



Review Articles

Volume-06|Issue-04|2026

Generative Adversarial Networks for Cross-Modality Medical Image Translation: A Comprehensive Review of Architectures, Applications, and Future Directions

Abhinav Pandey¹, Abhinab Das², Abhishek Suresh Ullegaddi³, Priyanshu Thakur⁴, and Nayana H. J.⁵

^{1,2,3,4}UG Student, Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

⁵Assistant Professor, Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History

Received: 05.05.2026

Accepted: 22.06.2026

Published: 25.07.2026

Citation

Pandey, A., Das, A., Ullegaddi, A. S., Thakur, P., Nayana, H. J. (2026) Generative Adversarial Networks for Cross-Modality Medical Image Translation: A Comprehensive Review of Architectures, Applications, and Future Directions. *Indiana Journal of Multidisciplinary Research*, 6(4), 446-451.

Abstract: Generative Adversarial Networks (GANs), introduced in 2014, have become a transformative tool in medical imaging. They support diverse applications, including cross-modal image synthesis, enhancement of visual clarity, and improved segmentation outcomes. This paper chronicles the progression of Generative Adversarial Network (GAN) models over time. It highlights their various applications, such as transforming MRI to CT images, generating high-resolution medical imagery, and identifying anomalies. Advanced architectures like CycleGAN and pix2pix are discussed in the context of their utility and performance in medical scenarios. Moreover, the paper provides a critical review of datasets and evaluation metrics, highlighting ongoing challenges related to training stability and anatomical precision. It concludes by proposing a comprehensive research roadmap to guide the field's transition to clinical practice.

Keywords: Generative Adversarial Networks · Medical Imaging · Cross-Modality Translation · Image Synthesis · Segmentation · Deep Learning.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0).

INTRODUCTION

Medical imaging stands as a cornerstone of modern medicine. Modalities like Magnetic Resonance Imaging (MRI) give us an unparalleled view of soft tissues, while Computed Tomography (CT) provides exquisite detail of bone and lung structures. Yet, in the day-to-day reality of clinical practice, obtaining multiple types of scans for a single patient can be a significant hurdle. The reasons are many: high costs, logistical challenges, and the health risks associated with repeated exposure to ionizing radiation. This clinical dilemma has sparked a pressing need for a new kind of technological alchemy: cross-modality image translation. Imagine being able to generate a CT scan from an existing MRI, a technique that could revolutionize tasks like radiotherapy planning, where both soft-tissue and bone detail are crucial. At the forefront of this innovation are Generative Adversarial Networks (GANs), a powerful deep learning framework introduced in 2014 by Ian Goodfellow and his colleagues [2]. GANs have quickly become a transformative tool, enabling a diverse range of applications from synthesizing one scan type from another to enhancing image quality and augmenting limited datasets to train more powerful diagnostic models. The field has evolved at a breathtaking pace, moving from early models that wrestled with instability to highly specialized

architectures designed for specific clinical challenges. This review provides a rigorous, multi-faceted analysis of this dynamic landscape. We move beyond a simple chronological summary to offer a critical synthesis of the literature, uncovering the key architectural themes, the trade-offs between different applications, and a challenge that looms large over the entire field: the fundamental disconnect between technical scores and real-world clinical utility. Our goal is to provide a comprehensive map of the field's evolution, a robust framework for evaluating its progress, and a forward-looking roadmap to guide its journey into the heart of clinical practice.

METHODOLOGY: SEARCH STRATEGY AND INCLUSION CRITERIA

To ensure this review was both comprehensive and methodologically sound, we conducted a systematic search across major scientific databases, including PubMed, IEEE Xplore, and arXiv. Our search strategy was designed to cast a wide net while maintaining sharp focus. We began with a set of primary keywords to capture the core of the field: "Generative Adversarial Networks," "Medical Imaging," "Cross-Modality Translation," "Image Synthesis," "GANs for Medical

Data Augmentation," and "Deep Learning." To ensure we captured the latest architectural innovations, this initial search was supplemented with specific model names, such as "pix2pix," "CycleGAN," "WGAN," "ProGAN," "SAGAN," and the emerging "Diffusion Models." Our inclusion criteria targeted peer-reviewed articles and significant preprints published from 2014 to the present that directly addressed the use of GANs in medical imaging. We prioritized papers that offered clear discussions of their architectural designs, loss functions, and evaluation methods. Studies that did not focus on generative modeling or were not available in English were excluded. Finally, to ensure our review was grounded in the field's most influential work, we cross-referenced our included papers with top-cited review articles, confirming that our analysis incorporates the landmark publications that have shaped this technology's trajectory.

