

Synthetic Cardiac MRI Image Generation using Deep Generative Models

Ishan Kumarasinghe

*Department of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20211@eng.pdn.ac.lk*

Dasuni Kawya

*Department of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20197@eng.pdn.ac.lk*

Madhura Edirisooriya

*Department of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20093@eng.pdn.ac.lk*

Vajira Thambawita

*Department of Holistic Systems
SimulaMet
Oslo, Norway
vajira@simula.no*

Isuru Nawinne

*Department of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
isurunawinne@eng.pdn.ac.lk*



Abstract—Synthetic cardiac MRI (CMRI) generation has emerged as a promising strategy to overcome the scarcity of annotated medical imaging data. Recent advances in GANs, VAEs, diffusion probabilistic models, and flow-matching techniques aim to generate anatomically accurate images while addressing challenges such as limited labeled datasets, vendor variability, and risks of privacy leakage through model memorization. Mask-conditioned generation improves structural fidelity by guiding synthesis with segmentation maps, while diffusion and flow-matching models offer strong boundary preservation and efficient deterministic transformations. Cross-domain generalization is further supported through vendor-style conditioning and pre-processing steps like intensity normalization. To ensure privacy, studies increasingly incorporate membership inference attacks, nearest-neighbor analyses, and differential privacy mechanisms. Utility evaluations commonly measure downstream segmentation performance, with evidence showing that anatomically constrained synthetic data can enhance accuracy and robustness across multi-vendor settings. This review aims to compare existing CMRI generation approaches through the lenses of fidelity, utility, and privacy, highlighting current limitations and the need for integrated, evaluation-driven frameworks for reliable clinical workflows.

Index Terms—Cardiac MRI, Synthetic Data, Generative Models, Fidelity, Privacy

I. INTRODUCTION

Synthetic cardiac MRI (CMR) generation has gained significant attention as a promising solution to the longstanding problem of limited annotated medical imaging data [1], [2]. Manual segmentation of cardiac structures is time-consuming, requires expert clinicians, and remains difficult to scale, particularly for multi-frame cine MRI, where numerous slices and temporal phases must be labeled [3]. Alongside these constraints, strict privacy regulations, the rarity of certain cardiac pathologies, and substantial variation in acquisition protocols across scanner vendors further restrict the availability of high-quality, diverse training datasets [1]. These challenges collectively motivate the development of generative models capable of producing realistic, clinically

meaningful synthetic CMR data under limited labeled data settings [1][3].

Generative modelling techniques such as generative adversarial networks (GANs), variational autoencoders (VAEs), diffusion probabilistic models, and more recent flow-matching frameworks have all demonstrated varying degrees of success in capturing cardiac anatomy [4]. While GANs offer visually plausible results, they often suffer from instability and artifacts [5]; VAEs provide smooth latent representations but may lack fine structural detail [2]. Diffusion models, through their iterative noise-removal process, have recently emerged as state-of-the-art for high-fidelity medical image synthesis [6]. Flow-matching models present a newer alternative that models continuous transformations [7] between distributions, but their performance in mask-conditioned cardiac MRI generation remains under-explored.

Despite these advances, much of the existing work still relies on unconditional synthesis or simplistic augmentation strategies [5]. Such approaches fail to ensure anatomical precision, particularly regarding subtle structural variations [8] in the left ventricle (LV), right ventricle (RV), and myocardium (MYO). Conditioning on segmentation masks presents a more promising direction by explicitly guiding generative models to align with cardiac structure [3][9][10]. However, only a limited number of studies have rigorously evaluated conditional generative models across the three critical dimensions required for clinical usefulness: fidelity, downstream task utility [11], and privacy.

Privacy concerns remain especially relevant in medical imaging. Because cardiac MRIs inherently reflect individual anatomical characteristics, synthetic images may unintentionally reveal patient identity if generative models memorize training examples [12]. Privacy vulnerabilities

such as membership inference attacks and nearest-neighbour similarity analysis reveal the need for explicit privacy assessments[12]. Nevertheless, privacy evaluation is often absent or only superficially addressed in prior work [4]. Achieving a balance between generating realistic, clinically useful images and ensuring strong privacy protection therefore represents a core challenge[11].

Multi-vendor heterogeneity adds another layer of complexity[2]. Differences in scanner manufacturers, acquisition sequences, and reconstruction algorithms introduce substantial domain shifts. Models trained on a single vendor frequently degrade when tested on images from unseen scanners[13]. Synthetic augmentation has the potential to mitigate this by enriching training datasets with diverse, vendor-variant image styles, but only if the generative models are carefully validated through downstream tasks such as segmentation[9].

To navigate these challenges, a structured assessment of conditional generative models, with emphasis on diffusion and flow-matching techniques, is required. This review therefore examines existing CMRI synthesis approaches through the lenses of fidelity, utility, and privacy to identify methodological strengths, expose current gaps, and inform the development of dependable, privacy-conscious synthetic data frameworks for downstream clinical applications.

A. Background and Clinical Importance of Cardiac MRI

1) *Cardiac MRI in Clinical Diagnostics:* Cardiac magnetic resonance imaging (MRI) is a key modality for diagnosing and monitoring cardiovascular diseases because it provides high-resolution, high-contrast visualization of cardiac structure and function [14][8]. Its ability to delineate the left and right ventricles, myocardium, and atrial chambers enables accurate assessment of chamber volumes, myocardial mass, wall motion, and ejection fraction, critical biomarkers in conditions such as ischemic heart disease, cardiomyopathies, valvular disorders, and congenital abnormalities[15]. Clinical workflows benefit from a range of imaging sequences, each capturing different aspects of myocardial health. Cine MRI provides dynamic functional assessment across the cardiac cycle[16]; late gadolinium enhancement (LGE) reveals scar or fibrosis; and T1/T2 mapping quantifies tissue composition changes relevant to inflammation or edema[15]. Together, these sequences form comprehensive diagnostic protocols. However, extracting quantitative measures typically requires manual contouring by experts, which is time-consuming and limits scalability[13][14].

2) *Advantages Over Other Imaging Modalities:* Cardiac MRI (CMR) stands out among cardiac imaging techniques because it provides high-resolution, high-contrast images without ionizing radiation[17][4], unlike CT, making it ideal for repeated acquisitions and safe long-term follow-up. Compared to echocardiography, CMR offers consistent image quality that

does not depend on operator skill or patient anatomy, ensuring reliable datasets for machine learning[18]. A major advantage for generative modelling is CMR's ability to produce multiple standardized sequences, such as cine, T1w, T2w, and LGE[4], each capturing unique structural or functional information. This variety enables models to learn richer anatomical and tissue characteristics, supporting high-fidelity synthetic image generation[3]. MRI also provides flexible multi-planar acquisitions[8] with excellent anatomical coverage, enabling consistent short-axis, long-axis, and 4D cine sequences[16]. For generative modelling, this flexibility means that models can learn robust anatomical structures across varying slice orientations, important for 2D and 3D synthetic reconstruction tasks[9]. The clarity of myocardial borders in MRI improves segmentation ground truth quality, supporting more accurate supervised learning and ensuring synthetic images encode precise anatomical boundaries. Although multi-vendor variability introduces challenges, this heterogeneity benefits synthetic data research by allowing models to learn cross-domain variations and generalize better to real-world clinical environments[9]. These combined strengths make CMR the most suitable modality for mask-conditioned synthetic generation, fidelity-utility evaluation, and privacy-aware modelling in cardiac imaging.

B. Challenges in Cardiac MRI Data Availability and Usage

1) *Limited Annotated Data:* The scarcity of annotated cardiac MRI datasets remains one of the central limitations for both clinical AI and synthetic data research[4]. High-quality segmentation labels require time-consuming manual effort from expert clinicians[9], often across dozens of slices and multiple cardiac phases[13]. This workload makes it difficult to build large, diverse labeled datasets, especially for 4D cine MRI[15]. As a result, models are frequently trained on small samples that may not capture the full variability of cardiac anatomy, scanner vendors, or rare pathologies.

