

A Systematic Review on Synthetic Medical Images Generation—Recent Trends and Future Opportunities

Yumna Waheed ¹, Muhammad Nouman Noor ¹  and Imran Ashraf ^{2,*} 

¹ Department of Artificial Intelligence and Data Science, National University of Computer and Emerging Sciences, Islamabad 44000, Pakistan; nouman.noor@isb.nu.edu.pk (M.N.N.)

² Computer Engineering Lab, Quantum and Computer Engineering Department, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), TU Delft, 2628 CD Delft, The Netherlands

* Correspondence: i.ashraf-1@tudelft.nl

Abstract

Background/Objectives: Generative Models have revolutionized the synthesis of complex, realistic medical images. However, obtaining annotated, high-quality datasets is challenging and expensive due to privacy concerns, high human annotation costs, and data scarcity. This systematic literature review (SLR) provides a comparative overview of Generative AI models used for medical image generation, categorizing the research into GANs, Diffusion Models, and Autoencoders used for medical image generation and GAN loss functions. **Methods:** This systematic review followed the PRISMA 2020 guidelines. Studies published between 2021 and 2026 were retrieved from selected scientific databases and screened using predefined inclusion criteria. We selected 120 papers published in reputable databases and journals. The review examined generative modeling approaches aimed at addressing data scarcity and annotation limitations in medical imaging, including GAN-based adversarial synthesis, diffusion-based iterative denoising, and Autoencoder-based latent representation learning, along with an in-depth analysis of GAN loss functions. **Results:** GANs were the most widely used approach (36% of the reviewed studies), achieving the lowest FID scores (15.552–31.349). Diffusion Models (26% of the reviewed studies) showed superior structural fidelity, achieving SSIM values of up to 0.915, whereas GANs achieved 0.593 on brain MRI datasets. In addition, hybrid loss functions combining perceptual, structural, and pixel-level terms resulted in improved image quality performance (PSNR = 35.6; SSIM = 0.95). **Conclusions:** The analysis showed that GAN-based generative models produce visually realistic images with detailed textures, whereas Diffusion Models generate high-resolution images with superior structural fidelity but require substantial computational resources for training and image generation. Furthermore, the clinical implications, limitations, and challenges of Generative AI models were discussed. Despite these advances, the medical imaging field still faces several challenges, including the need for annotated medical datasets, high-quality realistic images for training, and limited data availability.



Academic Editors: Ghazanfar Latif and Jaafar Alghazo

Received: 18 May 2026

Revised: 15 June 2026

Accepted: 7 July 2026

Published: 10 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: medical image generation; GAN; diffusion models; variational autoencoders; loss functions; synthesis

1. Introduction

In the past few years, Generative AI models have been demonstrated in clinical imaging. In healthcare, the generation and enhancement of medical images play a significant role in image diagnosis and classification. Therefore, the ability to generate realistic high-resolution medical images is a crucial step in accurate treatment planning. Generative

modeling, as a form of unsupervised learning, is a field that researchers have applied to various imaging generation tasks. Recent contributions, including Variational Autoencoders (VAEs) [1], Generative Adversarial Networks (GANs) [2], and Diffusion Models [3], are based on probabilistic modeling, adversarial training, and iterative denoising processes. These models learn complex data distributions to generate realistic and high-quality medical images. Despite these advancements, medical imaging faces several challenges, such as low contrast, small lesion visibility, noisy images, structural variability, mode collapse, and data scarcity, reducing the ability to generate high-resolution images [4]. For example, wireless capsule endoscopy, an imaging modality, captures thousands of frames from the gastrointestinal tract [5]. The limited perspective and image quality lead to missing structural and anatomical information for analysis and generation, making it difficult to obtain high-quality images [6].

Generative models have been recognized as a pivotal innovation in AI research, offering unique capabilities that differentiate them from traditional approaches [7]. Medical image synthesis requires advanced techniques for realistic synthesis, reconstruction, and enhancement. These generative models have demonstrated significant potential in the field of medical image generation by learning complex patterns and data distribution. Diffusion Models have been used to produce quality images, preserving the global structure and semantics of the image [8–10]. GANs have been employed to produce 2D or 3D images, improving image sharpness and reducing artifacts in medical data [11,12].

Although publicly available annotated medical datasets such as Kvasir, BraTS, ChestX-ray, and Hyper-Kvasir have advanced medical image diagnosis, they exhibit structural constraints and lack diversity [13]. Improving the quality and applicability of medical image generation requires a structured understanding of how model architectures mitigate these challenges. Existing systematic reviews are fragmented in their evaluation of model optimization and provide limited evidence on how GAN loss formulations, mode collapse, vanishing gradients, architectural hyperparameters, and optimization inconsistencies collectively affect the reliability of synthetic medical image generation [14–22]. Although prior literature focuses on model architectures and datasets, providing a limited overview of various GAN loss functions, Wasserstein distances, least-squares formulations, hinge losses, and hybrid objective functions within a unified framework. The limited use of GAN loss functions across diverse datasets and the widespread absence of standardized reporting for training convergence profiles further obscure the true reproducibility of published synthetic frameworks. Furthermore, no prior review has systematically evaluated how different GAN loss functions affect image quality, training stability, and convergence in the context of medical imaging.

To examine GANs, Diffusion Models, and VAEs in image generation, a systematic literature review is crucial. In this SLR, we review the generative models for medical image generation and their variants, the impact of the loss function in GAN, and the challenges in the existing literature. We also analyze various generative architectures and medical image modalities by conducting a survey on medical image generation models. Despite the significant success of generative models in image generation, which allows the model to produce high-quality results, generative models suffer from high computational cost, mode collapse, and unstable training. In this review, we also discuss the problems in GANs and associated loss functions. In this SLR, we present the following key contributions:

- We provide a comprehensive analysis of Generative AI models, including Variational Autoencoders, GANs, and Diffusion Models in medical imaging.
- An in-depth evaluation of GAN loss functions in image generation to improve quality, stability, and convergence.

- We discuss GAN architectures commonly used in medical imaging and their applications across various medical imaging modalities.
- We systematically examine key training and data challenges identified based on the synthesis of the 120 included studies. Specifically, mode collapse, vanishing gradients, and training instability in GAN-based models are analyzed in Sections 5.3.1, 5.3.2, and 5.3.3 respectively. Section 5.2 provides a relative comparison of computational cost over model categories. Data scarcity and imbalanced datasets are identified as recurring limitations across the included studies and are discussed in the context of Section 5. The challenges reported throughout this review were identified through systematic thematic synthesis of the limitations sections of all 120 included studies. These challenges were identified based on the systematic extraction and synthesis of reported limitations using the data extraction approach detailed in Section 2.5.

This paper is organized as follows. Section 2 outlines the methodology, Section 3 presents various generative models and hybrid approaches, Section 4 presents GAN loss functions applied in image generation and the strengths and weaknesses of commonly used loss functions, and Section 5 outlines the discussion. Section 6 outlines limitations and Section 7 concludes the review.

2. Methods

The methodology of this systematic literature review (SLR) was conducted in accordance with the PRISMA 2020 guidelines [23] and the completed PRISMA 2020 Checklist is provided in the Supplementary Materials. The methodology presents the technique and criteria to assess articles. It is divided into research questions, search strategy, search outcomes, quality assessment criteria, dataset overview, and data extraction strategy. The protocol for this review was not recorded in any other publicly accessible registry (such as PROSPERO); all methodological procedures were conducted in accordance with PRISMA guidelines.

2.1. Data Collection

The systematic search identified papers from selected databases such as IEEEExplore, Google Scholar, ScienceDirect, PubMed, Web of Science, and Scopus, and extracted a total of 3194 papers. After the removal of duplicates, a total of 2840 papers were screened. A total of 545 papers were retrieved based on title and abstract screening, then reduced to 176 papers via full-text review. In the final stage, 120 papers were included after applying all inclusion and exclusion criteria. PRISMA diagram outlining identification, screening, eligibility, and inclusion is presented in Figure 1.

2.2. Research Question

This SLR answers the following questions:

RQ1: What are the generative model techniques employed for the synthesis of medical images in different modalities?

RQ2: Which loss functions are commonly used in GANs for improving the quality and stability of synthetic image generation?

RQ3: Which generative models have been applied in clinical settings?

RQ4: How are generative models used to overcome data limitations?

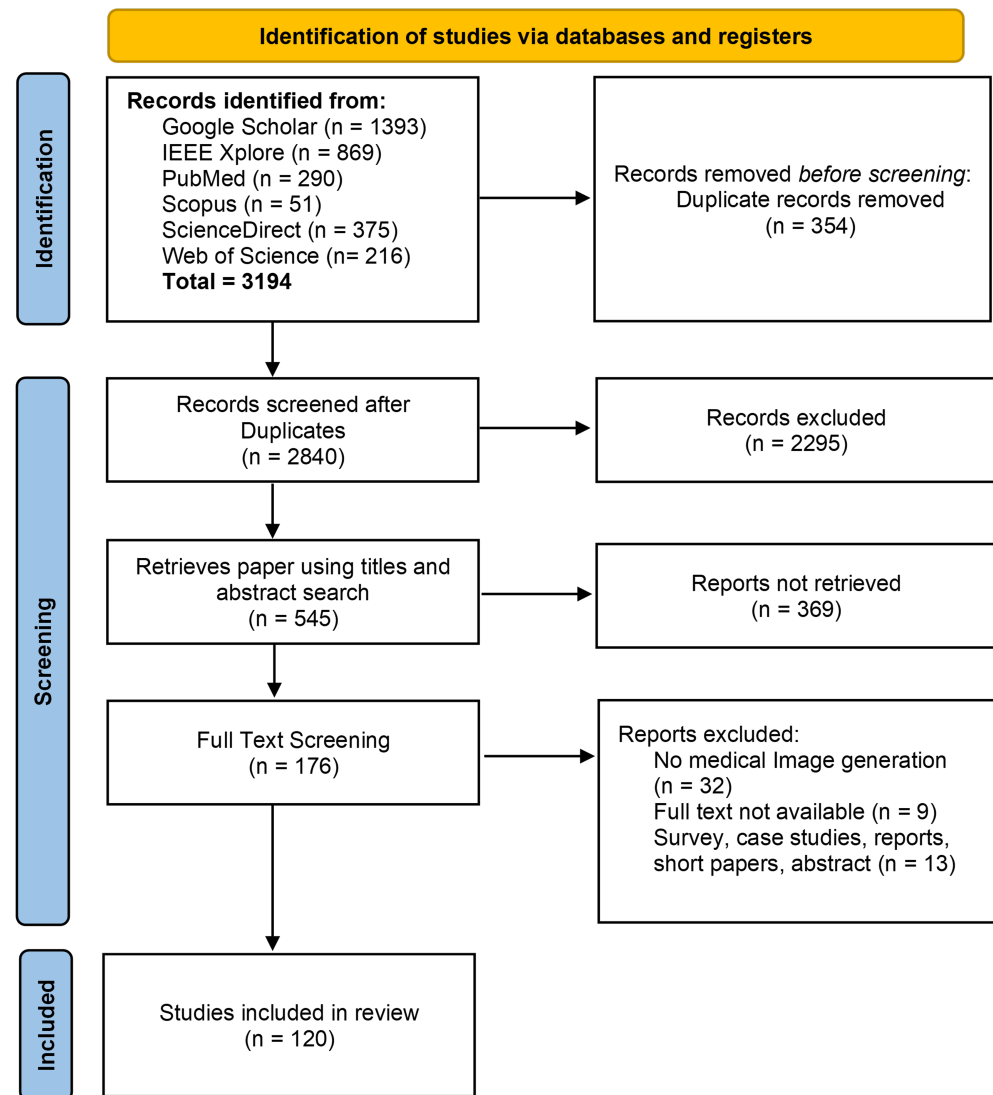


Figure 1. PRISMA diagram.

2.3. Formulate Search Query

The papers were identified from six well-known databases using a systematic search. Each database was accessed directly through official web interfaces. To identify relevant studies, we applied the following search queries: “synthetic image generation”, “GANs in medical imaging”, and “data augmentation”. We also used combined keywords to enhance the following search query:

- (“Autoencoder” OR “VAE” OR “Variational Autoencoder” OR “Variational Autoencoders” OR “GAN” OR “GANs” OR “Generative Adversarial Network” OR “Generative Adversarial Networks” OR “Diffusion Model” OR “Diffusion Models” OR “Latent Diffusion Model” OR “DDPM” OR “score-based generative model” OR “score matching”) AND (“medical image generation” OR “medical image synthesis” OR “clinical image synthesis” OR “synthetic medical images” OR “modality translation” OR “cross-modality synthesis”) [Filter: 2021–2026, English]
- (“GAN” OR “GANs” OR “Generative Adversarial Network”) AND (“loss function” OR “objective function” OR “adversarial loss” OR “adversarial training loss” OR “perceptual loss” OR “Wasserstein loss” OR “hinge loss” OR “cycle-consistency loss”) AND (“medical image synthesis” OR “medical image generation” OR “clinical image generation” OR “synthetic imaging”) [Filter: 2021–2026, English]

- (“score-based generative model” OR “score matching” OR “text-to-image” OR “Stable Diffusion” OR “privacy-preserving” OR “modality translation” OR “cross-modality synthesis” OR “domain adaptation”) AND (“medical image generation” OR “medical image synthesis” OR “clinical imaging” OR “radiology image synthesis”) [Filter: 2021–2026, English]

Database-specific search settings were applied to limit searches where applicable. The search strategy was designed using a PICOC-guided approach, focusing on generative AI-based medical image generation. Table 1 illustrates the overview of the literature selected using the PICOC framework.

Table 1. PICOC Search Strategy.

PICOC Elements	Keywords
Population (P)	“medical imaging”, “clinical imaging”
Intervention (I)	“image generation”, “image synthesis”, “synthetic image”, “GAN”, “variational autoencoder”, “diffusion model”, “DDPM”, “stable diffusion”, “latent diffusion models”
Comparison (C)	“Diffusion Model”, “Generative Adversarial Network”, “Variational Autoencoders”
Outcome (O)	“image quality”, “FID score”, “SSIM”, “PSNR”
Context (C)	“medical image synthesis”, “healthcare AI”, “clinical diagnosis”, “disease detection”, “rare disease imaging”

2.4. Eligibility Criteria

In this systematic literature review, we defined inclusion and exclusion criteria to ensure the selection of relevant and high-quality papers. Table 2 illustrates the eligibility criteria:

Table 2. Inclusion and Exclusion Criteria for Study Selection.