THE FOUNDATIONAL EVOLUTION OF GANS: FROM THEORETICAL CONCEPT TO STABLE FRAMEWORK

Although the original 2014 GAN framework [2] was groundbreaking, it proved extremely difficult to train. It often suffered from severe instability and a phenomenon known as mode collapse, in which the generator

repeatedly produces only a narrow range of images and thus fails to capture the full diversity of real data. The core theoretical problem persisted: the Jensen–Shannon divergence at the heart of the GAN framework provided a vanishingly small gradient when the real and generated data distributions failed to overlap—a common scenario in early training that effectively halted the generator's learning. A true solution required a theoretical leap, which came in 2017 with the Wasserstein GAN (WGAN) [8]. Researchers replaced the problematic loss with the Earth Mover's (Wasserstein-1) distance, a metric that provides a smooth and continuous gradient even when the distributions are disjoint, yielding a loss that correlated far better with image quality. The initial WGAN model, however, enforced its constraints with a crude weight clipping mechanism that could introduce new instabilities. This was resolved by its successor, WGAN with Gradient Penalty (WGAN-GP) [10], which used a more principled gradient-norm penalty to enforce the necessary Lipschitz condition. This journey, from early instability to the theoretical rigor of WGAN-GP, marks the critical maturation of GANs from a brilliant but fragile concept into a robust and reliable generative framework. As shown in Fig. 1, subsequent architectures evolved from the original GAN framework toward more stable and specialized variants.

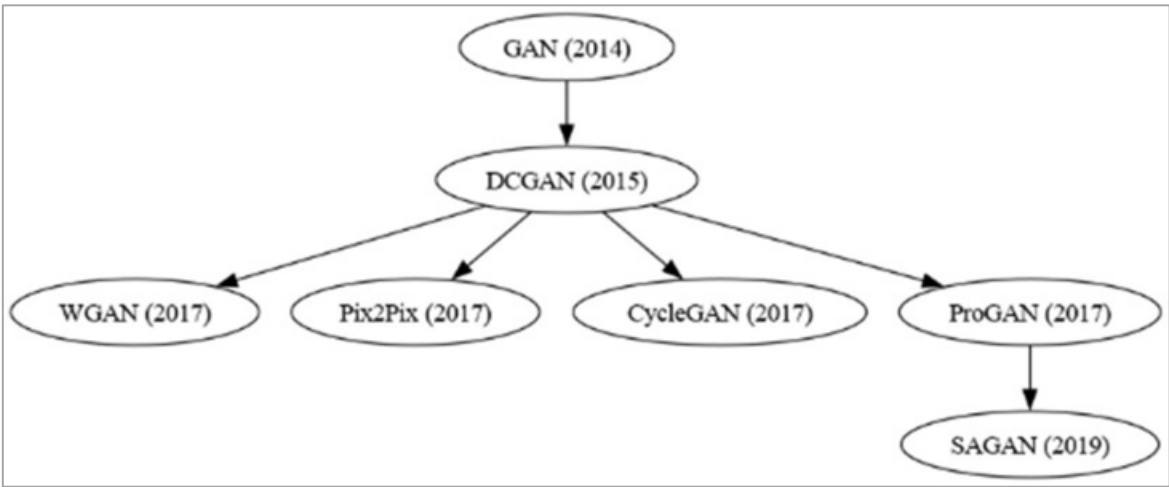


Fig. 1: Evolution of GAN architectures from 2014 to 2019.