Privacy regulations such as GDPR further restrict multi-institutional data sharing[19], creating fragmented data silos that cannot be easily pooled[1]. This exacerbates overfitting risks [4] and limits the representativeness of training, validation, and test sets. Overall, limited annotated data is not only a practical constraint but a methodological challenge influencing model design, evaluation, and privacy. Addressing this requires hybrid pipelines that maximize unlabeled data usage, incorporate synthetic augmentation, and include robust privacy assessments to prevent patient-specific leakage.

2) *Vendor and Scanner Variability:* Multi-center studies reveal significant performance degradation when models trained on one dataset are deployed on data from different scanner vendors, imaging protocols, or acquisition sites[9][13]. Even standardized cine sequences can differ subtly across vendors, causing models to misinterpret vendor-specific characteristics as anatomical changes[13]. The M&Ms (Multi-Center, Multi-Vendor, Multi-Disease) challenge demonstrated that models

achieving 93% Dice score on single-center data experience performance drops to 88% when tested on multi-center, multi-vendor datasets due to variations in image intensity, contrast, spatial resolution, and noise characteristics [13].

C. Motivation for Synthetic Data Generation in Medical Imaging

Synthetic data generation offers a promising solution to limited annotated datasets and strict privacy regulations by:

- 1) augmenting limited real datasets with diverse, anatomically plausible examples [11].
- 2) simulating rare pathologies underrepresented in training data [19] [20].
- 3) enabling cross-vendor domain adaptation without requiring additional acquisitions [9].
- 4) potentially preserving privacy by sharing trained generative models rather than patient data [19].

Traditional data augmentation techniques, including geometric transformations (rotation, scaling, flipping) and intensity modifications (brightness, contrast adjustment) provide only modest improvements and cannot introduce novel anatomical variations [8].

II. PREPROCESSING STRATEGIES IN CARDIAC MRI

Pre-processing profoundly influences synthetic image fidelity, cross-vendor generalization, and training stability. Pre-processing decisions represent the first critical control point for downstream generative model performance [21].

A. Rescaling and Intensity Normalization

Rescaling and intensity normalization are essential pre-processing steps in cardiac MRI pipelines, ensuring consistent input quality across diverse scanners, vendors, and acquisition protocols [9]. Because cardiac MRIs vary in voxel size, field of view, and slice count, spatial rescaling aligns all scans to a fixed resolution and geometry [14], enabling generative and segmentation models to operate on uniform inputs. Care is taken to preserve anatomical proportions typically by maintaining aspect ratio and using padding to avoid distortions that would affect mask-conditioned synthesis or downstream clinical measurements such as ventricular volume. Intensity normalization addresses the lack of an absolute MRI intensity scale. Common approaches include:

- 1) **Min-max normalization** - Rescaling intensity to [0,1] or [-1,1] range used extensively in cardiac synthesis studies [16].
- 2) **Z-score normalization** - Computing mean and standard deviation across foreground pixels, normalizing to zero-mean, unit-variance distributions [3] [8].
- 3) **Percentile clipping** - Clipping intensities to 2nd-98th percentiles before normalization reduces outlier artifacts from motion or flow artifacts [9].
- 4) **N4ITK bias field correction** - Correcting spatially varying signal inhomogeneities persistent across CMR sequences [7] [8].

B. Cropping and ROI Extraction

Region-of-interest extraction reduces computational burden and improves model focus:

- 1) **Bounding box cropping**: Automatically detecting cardiac regions using CNNs, then cropping to 128×128 or 256×256 pixels centered on the heart [9].
- 2) **Consistent field-of-view**: Standardizing FOV across subjects by resampling [9].

C. Slice Selection and Temporal Frames

CMR acquisitions contain multiple spatial and temporal dimensions [15] requiring careful selection:

- 1) **End-diastolic (ED) and end-systolic (ES) frames** - Most clinical segmentation studies use ED and ES frames representing maximum and minimum ventricular volume states [13].
- 2) **Short-axis (SAX) stack** - Standard acquisition spanning apex to base with 8–10 mm slice thickness and variable inter-slice distance (typically 6–10 mm) [14].
- 3) **Central slice selection** - Studies commonly select 2–4 central slices per volume to focus on regions with clear left ventricle (LV), right ventricle (RV), and myocardium (MYO) visibility [22].

Frame selection directly impacts synthesizing model training: including both ED and ES improves robustness to temporal variability, and including unlabeled apical/basal slices struggles with synthesizing plausible appearance of anatomy [9].

III. SYNTHETIC MEDICAL IMAGE GENERATION

Figure 1 provides an overview of the major generative model families used for CMRI synthesis, highlighting both unconditional and conditional architectures that are discussed in this section.

A. Unconditional Generative Models

Unconditional generative models learn to map random noise vectors directly to realistic images without any auxiliary conditioning variables, modeling the overall data distribution rather than specific anatomical or pathological subsets [1] [3]. In cardiac imaging, unconditional models primarily serve to expand datasets and investigate general appearance variability [1]. Unconditional generation provides baseline capability but lacks anatomical control. GANs learn mappings from noise vectors to images through adversarial training [3]. VAEs learn lower-dimensional latent representations enabling diverse generation [23]. Diffusion models reverse learned noise processes through iterative denoising, achieving superior fidelity and diversity [6].

- 1) **Unconditional GAN Models**: Unconditional generative adversarial networks (GANs) learn to map random noise vectors to realistic images by training a generator and discriminator in an adversarial framework, without requiring auxiliary conditioning information such as class labels or segmentation masks [24] [12] [23]. Early unconditional GAN architectures, including DCGAN and Progressive GAN

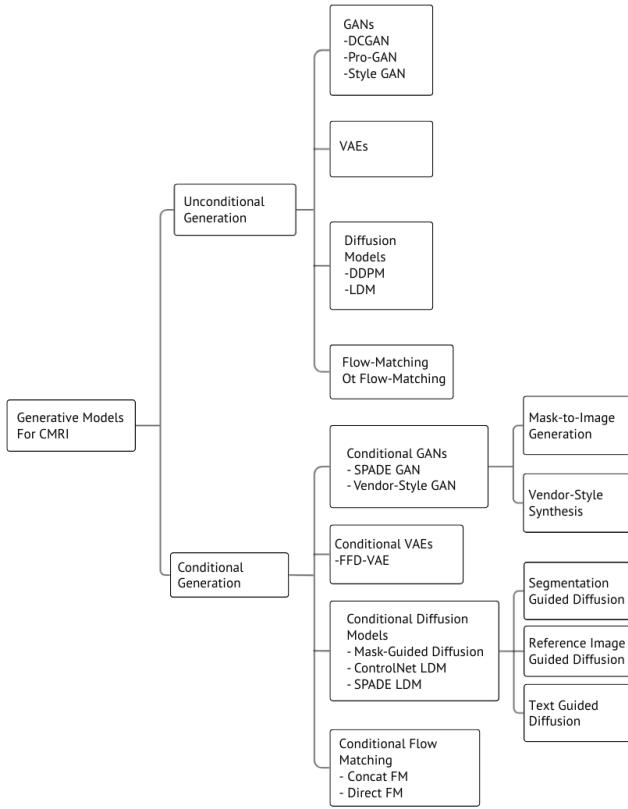


Fig. 1. Taxonomy of generative models used in CMRI synthesis

(ProGAN), demonstrated the ability to synthesize plausible medical images by progressively increasing resolution during training, thereby improving stability and output quality [1][25][3]. In cardiac imaging, unconditional GANs have been applied to generate synthetic MRI frames for data augmentation purposes, ProGAN-generated cardiac MR images can augment training sets for congenital heart disease applications. Unconditional GAN frameworks were used for brain MRI synthesis, highlighting their potential for expanding limited datasets [1][24].