Criteria Type	Description
Inclusion Criteria (IC)	IC1: Peer-reviewed full-text journal articles and conference proceedings. IC2: Studies focus on medical image generation or synthesis using GANs, Diffusion Models, or Autoencoders. IC3: Must be in English. IC4: Papers published between 2021 and 2026
Exclusion Criteria (EC)	EC1: Preprints (e.g., non-peer-reviewed arXiv papers), abstracts, short papers. EC2: Studies not focusing on clinical or medical imaging data generation tasks. EC3: Manuscripts with full text unavailable or not written in English. EC4: Literature published prior to 1 January 2021.

- Applying these criteria, the selection followed a PRISMA 2020 multi-stage workflow.
- Title and Abstract Screening: Initial screening of studies that were not related to medical imaging to remove irrelevant literature.
 - Full-Text Review: The remaining studies were assessed for data collection details, methodology clarity, and evaluation results.

- Backward and Forward Search: With PRISMA guidelines, a supplementary backward and forward reference search was conducted to ensure that all relevant studies were included.

Table 3 presents an overview of the generative model architectures, including GAN-based approaches, Diffusion-based methods, and Autoencoder-based architectures used in this review.

Table 3. Generative Models and Variants for Medical Imaging.

Model Type	Variants
GAN-based architecture	Deep Convolutional GAN, Denoising GANs, StyleGAN, Cycle GAN, Pix2Pix, conditional GAN, PGAN, super-resolution GAN
Diffusion Models	Stable Diffusion, Denoising Diffusion Probabilistic Model (DDPM), Latent Diffusion Model (LDM)
Autoencoder	Convolutional VAE (CVAE), Conditional VAE, VAE

The pie chart in Figure 2 highlights the proportion of studies that employed different generative models in the reviewed literature.

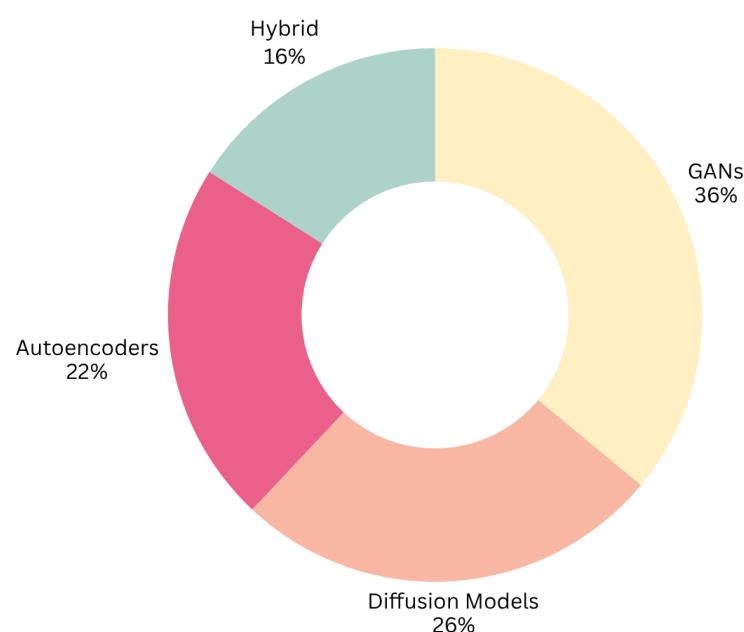


Figure 2. Distribution of Generative Model Architectures.

2.5. Data Extraction

A data extraction form was designed to maintain reliability across all 120 included studies. All relevant data fields were systematically and manually extracted from each included study by the primary reviewer (Y.W.). The extracted data were subsequently reviewed and independently validated by the second reviewer (M.N.N.) in a supervisory role and validation on a random subset of the included studies (20%, $n = 16$), and by the third reviewer (I.A.) as part of formal analysis and validation, to ensure consistency, completeness, and reproducibility. Disagreements between reviewers were resolved through structured discussion or by consulting a third reviewer. The following data fields were extracted for systematically capturing methodological practices:

- Generative model type: GAN-based, Diffusion Model-based, Autoencoder-based, or Hybrid architecture.

- Medical imaging modality: MRI, CT, X-ray, endoscopy, ultrasound, PET, OCT, dermoscopy, or multimodal imaging.
- Datasets used: Public benchmark datasets (e.g., BraTS, Kvasir-v2, NLST, MedicalMNIST) or private datasets, including image count and domain specificity.
- Loss functions applied: Adversarial, perceptual, structural similarity, pixel-level, textural, content, or hybrid loss formulations used during GAN training.
- Data augmentation strategies: Geometric transformations, Gaussian noise, motion blur, or other augmentation techniques employed to expand training datasets.
- Performance metrics: Image quality metrics used in medical image generation: Fréchet Inception Distance (FID), Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR).

The challenges reported in this systematic review, such as mode collapse, vanishing gradients, training instability, data scarcity, dataset imbalance, and computational cost, were identified through systematic thematic synthesis of the limitations sections of all 120 included studies and discussed in Sections 5.3.1–5.3.3 and 2.6.

Due to heterogeneity in generative model architectures, imaging modalities, datasets, and evaluation metrics across the included literature, a quantitative meta-analysis was not feasible. Instead, findings were grouped by generative model category (GAN, Diffusion Model, Autoencoder, and Hybrid). The synthesized results are available in the Supplementary Materials.

2.6. Datasets Overview

A number of studies have utilized small datasets, which impacts the ability to learn meaningful representations and complex data distributions in medical image synthesis. The medical images used in many studies are domain-specific. For model training, publicly available datasets were employed in many studies, and some studies used multimodal datasets. The summary of datasets used in this review is shown in Table 4:

Table 4. Overview of Datasets, Total Images, and Domain.

Dataset	Domain	Total Images	Year Released	References
Kvasir-v2	Gastrointestinal (Endoscopy)	8000	2017	[12,24–26]
KID	Gastrointestinal (Capsule Endoscopy)	2371	2017	[27,28]
Kvasir-Capsule	Gastrointestinal (Capsule Endoscopy)	47,238	2020	[5,28]
Hyper-Kvasir	Gastrointestinal (Endoscopy)	110,079	2020	[29]
Kvasir-SEG	Polyp Segmentation	1000	2020	
MedicalMNIST	Multi-domain	58,954	2017	[30]
NLST (National Lung Screening Trial)	Lung-CT	50,000	2014	[31]
BraTS	MRI	369 cases	2020	[32,33]
Cervical Spine MRI Dataset	MRI (Cervical Spine)	514	2022	[34]
Knee Osteoarthritis	Knee X-ray	5778	2018	[30]
TotalSegmentator	CT	1204	2023	[35]

The most common models for medical image synthesis are Diffusion Models, Generative Adversarial Networks (GANs), and Variational Autoencoders (VAE). Studies have also applied data augmentation techniques such as geometric and color transformations, Gaussian noise, and motion blur to expand the training dataset. The bar chart in Figure 3 presents a comparison of various datasets from the selected studies and shows multiple types of imaging data.

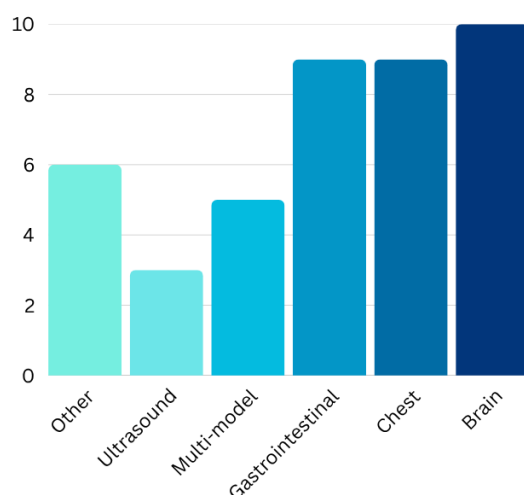


Figure 3. Distribution of datasets used in selected studies by imaging modality.

Figure 4 illustrates the distribution of Diffusion Models, Autoencoders, and GANs across different medical domains. The bar chart shows the Gastrointestinal dataset as the leading area in this literature.

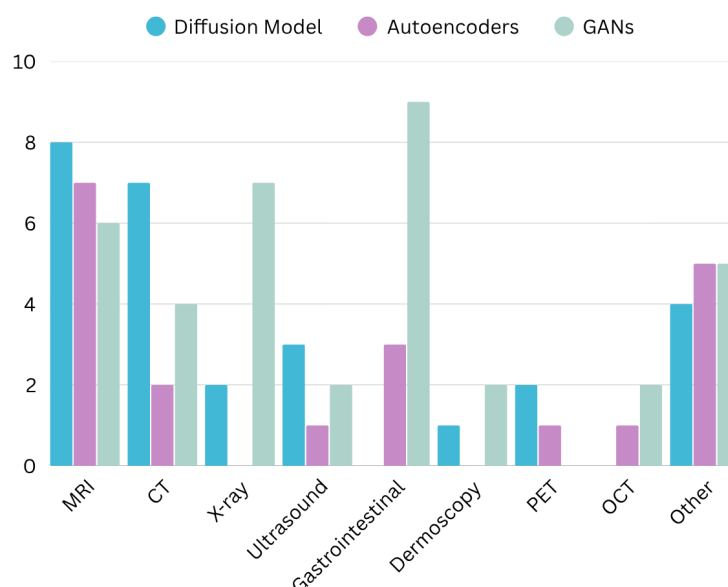


Figure 4. Selected studies across modalities and approaches.

2.7. Quality Assessment and Risk of Bias

To ensure consistency in the evaluation of the selected studies, we applied a structured quality assessment (QA) framework, as illustrated in Table 5. This framework assesses each study on five predefined criteria: (1) clarity of the proposed generative model objectives, (2) completeness of the methodological framework, (3) relevance to medical image generation, (4) quality of evaluation metrics (e.g., FID, SSIM, PSNR), and (5) clarity of reported limitations and conclusions. The risk of bias and methodological quality of the included studies were evaluated by two reviewers (Y.W. and M.N.N.) using the customized 5-point quality assessment (QA) framework, as illustrated in Table 5.

Each study was examined to determine whether it incorporated the key elements defined in the data extraction plan. Articles were selected based on the assigned scores of 1 (met), 0.5 (partially met), and 0 (not met) for the specified assessment criteria. As presented in Figure 5, the studies presented low risk of bias, with the majority of studies

scoring 1.0 (Met) across these criteria. ‘Standard Metrics Used’ and ‘Limitation Clarity’ exhibited higher proportions.

Table 5. Quality Assessment (QA) Criteria for Literature Selection.

QA ID	Quality Assessment Question	Scoring Criteria
QA 1	Is the proposed generative architecture and objective clearly described?	1: Objectives clearly defined.
		0.5: Partially specified.
		0: Not specified.
QA 2	Is the methodology (architecture, loss functions, datasets) defined clearly?	1: Methodology is defined.
		0.5: Partially elaborated.
		0: Not elaborated.
QA 3	Does this research focus on medical image generation?	1: Focuses on medical images.
		0.5: Partially elaborated.
		0: Non-medical.
QA 4	Are the results evaluated using standard metrics (FID, SSIM, PSNR)?	1: Results fully evaluated.
		0.5: Partially discussed.
		0: Not evaluated.
QA 5	Are the limitations and findings clearly concluded?	1: Concluded clearly.
		0.5: Partially concluded.
		0: No conclusion.

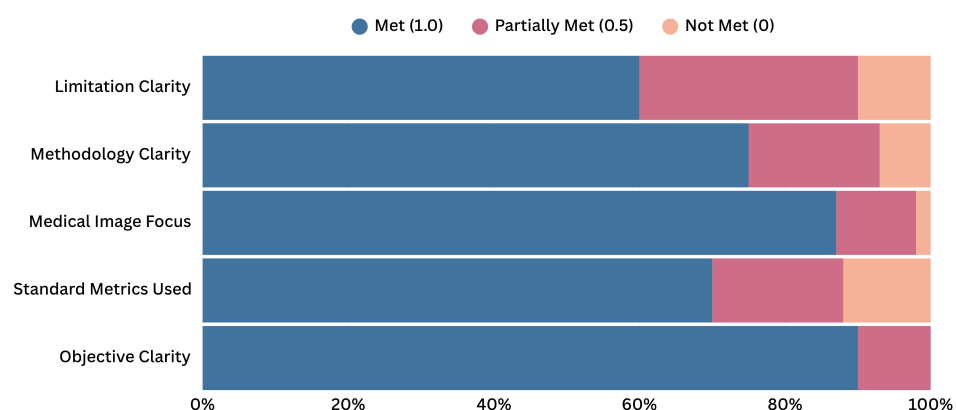


Figure 5. Risk of Bias Assessment across Selected Studies.

3. Medical Image Synthesis with Generative Models

Numerous researchers have employed generative models to perform various tasks, including generation, reconstruction, segmentation, and denoising. This section discusses some generative models such as Generative Adversarial Networks (GANs), Diffusion Models, and Autoencoders, their applications, and imaging modalities in detail. A complete summary of study characteristics, including methods, datasets, model types, and evaluation metrics for all included studies, is provided in Supplementary Table S1.

3.1. Generative Adversarial Networks in Medical Image Synthesis

GANs were first introduced in 2014 [2] and have been used to learn data distributions for image generation through adversarial training. These methods use a generator and a discriminator to produce synthetic data that resembles real data. The model generates complex, high-dimensional data by transforming random noise into realistic samples.

Various GAN architectures have been proposed, for instance, CycleGANs [36], which improve image translation by learning a mapping between two different image domains, such as in medical MRI-to-CT and CT-to-MRI, etc. By incorporating the cycle-consistency loss function, it ensures more accurate translation. StyleGAN can generate high-resolution synthetic medical images [37].

Numerous researchers have applied GAN frameworks to medical image generation. For instance, to improve image generation, various advanced techniques have been incorporated into GANs. Maguluri et al. [38] proposed a PGAN that synthesizes chest images using residual connections and progressive upsampling. Performance metrics included SSIM and FID, which achieved 22.4 and 0.9, respectively. Zhang et al. [24] aimed to generate quality data using a GenSeg framework. The framework used multi-level optimization to achieve better performance. Feng et al. [39] leveraged a knowledge distillation GAN to address the data scarcity for medical image generation. The KGAN achieved SSIM of 0.593 and SCD of 1.229. Khatun et al. [30] aimed to generate medical images using quantum learning, named QIGL. Liu et al. [40] aimed to propose a RadImageGAN, a radiologic image generation across multiple modalities. They combined with BigDatasetGAN to produce multiclass, high-resolution synthetic medical images. Shah et al. [41] leveraged SRGAN, a super-resolution GAN, to improve image quality. To improve image resolution, perceptual and adversarial losses are used to preserve the texture and sharpness of medical images. Figure 6 illustrates the GAN architectures used in medical image generation.