Table 1: Comparative Analysis of Key Generative Adversarial Network (GAN) Models

Model	Year	Primary Application	Core Architectural Innovation	Key Loss Function	Problem Solved
WGAN	2017	Enhanced training stability	Wasserstein-1 distance (Earth Mover's distance) based architecture	Wasserstein loss	Addressed training instability, vanishing gradients, and prevented mode collapse during training.
Pix2Pix	2017	Paired/Supervised image-to-image translation	U-Net-based generator (with skip connections) and PatchGAN discriminator	Conditional GAN loss combined with L1 reconstruction penalty	Achieved accurate source-to-target image mappings while maintaining visual realism.

Model	Year	Primary Application	Core Architectural Innovation	Key Loss Function	Problem Solved
CycleGAN	2017	Unpaired image-to-image translation	Cyclic generator-discriminator pairs enforcing cycle consistency	Cycle-consistency loss	Facilitated modality shifts without requiring paired training examples.
UNIT GAN	2017	Unpaired image-to-image translation	Shared latent space through coupled VAEs	VAE-GAN objective	Supported latent-space bridging for unsupervised cross-domain mappings.
ProGAN	2017	High-Resolution Image Synthesis	Progressive growing of the generator and discriminator (Progressive Training)	N/A	Produced intricate, high-fidelity visuals through phased resolution escalation.
SAGAN	2019	Data Augmentation	Self-attention mechanism	Hinge adversarial loss	Captured long-range dependencies, improving structural coherence and visual quality in generated outputs.

GAN ARCHITECTURES FOR MEDICAL IMAGE TRANSLATION: A THEMATIC ANALYSIS

The application of GANs to medical image translation is shaped by a fundamental real-world constraint: the availability of data. Acquiring perfectly aligned scans from different modalities for the same patient is often difficult and expensive. This challenge has given rise to two distinct families of GAN architectures, each tailored to a different data scenario.

Supervised Translation with Paired Data: The Pix2Pix Framework

When a dataset of paired images exists—for example, a collection of MRIs and their corresponding CT scans **for the same patients**—the pix2pix framework [12] provides a powerful, supervised solution. As a Conditional GAN (cGAN), it learns a direct mapping from an input image to a target output. Its success lies in two key architectural innovations: a U-Net generator with skip connections that excels at preserving fine-grained details, and a PatchGAN discriminator that evaluates the realism of the image at a local level, forcing the generator to produce sharp, high-frequency details. This makes pix2pix highly effective for tasks like denoising or super-resolution, where it can restore subtle anatomical structures in low-quality scans. Its primary limitation, however, is its strict need for paired data, which remains a bottleneck in many clinical settings.

Unsupervised Translation with Unpaired Data: CycleGAN and UNIT GAN

When paired data is scarce, models capable of handling unpaired data become essential. CycleGAN [13], introduced by Zhu et al. in 2017, is a leading model for unpaired image-to-image translation. Its central innovation is the cycle consistency loss, which enforces the assumption that an image translated to a different modality and **back** should be identical to the original. This bi-directional learning framework allows the model

to learn a meaningful translation without a direct one-to-one correspondence between images in the source and target domains. As an alternative, UNIT GAN [18], introduced by Liu et al., tackles the same problem from a different theoretical perspective, relying on a shared latent space assumption. While CycleGAN's cycle consistency is a brilliant general-purpose solution, its lack of domain-specific knowledge can lead to anatomical inaccuracies and distortions in the generated outputs. This fundamental failure mode has led to the development of specialized variants like Structure-Constrained CycleGAN [17] and DC-CycleGAN [7], which add extra, domain-specific losses to improve anatomical fidelity.

Specialized Architectures: Addressing Second-Order Problems

As the field matured, research moved beyond the basic paired/unpaired dichotomy to solve more nuanced challenges. ProGAN [20] tackled the problem of generating high-resolution images by progressively growing the generator and discriminator, starting with low-resolution images and gradually adding layers. This curriculum-based learning approach dramatically improved stability and output quality, making it ideal for creating the detailed scans needed for radiotherapy planning. The Self-Attention GAN (SAGAN) [14] addressed the challenge of modeling long-range dependencies. Incorporating a self-attention mechanism allows the model to ensure that features in one part of an image are consistent with features in distant parts, leading to more globally coherent and anatomically accurate results.