However, unconditional GANs exhibit notable limitations: they lack explicit control over anatomical structures or pathology, often suffer from mode collapse, and may produce images with unrealistic artifacts [23][19].

2) Unconditional Diffusion Models: Unconditional diffusion models have emerged as a robust alternative to GAN and VAE-based synthesis for medical imaging, offering improved stability and fidelity when modeling complex anatomical structures. These models operate by reversing a predefined forward Markov process that gradually corrupts data with Gaussian noise, enabling high-quality image generation through iterative denoising steps [8]. Unlike GANs, which depend on adversarial optimization and are

susceptible to mode collapse, diffusion models benefit from a well-posed likelihood-based objective that facilitates broader distribution coverage and more reliable convergence [8]. Their capacity to reconstruct images through progressive refinement has fueled widespread adoption across medical imaging tasks, with recent studies demonstrating high-fidelity cardiac textures and ventricular structures synthesized using Denoising Diffusion Probabilistic Models (DDPMs) implemented within frameworks such as MONAI [23][26].

Despite their strengths in realism and diversity, unconditional diffusion models face practical challenges. They require substantial computational resources due to the multi-step sampling process, though recent work exploring random latent embeddings has shown promise in accelerating inference without compromising quality [27]. Furthermore, privacy concerns persist, as nearest-neighbor leakage can occur when unconstrained diffusion models inadvertently reproduce features of specific training samples, prompting the need for formal safeguards such as differential privacy mechanisms [1][22].

3) Unconditional Flow Matching Models: Flow matching is an emerging class of generative models that learns continuous normalizing flows by regressing a velocity field to transform a simple prior distribution (e.g., Gaussian noise) into the target data distribution [28]. Unlike diffusion models, which require hundreds of denoising steps, flow matching models directly learn a transport map, enabling significantly faster sampling while maintaining competitive image quality. Flow matching bridges the gap between speed and quality in medical image synthesis, achieving inference times substantially lower than traditional diffusion approaches without sacrificing fidelity [28][7][29].

In unconditional settings, flow matching models generate images by sampling from noise and following the learned velocity field trajectory to produce realistic outputs. Applying flow matching for MRI-CT and CBCT-CT image synthesis, showing that unconditional flow-based generation can produce anatomically coherent medical images [29]. Flow matching achieves intermediate performance between GANs and diffusion models, offering faster inference than diffusion while providing better diversity than GANs.

However, unconditional flow matching lacks explicit anatomical control, limiting its applicability to pathology-specific or mask-conditioned generation tasks. This motivates the exploration of conditional flow matching frameworks for cardiac MRI synthesis [28][7].

B. Conditional Generative Models

Conditional generative models extend unconditional models by incorporating additional information such as labels, segmentation masks, or text to control synthesis,

enabling anatomy- or phenotype-specific image generation [20][3][23]. Recent advancements have moved beyond simple class labels, employing high-dimensional spatial masks to enforce anatomical consistency [16][30]. In cardiac imaging, conditional models improve fidelity, diversity, and clinical relevance of synthetic MR data for augmentation and downstream tasks [20][3][9]. Furthermore, the architectural landscape has evolved from Conditional Generative Adversarial Networks (cGANs), which require careful balancing of generator-discriminator dynamics to Conditional Diffusion Models, which offer stable training objectives and fine-grained control over the synthesis process through mechanisms like classifier guidance and cross-attention [31]. The following sections review these models, categorized by their conditioning strategies and architectures.

1) Deep Convolutional Generative Adversarial Network (DCGAN): A DCGAN architecture was implemented as a core component of proposed models for synthesizing medical images. Although the DCGAN concept is general, the network utilizes Convolutional Neural Networks (CNNs) with specific filters to detect and reproduce essential patterns from the data distribution. The generated images, intended to mimic sagittal plane brain MRIs, were successfully used to assess the efficacy of downstream classification tasks [21].

2) Conditional Generative Adversarial Network (Vendor-Specific Synthesis): Conditional GANs were implemented to facilitate domain adaptation in cardiac MR segmentation by synthesizing data in a vendor-specific style. The model involved training two separate generators for different scanner vendors (Vendor A and Vendor B). This allowed the system to synthesize images with the appearance of one vendor based on anatomical labels acquired from another, effectively utilizing conditional synthesis to address domain shift [9].

3) GauGAN (SPADE-based Architecture): A generative approach based on the GauGAN architecture was implemented to synthesize short-axis cardiac MR stacks. This model utilized a high-level conditioning mechanism to control the generated images, enabling the synthesis of data with a wide and dense distribution of cardiac functional metrics, such as ejection fraction (EF). This synthetic data was subsequently used for pre-training a regression network for cardiac function quantification [14].

4) Reconstruction Conditional GAN (RecCGAN): The RecCGAN framework was implemented for cardiac MR under-sampled image reconstruction. This specific Conditional GAN architecture integrates a hybrid spatial and frequency loss function to calculate the difference between the target and reconstructed images. The CGAN objective is supplemented with an L1 distance to enhance accuracy and reduce image blurring during the reconstruction process [17].

5) Wasserstein Generative Adversarial Network (Cardiac Biasing): A conditional generative model featuring a conditional biasing mechanism was implemented for cardiac imaging synthesis. This model employed residual and attentional blocks along with a sinusoidal embedding to transform scalar covariates (e.g., age or BMI) into vector conditions. This structure enabled the fine-grained control necessary to induce plausible morphological variations, such as synthesizing images representing cardiac aging or modifying the apparent BMI of subjects [32].

6) Rényi Differential Privacy-Conditional GAN (RDP-CGAN): The RDP-CGAN framework was proposed and implemented to ensure Rényi Differential Privacy (RDP) during synthetic data generation, especially for tabular/sequential medical records. This model uses a Convolutional GAN to capture temporal and correlated information, and crucially, enforces RDP exclusively on the discriminator training while utilizing a pre-trained convolutional autoencoder for handling mixed-type data features [22].

7) Feature Factorization and Distillation VAE (FFD-VAE): The Feature Factorization and Distillation (FFD-VAE) architecture was implemented to enhance cross-domain generalization. This VAE employs two separate encoders to disentangle the latent space, learning independent representations for segmentation knowledge (z_{seg}) and domain-related knowledge (z_{dom}). This factorization strategy ensures that the essential anatomical structure is preserved while allowing the extracted domain features to be used as a condition for subsequent diffusion models [5].

8) Segmentation-Guided Diffusion Model: A Segmentation-Guided Diffusion Model was implemented to condition image generation on a multi-class anatomical mask (m), aiming to sample from the distribution $p(x|m)$. This model demonstrated high faithfulness, accurately reproducing anatomical realism required for downstream tasks, like training segmentation networks, with performance nearly matching models trained purely on real data [30].

9) Source Domain-Trained Diffusion Model (SD-DM): The Source Domain-Trained Diffusion Model (SD-DM) was implemented to mitigate domain shift in cardiac MR imaging by performing a reference-based adaptation. The sampling process is iteratively refined, conditioned on a reference image, to ensure the synthetic data maintains structural consistency with anatomical features while adopting the visual characteristics of the source training domain [33].

10) Controllable Mask Diffusion Framework: A multi-label conditioned diffusion framework was implemented, involving several complex steps to generate pseudo-samples. The process includes training a pseudo-label mask diffusion model, followed by a cardiac MR image implicit diffusion

model. This second stage integrates the SPADE module into its decoding process to enforce pixel-level alignment with the pseudo-labels, coupled with a label screening strategy to filter noisy inputs [16].