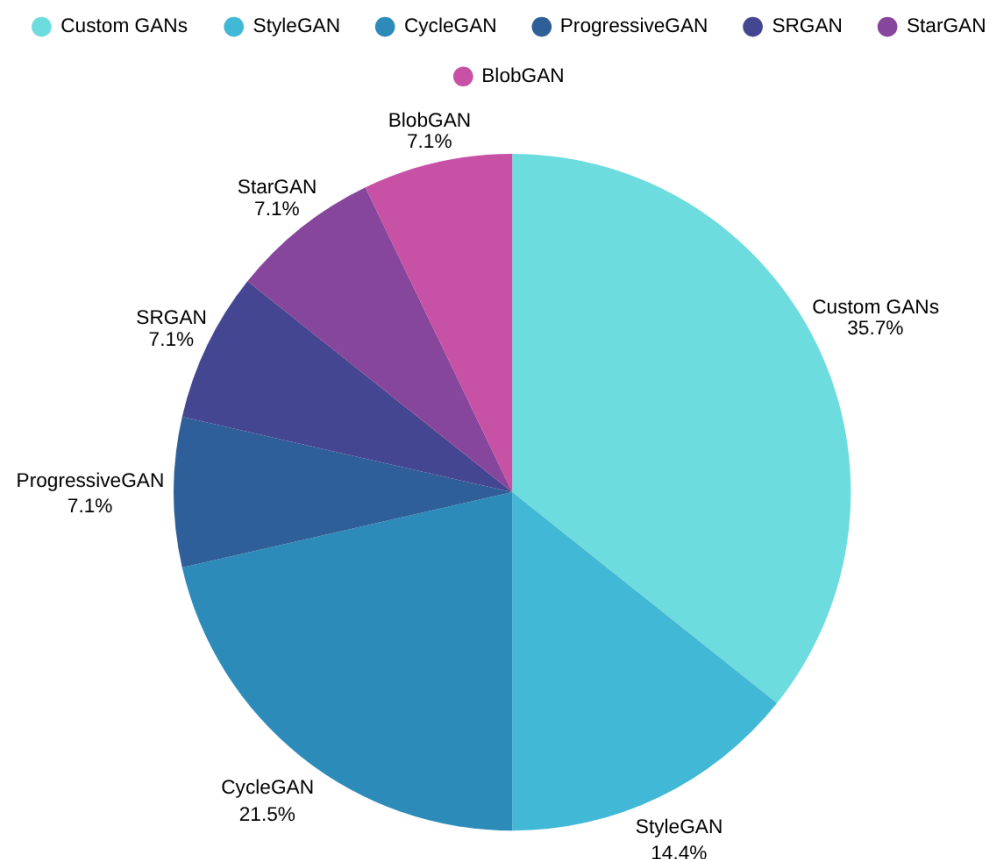


Figure 6. Distribution of GAN architectures.

In [12], the authors Janutenas and Sesok applied the Viewpoint Attention Module and Perspective Transformation Module within StarGAN to augment endoscopic images. The goal was to improve multiclass classification accuracy against InceptionNet-V3, DenseNet-121, VGG-16, ResNet-50, and EfficientNetB7. Similarly, Magalhães et al. [11] applied a Deep

convolutional GAN for endoscopic image augmentation. Their proposed approach was evaluated using ResNet50 and Vision Transformer (ViT) architectures for classification. A multiscale generative network was designed by Liu et al. [26] to generate super-resolution images. They combined five loss functions to improve endoscopic image quality and achieved 0.83 SSIM and 31.86 PSNR on the Kvasir dataset. Ayubi et al. [25] employed BlobGAN with self-attention for GI image generation. A self-attention block improves feature representation and image realism, and with an FID of 15.5, it enhances the quality of gastrointestinal images and improves classification systems. In another study, Prabu et al. [42] introduced a method to improve the classification of endoscopic images by utilizing a machine learning (ML) algorithm and a GAN, named EIC-MLA. The proposed method enhances image quality and GI disease classification, unlike traditional methods that achieve lower classification accuracy. Mao et al. [43] employed CycleGAN, a GAN-based framework, to transform low-normal-light and normal-low-light gastroscopic images. The study improved image brightness for clinical diagnosis using an illumination-aware attention module. Numerous studies have utilized CNN architectures to capture complex patterns and hierarchical features, improving classification, diagnosis, and image generation tasks. Xiao et al. [5] aimed to design a capsule endoscopy generation method using a special common attention GAN (SCAGAN). This method was used to capture external and internal features. To improve GAN performance, a DiffSeg method was employed. The experiment showed that SCAGAN enhanced model quality and achieved an FID of 31.349. Yoon et al. [44] employed a GAN-based system to learn SSL-specific features and generate realistic endoscopic images. The proposed method based on a CAdE configuration with augmentation methods achieved an FID of 42.19 and a sensitivity of 93.98%. Lee et al. [45] improved detection of gastric cancer by integrating YOLOv8 with synthetic image generation using StyleGAN3. It achieves this by generating high-resolution gastric cancer images. Figure 7 shows the usage of different CNN models in the selected review:

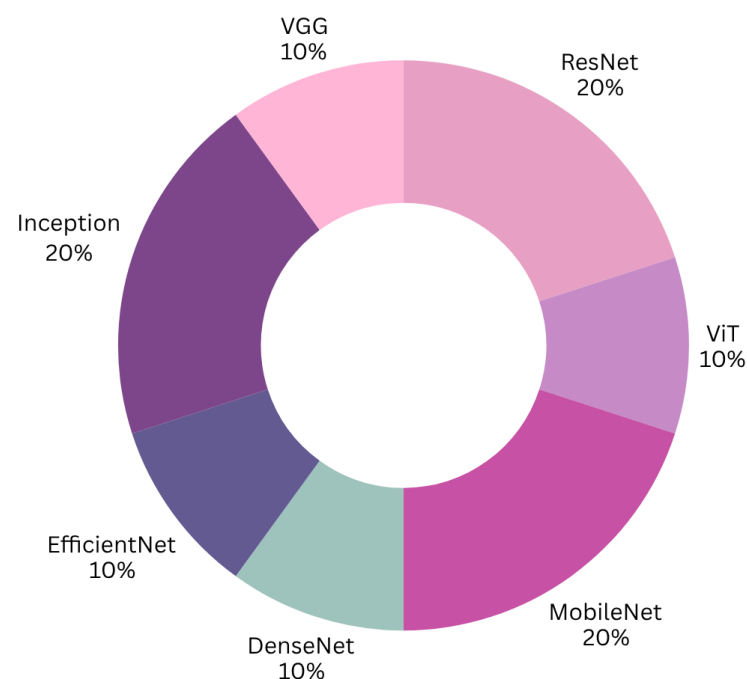


Figure 7. Distribution of CNN Architecture.

In brain MRI synthesis approaches, researchers have utilized GAN-based methods, possibly due to their ability to generate sharp, highly realistic anatomical details. Mukherjee et al. proposed an aggregated GAN-based framework, denoted as AGGrGAN, designed to synthesize brain tumor MRI generation. Specifically, two variants of DCGAN [46] and

a Wasserstein GAN (WGAN) [47] were incorporated into a unified framework to capture diverse feature representations from the input data. Furthermore, a style transfer mechanism was integrated into the generation pipeline to enhance the visual realism and structural resemblance of the synthesized images [48]. Onakpojeruo et al. proposed a novel Pix2Pix GAN-based framework for synthetic brain MRI generation, leveraging conditional image-to-image translation. The framework generates realistic brain tumor images through a conditional adversarial learning process, providing augmented training data for a Conditional Deep Convolutional Neural Network for multiclass brain tumor classification [49]. Hamghalam et al. proposed a conditional Generative Adversarial Network (cGAN)-based framework for brain tumor segmentation, employing two complementary Enhancement and Segmentation GAN (ESGAN) and Enhancement GAN (EnhGAN) models. To enhance the contrast of tumor sub-regions within the image, the authors utilized a contrast-enhancement process that adaptively calibrates voxel intensities within input patches [50]. Zhou et al. introduced a vector quantization GAN (VQ-GAN), designed to generate 3D brain tumor ROIs [51]. Wang et al. proposed a Selective State-Space Module, denoted as SelSSM, designed to synthesize multimodal MRI images across diverse GAN architectures. To improve anatomical consistency, the authors utilized a gradient-constrained adaptation that improves spatial alignment. The experiments demonstrated that SelSSM improves both image synthesis quality and segmentation performance [52]. Saimon et al. employed a conditional GAN (cGAN)-based framework, designed to generate synthetic brain tumor images. The study utilized these generated images for data augmentation by training multiple deep learning models on synthetic data to address class imbalance in brain tumor classification [53]. Lai et al. proposed a StyleGAN2-ADA-based framework, named StyleGAN2-ADA, designed to generate high-quality synthetic brain MRI images for data augmentation and privacy-preserving medical image sharing. Specifically, training set size variations were systematically incorporated into the model training process to analyze the impact on image fidelity and diversity metrics [54].

3.2. Diffusion Model in Medical Image Synthesis

Diffusion-based models are designed to generate realistic, high-quality, and diverse data by learning to reverse a noise-corruption process [3]. Many diffusion-based methods have been proposed to enhance reconstruction, segmentation, generation, and denoising from complex data distributions. Wang et al. [55] primarily discussed two models for medical image restoration. In this study, the authors categorized Diffusion Models into score-based models and denoising diffusion probabilistic models and utilized these to generate CT, MRI, endoscopic, and PET data. In the conditional DDPM method, Wang et al. [35] generated high-quality CT images using masks and labels. Better recovery of anatomical details was achieved using a super-resolution GAN, achieving a PSNR of 19.51.

Krishna et al. [31] proposed a denoising diffusion probabilistic model with multi-condition guidance for lung CT and mammography image synthesis. Performance metrics included AUC and F1 Score, achieving an AUC of 0.599 and an F1 Score of 0.319 on 10,000 synthetic images. In another study [8], Ma et al. introduced MDDPM, which used a multi-level UNet and a cross-attention mechanism to generate super-resolution MRI images. Approximate SSIM for Alzheimer MRI, Knee MRI, and Brain MRI datasets were 0.888, 0.899, and 0.915, respectively. Pan et al. [34] employed DDPM for the generation of cervical MRI images, named CSM-DPM. The method enhances the DDPM by incorporating Gaussian and point noise. Conditional diffusion-based mammogram synthesis, introduced by Montoya et al. [10], allows the controlled generation of images. Instead of relying on multiple separate models, which increase computational complexity, the proposed framework fine-tunes a pretrained Stable Diffusion Model to learn domain-specific and structural features, while

Hung et al. [56] applied a conditional Denoising Diffusion Probabilistic Model (DDPM) for MRI super-resolution, inpainting, and CT/X-ray image synthesis. Liu et al. [57] presented a diffusion-based framework for fully automatic polyp image generation, named Polyp-Gen. This method employed a spatial-aware diffusion training to improve polyp sharpness and structure. Freiche et al. [58] fine-tuned Stable Diffusion to generate high-quality ultrasound images. Similarly, Kaleta et al. [59] employed stable diffusion for endoscopic image synthesis. This method is based on image-to-image translation and achieved an IOU of 69.76% in another study. Figure 8 summarizes the usage of different diffusion model architectures in the reviewed studies.

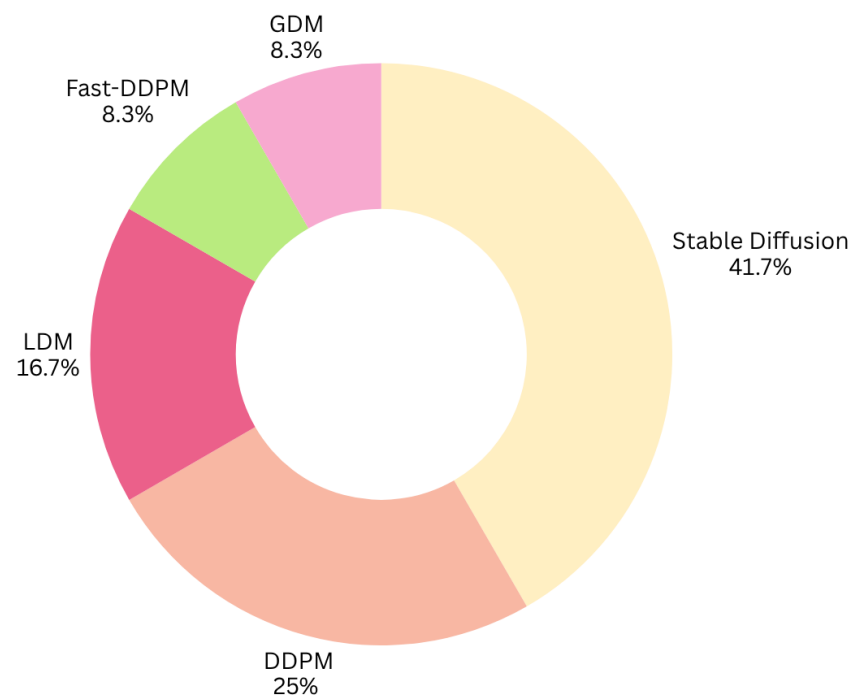


Figure 8. Diffusion Model Architectures Distribution.

Onakpojeruo et al. [60] introduced a denoising diffusion model (DDM) based framework designed for synthetic data generation in brain tumor classification. Leveraging a denoising Diffusion Model (DDM) and a deep convolutional neural network for synthetic image generation and brain tumor classification, the method aims to enhance the data diversity and improve classification performance. Moshe et al. [61] presented a diffusion-based framework for brain tumor MRI synthesis and classification generalization, employing Denoising Diffusion Probabilistic Models (DDPM) and mutual information (MI) regularization to generate realistic and diverse synthetic MR images. Kebali et al. [62] proposed a slice-based latent diffusion framework for 3D multimodal medical image and mask synthesis. The proposed approach enables simultaneous generation of volumetric images and corresponding multi-label tumor masks through a diffusion-driven generative process, providing enhanced structural consistency and diversity for downstream tumor segmentation tasks. Similarly, Mirowski et al. [63] introduced a diffusion-based framework for brain lesion segmentation, which leverages ControlNet [64] and a custom-designed Diffusion Model to synthesize realistic MRI brain images. In the proposed framework, a diffusion-based generative network is employed to synthesize realistic MRI brain images for lesion segmentation, and a U-Net segmentation network is utilized to learn robust feature representations.

Several methods have been proposed to reduce computational cost and training time. For instance, Jiang et al. [33] introduced Fast-DDPM with the aim of reducing

training for realistic image-to-image generation. The author designed a noise scheduler by incorporating uniform and non-uniform sampling. The results show that the model achieved a high PSNR and SSIM across different tasks. Similarly, Zhang et al. [65] proposed a GDM, a geodesic path to generate medical images. The model in this paper reduced 50× and 10× training times compared with DDPM and Fast-DDPM, respectively.

