: 87.22 to 44.84), with the sampling strategies providing further gains. MF-CFG + CB achieved the lowest FID (37.08), while MFCFG + IFS achieved the lowest FRD (4.27). As the margin between the combined model and the best single-strategy variants is small, we adopt the combined configuration as the representative model for the subgroup and diversity analyses that follow, since it balances distributional and paired fidelity rather than optimizing either metric in isolation.

In the combined model, paired fidelity metrics, with MAE equal to 0.260 and MS-SSIM equal to 0.171, indicate that it preserves the short-axis CMR appearance without behaving as an exact reconstruction model. As it conditions on metadata alone, it is not intended to reproduce a specific reference image; the higher paired similarity of Skorupko et al. [15], which additionally uses cardiac geometry, reflects that extra spatial information. Distributional fidelity (FID, FRD) is therefore better suited to this metadata-conditioned setting. Moreover, the near-unity results of both diversity ratios indicate that the

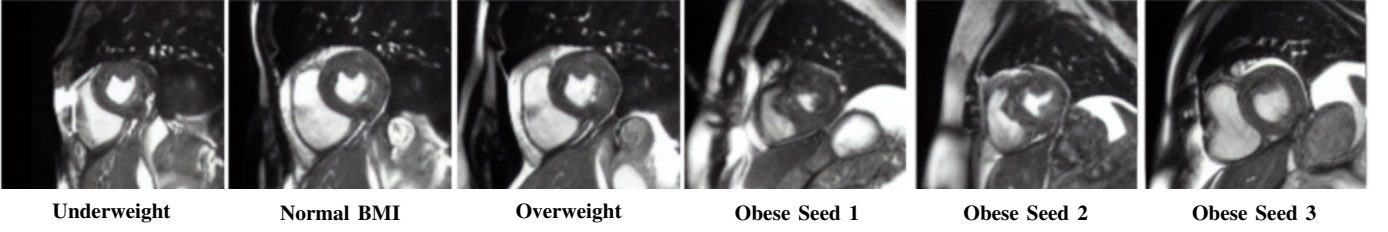


Fig. 2. Representative generated CMR images by BMI category showing the conditioning effect on generation. From underweight to obese, the generated images qualitative show progressively greater adipose tissue, consistent with increasing BMI. The three rightmost images were generated from the same obese conditioning prompt using different random seeds, demonstrating the diversity achievable under identical conditioning.

generated set preserves realistic variation without collapsing into a narrow set of similar images. Representative generated samples for different BMI classes are shown in Fig. 2.

B. Subgroup Performance Analysis

The distributional FID matrices in Fig. 3 summarizes subgroup-level performance across metadata attributes for the combined model, with lower diagonal values indicating better agreement between the generated and real distributions. Diagonal mean subgroup FID ranged from 44.90 for sex and 49.48 for slice position to 80.55 for BMI and 92.45 for age, while pathology exhibited the highest value (118.87), indicating that even same-subgroup generation is least accurate for disease categories. For every attribute, the diagonal mean FID was lower than the corresponding off-diagonal value (61.83 for sex, 57.25 for slice position, 129.53 for BMI, 115.17 for age, and 138.68 for pathology), confirming that generated subgroups align more closely with their matching real subgroups than with different ones. These results indicate that conditioning is most reliable for well-balanced metadata attribute. However, pathology is also the most underrepresented attribute, so part of the elevated FID may reflect the known sensitivity of FID to small sizes [24] rather than conditioning difficulty alone. Overall, subgroup performance closely followed the underlying data distribution, with the poorest results consistently corresponding to the most underrepresented metadata values, indicating that the applied strategies substantially improve generation but do not fully overcome class imbalance.

V. CONCLUSION

Building on the MINIM text-conditioning paradigm and public Stable Diffusion weights, we proposed a metadata-

aware framework for conditional CMR synthesis that conditions image generation on structured patient metadata and slice position without requiring cardiac geometry as input. The framework improved distributional fidelity over both a fine-tuned Stable Diffusion baseline and a previously published CMR text-conditioned diffusion baseline while generating realistic and diverse short-axis CMR images. Subgroup analyses further showed that conditioning is more reliable for balanced metadata attributes, such as sex and slice position, than for highly imbalanced clinical variables, particularly pathology. The ablation results showed that MF-CFG was the main driver of distributional improvement, while the balancing strategies provided complementary refinements. These results demonstrate the potential of latent diffusion models for metadata-aware CMR synthesis while highlighting that, although the proposed strategies improve distributional fidelity, robust and fully controllable metadata conditioning under class imbalance remains an open challenge. Future work will assess the effectiveness of the generated images in downstream CMR analysis tasks, particularly for improving the performance and fairness of models trained on underrepresented patient subgroups.