11) Latent Diffusion Model (Pathology and Modality Conditioning): A Latent Diffusion Model (LDM) was implemented for conditioned brain MRI generation. The model generates images in a compressed latent space and is specifically conditioned on auxiliary inputs specifying both pathology (e.g., Glioblastoma, Sclerosis) and the target acquisition modality (e.g., T1w, T2w). This allows the model to generate synthetic images representing configurations not explicitly present in the training data, demonstrating extrapolation capability [19].

12) ControlNet (Multi-Conditional LDM): ControlNet, built upon a Stable Diffusion 2.1 base model, was implemented and fine-tuned on UK Biobank cardiac MR data. This model allows for multi-modal conditioning using both text prompts and segmentation masks, leveraging the LDM's cross-attention mechanisms for flexible control over synthesis, and demonstrating that complex conditional diffusion can be achieved with modest computational resources [32].

13) MOTFM (Medical Optimal Transport Flow Matching): The Medical Optimal Transport Flow Matching (MOTFM) framework was implemented for efficient, high-quality medical image synthesis, particularly for echocardiographic images. This model uses a UNet backbone enhanced with attention layers to estimate the optimal velocity field. MOTFM demonstrated superior efficiency, achieving performance comparable to 50-step DDPM in significantly fewer steps (e.g., 10 steps) for mask-conditioned synthesis [28].

14) Conditional Flow Matching (Concat. FM and Direct FM): Conditional Flow Matching (CFM) was implemented with two distinct conditioning strategies for T1w-to-T2w MRI translation. Concat. FM conditions the flow by concatenating the T1w slice with the intermediate flow sample. Direct FM utilizes the T1w slice directly as the source sample (z_0). Both strategies train the model to match the constant conditional velocity field (z_1-z_0) to learn the flow dynamics [7].

IV. FIDELITY EVALUATION IN GENERATIVE MODELS

The correct evaluation of synthetic medical images is crucial, as the quality of these images can impact human life and health [18]. Since medical imaging modalities, such as MRI, CT, and Echo images, are typically noisier, blurrier, and possess less defined edges compared to common images (e.g. cats or faces), specialized and accurate evaluation metrics are necessary [18]. While metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Structural Similarity Index Measure (SSIM) are traditionally used, they may not

always accurately reflect the true visual quality of the medical image [23].

Fidelity generally refers to the realism of the synthetic image, quantified by how closely its feature distribution and structural characteristics align with those of real data [11][34]. The Fréchet Inception Distance (FID) is widely adopted as the de facto standard metric for overall sample quality in generative models because it captures aspects of both fidelity and diversity [31]. A lower FID score indicates that the synthetic images' feature distribution is closer to that of the real images [35]. In addition to FID, quantitative measures used frequently in cardiac MRI synthesis include Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and its multi-scale variant, MS-SSIM [8][36][15]. For cardiac images specifically, assessing fidelity often goes beyond simple pixel comparisons to include measuring clinically relevant anatomical parameters, such as the Left Ventricle (LV)/Myocardium (Myo) Ratio, Right Ventricle (RV)/LV Ratio, and Myocardium Thickness, ensuring the synthetic data maintains structural integrity [8].

However, reliance on distribution-based metrics such as the FID alone is insufficient for evaluating synthetic medical images. A low FID score may still occur in cases of mode collapse or overfitting, where the model generates near-identical replicas of training samples rather than truly diverse images [18]. For this reason, application-driven utility evaluation is widely regarded as the most reliable assessment strategy [23][9][5].

In terms of generative model comparison, recent work indicates that Diffusion Models often achieve superior fidelity, resulting in the lowest FID scores during unconditional image generation compared to GANs and VAEs [8][37]. Conversely, for image-to-image translation tasks (e.g., T1w-to-T2w MRI translation), GAN-based methods like Pix2Pix have sometimes demonstrated superior structural fidelity and computational efficiency over Diffusion Models and Flow Matching techniques [7]. This complexity highlights the ongoing challenge of navigating the inevitable trade-off that exists among image fidelity, sample variety, and data privacy [11].

A. Summary of Fidelity Metrics

Assessing the quality of artificial medical images is tricky because there are many ways to measure success. **Table I** summarizes these key metrics, which are essential to determine whether the generated images are realistic and useful for heart segmentation.

V. PRIVACY EVALUATION IN GENERATIVE MODELS

A. Balancing Privacy and Image Fidelity

Evaluating privacy in generative cardiac MRI systems requires navigating a core tension: protecting patient identity while preserving clinically useful image fidelity. Differential Privacy (DP) offers formal guarantees by injecting noise to

TABLE I
SUMMARY OF FIDELITY METRICS USED IN CMRI SYNTHETIC IMAGE EVALUATION

Metric	Definition	Level (Pixel / High)	Application (Generation / Segmentation)	Frequency in Literature
Fréchet Inception Distance (FID)	Measures distribution similarity between real and generated image features; captures fidelity and diversity.	High (Feature / Distribution)	Image Generation (Standard metric for GANs / Diffusion models)	Very High
Structural Similarity Index Measure (SSIM)	Measures structural similarity and perceptual quality between two images.	Pixel-Level (Structural)	Image Generation; MRI Reconstruction	High
Multi-Scale SSIM (MS-SSIM)	Multi-scale variant of SSIM assessing structure at various resolutions.	Pixel / High (Multi-scale)	Image Generation (Structural fidelity)	High
Peak Signal-to-Noise Ratio (PSNR)	Measures pixel-level fidelity based on error magnitude.	Pixel-Level (Error)	MRI Reconstruction; Image Generation	High
Dice Similarity Coefficient (DSC)	Measures overlap between segmentation prediction and ground truth.	High (Spatial Overlap)	Segmentation Models; Utility of Synthetic Data	Very High
Hausdorff Distance (HD)	Measures boundary distance between two shapes/segmentations.	Pixel / High (Boundary Distance)	Segmentation Models (Boundary accuracy)	Medium
Kernel Inception Distance (KID)	Evaluates generative performance using Inception features, unbiased alternative to FID.	High (Feature / Distribution)	Image Generation	Low
Inception Score (IS)	Measures quality and diversity of generated images using classification confidence.	High (Feature / Distribution)	Image Generation	Low
Fréchet ResNet Distance (FRD)	Alternative distance metric using ResNet-based features for generative evaluation.	High (Feature / Distribution)	Image Generation	Low
Reconstruction Fréchet Inception Distance (rFID)	Measures perceptual quality for image reconstruction relative to ground truth.	High (Feature / Distribution)	Image Reconstruction / Inpainting	Low

limit the influence of any individual patient [34]. However, for high-dimensional medical images, strict privacy budgets often degrade anatomical detail, producing blurry or structurally unreliable outputs [22]. Because DP tends to conflict with the fine-grained realism required in clinical imaging, recent work increasingly favors attack-based, post-training privacy evaluation, where models are stress-tested like adversaries to identify potential leakage [12].

B. Membership Inference and Frequency-Calibrated Attacks

A common attack-based technique is the Membership Inference Attack (MIA), which attempts to determine whether a specific real image was part of the training set by analyzing reconstruction error patterns [12]. Privacy quantification here is twofold: attack-based metrics measure adversarial success rates, while overfitting-based metrics estimate information leakage by analyzing the generalization gap between training and validation performance [38]. However, standard MIAs often perform poorly on medical images due to “inherent image difficulty” - cardiac MRI contains polarized frequency components that make high-frequency details equally challenging for both members and non-members [39].

To address this, Frequency-Calibrated Reconstruction Error (FCRE) has been introduced. FCRE isolates mid-frequency components removing noisy high frequencies and uninformative low frequencies to better differentiate true memorization from general reconstruction limitations [39].