This progression—from foundational models to task-specific frameworks and finally to specialized architectures—shows a research field that is continually evolving to meet the complex, domain-specific demands of medical imaging.

A CRITICAL APPRAISAL OF APPLICATIONS

Cross-modality image synthesis stands as a central use case for GANs in medical imaging. For example, generating a CT scan from an existing MRI can be invaluable in radiotherapy planning: MRI provides superior soft-tissue contrast, while CT offers essential information about bone structure and density for dose calculations. Importantly, such translations often lack paired datasets of aligned images. Models like CycleGAN and UNIT GAN are particularly well-suited for this scenario, as they can learn to translate between MRI and CT without requiring perfectly aligned image pairs, thereby overcoming a major practical barrier in clinical datasets.

In addition to modality translation, GANs are widely used for image enhancement and denoising. Architectures such as pix2pix, which employ a PatchGAN discriminator, excel at improving image quality by recovering fine details. In practice, pix2pix can restore subtle anatomical structures in low-resolution or noisy scans, refining these images to higher quality levels. These enhancements are crucial, since preserving minute details in a scan can significantly affect diagnostic accuracy and the reliability of downstream tasks like segmentation.

THE EVALUATION CONUNDRUM: BRIDGING THE TECHNICAL-CLINICAL GAP

The most significant obstacle to the widespread clinical adoption of GANs is a fundamental disconnect between what is considered a success in a computer science lab and what is useful in a hospital. The field's reliance on generic computer vision metrics, which measure realism but not correctness, creates a dangerous blind spot. A GAN can be a technical triumph and a clinical failure at the same time. Standard metrics like Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Fréchet Inception Distance (FID) are insufficient for clinical validation. High SSIM and PSNR scores do not guarantee that subtle but diagnostically critical features, like small lesions or variations in bone density, have been preserved. A low FID score indicates that an image looks realistic and diverse, but it says nothing about whether it is anatomically or pathologically correct. This evaluation gap is a higher-order problem that persists regardless of a model's architectural sophistication. The true value of a generated medical image can only be measured by its impact on a clinical outcome, and the field currently lacks standardized methods to do so.

To bridge this gap, a rigorous, multi-tiered clinical evaluation framework is essential. This framework must

move beyond technical scores to a task-level assessment of clinical impact.

Table 2: Public Datasets for GAN-Based Medical Imaging Research

Dataset Name	Modalities Included	Primary Applications
BraTS	T1, T1c, T2, FLAIR MRI	Brain tumor synthesis and segmentation
IXI	MRI, CT, PET	Cross-modality translation (e.g., MRI-to-CT)
MM-WHS	MRI, CT	Cardiac segmentation and cross-modality translation
LIDC-IDRI	Lung CT scans	Lung segmentation and anomaly detection

Quantitative Analysis

Initial assessment is performed using standard metrics to quantify basic technical quality. This includes per-pixel and structural similarity metrics like SSIM and PSNR, distribution metrics like FID to assess realism and diversity, and task-specific metrics like the Dice Similarity Coefficient (DSC) to quantify segmentation accuracy.

Expert Qualitative Review

A more rigorous assessment requires a blinded A/B protocol with a panel of certified medical experts. A panel of at least three radiologists should independently evaluate images without knowledge of their source (real vs. synthetic) to minimize confirmation bias. Raters should use a 5-point Likert scale to assess key criteria such as anatomical correctness, the presence of artifacts, and diagnostic confidence. The consistency of this expert opinion must be quantitatively measured using Cohen's kappa, which assesses inter-rater agreement and accounts for chance agreement.