ACKNOWLEDGMENT

This work is part of the project TrustAI-ES (PID2023-146751OA-I00), funded by MICIU/AEI/10.13039/501100011033. This work received funding from the European Union’s Horizon Research and Innovation program under Grant Agreement No. 101080430 (AI4HF project). This work is also supported by the European Union’s Horizon Europe research and innovation program under Grant Agreement No. 101057849 (DataTools4Heart project). This work was conducted using the UK Biobank resource under access application 2964.

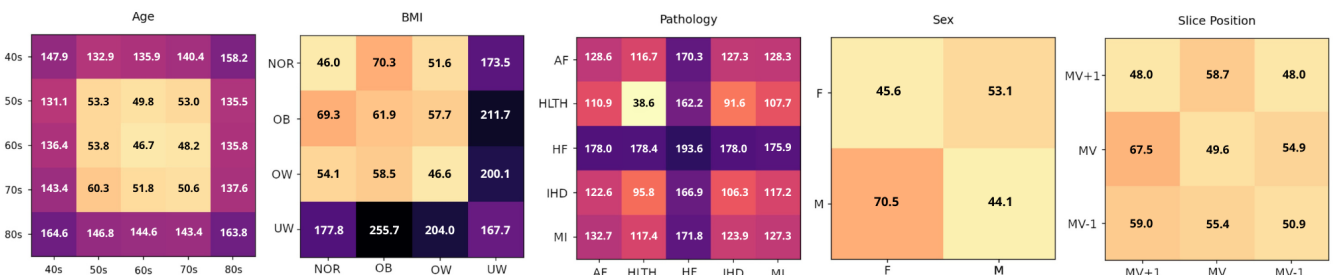


Fig. 3. Subgroup distributional fidelity (FID) matrices for each metadata attribute. Columns correspond to generated subgroups and rows to reference subgroups. Abbreviations: NOR, normal; OB, obese; OW, overweight; UW, underweight; AF, atrial fibrillation; HLTH, healthy; HF, heart failure; IHD, ischemic heart disease; MI, myocardial infarction; F, female; M, male; MV, mid-ventricular.