C. Structural Similarity Audits and Mode Collapse Checks

Beyond membership inference, privacy auditing must also examine whether synthetic samples structurally replicate real patients. Pairwise similarity attacks compute the minimum pixel or latent-space distance between synthetic outputs and real images to detect nearest-neighbor leakage [11]. This risk

is amplified in small datasets or rare-pathology settings [16].

Complementary distribution-based attacks check whether many synthetic samples cluster around a single real patient, indicating mode collapse–driven memorization. Such multi-level auditing is especially important in style-conditioned models, where scanner-specific artifacts may act as unintended identifiers [33], and in deterministic generative frameworks like flow matching, where fixed noise-to-sample mappings require additional privacy stress tests [7].

D. Practical Safeguards and Regulatory Alignment

Regulatory discussions increasingly emphasize that anonymization cannot rely solely on transformations; instead, layered safeguards are required. To standardize these evaluations, recent frameworks propose optimized metrics suites that weigh and compose diverse privacy and utility metrics (e.g., fidelity, specific utility) to rigorously assess trade-offs [38]. Recommended measures include:

- 1) **Technical transformations** - resolution / intensity harmonization and removal of non-essential anatomy [33],
- 2) **Generative training constraints** - DP-enforced optimization [22], diversity maximization to reduce mode collapse-linked memorization [11],
- 3) **Post-generation audits** - MIAs with anatomically matched conditions [12], nearest-neighbor searches across entire protected archives [11],
- 4) **Policy controls** - usage contracts forbidding re-identification attempts and outlining permissible applications.

VI. UTILITY EVALUATION IN GENERATIVE MODELS

A. Segmentation Performance and Fidelity

The primary measure of synthetic data utility in cardiac MRI is its ability to improve segmentation performance,

TABLE II
SUMMARY OF CARDIAC MRI (CMRI) DATASETS USED IN GENERATIVE MODELLING AND SEGMENTATION STUDIES

Dataset Name	Vendors / Sites	Number of Samples	Label Availability	Usefulness
M&Ms Challenge (Multi-Center, Multi-Vendor, Multi-Disease)	6 hospitals across Spain, Canada, Germany; Vendors: Siemens, Philips, GE, Canon	350 images [9]; 150 annotated, 25 unannotated	Open-access; LV/RV/MYO labels at ED/ES annotated by clinicians	Most diverse CMR dataset; benchmark for generalization across vendors [13]; used in image synthesis and multi-tissue segmentation studies [9].
UK Biobank (UKB)	Single center; Siemens MAGNETOM Aera 1.5T	Over 5,000 subjects; some studies use 43,352 participants or 25,480 SAX cine sequences	Restricted access; LV/RV annotations at ED/ES	Used for large-scale cardiac function studies, phenotype-guided generation [40], fairness correction, debiasing, and conditional CMR synthesis [41].
ACDC Challenge	Single clinical center	150 subjects	Public; LV/RV/MYO segmentation at ED/ES; includes NOR, DCM, HCM pathologies	Established segmentation benchmark [13]; widely used for domain shift [33] and deep learning evaluation.
OCMR Dataset	Siemens Prisma 3T, Avanto 1.5T, Sola 1.5T	53 fully sampled + 212 under-sampled (total 1383 k-space entries)	Open-access; full multi-coil k-space data (healthy subjects)	Used for DL reconstruction, acceleration studies [17], and unconditional diffusion model training [8].
M&Ms-2 Challenge	Multi-vendor; testing on Siemens, Philips, GE, Canon	50 studies per vendor; extra 50 from unseen vendor (Canon)	Segmentation labels for LV/RV/MYO	Rigorous benchmark for generalizing to unseen moderate/severe cardiac pathologies [20].
MS-CMRSeg Challenge	Multi-sequence challenge (LGE focus)	45 subjects	Manual LV/RV/MYO labels by three observers; LGE-CMR images	Used for domain transfer, LGE-focused synthesis and segmentation methods [5].
MM-WHS Challenge (Multi-Modality Whole Heart Segmentation)	Multi-center, multi-modality (MRI + CT)	120 images (60 CT, 60 MRI)	Labels for 7 structures (LV, RV, LA, RA, Myo, Ao, PA)	Benchmark for whole-heart 3D segmentation [5].
Sunnybrook Dataset	N/A	45 cases	Public; LV segmentation labels [4]	Used for segmentation algorithm evaluation and comparison.

especially in accurately delineating cardiac structures [9][16]. A standard evaluation strategy involves training segmentation networks solely on synthetic datasets generated using GANs, VAEs, or diffusion models and comparing their accuracy with models trained on real images [1][11]. Methods incorporating segmentation-informed conditional GANs have reported measurable gains, including up to a 4% increase in Dice score and a 40% reduction in Hausdorff distance when synthetic samples are added to real training data across multi-site datasets [9].

Mask-conditioned diffusion frameworks further enhance anatomical precision, demonstrating superior myocardium boundary reconstruction relative to unconditioned generators, which often fail to retain endocardial and epicardial detail [30]. Notably, diffusion-generated cardiac MRIs achieve left ventricle segmentation Dice scores that closely match those obtained from real data and consistently outperform unconditional GAN-based synthesis [8][16]. Results from Progressive GAN studies indicate negligible differences (below 1%) in segmentation metrics between models trained on real versus synthetic sets, reinforcing the clinical promise of these approaches [1].

B. Cross-Vendor Robustness and Dataset Diversity

Clinical deployment requires models that generalize across MRI scanners, where hardware differences introduce domain shifts affecting contrast and noise characteristics [20]. Synthetic data helps mitigate this variability through vendor-style conditioned generation, enabling models to learn scanner-specific visual properties without altering

underlying anatomical structures [9][33]. However, such style adaptation must be carefully interpreted to ensure that subtle pathology-related features remain detectable [20].

Synthetic generation also increases dataset diversity by exposing models to a broader spectrum of anatomical and stylistic variations [7][9], which is particularly valuable for downstream phenotype prediction. For instance, cardiac phenotype-guided generative models have improved performance on metrics such as Left Ventricular End-Diastolic Volume (LVEDV) and Left Ventricular Ejection Fraction (LVEF), demonstrating that synthetic data can strengthen clinical outcome prediction tasks [40].

C. Architectural Trade-offs and Privacy

Utility varies significantly by generative architecture. GANs can introduce spurious high-frequency artifacts that segmentation networks mistakenly interpret as boundaries [16]. Diffusion models, by contrast, gradually refine images through iterative denoising, yielding smoother textures and more stable volumetric estimations across heterogeneous test datasets [23].

Clinical evaluations further highlight the realism of diffusion-generated cardiac MRI, with radiologists achieving only 60% accuracy when asked to distinguish synthetic from real images close to chance-level discrimination [8].

Regarding privacy, integrating differential privacy mechanisms may degrade anatomical quality through noise-induced blurring [22]. Therefore, complementary assessments such as

membership inference attacks remain necessary to ensure that synthetic samples do not encode identifiable patient-specific features, maintaining low detection accuracy to minimize leakage risk [12][34].

D. Augmentation Efficiency and Practical Constraints

The volume and quality of synthetic augmentation must be carefully balanced. Excessive synthetic samples can reduce performance if the generator collapses toward narrow stylistic modes, reducing dataset diversity [11]. This issue is especially relevant for rare pathology augmentation, where small but targeted synthetic additions often provide more substantial improvements than large volumes of normal-structure images [9][20].

Computational cost is another constraint: diffusion-based generation remains significantly slower than GANs and VAEs. For example, synthesizing a 128×128 cardiac MRI slice using diffusion takes ~ 2.31 seconds, compared to 0.12 seconds for GANs and 0.08 seconds for VAEs [8]. Such disparities influence feasibility for large-scale or real-time clinical pipelines.

VII. COMPARATIVE ANALYSIS

A. Summary of Datasets

To contextualize the diversity and limitations of existing cardiac MRI resources, we summarize the major datasets used in generative modelling and segmentation studies, highlighting sample size, vendor variability, and available annotations **Table III**.