Clinical Utility Analysis

The ultimate validation is a task-level assessment of clinical impact. This can be quantified using a Diagnostic Accuracy Improvement (DAI) metric, which measures the change in a clinician's diagnostic performance when using GAN-generated images compared to a control group. To enable this framework, the field must assemble a new class of "Clinically-Vetted Benchmarks" that are multi-center, include rich metadata (e.g., scanner type and acquisition parameters), and have ground-truth annotations and quality scores from multiple certified radiologists.

Table 3: Evaluation Metrics and Their Clinical Limitations

Metric	Technical Purpose	Clinical Limitation
SSIM & PSNR	Measure pixel-wise and structural similarity to the ground truth.	High scores do not guarantee preservation of subtle diagnostic features.
FID	Assesses realism by comparing feature distributions with real data.	A low FID score does not ensure anatomical accuracy or the presence of critical pathological features.
DSC	Quantifies the overlap between predicted and ground-truth segmentations.	Does not account for the clinical significance of misclassifying small but critical lesions.

THE ROAD AHEAD: CRITICAL CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Despite notable technical progress, GANs face important barriers to safe, routine clinical adoption. Chief among them is the mismatch between standard engineering metrics and real clinical value. Addressing this—and several other challenges—will determine how the field matures alongside emerging generative paradigms.

The Rise of Diffusion Models

Diffusion models offer greater training stability and often produce very high-fidelity images by iteratively denoising samples. However, they are computationally heavy and slow at inference. GANs remain preferable when fast synthesis and lower compute are required, and in low-data regimes, some GAN variants still outperform diffusion models. Choice of method, therefore, depends on the clinical task and resource constraints.

Overcoming Anatomical Inaccuracies

Generative models in unpaired settings can sometimes produce images with anatomically unrealistic features or artifacts, which could mislead clinical interpretations. Physics-Informed GANs (PI-GANs) seek to address this problem by embedding domain-specific knowledge into the generation process. For instance, in MRI-to-CT translation, a PI-GAN could include a constraint in its loss function that penalizes deviations from known Hounsfield Unit (HU) values for different tissues. By incorporating such physical constraints directly into training, PI-GANs enforce anatomical plausibility that purely data-driven GANs might miss. In parallel, there is growing interest in quantifying uncertainty in generated images. Methods that produce per-voxel uncertainty or confidence maps give clinicians insight into which regions of a synthetic image are most reliable and which should be interpreted with caution. These approaches together aim to improve the trustworthiness and safety of GAN-generated medical images by ensuring both anatomical fidelity and clear uncertainty estimates.

A Practical Research Roadmap

To accelerate translation into clinical practice, the community should pursue several coordinated actions:

- Develop and publish open-source PI-GAN baselines that make physics constraints standard practice.
- Integrate uncertainty estimation into synthesis models to provide per-voxel confidence.
- Assemble large, multi-center benchmark datasets with rich metadata and radiologist annotations.
- Run blinded, multi-rater clinical validation studies that evaluate models at the task level rather than solely by low-level similarity metrics.

CONCLUSION

The trajectory of GAN research in medical imaging is one of steady refinement: early architectural advances (e.g., DCGAN) improved empirical stability, theoretical innovations (e.g., WGAN/WGAN-GP) supplied more reliable training objectives, and task-specific architectures (e.g., pix2pix, CycleGAN) tailored solutions for paired and unpaired translation. More recent models, such as ProGAN and SAGAN, have extended capabilities for high fidelity and global consistency. Yet, substantial work remains to bridge the divide between high scores on technical benchmarks and demonstrable clinical benefit. Progress will require not only better models but also new evaluation frameworks, hybrid approaches that integrate domain knowledge, and rigorous clinical studies. Only through coordinated methodological, dataset, and validation advances can the promise of generative methods be safely and usefully realized in clinical practice.