3.3. Autoencoder in Medical Image Synthesis

Numerous researchers have employed Autoencoders to reconstruct and generate medical images. Diamantis et al. [27] proposed an EndoVAE method, which utilized Variational Autoencoders to generate endoscopic images. The model was evaluated on the KID dataset and generated images with improved diversity and quality, enhancing the performance of the endoscopic classifier. Figure 9 provides an overview of the distribution of various Autoencoders used in the selected studies.

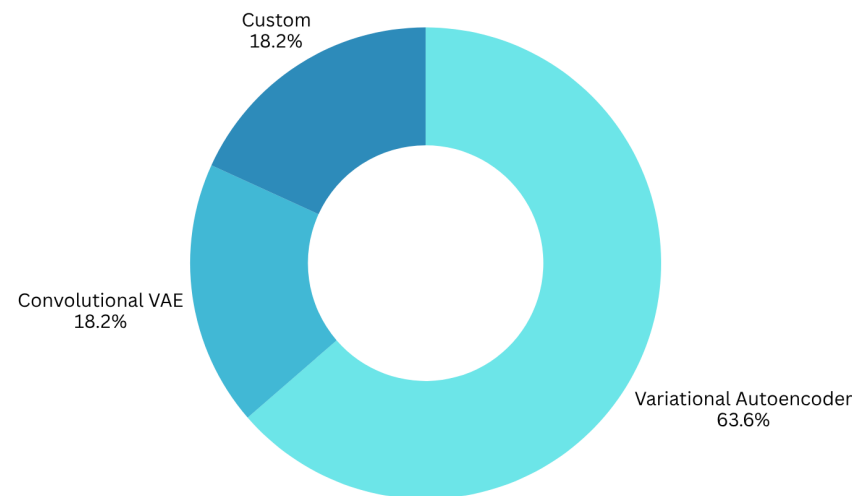


Figure 9. Distribution of Autoencoder types.

Fuente et al. [66] introduced a class-specific image generation Variational Autoencoder method with the aim of generating realistic images for classification using Variational Autoencoders. The proposed method increased accuracy by 18% and improved performance on imbalanced datasets. Diamantis et al. [28] proposed TIDE, a unique VAE that generates diverse and high-quality images, incorporating residual connections. TIDE was evaluated on the KID and Kvasir-Capsule datasets and achieved AUCs of 0.894 and 0.802, respectively. Jaen et al. [67] combined two techniques: style transfer and convolutional VAE to produce cardiac MR images. The combined technique enhances image quality by reducing blurring effects. Hangaragi et al. [68] presented a method to generate lung cancer images by using CAE. The model learns a latent representation for generating realistic chest images, improving classification and augmentation. Huang et al. [69] introduced a L-VAE method, a unique Variational Autoencoder for lumbar spine generation. Rais et al. [32] introduced CAT-VAE, a combination of a Cross-Attention Transformer and a Variational Autoencoder to enhance image quality and capture both local and global details. Their findings showed that CAT-VAE achieved an accuracy of 97.50%, which outperformed in classification tasks. Similarly, Dorent et al. [70] proposed MHVAE, a unique multimodal hierarchical VAE for MR/iUS synthesis, outperforming existing models in SSIM, PSNR, and LPIPS. Gan and Wang et al. [71] introduced adversarial learned VAE for esophageal OCT synthesis. It employed GAN and VAE to capture tissue details and reduce noise, resulting in higher image quality than existing models. Choudhury et al. [72] introduced a new method for TTE to CMR image synthesis. The method utilizes an Autoencoder with

a vision transformer to generate CMR from TTE by learning patterns from Transthoracic Echocardiography (TTE) data.

3.4. Hybrid Generative Model in Medical Image Synthesis

There are several studies in the medical imaging domain that explore the use of Variational Autoencoders and GANs for image synthesis. Liang et al. [73] achieved improved quantitative results in thyroid ultrasound image generation using an alpha-WGAN-GP framework. Furthermore, the framework also trains a UNet to achieve better segmentation results. Yang et al. [74] presented a hybrid GAN-guided conditional diffusion for enhanced CT images. GANs-CDM use image-based conditioning to guide the Diffusion Model for refined image synthesis. Experimental results show that images generated by GANs-CDM can simulate realistic contrast-enhanced CT images and anatomical structure for conditional medical image synthesis. In [75–77], the authors combined Variational Autoencoders and GANs for image synthesis. Rguibi et al. [75] improved feature learning and image generation quality by integrating GANs with VAE. In another study, Sa et al. [77] used a VAE-GAN for knee MRI image synthesis using hybrid generative models, improving FID and MSE for image synthesis. Feihong et al. [76] used VAE in the GANs generator for ASL image synthesis. A detailed experiment of dementia disease diagnosis is provided, along with a comprehensive comparison of multiple models. Cao et al. [78] designed a multimodal framework for generating modality-specific medical images, namely AE-GAN, achieving high PSNR and SSIM in both generation and segmentation tasks. A detailed experiment of various models on three different datasets is provided, along with an analysis of the effect of the hybrid loss function. Afnan et al. [79] integrated UNet, VAE, and CycleGAN for brain CT and MRI images using bidirectional learning, namely VAE-CycleUNet. Performance metrics including MAE, PSNR, and SSIM were employed, assessing the model on brain MRI and CT dataset. In a related study, Fatima et al. [80] combined GAN with Autoencoder for lung ultrasound image synthesis using a conditional latent space to guide GAN training.

3.5. Emerging Trends in Generative Medical Imaging

Numerous researchers have explored emerging trends in generative medical imaging and discussed three prominent directions, including Latent Diffusion Models (LDMs) in healthcare, multimodal Generative AI for text-to-medical image synthesis, and privacy-preserving generative frameworks.

3.5.1. Latent Diffusion Models in Healthcare

Latent Diffusion Models (LDM) [81] are applied in healthcare for image synthesis. Unlike conventional Diffusion Models, which operate directly on high-resolution image pixels, LDMs perform the forward and reverse diffusion process within a compressed latent space. In another study, Wang et al. [82] presented a unified text-to-medical image generation framework (MINIM). A latent Stable Diffusion Model was used to generate OCT, fundus, chest CT, and chest X-ray images. Performance metrics including FID, MS-SSIM, and IS were employed, achieving an FID of 65.3 for OCT. Similarly, Martyniak et al. [83] introduced a multi-stage pipeline, which utilized stable diffusion and Low-Rank Adaptation for realistic synthetic surgical endoscopic image generation. This method was used to capture complex details and prevent the gap between real and generated data. Kim and Park [84] designed a Multiple Switchable Spatially Adaptive Normalization, denoted as MS-SPADE, which enhances and integrates into a latent Diffusion Model-based framework for 3D multimodal medical image-to-image translation, addressing the challenge of volumetric cross-modality synthesis. Mahdi et al. [85] provided a stable and high-accuracy 3D latent diffusion framework for MR-only radiotherapy, focusing on robust

MRI-to-CT synthesis for clinically reliable synthetic CT generation. Kui et al. [86] proposed a novel medical latent variational Diffusion Model, denoted as Med-LVDM, designed for efficient medical image translation in multimodal and cross-modality settings. Specifically, a variational diffusion formulation was introduced based on weighted mean squared error to enable a more stable and interpretable optimization of the image translation process. Machacek et al. [87] proposed a mask-conditional latent Diffusion Model for gastrointestinal (GI) polyp image synthesis, where the generation process is guided by segmentation masks. Applying LDM-based methods to medical image generation has been extensively explored to enhance data diversity and improve model generalization. Existing studies suggest that latent Diffusion Models can learn complex anatomical representations and generate realistic volumetric medical images that preserve structural fidelity, which can strengthen downstream learning tasks when such synthetic data are integrated into training pipelines.

3.5.2. Multimodal Generative AI: Text-to-Medical Image

The application of text-to-image synthesis is an emerging research direction in medical imaging. Recent research on the advancement of diffusion-based generative techniques such as Stable Diffusion and text-guided latent diffusion frameworks has led to increased interest in text-to-image synthesis. Text-conditional MR image generation was investigated by Kim et al. [88] using magnetic resonance images, enabling the generation of images directly from medical text descriptions. Furthermore, cross-attention map analysis was performed to interpret textual conditions and generated image features. Xu et al. [89] designed a hierarchical text-guided diffusion framework, denoted as MedSyn, which enhances low-resolution text-conditioned synthesis into a high-resolution 3D lung CT image generation system, while Khader et al. [90] applied the DDPM method for the synthesis of 3D CT/MRI medical images. Huang et al. [91] designed a lightweight transformer-based diffusion framework, denoted as Chest-Diffusion, that transforms report-guided text representations into a text-to-chest X-ray generation model for authentic and computationally efficient medical image synthesis, improving the quality of generated images. Kidder et al. [9] employed DreamBooth with improved text-to-image prompts for the generation of MRI and mammography images. In another study, Li et al. [92] applied controllable Diffusion Models to improve generative data augmentation in medical imaging, integrating lesion masks and text descriptions to generate anatomically consistent and semantically diverse synthetic images. Anaya et al. [93] highlighted the importance of prompt-conditioned Diffusion Models in medical image generation, showing that a Stable Diffusion framework trained on chest X-ray data and guided by ClinicalBERT embeddings achieves clinically meaningful image synthesis, while Xie et al. [94] fine-tuned Stable Diffusion Models, denoted as MedDiff-FT, designed to generate images using guided prompts. Montoya et al. [10] fine-tuned a pretrained Stable Diffusion Model for mammographic imaging, which generates quality images through text prompts and latent space processing.

3.5.3. Privacy-Preserving Generative Models

Privacy regulations such as HIPAA and GDPR have strict restrictions on the acquisition, storage, and sharing of medical imaging data, which directly restrict access to large labeled datasets that are essential for training generative models used in clinical AI. Synthetic data generation aims to produce images that closely mimic real-world data distributions while ensuring privacy preservation. Several existing studies have explored privacy-preserving techniques in generative models for medical imaging, examining approaches such as differential privacy and federated learning [95–97]. The limitation of medical dataset availability was addressed by proposing “Privacy-Preserving Latent Diffusion-Based Synthetic Medical Image Generation”. A novel technique was introduced

that integrates PF-VGG, a privacy filter, which removes generated images that are similar to real images while maintaining data privacy [98]. A differentially private WGAN-based method for synthetic radiology image generation was implemented, enabling the creation of high-quality and privacy-preserving medical images to support deep learning diagnosis tasks across distributed healthcare institutions without sharing or storing patient data on a central server [99]. Deepee extends differentially private deep learning to medical imaging applications for image classification and segmentation tasks while ensuring privacy. In Deepee, the training process is designed not only to learn accurate diagnostic representations but also to protect patient information [100].

4. Loss Functions in GANs

GANs are the most utilized generative approach in the reviewed studies. GAN loss function is important in guiding the training process for image synthesis. Several researchers have investigated GAN loss functions to enhance the visual quality. The choice of loss function in GANs impacts model training and performance, guiding the network in capturing relevant features, reducing instability, mode collapse, and non-convergence by stabilizing adversarial learning [101]. It becomes possible to achieve stable training and generate higher-quality images by improving GAN loss functions. The purpose of the vanilla GAN loss function is to minimize the difference between the generated and real images and is defined as in Equation (1):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_d(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

GANs [2] have several common failure modes, such as mode collapse and vanishing gradients caused by minimizing JSD between the generated and true data distributions. Nowozin et al. [102] replaced the standard GAN objective with formulations based on divergence to measure the difference between real and generated probability distributions. The divergence is defined in Equation (2):

$$D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad (2)$$

Mao et al. [103] replaced the standard GAN sigmoid cross-entropy loss with a least squares loss function for the discriminator to reduce the vanishing gradient problem, named LSGAN. It is defined in Equations (3) and (4):

$$\begin{aligned} \min_D V_{\text{LSGAN}}(D) = & \frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}(x)} [(D(x) - y_{\text{real}})^2] \\ & + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - y_{\text{fake}})^2] \end{aligned} \quad (3)$$

$$\min_G V_{\text{LSGAN}}(G) = \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - y_{\text{gen}})^2] \quad (4)$$

To address the problems of vanishing gradients and mode collapse, Arjovsky et al. [47] introduced Wasserstein GAN, which uses Wasserstein distance to assess the difference between synthetic and real probability distributions. Unlike JS-divergence and KL-divergence, which fail to provide meaningful gradients for disjoint distributions, earth mover distance maintains meaningful gradient information, as defined in Equation (5):

$$W(P_{\text{data}}, P_{\text{gen}}) = \inf_{\gamma \in \Pi(P_{\text{data}}, P_{\text{gen}})} \mathbb{E}_{(u,v) \sim \gamma} [\|u - v\|] \quad (5)$$

Gulrajani et al. [104] introduced WGAN-GP, which used a different technique compared with the original WGAN. In WGAN, the authors faced Lipschitz constraint problems. To overcome this, WGAN-GP employs a gradient penalty in the loss function, optimizing the gradient penalty term, which is defined in Equation (6):

$$L = \mathbb{E}_{\tilde{x} \sim P_g}[D(\tilde{x})] - \mathbb{E}_{x \sim P_r}[D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (6)$$

Table 6 provides the advantages and limitations of loss functions commonly used in research.

Table 6. Pros and Cons of Commonly Used GAN Loss Functions.