B. Summary of CMRI Generative Models

We further provide a consolidated overview of the generative models applied to CMRI synthesis, comparing their conditioning strategies, backbone architectures, and reported performance across fidelity, utility, and privacy dimensions **Table III**.

C. Summary of CMRI Evaluations

Figure 2 summarizes the proportion of studies that report fidelity, utility, and privacy evaluations, showing that while fidelity is widely assessed, utility is less consistently reported and privacy evaluation remains largely neglected in current CMRI generative modeling literature.

VIII. RESEARCH GAPS AND FUTURE DIRECTIONS

A. What the Literature Agrees On

Generative model progression: GANs established early mask-to-image translation capabilities [3]. Diffusion models now offer state-of-the-art fidelity, stability, and mode coverage [31], [36], [23], whereas flow-matching shows promise but overfits on small medical datasets [7]. SPADE-based conditioning achieves strong fidelity–efficiency trade-offs for cardiac MRI [16], [3].

Value of mask conditioning: Studies consistently show that multi-structure anatomical masks outperform cavity-only

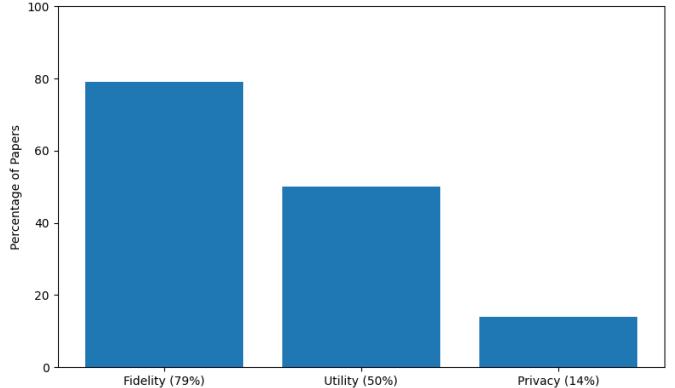


Fig. 2. Summary of Fidelity, Utility, and Privacy Reporting

labels, improving realism, shape diversity, and structural accuracy [3]. High-quality semantic guidance often yields greater gains than simply increasing dataset size [3].

Utility of pathology synthesis: Although underexplored, incorporating synthetic pathological variations (e.g., altered geometries indicating hypertrophy or dilation) significantly improves segmentation robustness for rare diseases [20].

Lack of privacy evaluation: Despite motivations rooted in data confidentiality, most CMRI synthesis studies omit quantitative privacy assessment, including membership inference or nearest-neighbor leakage tests.

Insufficient cross-dataset evaluation: Most works remain limited to within-dataset validation, and the few cross-dataset evaluations (e.g., ACDC \rightarrow M&Ms) report substantial performance drops [9].

B. Identified Research Gaps

1) *Limited pathology-specific mask-conditioned generation:* Existing models primarily synthesize healthy or generic cardiac shapes. Missing capabilities include controlled synthesis of hypertrophy, DCM, ischemic defects, and congenital abnormalities [20], [3].

Future direction: extend latent-space pathology modelling, enable continuous severity control, and validate realism through expert clinical review.

2) *Underexplored diffusion and flow-matching performance on small CMRI datasets:* Flow-matching models overfit under small-sample regimes [7], and diffusion inference remains slow (40s per sample).

Future direction: systematically evaluate these architectures under limited-data conditions, develop accelerated cardiac-specific variants, and compare all models under identical fidelity–utility–privacy protocols.

3) *Absence of systematic privacy evaluation:* Cardiac MRI generative studies rarely include quantitative privacy assessment [3], [36].

Future direction: integrate MIAs with calibrated thresholds, differential privacy budgets, and privacy–utility trade-

TABLE III
SUMMARY OF CMRI GENERATIVE MODELS FROM LITERATURE

Reference	Generative Backbone	Dataset (Size / Limitation)	Conditioning (Mask / Covariate / Other)	Fidelity Evaluation	Utility Evaluation	Privacy Evaluation
UNCONDITIONAL GENERATION						
Bubeck et al. (2022) [37]	UViT-B2 / DiT-B2 (Diffusion), MaskGIT (Autoregressive)	UK Biobank (11,360 subjects; large cohort)	Unconditional generation	FID (38.78), KID	Benchmarking across spectrum; MaskGIT robust to masking	Not reported
Özdemir & Eroğlu (2024) [8]	DDPM with asymmetric attention-enhanced U-Net	OCMR dataset (limited data , restricted views)	Unconditional	FID (77.78), SSIM, MS-SSIM; radiologist accuracy 60%	Improves clinical downstream tasks; anatomical fidelity confirmed	Not reported
Tetralogy of Fallot Study (PG-GAN) [10]	Progressive GAN (PG-GAN), 256×256 output	Tetralogy of Fallot dataset (303 patients; limited)	Unconditional latent-based generation	Expert review; Dice comparison against GT masks	Synthetic training achieves U-Net accuracy within 1% of real-data baseline	Synthetic images free of privacy constraints
CONDITIONAL GENERATION						
Al Khalil / Amirrajab (2022–2023) [9][13]	Conditional GAN with SPADE layers	M&Ms Challenge (multi-vendor/multi-disease; data scarcity)	Multi-tissue segmentation maps (LV, RV, Myo, lung, fat, muscle)	SSIM, PSNR, NRMSE; SPADE superior to Pix2Pix/Pix2PixHD	Up to 4% Dice improvement; 40% HD reduction; replacement viability	Addresses privacy constraints in data sharing
Amirrajab et al. (2023) [20]	Two-stage: VAE (deformation), GAN (label-conditioned)	ACDC + M&Ms (pathology-limited datasets)	Mask latent space manipulation for pathology synthesis	EDV/ESV distribution comparison, Dice, HD	Enhances generalization across vendors and unseen pathologies	Not reported
Gheorghita et al. (2022) [14]	Conditional GAN (GauGAN-based mask-to-image)	Kaggle + UK Biobank (scarce, unbalanced)	Binary masks with uniform EF distribution	Clinical metric accuracy focused (not explicit fidelity metrics)	Pretraining improves regression accuracy for EF/EDV/ESV	Not reported
Campello et al. (2022) [13]	Conditional Generative Adversarial Network (cGAN)	UK Biobank (43,352 subjects; cross-sectional)	Covariates : age, BMI, ED/ES cardiac phase	Qualitative; MAE from pretrained ResNet18 regressor	Extracts longitudinal patterns; debiases datasets	Not reported
Cheng et al. (2023) [36]	Two-stage 3D Latent Diffusion (VAE + DDIM)	UK Biobank (927 sequences; large)	3D segmentation masks stacked with images	FID (28.37), FRD (2.92), MS-SSIM, Precision/Recall	Synthetic images/masks viable as real-data replacements	Not reported
Li et al. (2024) [40]	Masked Autoregressive Diffusion (CPGG)	UK Biobank (32,444 cine), private CMDS dataset (large-scale)	82 cardiac phenotypes (LVEF, LVEDV, etc.)	FID, Fréchet Video Distance (FVD)	Improves downstream disease classification and phenotype prediction	Not reported
Skorupko et al. (2023) [32]	Latent Diffusion (Stable Diffusion) + ControlNet	UK Biobank (biased, underrepresented groups)	Textual attributes + segmentation masks (shape)	FID, qualitative analysis of BMI features	Targeted augmentation reduces algorithmic bias	Mitigates privacy & fairness issues
Ma et al. (2024) [16]	Two-stage Controllable Mask Diffusion (LDM + SPADE)	M&Ms Challenge (Canon/Siemens; data scarcity)	Multi-category spatial masks (LV, RV, septum)	Dice, IoU (segmentation proxy for fidelity)	5–10% DSC improvement over traditional augmentation	Not reported
Zou et al. (2023) [42]	Conditional VAE + Generative SToRM (manifold reconstruction)	IRB-approved contrast-sequence dataset	Motion signals + inversion time (contrast)	Qualitative cine comparison vs bSSFP CINE	Generates T1 maps + CINE MRI simultaneously; improves efficiency	Not reported

off curves. Establish minimum privacy reporting standards for medical generative models.