REFERENCES

- [1] D. J. Pennell, "Cardiovascular magnetic resonance," *Circulation*, vol. 121, no. 5, pp. 692–705, Feb. 2010.
- [2] Y. Fu, W. Bai, W. Yi, C. Manisty, A. N. Bhuva, T. A. Treibel, J. C. Moon, M. J. Clarkson, R. H. Davies, and Y. Hu, "Development and validation of a versatile foundation model for cine cardiac magnetic resonance image analysis," *Communications Medicine*, 2026.
- [3] Y.-R. Wang, K. Yang, Y. Wen, P. Wang, Y. Hu, Y. Lai, Y. Wang, K. Zhao, S. Tang, A. Zhang *et al.*, "Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging," *Nature Medicine*, vol. 30, no. 5, pp. 1471–1480, 2024.
- [4] R. Shad, C. Zakka, D. Kaur, M. Mathur, R. Fong, J. Cho, R. W. Filice, J. Mongan, K. Kallianos, N. Khandwala *et al.*, "A generalizable deep learning system for cardiac mri," *Nature Biomedical Engineering*, pp. 1–16, 2026.
- [5] E. Puyol-Antón, B. Ruijsink, J. Mariscal Harana, S. K. Piechnik, S. Neubauer, S. E. Petersen, R. Razavi, P. Chowiecnyk, and A. P. King, "Fairness in cardiac magnetic resonance imaging: Assessing sex and racial bias in deep learning-based segmentation," *Frontiers in Cardiovascular Medicine*, vol. 9, 2022.
- [6] A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proceedings of the National Academy of Sciences*, vol. 117, no. 23, pp. 12 592–12 594, 2020.
- [7] E. Petersen, A. Feragen, M. L. da Costa Zemsch, A. Henriksen, O. E. Wiese Christensen, and M. Ganz, "Feature robustness and sex differences in medical imaging: A case study in mri-based alzheimer's disease detection," in *Medical Image Computing and Computer Assisted Intervention*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds. Cham: Springer Nature Switzerland, 2022, pp. 88–98.
- [8] A. Zhang *et al.*, "Shifting machine learning for healthcare from development to deployment and from models to data," *Nature Biomedical Engineering*, vol. 6, no. 12, pp. 1330–1345, 2022. [Online]. Available: <https://doi.org/10.1038/s41551-022-00898-y>
- [9] L. R. Koetzier, J. Wu, D. Mastrodicasa, A. Lutz, M. Chung, W. A. Koszek, J. Pratap, A. S. Chaudhari, P. Rajpurkar, M. P. Lungren, and M. J. Willemink, "Generating synthetic data for medical imaging," *Radiology*, vol. 312, no. 3, p. e232471, Sep. 2024.
- [10] A. DuMont Schütte, J. Hetzel, S. Gatidis, T. Hepp, B. Dietz, S. Bauer, and P. Schwab, "Overcoming barriers to data sharing with medical image generation: a comprehensive evaluation," *NPJ Digit. Med.*, vol. 4, no. 1, p. 141, Sep. 2021.
- [11] I. Ktena, O. Wiles, I. Albuquerque, S.-A. Rebuffi, R. Tanno, A. G. Roy, S. Azizi, D. Belgrave, P. Kohli, T. Cemgil, A. Karthikesalingam, and S. Goyal, "Generative models improve fairness of medical classifiers under distribution shifts," *Nat. Med.*, vol. 30, no. 4, pp. 1166–1173, Apr. 2024.
- [12] J. Kaleta, D. Dall'Alba, S. Plotka, and P. Korzeniowski, "Minimal data requirement for realistic endoscopic image generation with stable diffusion," *International journal of computer assisted radiology and surgery*, vol. 19, no. 3, pp. 531–539, 2024.
- [13] J. Wang, K. Wang, Y. Yu, Y. Lu, W. Xiao, Z. Sun, F. Liu, Z. Zou, Y. Gao, L. Yang *et al.*, "Self-improving generative foundation model for synthetic medical image generation and clinical applications," *Nature Medicine*, vol. 31, no. 2, pp. 609–617, 2025.
- [14] W. H. L. Pinaya, P.-D. Tudosiu, J. Dafflon, P. F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin, and M. J. Cardoso, "Brain imaging generation with latent diffusion models," in *Deep Generative Models*, ser. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022, pp. 117–126.
- [15] G. Skorupko, R. Osuala, Z. Szafranowska, K. Kushibar, V. N. Dang, N. Aung, S. E. Petersen, K. Lekadir, and P. Gkontra, "Fairness-aware data augmentation for cardiac mri using text-conditioned diffusion models," in *MICCAI Workshop on Fairness of AI in Medical Imaging*. Springer, 2025, pp. 63–73.
- [16] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.
- [17] P. Zhao and T. Zhang, "Accelerating minibatch stochastic gradient descent using stratified sampling," *arXiv preprint arXiv:1405.3080*, 2014.
- [18] Z. Yang, T. Huang, M. Ding, Y. Dong, R. Ying, Y. Cen, Y. Geng, and J. Tang, "Batchsampler: Sampling mini-batches for contrastive learning in vision, language, and graphs," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 3057–3069.
- [19] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins, "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS Med.*, vol. 12, no. 3, p. e1001779, Mar. 2015.
- [20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 10 684–10 695.
- [21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The thirty-seventh asilomar conference on signals, systems & computers*, 2003, vol. 2. Ieee, 2003, pp. 1398–1402.
- [22] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6629–6640.
- [23] N. Konz, R. Osuala, P. Verma, Y. Chen, H. Gu, H. Dong, Y. Chen, A. Marshall, L. Garrucho, K. Kushibar, D. M. Lang, G. S. Kim, L. J. Grimm, J. M. Lewin, J. S. Duncan, J. A. Schnabel, O. Diaz, K. Lekadir, and M. A. Mazurowski, "Fr chet radiomic distance (frd): A versatile metric for comparing medical imaging datasets," *Medical Image Analysis*, vol. 110, p. 103943, May 2026. [Online]. Available: <http://dx.doi.org/10.1016/j.media.2026.103943>
- [24] M. J. Chong and D. A. Forsyth, "Effectively unbiased FID and inception score and where to find them," *CoRR*, vol. abs/1911.07023, 2019. [Online]. Available: <http://arxiv.org/abs/1911.07023>