Loss Function	Pros	Cons
Minimax GAN	Produce quality synthetic images; Easy to implement	Mode Collapse; Training Instability; Vanishing Gradient
WGAN	Improve training stability and reduced mode collapse	Weight Clipping; Computational Complexity; Sensitive Hyperparameters
WGAN-GP	Generates higher-fidelity images than WGAN; Improves training stability and convergence	Slow Convergence; Longer Processing Time
LSGAN	Generates quality images; Faster Convergence	Limited output diversity
EBGAN	Prevents mode collapse; Stabilizes training	Lower image quality: Autoencoder-dependent performance

Qi et al. [105] introduced a loss-sensitive GAN by incorporating Lipschitz regularity to deal with the limitations of unstable training and poor generalization in standard GAN, and it is defined in Equation (7):

$$\begin{aligned} \min_D V(D) = & \mathbb{E}_{x \sim P_{\text{data}}(x)}[L_{\theta}(x)] \\ & + \lambda \mathbb{E}_{x \sim P_{\text{data}}(x)}[\Delta(x, G(z)) + L_{\theta}(x) - L_{\theta}(G(z))] \end{aligned} \quad (7)$$

Li et al. [106] designed a GAN-based framework by incorporating a hybrid loss function, including sharpness, perceptual, adversarial, and structural similarity loss for low-dose CT images, named SSWGAN. This study used WGAN for denoising and showed that SSWGAN achieved PSNR of 35.6 and SSIM of 0.95, and it is represented in Equation (8):

$$\begin{aligned} L_{\text{SSWGAN}} = & \alpha L_{\text{WGAN}}(G, D) + \beta L_{\text{Perceptual}}(G) \\ & + \gamma L_{\text{Sharp}}(G) + \omega L_{\text{SSIM}}(G) \end{aligned} \quad (8)$$

The proposed hybrid loss function help to preserve sharpness, texture, and anatomical detail and maintains structural similarity between normal and denoised CT images. Chen et al. [107] introduced an ICycle-GAN with an aim to maintain CT and MRI image sharpness in generated images for the segmentation task, incorporating dual generators, dual discriminators, and a correction network. They employed an encoder and decoder

architecture as a correlation network to reduce feature blurring and preserve fine details. The ICycle-GAN loss function is defined in Equation (9):

$$\begin{aligned} L_{\text{GAN}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) = & L_{\text{GAN}}(G_{X \rightarrow Y}, D_Y, X, Y) \\ & + L_{\text{GAN}}(G_{Y \rightarrow X}, D_X, Y, X) \\ & + \lambda L_{\text{cyc}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}) \end{aligned} \quad (9)$$

In CT $\rightarrow$ MRI stage, the loss $L_{\text{GAN}}(G, D_Y, X, Y)$ evaluates the ability of the discriminator D_Y to distinguish real MRI images and generated MRI images translated from the CT modality. Similarly, in MRI $\rightarrow$ CT stage, the loss $L_{\text{GAN}}(G, D_X, Y, X)$ trains the discriminator D_X to differentiate real CT images from the synthesized CT image generated from the MRI modality. Heng et al. [108] introduced HLSNC-GAN by combining adversarial loss, identity loss, cycle-consistency loss, and hinge loss function for CT and MRI image synthesis. Furthermore, switchable normalization and the RMSprop optimizer are employed to improve the quality and diversity of the image. It is formulated in Equation (10):

$$\begin{aligned} L_{\text{total}}(G, F, D_1, D_2) = & L_{\text{adv}}(G, D_1) + L_{\text{adv}}(F, D_2) \\ & + \lambda \cdot L_{\text{cycle}}(G, F) + \lambda \cdot L_{\text{identity}}(G, F) \\ & + \lambda \cdot L_D + \lambda \cdot L_G \end{aligned} \quad (10)$$

Their proposed framework increased PSNR and SSIM, and eliminated inconsistencies between multiple modalities. Zheng et al. [109] introduced LFGAN for synthesizing CT liver tumors by reducing noise and distortion in images, while maintaining texture details. Moreover, GLCM is applied to measure correlation, contrast, and dissimilarity and employs reconstruction and hinge loss; it is formulated as stated in Equations (11) and (12):

$$\begin{aligned} L_{\text{GLCM}} = & \sum \left(|\text{CON}(f) - \text{CON}(x)| \right. \\ & + |\text{DIS}(f) - \text{DIS}(x)| \\ & \left. + |\text{COR}(f) - \text{COR}(x)| \right) \end{aligned} \quad (11)$$

$$\begin{aligned} \mathcal{L}_D = & -\mathbb{E}_{x \sim P_{\text{real}}} [\min(0, -1 + D(x))] \\ & - \mathbb{E}_{\hat{x} \sim G(z)} [\min(0, -1 - D(\hat{x}))] \\ & + \mathcal{L}_{\text{recons}} + \mathcal{L}_{\text{GLCM}} \end{aligned} \quad (12)$$

In addition, the authors compared classification models, including MobileNetV2, SqueezeNet, ResNet34, ConvNeXt, and ShuffleNetV2, to assess their effectiveness using accuracy, sensitivity, specificity, and AUC. Gajera et al. [110] introduced a GAN-based network by incorporating the Charbonnier loss for CT-scan denoising, named CL-GAN, and it is formulated as in Equation (13):

$$L_{\text{CL}} = C(\hat{x}, x) = \sum_i \sum_j \sqrt{(\hat{x}_{i,j} - x_{i,j})^2 + \epsilon^2} \quad (13)$$

Furthermore, the proposed hybrid loss function combined WGAN loss, perceptual loss, and Charbonnier loss, and is formulated in Equation (14):

$$\min_G \left\{ \lambda_1 \left[\max_D L_{\text{WGAN}}(D, G) \right] + \lambda_2 L_{\text{PL}}(G) + (1 - \lambda_1 - \lambda_2) L_{\text{CL}}(G) \right\} \quad (14)$$

The results demonstrate superior performance compared with baseline synthetic methods. Li and Jiang et al. [111] utilized SRGAN by combining Texture loss for super-resolution image synthesis, named TSRGAN. This texture loss used the Gram matrix, a statistical representation of feature correlations, that captures texture information. In addition, WGAN-GP was employed to enhance adversarial training. They combined perceptual loss, content loss, and texture loss to refine image quality, and it is formulated in Equation (15):

$$L_G = \lambda_{\text{per}} L_{\text{per}} + \lambda_{\text{tex}} L_{\text{tex}} + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{con}} L_{\text{con}} \quad (15)$$

According to their experiments, the TSRGAN method achieved promising results, with an average PSNR of 27.99 dB and an SSIM of 0.77. Similarly, Zhao et al. [112] combined a denoising CNN with SRGAN for super-resolution image reconstruction. By incorporating MSE loss, VGG loss, and WGAN loss, they achieved high-quality images while reducing noise in the generated outputs, and the total loss is formulated in Equation (16):

$$L_{SR} = L_{SR}^{\text{C-MSE}} + \lambda_1 L_{SR}^{\text{VGG}} + \lambda_2 L_{SR}^{\text{WGAN}} \quad (16)$$

The proposed texture loss is formulated in Equation (17):

$$L_{\text{tex}} = \left\| G(\phi(I^{\text{gen}})) - G(\phi(I^{\text{HR}})) \right\|_2^2 \quad (17)$$

Performance metrics, including SSIM and PSNR, were employed, yielding an SSIM of 0.9 and a PSNR of 30.4. Wang et al. [113] utilized an encoder–decoder architecture as a generator network to enhance image quality in a low-light environment and designed a multiscale feature extraction method to collect more informative details. The multiscale feature module comprises a convolution layer, max, and average pooling. In addition, a dual-discriminator is introduced comprising local and global discriminators to improve generator performance. By incorporating perceptual and color loss in the generator, the visual quality, perceptual similarity, and color correctness of the images are enhanced. It is formulated in Equation (18):

$$\text{Loss} = L_{\text{GAN}} + L_{\text{per}} + L_{\text{color}} \quad (18)$$

Akhmedova and Korber [29] presented GANetic loss for improving GAN training using Genetic programming, which are employed as an optimization problem. The loss function in their approach consists of Wasserstein, Least Squares, Binary Cross-Entropy, and Hinge loss, improving image generation and anomaly detection tasks. The proposed loss is formulated in Equation (19):

$$\mathcal{L}_{\text{GANetic}} = \frac{1}{N} \sum_{i=1}^N \left(\left(y_{\text{pred}}^{(i)} \right)^3 + \sqrt{\frac{\alpha y_{\text{real}}^{(i)}}{y_{\text{pred}}^{(i)} + \epsilon}} + \epsilon \right) \quad (19)$$

The model was trained on the CIFAR-10 and MNIST datasets using the DCGAN architecture. Furthermore, it is validated on Hyper-Kvasir, BreCaHAD, and LAG for image synthesis. Xiao et al. [114] presented a Monte Carlo GAN (MCGAN) algorithm to improve generator training. Regression loss is introduced to minimize MSE between real and generated samples. The performance of MCGAN is evaluated on diverse data types, including image generation, video generation, and time series data, and demonstrates

superior performance in both conditional and unconditional training. It is formulated in Equation (20):

$$\mathcal{L}_R(\theta, \phi) = \mathbb{E}_{(a,b) \sim \mu} \left[D_\phi(a) - \mathbb{E}_{\hat{a} \sim v_\theta(b)} [D_\phi(\hat{a})] \right]^2 \quad (20)$$

where μ denotes the true data distribution, v_θ represents the generated data distribution and D_ϕ is the discriminator. Guo et al. [115] designed an adaptive GAN to address the complexity in generating RGB images and employed Wasserstein loss to measure convergence, named MedGAN. Their findings show that the MedGAN model, which is trained adaptively, achieves better visual quality and faster training of generated images, outperforming WGAN. Gonzalez et al. [116] designed an evolutionary technique by identifying optimal parameters to improve the quality of generated images and mode collapse in GANs, named TaylorGAN. Jin et al. [117] designed a model for accurate and high-visual-quality image synthesis, named HMS-MambaGAN. The author initially trained HMS-MambaGAN using multimodal datasets and simultaneously employed a novel Gray-Level Gradient Co-occurrence Matrix (GLGCM)-based loss function to enhance the textures and realism of the images. Experimental results demonstrate better performance compared with existing medical image synthesis models and are formulated in Equation (21):

$$\begin{aligned} L_{G_1} &= L_{\text{GAN}} + \lambda_1 L_{L1} + L_{\text{HisGDL}} + L_{\text{GLGCM}}, \\ L_{G_2} &= L_{\text{GAN}} + \lambda_1 L_{L1} + L_{\text{HisGDL}}, \\ L_G &= L_{G_1} + L_{G_2}. \end{aligned} \quad (21)$$

Table 7 provides a review of loss functions employed in GAN-based image generation.

Table 7. Summary of Loss Functions.

Category	Purpose	Loss Functions
Adversarial Loss [2,108,109]	Improves the realism of generated samples through adversarial training	Vanilla GAN, WGAN, WGAN-GP, LSGAN, Hinge Loss
Perceptual Loss [106,110–113]	Preserves high-level semantic features and visual textures	VGG Perceptual Loss, Perceptual Loss
Structural Loss [106–108]	Maintains anatomical structures and spatial relationships	SSIM Loss, Cycle Consistency Loss
Color Loss [108,113]	Ensures accurate color representation in generated images	Color Loss, Identity Loss
Textural Loss [109,111,117]	Captures fine-grained textural features in generated images	GLCM Loss, GLGCM Loss, Texture Loss
Pixel-Level Loss [110–112]	Minimizes pixel-level differences between images	L1 Loss, L2 Loss, Charbonnier Loss
Edge Loss [106,117]	Preserves image boundaries and fine details	Sharpness Loss, Gradient Loss
Content Loss [111,112]	Preserves important content-related features	L1 Loss, L2 Loss

Goceri [118] employed a conditional GAN to augment dermoscopy images for classification tasks. A hybrid loss function, including content loss, L1 loss, and structural similarity loss, was introduced to improve the generator's ability, improving classification accuracy to 93.12%. Veiner et al. [119] introduced a unifying α -parametrized generator loss function

to address the unstable training in GANs. This model employed a canonical discriminator loss and a generator loss based on a function L_α . Under an optimal discriminator, the generator's objective reduces to minimizing a Jensen– f_α divergence, which generalizes the Jensen–Shannon divergence.

5. Discussion

This review examines the recent developments in medical image generation using generative models. The generative frameworks, including Diffusion Models, GANs, and VAEs, various GAN loss functions, medical image modalities, and datasets were examined.

5.1. Generative Imaging Modalities

Generative models (VAE, GAN, and Diffusion Models) have been applied to various imaging modalities and are capable of producing high-resolution, realistic images for classification, diagnosis, and segmentation [6,26,30,38,43]. In our analysis, most studies applied GANs to enhance and generate diverse medical datasets. Additionally, generative models have been applied across multiple imaging modalities to enhance low-dose CT and MRI scans, generate realistic chest X-rays, and improve endoscopic and dermoscopic images, providing a consistent framework for evaluation and comparison [27,32,67,68].

A large number of studies have been published in the domain of gastrointestinal endoscopy, focusing on high-quality image synthesis, lesion detection, and classification. Endoscopy procedures provide high-resolution visualization of mucosal surfaces, lesions, and polyps for detection. Using multi-model endoscopy imaging, assessment is performed by integrating information from multiple imaging sources, which can introduce systematic errors due to differences in acquisition, lighting, and imaging angle. Several researchers have employed generative models, including cGAN, StyleGAN, Pix2Pix, BlobGAN, VAE-GAN, and Diffusion Models for endoscopy image synthesis [12,24–29].

MRI and CT are used in radiotherapy for treatment planning. MRI is a widely used imaging modality applied in brain, cancer, pelvis, and soft tissue imaging, and existing techniques fail to restore fine details. Numerous researchers have investigated MRI image synthesis to achieve high-quality MRI scans for better diagnosis [8,32–34,77]. Generative models can produce diverse MRI images, improving image quality, reducing noise, and enhancing resolution. Various generative models have been employed for brain MRI synthesis; for example, Feng et al. [39] employed a deep CNN as a generator, while Dorent et al. introduced a multimodal hierarchical VAE. Various CycleGAN-based imaging techniques have been proposed for transforming CT images into MRI and other imaging modalities, such as ultrasound, Positron Emission Tomography (PET), Optical Coherence Tomography (OCT), and cervical imaging, where images are synthesized by learning the cross-modality [5,67].

The findings suggest that generative models, particularly GAN-based methods, have higher fidelity and realism than Diffusion Models and Variational Autoencoders on the same diverse medical imaging modalities. They found that generative-based methods were more accurate in improving low-dose and noisy scans and enabled realistic augmentation of chest X-rays, MRI, CT, and endoscopy images than diffusion-based methods, while requiring fewer computational resources and time.

5.2. Comparative Analysis

GANs and Diffusion Models provide complementary advantages in order to achieve high-quality and realistic images. GANs provide high-quality, realistic, and sharp image details due to the discriminator–generator training process, while Diffusion Models focus on the iterative denoising process, and the generated results improve structural consistency

and finer detail preservation. VAEs provide a structured probabilistic framework with a continuous latent space but produce lower image sharpness. Numerous researchers have applied generative models for image generation tasks, for instance, Refs. [67–69] have applied Variational Autoencoders to multiple medical imaging modalities, such as lung cancer, cardiac MRI, and lumbar spine image synthesis.

From a speed perspective, the authors of [120] compared GANs and Diffusion Models within a similar training setup. They generated 100,000 images for model training with the diffusion-based method and the GAN-based method. The Diffusion Model required approximately 1.5 days, whereas the GAN required approximately 10 min. From a stability perspective, GANs are sensitive to training dynamics due to their adversarial networks; although Diffusion Models provide improved stability, they involve more complex architectures and require longer processing times compared with GAN-based approaches. In contrast, VAEs tend to exhibit stable training, indicating a trade-off in image sharpness. Table 8 summarizes the evaluations of image quality, training stability, computational cost, output diversity, and controllability.

Table 8. Comparison of GANs, Diffusion Models and VAEs.