4) *Insufficient cross-dataset utility evaluation:* Most segmentation utility tests rely on single datasets and lack pathology-stratified or vendor-aware validation [9].

Future direction: adopt standardized cross-dataset protocols (e.g., ACDC → MMs), evaluate by pathology subtype, and use fixed segmentation backbones for reproducible comparisons.

IX. CONCLUSION

Synthetic cardiac MRI generation has emerged as a powerful strategy to mitigate data scarcity, vendor variability, and privacy restrictions that limit real-world medical imaging datasets. Contemporary generative models, particularly mask-conditioned diffusion frameworks, demonstrate strong capability in producing anatomically faithful MRIs that enhance downstream segmentation accuracy and support rare-pathology augmentation. These models outperform

unconditional generators by enforcing structural fidelity through segmentation-guided synthesis while preserving fine-grained cardiac morphology. However, achieving robust generalization remains challenging due to multi-vendor acquisition differences, motivating the use of vendor-style conditioning and harmonization to maintain anatomical integrity across domains. Privacy remains a critical concern, with risks of memorization and nearest-neighbor leakage requiring comprehensive audits, including membership inference, structural similarity checks, and differential privacy-aware training. Current research gaps, such as limited pathology-specific conditioning, inadequate multi-dimensional privacy evaluation, and insufficient head-to-head model comparisons highlight the need for integrated pipelines that jointly optimize fidelity, diversity, and privacy for clinically safe synthetic cardiac MRI generation.

GLOSSARY TERMS

Imaging & Anatomy Abbreviations

- CMR / CMRI** — Cardiac Magnetic Resonance Imaging.
MRI — Magnetic Resonance Imaging
LV — Left Ventricle
RV — Right Ventricle
MYO — Myocardium
ED — End-Diastolic phase
ES — End-Systolic phase
SAX — Short-Axis view
FOV — Field of View (extent of MRI acquisition)
ROI — Region of Interest (cropped region containing the heart)

Generative Modelling Terms

- GAN** — Generative Adversarial Network
VAE — Variational Autoencoder
Diffusion Model — A generative model trained to iteratively denoise samples
DDPM — Denoising Diffusion Probabilistic Model (canonical diffusion model)
Forward Markov Process — The noise-adding process in diffusion models that gradually destroys image structure
Flow-Matching Model — A model learning continuous, deterministic transformations between distributions
Mask-Conditioned Generation — Using segmentation masks as structural guidance during image synthesis
Latent Space — Compressed representation where models encode high-level features
de facto — A standard method or practice that is widely used in practice, though not formally designated
MONAI — Medical Open Network for AI; an open-source framework for training medical imaging models

Evaluation Metrics

- MAE** — Mean Absolute Error
MSE — Mean Squared Error
SSIM — Structural Similarity Index Measure
FID — Fréchet Inception Distance (measures realism of generated images)
FCRE — Feature Consistency and Reconstruction Error (used in some generative evaluations)

Privacy-Related Terms

- MIA** — Membership Inference Attack (tests whether model reveals if a sample was in training data)
Nearest-Neighbour Leakage — Privacy risk where generated data resembles real training samples

- Differential Privacy (DP)** — Formal framework adding controlled noise to ensure privacy guarantees

Data & Domain Shift Terms

- Vendor Variability** — Differences in scanners, protocols, reconstruction methods across manufacturers
Domain Shift — Performance drop when model is tested on data from distributions different from training
Cross-Vendor Generalization — Ability of a model to perform consistently across different scanner vendors

Dataset Names

- M&Ms Dataset** — Multi-Center, Multi-Vendor, Multi-Disease cardiac MRI dataset
ACDC Dataset — Automated Cardiac Diagnosis Challenge dataset

REFERENCES

- [1] G. P. Diller, J. Vahle, R. Radke, M. L. B. Vidal, A. J. Fischer, U. M. Bauer, S. Sarikouch, F. Berger, P. Beerbaum, H. Baumgartner, and S. Orwat, "Utility of deep learning networks for the generation of artificial cardiac magnetic resonance images in congenital heart disease," *BMC Medical Imaging*, vol. 20, 10 2020.
- [2] J. M. Jaen-Lorites, M. Perez-Pelegrí, V. Laparra, M. P. Lopez-Lereu, J. V. Monmeneu, A. M. Maceira, and D. Moralat, "Synthetic generation of cardiac mr images combining convolutional variational autoencoders and style transfer," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2022-July. Institute of Electrical and Electronics Engineers Inc., 2022, pp. 2084–2087.
- [3] S. Amirrajab, Y. A. Khalil, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "Label-informed cardiac magnetic resonance image synthesis through conditional generative adversarial networks," *Computerized Medical Imaging and Graphics*, vol. 101, 10 2022.
- [4] Z. Liu, K. Kainth, A. Zhou, T. W. Deyer, Z. A. Fayad, H. Greenspan, and X. Mei, "A review of self-supervised, generative, and few-shot deep learning methods for data-limited magnetic resonance imaging segmentation," 8 2024.
- [5] L. Zhang, F. Wu, K. Bronik, and B. W. Papiez, "Diffuseg: Domain-driven diffusion for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, pp. 3619–3631, 2025.
- [6] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," Tech. Rep. [Online]. Available: <https://github.com/CompVis/latent-diffusion>
- [7] A. Moschetto, L. Puglisi, A. Sargood, P. Dell'Acqua, F. Guarnera, S. Battiatto, and D. Ravì, "Benchmarking gans, diffusion models, and flow matching for t1w-to-t2w mri translation," 7 2025, preprint. [Online]. Available: <http://arxiv.org/abs/2507.14575>
- [8] M. Strzelecki, R. Obuchowicz, K. Nurzynska, M. Özdemir, and O. Ero, "Diffusion model-based augmentation using asymmetric attention mechanisms for cardiac mri images," *Diagnostics* 2025, Vol. 15, Page 1985, vol. 15, p. 1985, 8 2025. [Online]. Available: <https://www.mdpi.com/2075-4418/15/16/1985>
- [9] Y. A. Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images," *Medical Image Analysis*, vol. 84, p. 102688, 2 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841522003164>
- [10] C. Heo and J. Jung, "Controllable mask diffusion model for medical annotation synthesis with semantic information extraction," *Computers in Biology and Medicine*, vol. 196, p. 110807, 9 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0010482525011588?via%3Dihub>
- [11] X. Xing, F. Felder, Y. Nan, G. Papanastasiou, W. Simon, and G. Yang, "You don't have to be perfect to be amazing: Unveil the utility of synthetic images," 5 2023. [Online]. Available: <http://arxiv.org/abs/2305.18337>
- [12] R. V. Bergen, J.-F. Rajotte, F. Yousefizadeh, A. Rahmim, and R. T. Ng, "Assessing privacy leakage in synthetic 3-d pet imaging using transversal gan," Tech. Rep. [Online]. Available: <https://www.github.com/robbergen/TrGAN>
- [13] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma, M. Parreno, A. Albiol, F. Kong, S. C. Shadden, J. C. Acero, V. Sundaresan, M. Saber, M. Elattar, H. Li, B. Menze, F. Khader, C. Haarburger, C. M. Scannell, M. Veta, A. Carscadden, K. Punithakumar, X. Liu, S. A. Tsafaris, X. Huang, X. Yang, L. Li, X. Zhuang, D. Vilades, M. L. Descalzo, A. Guala, L. L. Mura, M. G. Friedrich, R. Garg, J. Lebel, F. Henriquez, M. Karakas, E. Cavus, S. E. Petersen, S. Escalera, S. Segui, J. F. Rodriguez-Palomares, and K. Lekadir, "Multi-centre, multi-vendor and multi-

- disease cardiac segmentation: The mms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, pp. 3543–3554, 12 2021.
- [14] B. A. Gheorghita, L. M. Itu, P. Sharma, C. Suciu, J. Wetzel, C. Geppert, M. A. A. Ali, A. M. Lee, S. K. Piechnik, S. Neubauer, S. E. Petersen, J. Schulz-Menger, and T. Chitiboi, "Improving robustness of automatic cardiac function quantification from cine magnetic resonance imaging using synthetic image data," *Scientific Reports*, vol. 12, pp. 1–12, 12 2022. [Online]. Available: <https://www.nature.com/articles/s41598-022-06315-3>
- [15] J. A. Oscanoa, M. J. Middione, C. Alkan, M. Yurt, M. Loecher, S. S. Vasanawala, and D. B. Ennis, "Deep learning-based reconstruction for cardiac mri: A review," 3 2023.
- [16] J. Li, X. Ma, and Y. Shi, "Multi-label conditioned diffusion for cardiac mr image augmentation and segmentation," *Bioengineering*, vol. 12, 8 2025.
- [17] W. Al-Haidri, I. Matveev, M. A. Al-antari, and M. Zubkov, "A deep learning framework for cardiac mr under-sampled image reconstruction with a hybrid spatial and k-space loss function," *Diagnostics*, vol. 13, 3 2023.
- [18] A. B. Abdusalomov, R. Nasimov, N. Nasimova, B. Muminov, and T. K. Whangbo, "Evaluating synthetic medical images using artificial intelligence with the gan algorithm," *Sensors*, vol. 23, 4 2023.
- [19] M. H. G. del Castillo, R. M. Garcia, M. J. C. Mazón, E. A. Garcia, and P. M. Fernández-Miranda, "Diffusion models for conditional mri generation," 2 2025, preprint. [Online]. Available: <http://arxiv.org/abs/2502.18620>
- [20] S. Amirrajab, Y. A. Khalil, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "Pathology synthesis of 3d-consistent cardiac mr images using 2d vaes and gans," *Machine Learning for Biomedical Imaging*, vol. 2, pp. 288–311, 6 2023.
- [21] D. Mourad and K. O. Oseni, "Synthetic brain images: Bridging the gap in brain mapping with generative adversarial model," *International Journal of Managing Information Technology*, vol. 16, pp. 01–14, 2 2024.
- [22] A. Torfi, E. A. Fox, and C. K. Reddy, "Differentially private synthetic medical data generation using convolutional gans," Tech. Rep., 2021.
- [23] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," 4 2023.
- [24] H. H. N. Alrashedy, A. F. Almansour, D. M. Ibrahim, and M. A. A. Hammoudeh, "Braingan: Brain mri image generation and classification framework using gan architectures and cnn models," *Sensors*, vol. 22, 6 2022.
- [25] Y. Skandarani, A. Lalande, J. Afifalo, P.-M. Jodoin, and P.-M. J. G. Adver, "Generative adversarial networks in cardiology," *Canadian Journal of Cardiology*, vol. 2022, pp. 196–203, 2021. [Online]. Available: <https://hal.science/hal-03612001v1>
- [26] J. Urcia-Vázquez, M. Pérez-Pelagrí, J. V. Monmeneu, M. P. López-Lereu, J. M. Santabarbara, A. M. MacEira, and D. Moratal, "Generating realistic cardiac mr images using diffusion models," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., 2024.
- [27] B. Hejrati, S. Banerjee, C. Glide-Hurst, and M. Dong, "Conditional diffusion model with spatial attention and latent embedding for medical image segmentation," 2 2025. [Online]. Available: <http://arxiv.org/abs/2502.06997>
- [28] M. Yazdani, Y. Medghalchi, P. Ashrafiyan, I. Hacihaliloglu, and D. Shahriari, "Flow matching for medical image synthesis: Bridging the gap between speed and quality," 3 2025. [Online]. Available: <http://arxiv.org/abs/2503.00266>
- [29] A. Hadzic, S. J. Joham, and M. Urschler, "Flow matching for conditional mri-ct and cbct-ct image synthesis," 10 2025, preprint. [Online]. Available: <http://arxiv.org/abs/2510.04823>
- [30] N. Konz, Y. Chen, H. Dong, and M. A. Mazurowski, "Anatomically-controllable medical image generation with segmentation-guided diffusion models," 6 2024. [Online]. Available: <http://arxiv.org/abs/2402.05210>
- [31] P. Dhariwal, Openai, and A. Nichol, "Diffusion models beat gans on image synthesis," 5 2021.
- [32] 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," 9 2025. [Online]. Available: <http://arxiv.org/abs/2403.19508>
- [33] X. C. Wong, D. Sarikaya, K. Zucker, M. D. Kamps, and N. Ravikumar, "Can diffusion models bridge the domain gap in cardiac mr imaging?" 8 2025. [Online]. Available: <http://arxiv.org/abs/2508.06327>
- [34] H. Sun, J. Plawinski, S. Subramaniam, A. Jamaludin, T. Kadir, A. Readie, G. Ligozio, D. Ohlsson, M. Baillie, and T. Coroller, "A deep learning approach to private data sharing of medical images using conditional generative adversarial networks (gans)," *PLoS ONE*, vol. 18, 7 2023.
- [35] R. Sahoo, V. Naik, S. Singh, and S. Malik, "Gans and vaes as methods of synthetic data generation and augmentation to enhance heart disease prediction," *International Journal of Engineering and Advanced Technology*, vol. 11, pp. 17–23, 12 2021. [Online]. Available: <https://www.ijeat.org/wp-content/uploads/papers/v11i2/B32631211221.pdf>
- [36] N. Cheng, Z. Liu, Y. Deo, H. Dou, N. Bi, K. Wu, F. Lin, Z. A. Taylor, N. Ravikumar, and A. F. Frangi, "Synthesising 3d cardiac cine-mr images and corresponding segmentation masks using a latent diffusion model," in *Proceedings - International Symposium on Biomedical Imaging*. IEEE Computer Society, 2024.
- [37] N. Bubeck, Y. Zhang, S. Shit, D. Rueckert, and J. Pan, "Reconstruct or generate: Exploring the spectrum of generative modeling for cardiac mri."
- [38] D. R. Padariya, "A privacy-preserving framework for generative model-driven synthetic datasets," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 29 289–29 290, 4 2025. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/35222>
- [39] X. Zhao, Y. Tokuoka, J. Iwasawa, and K. Oda, "Frequency-calibrated membership inference attacks on medical image diffusion models," 6 2025. [Online]. Available: <http://arxiv.org/abs/2506.14919>
- [40] Z. Li, Y. Hu, Z. Ding, Y. Mao, H. Li, F. Yi, H. Zhang, and Z. Huang, "Phenotype-guided generative model for high-fidelity cardiac mri synthesis: Advancing pretraining and clinical applications," 5 2025. [Online]. Available: <http://arxiv.org/abs/2505.03426>
- [41] X. Li, J. Yang, S. Jiao, T. University, C. Eung, J. Lee, and V. M. Campello, "Cardiac aging synthesis from cross-sectional data with conditional generative adversarial networks," Tech. Rep.
- [42] Q. Zou, S. Priya, P. Nagpal, and M. Jacob, "Deep generative model for joint cardiac t1 mapping and cardiac cine," in *Proceedings - International Symposium on Biomedical Imaging*, vol. 2023-April. IEEE Computer Society, 2023.