Comparison Criterion	GANs	Diffusion Models	VAEs
Image Realism	Produces high-fidelity images with enhanced realism through adversarial learning.	Achieves higher overall image quality and better preserves global anatomical structure.	VAEs produce quality images, but image sharpness can be affected due to pixel-wise reconstruction loss.
Training Stability	Unstable during training compared with VAEs, and can suffer from issues such as mode collapse and sensitivity to hyperparameters.	More stable than GANs during training, but requires many denoising steps for image generation.	Generally stable with good convergence behavior but may suffer from posterior collapse due to the loss formulation.
Computational Cost	Training GANs requires moderate computational cost due to adversarial learning.	High computational cost due to fundamental constraints in their multi-step sampling process required for image generation.	Requires moderate computational cost due to the encoder–decoder architecture and sampling from a latent space.
Output Diversity	Produces diverse and high-quality images but may suffer from mode collapse, leading to reduced diversity.	Produces diverse and high-quality images with strong representation of varied data features.	Generate diverse images but may struggle with capturing high diversity and rare patterns in complex datasets.

Diffusion-based models rely on an incremental denoising process that preserves structural similarity and texture details from Gaussian noise; however, they are computationally intensive and exhibit slower inference times. GAN-based architectures focus on one-shot generation of high-quality images. By employing an adversarial training process, GANs can generate realistic, high-quality, and sharp images; however, they fail to capture data diversity due to mode collapse and are sensitive to hyperparameters and network design. Within a domain of generative models for medical image synthesis, GANs stand out for their ability to generate sharp and visually realistic images, but Diffusion Models generally achieve higher image fidelity and improved structural consistency. In contrast, VAEs produce blurrier outputs and can struggle with capturing high variability in complex datasets. Lastly, in terms of computational cost, VAEs and GANs generally require

moderate resources due to their architectural and training requirements, while Diffusion Models are computationally very expensive and require substantial resources. GANs also maintain realistic anatomical characteristics and fine-grained details but introduce training instability and mode collapse. First, the GAN discriminator provides pixel-level feedback, which is essential for accurately representing mucosal surfaces. BlobGAN with self-attention achieved FID 15.5 for gastrointestinal image generation [25], and SCAGAN achieved FID 31.349 for wireless capsule endoscopy [5], representing the lowest FID scores across all models for the gastrointestinal imaging task in the reviewed literature. GANs showcase high computational efficiency capabilities, generating high-quality realistic images with low computational cost [120], which is key for large-scale data augmentation in resource-constrained settings. GANs can generate fewer synthetic tumor images for brain MRI, leading to reduced downstream performance, which requires more diverse and accurate data synthesis. Diffusion Models outperform GANs in SSIM. Ma et al. [8] used cross-attention and a multi-level U-Net for Brain MRI image synthesis, achieving an SSIM of 0.915, while Feng et al. [39] used knowledge distillation GANs for Brain MRI image synthesis and achieved an SSIM of 0.593. In contrast with GANs, Diffusion Models generate complex and rare anatomical images using an iterative denoising process. For cervical spine MRI synthesis [34], it learns complex patterns and integrates Gaussian–point noise, which improves inter-vertebral disc boundary preservation, while GANs may not fully capture these structures due to mode collapse. Stable Diffusion has also been widely used, and when fine-tuned with ControlNet, it produces clinically viable high-resolution ultrasound images. Table 9 presents advantages and limitations of generative models.

Table 9. Pros and Cons of Generative Models.

Architecture	Pros	Cons
GAN Adversarial generator-discriminator training	Produces sharp, high-quality, and realistic images; fast generation	Mode collapse and limited diversity; Sensitive to hyperparameters; training instability; Vanishing Gradients
Diffusion Models Iterative denoising forward and reverse process	Preserves structural and semantic details; captures diverse modes; stable training	High computational cost; complex architecture; longer training time; Slow inference
Variational Autoencoders Encoder–Decoder architecture	Stable training; structured latent space	Produce blurry images, risk of posterior collapse, and limited fine-grained detail generation.

The quantitative synthesis of FID, SSIM, and PSNR [121] values presented in Table 10 shows clear performance across generative model categories and target organs.

Table 10. Reported FID, SSIM, PSNR, and KID Values Across Imaging Modalities and Generative Models.

Modality	Model	Metric(s) Reported	Refs.
MRI			
Brain MRI	GAN	SSIM = 0.593	[39]
Brain MRI	Diffusion	SSIM = 0.915, PSNR = 36.16	[8]
Knee MRI	Diffusion	SSIM = 0.899, PSNR = 36.24	[8]
Alzheimer’s MRI	Diffusion	SSIM = 0.888, PSNR = 31.03	[8]
Cardiac MR (SR)	GAN	SSIM = 0.90, PSNR = 30.4	[112]

Table 10. Cont.

Modality	Model	Metric(s) Reported	Refs.
CT Imaging			
Abdominal CT	Diffusion	SSIM = 0.576, PSNR = 19.51	[35]
Low-dose CT	GAN	SSIM = 0.85, PSNR = 28.89	[106]
Endoscopy/GI			
Endoscopy (Kvasir)	GAN	SSIM = 0.83, PSNR = 31.86	[26]
Endoscopic	Diffusion	FID = 79.80	[83]
Endoscopy/GI	GAN	FID = 15.552	[25]
Capsule Endoscopy	GAN	FID = 31.349	[5]
Colonoscopy	GAN	FID = 42.19	[44]
WCE Capsule	VAE	FID = 81.9	[27]
Ultrasound			
Ultrasound	VAE	FID = 55.95	[122]
Wireless Capsule Endoscopy			
WCE Capsule	VAE	KID = 89.4	[28]

5.3. Limitations of GAN in Image Generation

Numerous researchers have implemented various approaches in order to address the challenges associated with GAN training, including non-convergence, vanishing gradient, mode collapse, and training instability.

5.3.1. Model Collapse

Mode collapse is one of the most important problems in GAN training. Mode collapse occurs in GANs when the generator produces the same type of images and fails to capture the diversity of the training data, while the discriminator can still distinguish between real and generated samples. As a result, improving the diversity of synthetic medical images remains a challenging task for researchers. This is a common issue in GAN-based medical image generation. This limitation reduces GAN use in applications such as MRI, CT, or endoscopy image synthesis. Several researchers have proposed loss functions and training strategies in order to reduce mode collapse [47,123]. Despite these efforts, mode collapse still occurs in many GAN-based medical image generation models.

5.3.2. Vanishing Gradient

A significant challenge in GAN training is that the generator receives a small gradient signal and fails to update its parameters when the discriminator is too strong. The generator updates parameters based on discriminator feedback during training. As a result, achieving stable and effective training of GANs remains a challenging task for researchers. This limitation reduces image quality, diversity, and impact on clinical applications. In order to prevent vanishing gradient, researchers have designed various methods, including loss function, gradient penalty, and normalization techniques [47,103,104].

5.3.3. Training Instability

GANs suffer from training instability due to adversarial training. When the generator and discriminator fail to converge, training instability arises. Additionally, hyperparameter selection, with its impact on learning rates, batch sizes, and optimizer settings, is critical when stable and effective GAN training is required. As a result, maintaining diversity in the generated images remains a significant challenge for researchers. Researchers have applied spectral normalization, gradient penalty, and other regularization techniques to improve the quality of the generated image. In order to overcome instability and

convergence issues, Arjovsky et al. [47] have proposed the Wasserstein GAN, and Mao et al. [103] introduced the Least Squares GAN. Miyato et al. [124] introduced a stabilizing framework with a spectral normalization technique that constrains the Lipschitz constant of the discriminator. In a related study, Zhao et al. [125] introduced a Differentiable Augmentation technique to address the instability due to data scarcity. The Differentiable Augmentation model applied a consistent set of differentiable transformations, including translation, color jittering, and cutout, to both real and generated samples. This approach delivered more stable convergence and improved generator performance than previously employed augmentation strategies.

5.4. Addressing the Research Questions

This systematic review addresses the four formulated research questions as follows:

- RQ1 (Generative Techniques vs. Modalities): This question is answered in Section 3 (Medical Image Synthesis with Generative Models). We have categorized all three generative architectures, such as GAN, VAE, and Diffusion Models, across MRI, CT, X-ray, endoscopic modalities, and other medical image modalities. A comprehensive breakdown of these techniques, including specific author contributions from 2021–2026.
- RQ2 (Loss Functions in GANs): The specific objective functions, adversarial losses, and hybrid loss formulations used to stabilize training and improve structural similarity or pixel-level accuracy are detailed in Section 4 (Loss Functions in GANs).
- RQ3 (Clinical Applications): The translation of these models into clinical settings such as gastrointestinal lesion detection, gastric cancer synthesis, and multimodal diagnostics is explored throughout the subsections of Section 3. Lee et al. [45] deployed StyleGAN3, designed to synthesize gastric cancer images and further integrated with YOLOv8 detection, achieving 93.98% sensitivity in clinical-grade evaluation. Yoon et al. [44] applied GAN-generated colonoscopy images to improve detection performance. Wang et al. [82] proposed MINIM, a unified text-to-medical-image framework evaluated across OCT, fundus, chest CT, and chest X-ray in clinical imaging contexts. Shi et al. [98] designed a privacy-preserving filter with the aim of addressing clinical deployment constraints in real-world medical settings. Diffusion-based augmentation exhibits high-quality image generation performance for mammography (Montoya et al. [10]) and lung CT (Krishna et al. [31]), demonstrating its applicability to diagnostic decision support. Despite growing adoption of medical image synthesis approaches, a fundamental gap remains between automated image quality metrics and clinically meaningful radiological assessment, and their ability to match real clinical evaluation standards is limited. This is an important limitation, as commonly used evaluation metrics such as FID, SSIM, and PSNR measure statistical or pixel-level similarity between generated and real images rather than true clinical diagnostic value. Lee et al. [45] validated a gastric cancer generated images using YOLOv8 detection pipeline. The results demonstrate that StyleGAN3-generated images achieve 93.98% sensitivity for downstream clinical task performance. Similarly, Yoon et al. [44] validated GAN-generated colonoscopy images through detection performance rather than radiologist scoring. Wang et al. [82] evaluated MINIM across four clinical imaging modalities using AUROC as a downstream clinical metric, which is closer to diagnostic utility than FID alone.
- RQ4 (Overcoming Data Limitations): The methodologies for utilizing synthetic data are addressed in Section 2 (Datasets Overview, Table 4). GAN-based methods are used to generate synthetic images from small and imbalanced training datasets (Saimon et al. [53], Ding et al. [126], Feng et al. [39]). Transfer learning and fine-tuning of

pretrained models on small medical datasets was employed in diffusion-based studies, including DreamBooth fine-tuning (Kidder et al. [9]) and Stable Diffusion fine-tuning (Montoya et al. [10]). Multimodal image synthesis was used to generate missing imaging modalities from available datasets, reducing the need for acquiring additional data (Wang et al. [26], Afnan et al. [79]).

6. Limitations

Despite providing a comprehensive comparative synthesis of GANs, Diffusion Models, Variational Autoencoders, and Hybrid architectures for synthetic medical image generation, this review still has several limitations. The review protocol was not pre-registered in PROSPERO or a similar public registry, which may introduce a minor risk of bias. The included studies are representative but demonstrate relatively limited diversity in dataset size and imaging modality. In addition, most included studies only reported clinical findings based on a single automated image quality metric (eg, FID, SSIM, PSNR) compared with evaluations performed by radiologists, without further robust downstream clinical validation, which limits the strength of the conclusions. These limitations should be considered when interpreting the synthesized findings.

7. Conclusions

This SLR provides a comprehensive overview of Generative AI models for medical image generation and further examines GAN loss functions for producing high-quality images. Researchers have incorporated these models for synthesizing rare medical images, the complexity of medical structures and the scarcity of data are still major challenges. While the Diffusion Models can generate high-quality images, they are computationally very expensive, and VAEs face challenges in generating complex medical images. However, GANs offer faster image generation and lower computational requirements while maintaining high visual quality. This review provides an in-depth analysis of various generative models and GAN loss functions commonly used in the domain of medical image generation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics16142166/s1>, Table S1: Full study characteristics and generative models grouped by application domain. File S1: PRISMA 2020 checklist.

Author Contributions: Conceptualization, Y.W.; methodology, Y.W. and M.N.N.; data curation, Y.W.; formal analysis, Y.W. and I.A.; visualization, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, M.N.N. and I.A.; supervision, M.N.N.; validation, M.N.N. and I.A.; resources, I.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The study was conducted independently, without financial support from governmental, commercial, or non-profit funding agencies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This systematic review did not generate or analyze any new datasets.

Acknowledgments: We are thankful to everyone who participated in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GAN	Generative Adversarial Network
VAE	Variational Autoencoder
DDPM	Denoising Diffusion Probabilistic Model
LDM	Latent Diffusion Model
FID	Fréchet Inception Distance
SSIM	Structural Similarity Index Measure
PSNR	Peak Signal-to-Noise Ratio
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
PET	Positron Emission Tomography
OCT	Optical Coherence Tomography
WCE	Wireless Capsule Endoscopy
SLR	Systematic Literature Review
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WGAN	Wasserstein GAN
LSGAN	Least Squares GAN
CNN	Convolutional Neural Network
SRGAN	Super-Resolution GAN

References

- Kingma, D.P.; Welling, M. An introduction to variational autoencoders. *Found. Trends Mach. Learn.* **2019**, *12*, 307–392. [[CrossRef](#)]
- Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 139–144.
- Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
- Guo, P.; Zhao, C.; Yang, D.; Xu, Z.; Nath, V.; Tang, Y.; Simon, B.; Belue, M.; Harmon, S.; Turkbey, B.; et al. Maisi: Medical AI for synthetic imaging. In *Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: Piscataway, NJ, USA, 2025; pp. 4430–4441.
- Xiao, Z.; Zhang, D.; Chen, X.; Li, D. SCAGAN: Wireless capsule endoscopy lesion image generation model based on GAN. *Electronics* **2025**, *14*, 428. [[CrossRef](#)]
- Ali, M.; Ali, M.; Hussain, M.; Koundal, D. Generative adversarial networks (GANs) for medical image processing: Recent advancements. *Arch. Comput. Methods Eng.* **2025**, *32*, 1185–1198.
- Johari, M.F.; Chiew, K.L.; Tan, C.C.L. Generative Models in Synthetic 2D Medical Image Generation: A Systematic Review. *SSRN* **2025**. [[CrossRef](#)]
- Ma, J.; Jian, G.; Chen, J. Diffusion Model-Based MRI Super-Resolution Synthesis. *Int. J. Imaging Syst. Technol.* **2025**, *35*, e70021. [[CrossRef](#)]
- Kidder, B.L. Advanced image generation for cancer using diffusion models. *Biol. Methods Protoc.* **2024**, *9*, bpae062. [[CrossRef](#)] [[PubMed](#)]
- Montoya-del Angel, R.; Sam-Millan, K.; Vilanova, J.C.; Martí, R. MAM-E: Mammographic synthetic image generation with diffusion models. *Sensors* **2024**, *24*, 2076. [[CrossRef](#)] [[PubMed](#)]
- Magalhães, B.; Neto, A.; Almeida, E.; Libânio, D.; Chaves, J.; Dinis-Ribeiro, M.; Coimbra, M.; Cunha, A. Using generative adversarial networks for endoscopic image augmentation of stomach precancerous lesions. *Procedia Comput. Sci.* **2025**, *256*, 1092–1099. [[CrossRef](#)]
- Janutėnas, L.; Šešok, D. Perspective transformation and viewpoint attention enhancement for generative adversarial networks in endoscopic image augmentation. *Appl. Sci.* **2025**, *15*, 5655. [[CrossRef](#)]
- Li, J.; Zhu, G.; Hua, C.; Feng, M.; Bennamoun, B.; Li, P.; Lu, X.; Song, J.; Shen, P.; Xu, X.; et al. A systematic collection of medical image datasets for deep learning. *ACM Comput. Surv.* **2023**, *56*, 116. [[CrossRef](#)]
- Skandarani, Y.; Jodoin, P.M.; Lalande, A. Gans for medical image synthesis: An empirical study. *J. Imaging* **2023**, *9*, 69. [[CrossRef](#)] [[PubMed](#)]
- Akpınar, M.H.; Sengur, A.; Salvi, M.; Seoni, S.; Faust, O.; Mir, H.; Molinari, F.; Acharya, U.R. Synthetic data generation via generative adversarial networks in healthcare: A systematic review of image-and signal-based studies. *IEEE Open J. Eng. Med. Biol.* **2024**, *6*, 183–192. [[PubMed](#)]

16. Azad, M.; Fahad, N.M.; Raiaan, M.A.K.; Anik, T.R.; Khan, M.F.K.; Toyé, H.M.K.; Muhammad, G. A Systematic Review of Diffusion Models for Medical Image-Based Diagnosis: Methods, Taxonomies, Clinical Integration, Explainability, and Future Directions. *Diagnostics* **2026**, *16*, 211. [[CrossRef](#)] [[PubMed](#)]
17. Goyal, M.; Mahmoud, Q.H. A systematic review of synthetic data generation techniques using generative AI. *Electronics* **2024**, *13*, 3509. [[CrossRef](#)]
18. Sordo, Z.; Chagnon, E.; Hu, Z.; Donatelli, J.J.; Andeer, P.; Nico, P.S.; Northen, T.; Ushizima, D. Synthetic scientific image generation with VAE, GAN, and diffusion model architectures. *J. Imaging* **2025**, *11*, 252. [[CrossRef](#)] [[PubMed](#)]
19. Oulmalme, C.; Nakouri, H.; Jaafar, F. A systematic review of generative AI approaches for medical image enhancement: Comparing GANs, transformers, and diffusion models. *Int. J. Med. Inform.* **2025**, *199*, 105903. [[CrossRef](#)] [[PubMed](#)]
20. Weerasinghe, S. Generative Models for Synthesis of Medical Imaging Data: A Review of VAEs, GANs, and the Hybrid VAE-GANs. In Proceedings of the 2nd International Conference on Transformative Applied Research, Homagama, Sri Lanka, 30 January 2026; p. 23.
21. Patel, S.; Makwana, A. A Review on Medical Image Generation Generative Adversarial Networks (GANs). In *Proceedings of the 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1266–1271.
22. Tavse, S.; Varadarajan, V.; Bachute, M.; Gite, S.; Kotecha, K. A systematic literature review on applications of GAN-synthesized images for brain MRI. *Future Internet* **2022**, *14*, 351. [[CrossRef](#)]
23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, L.; Jindal, B.; Alaa, A.; Weinreb, R.; Wilson, D.; Segal, E.; Zou, J.; Xie, P. Generative AI enables medical image segmentation in ultra low-data regimes. *Nat. Commun.* **2025**, *16*, 6486. [[CrossRef](#)] [[PubMed](#)]
25. Ayubi, A.; Utama, N.P. Enhanced Generation of Gastrointestinal Images for Data Augmentation Using a Modified BlobGAN. In *Proceedings of the 2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*; IEEE: Piscataway, NJ, USA, 2024; pp. 161–166.
26. Liu, W.; Bao, R.; Wang, Y.; Zhu, H.; Yang, H.; Zhang, J. Multi-scale adversarial generative network for super-resolution of endoscopic images. In *Proceedings of the 2025 Joint International Conference on Automation-Intelligence-Safety (ICAIS) & International Symposium on Autonomous Systems (ISAS)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
27. Diamantis, D.E.; Gatoula, P.; Iakovidis, D.K. Endovae: Generating endoscopic images with a variational autoencoder. In *Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
28. Diamantis, D.E.; Gatoula, P.; Koulaouzidis, A.; Iakovidis, D.K. This intestine does not exist: Multiscale residual variational autoencoder for realistic wireless capsule endoscopy image generation. *IEEE Access* **2024**, *12*, 25668–25683. [[CrossRef](#)]
29. Akhmedova, S.; Körber, N. Ganetic loss for generative adversarial networks with a focus on medical applications. *Neural Comput. Appl.* **2025**, *37*, 19177–19197. [[CrossRef](#)]
30. Khatun, A.; Yeter Aydeniz, K.; Weinstein, Y.S.; Usman, M. Quantum generative learning for high-resolution medical image generation. *Mach. Learn. Sci. Technol.* **2025**, *6*, 025032. [[CrossRef](#)]
31. Krishna, A.; Papenhausen, E.; Coelho, D.; Mueller, K. Synthesis of realistic medical images with pathologies using diffusion models with application to lung CT and mammography. In *Medical Imaging 2025: Physics of Medical Imaging*; SPIE: Bellingham, WA, USA, 2025; Volume 13405, pp. 533–538.
32. Rais, K.; Amroune, M.; Haouam, M.Y.; Benmachiche, A. Cat-vae: A cross-attention transformer-enhanced variational autoencoder for improved image synthesis. *Stat. Optim. Inf. Comput.* **2025**, *14*, 2534–2558. [[CrossRef](#)]
33. Jiang, H.; Imran, M.; Zhang, T.; Zhou, Y.; Liang, M.; Gong, K.; Shao, W. Fast-DDPM: Fast denoising diffusion probabilistic models for medical image-to-image generation. *IEEE J. Biomed. Health Inform.* **2025**, *29*, 7326–7335. [[CrossRef](#)] [[PubMed](#)]
34. Pan, E.; Zhong, Y.; Li, P.; Yang, Y.; Zhou, J. Hybrid-noise generative diffusion probabilistic model for cervical spine MRI image generation. *Comput. Methods Progr. Biomed.* **2025**, *262*, 108639. [[CrossRef](#)]
35. Wang, X.; He, D.; Zhang, B.; Hao, Y.; Yang, D.; Duan, Y. Conditional Diffusion Model for Abdominal CT Image Synthesis. In *Proceedings of the 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–5.
36. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
37. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
38. Maguluri, K.K.; Ganti, V.K.A.T.; Yasmeen, Z.; Pandugula, C. Progressive GAN framework for realistic chest X-ray synthesis and data augmentation. In *Proceedings of the 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 755–760.

39. Feng, Y.; Zhang, B.; Xiao, L.; Yang, Y.; Gegen, T.; Chen, Z. Enhancing medical imaging with gans synthesizing realistic images from limited data. In *Proceedings of the 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1192–1197.
40. Liu, Z.; Smith, P.; Lautin, A.; Zhou, J.; Yoo, M.; Sullivan, M.; Li, H.; Deyer, L.; Zhou, A.; Yang, A.; et al. RadImageGAN—A Multi-modal Dataset-Scale Generative AI for Medical Imaging. In *Proceedings of the International Workshop on Applications of Medical AI*, Marrakesh, Morocco, 6 October 2024; pp. 173–185.
41. Shah, V.H.; Mishra, A.K.; Chandra, N.; Kumar, S.; Dash, P.P. Low Resolution Medical Image Enhancement Using Generative Artificial Intelligence. In *Proceedings of the 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1813–1818.
42. Prabu, S.; Saravanan, R.; Gomathi, T.; Surendran, R.; Deepa, R. Generative Adversarial Networks for Endoscopic Image Classification. In *Proceedings of the 2025 International Conference on Computing Technologies & Data Communication (ICCTDC)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–7.
43. Mao, R.; Zhao, R.; Zhang, L. Low-Light Gastroscopic Image Enhancement Network Based on CycleGAN. In *Proceedings of the 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*; IEEE: Piscataway, NJ, USA, 2025; pp. 2370–2373.
44. Yoon, D.; Kong, H.J.; Kim, B.S.; Cho, W.S.; Lee, J.C.; Cho, M.; Lim, M.H.; Yang, S.Y.; Lim, S.H.; Lee, J.; et al. Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. *Sci. Rep.* **2022**, *12*, 261. [[CrossRef](#)] [[PubMed](#)]
45. Lee, G.P.; Kim, Y.J.; Kim, K.G. Generative models for high-resolution synthetic image of gastric cancer for improved gastric cancer detection. In *Medical Imaging 2025: Computer-Aided Diagnosis*; SPIE: Bellingham, WA, USA, 2025; Volume 13407, pp. 21–27.
46. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2016**, arXiv:1511.06434.
47. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In *Proceedings of the International Conference on Machine Learning*, Sydney, Australia, 6–11 August 2017; pp. 214–223.
48. Mukherjee, D.; Saha, P.; Kaplun, D.; Sinitca, A.; Sarkar, R. Brain tumor image generation using an aggregation of GAN models with style transfer. *Sci. Rep.* **2022**, *12*, 9141. [[CrossRef](#)] [[PubMed](#)]
49. Onakpojeruo, E.P.; Mustapha, M.T.; Ozsahin, D.U.; Ozsahin, I. Enhanced MRI-based brain tumour classification with a novel Pix2pix generative adversarial network augmentation framework. *Brain Commun.* **2024**, *6*, fcae372. [[CrossRef](#)] [[PubMed](#)]
50. Hamghalam, M.; Simpson, A.L. Medical image synthesis via conditional GANs: Application to segmenting brain tumours. *Comput. Biol. Med.* **2024**, *170*, 107982. [[CrossRef](#)] [[PubMed](#)]
51. Zhou, M.; Wagner, M.W.; Tabori, U.; Hawkins, C.; Ertl-Wagner, B.B.; Khalvati, F. Generating 3D brain tumor regions in MRI using vector-quantization Generative Adversarial Networks. *Comput. Biol. Med.* **2025**, *185*, 109502. [[CrossRef](#)] [[PubMed](#)]
52. Wang, M.; Chen, J.; Li, Q.; He, Q.; Mou, J.; Wang, W. SelSSM: Selective State Space Module Empowered Generative Adversarial Networks for Multi-modal MR Image Synthesis. *Inf. Fusion* **2026**, *135*, 104433. [[CrossRef](#)]
53. Saimon, A.M.; Abdullah, S.; Hossain, R.; Arefin, M.S.; Hussain, M.M. Improving Brain Tumor Detection Through Synthetic Data Augmentation using Conditional GANs. In *Proceedings of the 2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*; IEEE: Piscataway, NJ, USA, 2026; pp. 1029–1035.
54. Lai, M.; Mascalchi, M.; Tessa, C.; Diciotti, S. Generating Brain MRI with StyleGAN2-ADA: The Effect of the Training Set Size on the Quality of Synthetic Images. *J. Imaging Inform. Med.* **2025**, *39*, 2319–2329. [[CrossRef](#)] [[PubMed](#)]
55. Wang, W.; Xia, J.; Luo, G.; Dong, S.; Li, X.; Wen, J.; Li, S. Diffusion model for medical image denoising, reconstruction and translation. *Comput. Med. Imaging Graph.* **2025**, *124*, 102593. [[CrossRef](#)] [[PubMed](#)]
56. Hung, A.L.Y.; Zhao, K.; Zheng, H.; Yan, R.; Raman, S.S.; Terzopoulos, D.; Sung, K. Med-cDiff: Conditional medical image generation with diffusion models. *Bioengineering* **2023**, *10*, 1258. [[CrossRef](#)] [[PubMed](#)]
57. Liu, S.; Chen, Z.; Yang, Q.; Yu, W.; Dong, D.; Hu, J.; Yuan, Y. Polyp-gen: Realistic and diverse polyp image generation for endoscopic dataset expansion. In *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: Piscataway, NJ, USA, 2025; pp. 15776–15782.
58. Freiche, B.; El-Khoury, A.; Nasiri-Sarvi, A.; Hosseini, M.S.; Garcia, D.; Basarab, A.; Boily, M.; Rivaz, H. Ultrasound image generation using latent diffusion models. In *Proceedings of the Medical Imaging 2025: Ultrasonic Imaging and Tomography*; SPIE: Bellingham, WA, USA, 2025; Volume 13412, pp. 287–292.
59. Kaleta, J.; Dall’Alba, D.; Płotka, S.; Korzeniowski, P. Minimal data requirement for realistic endoscopic image generation with stable diffusion. *Int. J. Comput. Assist. Radiol. Surg.* **2024**, *19*, 531–539. [[PubMed](#)]
60. Onakpojeruo, E.P.; Ozsahin, D.U.; Ozsahin, I. Enhancing brain tumor classification with a diffusion denoising model and a conditional deep convolutional neural network. *Adv. Neurol.* **2025**, *4*, 88–100. [[CrossRef](#)]
61. Moshe, Y.H.; Teicher, M.; Artzi, M. Enhancing Brain Tumor Classification and Generalization Using DDPM-Generated MRI, Mutual Information and Ensemble Learning. *Technol. Cancer Res. Treat.* **2026**, *25*, 15330338251405180. [[CrossRef](#)] [[PubMed](#)]

62. Kebaili, A.; Lapuyade-Lahorgue, J.; Vera, P.; Ruan, S. Multi-modal MRI synthesis with conditional latent diffusion models for data augmentation in tumor segmentation. *Comput. Med. Imaging Graph.* **2025**, *123*, 102532. [[CrossRef](#)] [[PubMed](#)]
63. Mirowski, P.; Fabijańska, A. Diffusion model-based synthesis of brain images for data augmentation. *Biomed. Signal Process. Control* **2026**, *113*, 108940. [[CrossRef](#)]
64. Zhang, L.; Rao, A.; Agrawala, M. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 3836–3847.
65. Zhang, T.; Jiang, H.; Gong, K.; Shao, W. Geodesic diffusion models for medical image-to-image generation. *arXiv* **2025**, arXiv:2503.00745.
66. De La Fuente, N.; Majó, M.; Luzko, I.; Córdova, H.; Fernández-Esparrach, G.; Bernal, J. Enhancing image classification in small and unbalanced datasets through synthetic data augmentation. In Proceedings of the Workshop on Clinical Image-Based Procedures, Marrakesh, Morocco, 6 October 2024; pp. 11–21.
67. Jaén-Lorites, J.M.; Pérez-Pelegrí, M.; Laparra, V.; López-Lereu, M.P.; Monmeneu, J.V.; Maceira, A.M.; Moratal, D. Synthetic Generation of Cardiac MR Images Combining Convolutional Variational Autoencoders and Style Transfer. In *Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*; IEEE: Piscataway, NJ, USA, 2022; pp. 2084–2087.
68. Hangaragi, S.; Neelima, N.; Venugopal, V.; Ganguly, S.; Mudi, J.; Choi, J.H. CAE SynthImgGen: Revolutionizing cancer diagnosis with convolutional autoencoder-based synthetic image generation. *Alex. Eng. J.* **2025**, *115*, 343–354. [[CrossRef](#)]
69. Huang, K.; Zhang, J. Three-dimensional lumbar spine generation using variational autoencoder. *Med. Eng. Phys.* **2023**, *120*, 104046. [[CrossRef](#)] [[PubMed](#)]
70. Dorent, R.; Haouchine, N.; Kogl, F.; Joutard, S.; Juvekar, P.; Torio, E.; Golby, A.J.; Ourselin, S.; Frisken, S.; Vercauteren, T.; et al. Unified brain MR-ultrasound synthesis using multi-modal hierarchical representations. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; pp. 448–458.
71. Gan, M.; Wang, C. Esophageal optical coherence tomography image synthesis using an adversarially learned variational autoencoder. *Biomed. Opt. Express* **2022**, *13*, 1188–1201. [[CrossRef](#)] [[PubMed](#)]
72. Choudhury, Z.Z.; McCane, B.; Coffey, S. Medical image synthesis using autoencoder with vision transformer. In *Proceedings of the 2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
73. Liang, J.; Chen, J. Data augmentation of thyroid ultrasound images using generative adversarial network. In *Proceedings of the 2021 IEEE International Ultrasonics Symposium (IUS)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4.
74. Yang, Y.; Liu, J.; Chen, Q.; Li, Y.; Han, X.H.; Hu, H.; Lin, L.; Chen, Y.W. GANs-guided conditional diffusion model for synthesizing contrast-enhanced computed tomography images. In *Proceedings of the 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–4.
75. Rguibi, Z.; Hajami, A.; Zitouni, D.; Maleh, Y.; Elqaraoui, A. Medical variational autoencoder and generative adversarial network for medical imaging. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *32*, 494–505. [[CrossRef](#)]
76. Li, F.; Huang, W.; Luo, M.; Zhang, P.; Zha, Y. A new VAE-GAN model to synthesize arterial spin labeling images from structural MRI. *Displays* **2021**, *70*, 102079. [[CrossRef](#)]
77. SA, R.; Babu, B.S. Synthesizing Realistic Knee MRI Images: A VAEGAN Approach for Enhanced Medical Data Augmentation. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 792–799. [[CrossRef](#)]
78. Cao, B.; Bi, Z.; Hu, Q.; Zhang, H.; Wang, N.; Gao, X.; Shen, D. Autoencoder-driven multimodal collaborative learning for medical image synthesis. *Int. J. Comput. Vis.* **2023**, *131*, 1995–2014. [[CrossRef](#)]
79. Afnaan, K.; Singh, T.; Duraisamy, P. Hybrid deep learning framework for bidirectional medical image synthesis. In *Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
80. Fatima, N.; Mento, F.; Afrakhteh, S.; Perrone, T.; Smargiassi, A.; Inchingolo, R.; Demi, L. Synthetic lung ultrasound data generation using autoencoder with generative adversarial network. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2025**, *72*, 624–635. [[CrossRef](#)] [[PubMed](#)]
81. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
82. Wang, J.; Wang, K.; Yu, Y.; Lu, Y.; Xiao, W.; Sun, Z.; Liu, F.; Zou, Z.; Gao, Y.; Yang, L.; et al. Self-improving generative foundation model for synthetic medical image generation and clinical applications. *Nat. Med.* **2025**, *31*, 609–617. [[PubMed](#)]
83. Martyniak, S.; Kaleta, J.; Dall’Alba, D.; Naskręt, M.; Płotka, S.; Korzeniowski, P. Simuscope: Realistic endoscopic synthetic dataset generation through surgical simulation and diffusion models. In *Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: Piscataway, NJ, USA, 2025; pp. 4268–4278.

84. Kim, J.; Park, H. Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2024; pp. 7604–7613.
85. Mahdi, M.A.; Al-Shalabi, M.; Alnfwawy, E.T.; Elbarougy, R.; Hadi, M.U.; Ali, R.F. 3D Latent Diffusion Model for MR-Only Radiotherapy: Accurate and Consistent Synthetic CT Generation. *Diagnostics* **2025**, *15*, 3010. [[CrossRef](#)] [[PubMed](#)]
86. Kui, X.; Liu, B.; Sun, Z.; Li, Q.; Zhang, M.; Liang, W.; Zou, B. Med-LVDM: Medical latent variational diffusion model for medical image translation. *Biomed. Signal Process. Control* **2025**, *106*, 107735. [[CrossRef](#)]
87. Macháček, R.; Mozaffari, L.; Sepasdar, Z.; Parasa, S.; Halvorsen, P.; Riegler, M.A.; Thambawita, V. Mask-conditioned latent diffusion for generating gastrointestinal polyp images. In Proceedings of the 4th ACM Workshop on Intelligent Cross-Data Analysis and Retrieval, Thessaloniki, Greece, 12–15 June 2023; pp. 1–9.
88. Kim, K.; Na, Y.; Ye, S.J.; Lee, J.; Ahn, S.S.; Park, J.E.; Kim, H. Controllable Text-to-Image Synthesis for Multi-Modality MR Images. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 7936–7945.
89. Xu, Y.; Sun, L.; Peng, W.; Jia, S.; Morrison, K.; Perer, A.; Zandifar, A.; Visweswaran, S.; Eslami, M.; Batmanghelich, K. MedSyn: Text-guided anatomy-aware synthesis of high-fidelity 3-D CT images. *IEEE Trans. Med. Imaging* **2024**, *43*, 3648–3660. [[CrossRef](#)] [[PubMed](#)]
90. Khader, F.; Müller-Franzes, G.; Tayebi Arasteh, S.; Han, T.; Haarburger, C.; Schulze-Hagen, M.; Schad, P.; Engelhardt, S.; Baeßler, B.; Foersch, S.; et al. Denoising diffusion probabilistic models for 3D medical image generation. *Sci. Rep.* **2023**, *13*, 7303. [[CrossRef](#)] [[PubMed](#)]
91. Huang, P.; Gao, X.; Huang, L.; Jiao, J.; Li, X.; Wang, Y.; Guo, Y. Chest-diffusion: A light-weight text-to-image model for report-to-cxr generation. In Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI); IEEE: Piscataway, NJ, USA, 2024; pp. 1–5.
92. Li, Y.; Liu, J.; Jain, R.K.; Chen, Y.W. Multimodal Brownian bridge diffusion model for controllable synthetic medical image generation. *Biomed. Signal Process. Control* **2026**, *117*, 109584. [[CrossRef](#)]
93. Ananya, J.; Udupa, S.; Avasthi, P.; Sanjana, M.; Likitha, P. Exploring the Robustness of Stable Diffusion in Chest X-ray Generation from Text Prompts. In Proceedings of the 2025 International Conference on Computer Technology Applications (ICCTA); IEEE: Piscataway, NJ, USA, 2025; pp. 146–153.
94. Xie, J.; Zhang, Z.; Weng, Z.; Zhu, Y.; Luo, G. MedDiff-FT: Data-Efficient Diffusion Model Fine-Tuning with Structural Guidance for Controllable Medical Image Synthesis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Daejeon, Republic of Korea, 23–27 September 2025; pp. 306–316.
95. Khalid, N.; Qayyum, A.; Bilal, M.; Al-Fuqaha, A.; Qadir, J. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Comput. Biol. Med.* **2023**, *158*, 106848. [[CrossRef](#)] [[PubMed](#)]
96. Zhu, Y.; Yin, X.; Wee-Chung Liew, A.; Tian, H. Privacy-preserving in medical image analysis: A review of methods and applications. In Proceedings of the International Conference on Parallel and Distributed Computing: Applications and Technologies, Hong Kong, China, 13–15 December 2024; pp. 166–178.
97. Shodiev, D.; Ushakov, E.; Litvinov, A.; Markin, Y. Privacy-Preserving Synthetic Mammograms: A Generative Model Approach to Privacy-Preserving Breast Imaging Datasets. *Informatics* **2025**, *12*, 112. [[CrossRef](#)]
98. Shi, Y.; Xia, W.; Niu, C.; Wiedeman, C.; Wang, G. Privacy-Preserving Latent Diffusion-Based Synthetic Medical Image Generation. *IEEE Trans. Med. Imaging* **2025**, *45*, 1140–1151.
99. Faisal, F.; Mohammed, N.; Leung, C.K.; Wang, Y. Generating privacy preserving synthetic medical data. In Proceedings of the 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA); IEEE: Piscataway, NJ, USA, 2022; pp. 1–10.
100. Ziller, A.; Usynin, D.; Braren, R.; Makowski, M.; Rueckert, D.; Kaissis, G. Medical imaging deep learning with differential privacy. *Sci. Rep.* **2021**, *11*, 13524. [[CrossRef](#)] [[PubMed](#)]
101. Saqlain, A.S.; Fang, F.; Ahmad, T.; Wang, L.; Abidin, Z.U. Evolution and effectiveness of loss functions in generative adversarial networks. *China Commun.* **2021**, *18*, 45–76. [[CrossRef](#)]
102. Nowozin, S.; Cseke, B.; Tomioka, R. f-gan: Training generative neural samplers using variational divergence minimization. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 271–279.
103. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.; Wang, Z.; Smolley, S.P. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2794–2802.
104. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5769–5779.
105. Qi, G.J. Loss-sensitive generative adversarial networks on lipschitz densities. *Int. J. Comput. Vis.* **2020**, *128*, 1118–1140.
106. Li, Z.; Shi, W.; Xing, Q.; Miao, Y.; He, W.; Yang, H.; Jiang, Z. Low-dose CT image denoising with improving WGAN and hybrid loss function. *Comput. Math. Methods Med.* **2021**, *2021*, 2973108. [[CrossRef](#)] [[PubMed](#)]

107. Chen, Y.; Lin, H.; Zhang, W.; Chen, W.; Zhou, Z.; Heidari, A.A.; Chen, H.; Xu, G. ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation. *Biomed. Signal Process. Control* **2024**, *92*, 106100. [[CrossRef](#)]
108. Heng, Y.; Yinghua, M.; Khan, F.G.; Khan, A.; Hui, Z. HLSNC-GAN: Medical image synthesis using hinge loss and switchable normalization in CycleGAN. *IEEE Access* **2024**, *12*, 55448–55464. [[CrossRef](#)]
109. Zheng, Z.; Wang, M.; Fan, C.; Wang, C.; He, X.; He, X. Light&fast generative adversarial network for high-fidelity CT image synthesis of liver tumor. *Comput. Methods Programs Biomed.* **2024**, *254*, 108252. [[CrossRef](#)] [[PubMed](#)]
110. Gajera, B.; Kapil, S.R.; Ziaei, D.; Mangalagiri, J.; Siegel, E.; Chapman, D. CT-scan denoising using a charbonnier loss generative adversarial network. *IEEE Access* **2021**, *9*, 84093–84109. [[CrossRef](#)]
111. Jiang, Y.; Li, J. Generative adversarial network for image super-resolution combining texture loss. *Appl. Sci.* **2020**, *10*, 1729. [[CrossRef](#)]
112. Zhao, M.; Wei, Y.; Wong, K.K. A generative adversarial network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images. *Magn. Reson. Imaging* **2022**, *85*, 153–160. [[CrossRef](#)] [[PubMed](#)]
113. Wang, L.; Zhao, L.; Zhong, T.; Wu, C. Low-light image enhancement using generative adversarial networks. *Sci. Rep.* **2024**, *14*, 18489. [[CrossRef](#)] [[PubMed](#)]
114. Xiao, B.; Ni, H.; Yang, W. MCGAN: Enhancing GAN training with regression-based generator loss. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 21644–21652. [[CrossRef](#)]
115. Guo, K.; Chen, J.; Qiu, T.; Guo, S.; Luo, T.; Chen, T.; Ren, S. MedGAN: An adaptive GAN approach for medical image generation. *Comput. Biol. Med.* **2023**, *163*, 107119. [[CrossRef](#)] [[PubMed](#)]
116. Gonzalez, S.; Kant, M.; Miikkulainen, R. Evolving gan formulations for higher-quality image synthesis. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 289–305.
117. Jin, L.; Peng, Y.; Wang, J.; Jiang, Y.; Zhang, K. Hierarchical multi-scale Mamba generative adversarial network for multi-modal medical image synthesis. *Expert Syst. Appl.* **2025**, *280*, 127451. [[CrossRef](#)]
118. Gocer, E. GAN based augmentation using a hybrid loss function for dermoscopy images. *Artif. Intell. Rev.* **2024**, *57*, 234. [[CrossRef](#)]
119. Veiner, J.; Alajaji, F.; Gharesifard, B. A unifying generator loss function for generative adversarial networks. *Entropy* **2024**, *26*, 290. [[CrossRef](#)] [[PubMed](#)]
120. Usman Akbar, M.; Larsson, M.; Blystad, I.; Eklund, A. Brain tumor segmentation using synthetic MR images-A comparison of GANs and diffusion models. *Sci. Data* **2024**, *11*, 259. [[CrossRef](#)] [[PubMed](#)]
121. Deo, Y.; Jia, Y.; Lassila, T.; Smith, W.A.; Lawton, T.; Kang, S.; Frangi, A.F.; Habli, I. Metrics that matter: Evaluating image quality metrics for medical image generation. *arXiv* **2025**, arXiv:2505.07175.
122. Mendez, M.; Castillo, F.; Probyn, L.; Kras, S.; Tyrrell, P.N. Leveraging domain knowledge for synthetic ultrasound image generation: A novel approach to rare disease AI detection. *Int. J. Comput. Assist. Radiol. Surg.* **2025**, *20*, 415–431. [[CrossRef](#)] [[PubMed](#)]
123. Zhao, J.; Mathieu, M.; LeCun, Y. Energy-based generative adversarial network. *arXiv* **2016**, arXiv:1609.03126.
124. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
125. Zhao, S.; Liu, Z.; Lin, J.; Zhu, J.Y.; Han, S. Differentiable augmentation for data-efficient gan training. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7559–7570.
126. Ding, H.; Huang, N.; Wu, Y.; Cui, X. Improving imbalanced medical image classification through GAN-based data augmentation methods. *Pattern Recognit.* **2025**, *166*, 111680. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.