REFERENCES

1. Wang, T., Lei, Y., Fu, Y., Wynne, J. F., Curran, W. J., Liu, T., & Yang, X. (2021). A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of Applied Clinical Medical Physics*, 22(1), 11–36.
2. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
3. Hussain, J., Bâth, M., & Ivarsson, J. (2025). Generative adversarial networks in medical image reconstruction: A systematic literature review. *Computers in Biology and Medicine*, 191, Article 110094.
4. Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised representation learning with deep convolutional generative adversarial networks*. arXiv. <https://arxiv.org/abs/1511.06434>
5. Rafiq, M., Ali, H., Mujtaba, G., Shah, Z., & Azmat, S. (2025). *Cross-modality medical image synthesis for improving liver segmentation*. arXiv. <https://arxiv.org/abs/2503.00945>
6. Czobit, C., & Samavi, R. (2024). *CycleGAN models for MRI image translation*. arXiv.

- <https://arxiv.org/abs/2401.00023>
7. Wang, J., Wu, Q. M. J., & Pourpanah, F. (2022). DC-CycleGAN: Bidirectional CT-to-MR synthesis from unpaired data. *Computerized Medical Imaging and Graphics*, 102, Article 102168.
 8. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 214–223).
 9. Nikbakhsh, S., Naghashyar, L., Valizadeh, M., & Amirani, M. C. (2023). *Enhanced synthetic MRI generation from CT scans using CycleGAN with feature extraction*. arXiv. <https://arxiv.org/abs/2310.20604>
 10. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. *Advances in Neural Information Processing Systems*, 30, 5769–5779.
 11. Wei, S., Si, L., Huang, T., Du, S., Yao, Y., Dong, Y., & Ma, H. (2023). Deep-learning-based cross-modality translation from Stokes image to bright-field contrast. *Journal of Biomedical Optics*, 28(10), Article 102911.
 12. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1125–1134).
 13. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2223–2232).
 14. Zhang, H., et al. (2019). Self-attention generative adversarial networks. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 7354–7363).
 15. Wang, C., Macnaught, G., Papanastasiou, G., MacGillivray, T., & Newby, D. (2018). Unsupervised learning for cross-domain medical image synthesis using deformation invariant cycle consistency networks. In A. Gooya, O. Goksel, I. Oguz, & N. Burgos (Eds.), *Simulation and Synthesis in Medical Imaging* (pp. 52–60). Springer.
 16. Heng, Y., Yinghua, M., Khan, F. G., Khan, A., & Hui, Z. (2024). HLSNC-GAN: Medical image synthesis using hinge loss and switchable normalization in CycleGAN. *IEEE Access*, 12, 55448–55462.
 17. Yang, Y., et al. (2020). Unpaired MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Transactions on Medical Imaging*, 39(12), 4249–4261.
 18. Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in Neural Information Processing Systems*, 30, 700–709.
 19. Nie, D., Trullo, R., Petitjean, C., Ruan, S., & Shen, D. (2016). *Medical image synthesis with context-aware generative adversarial networks*. arXiv. <https://arxiv.org/abs/1612.05362>
 20. Karras, T., Aila, T., & Laine, S. (2017). *Progressive growing of GANs for improved quality, stability, and variation*. arXiv. <https://arxiv.org/abs/1710.10196>
 21. Özbey, M., et al. (2024). Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 43(1), 1–13.
 22. Kearney, V., et al. (2020). Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks. *Radiology: Artificial Intelligence*, 2(2).
 23. Dou, Q., et al. (2019). PnP-AdaNet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
 24. Wolterink, J. M., Leiner, T., & Isgum, I. (2018). *Blood vessel geometry synthesis using generative adversarial networks*. arXiv. <https://arxiv.org/abs/1804.04381>
 25. Kunapinun, A., et al. (2023). Improving GAN learning dynamics for thyroid nodule segmentation. *Ultrasound in Medicine & Biology*, 49(2), 416–430.
 26. Rajak, R., Rathore, A., Ramteke, M., & Patil, S. (2025). MMIT-DDPM: Multilateral medical image translation with class and structure supervised diffusion-based model. *Computers in Biology and Medicine*, 169, Article 107931.
 27. Fan, Z., et al. (2024). Self-supervised learning based on StyleGAN for medical image classification on small labeled datasets. In *Medical Imaging 2024: Image Processing* (Vol. 12926, pp. 727–733).
 28. Arslan, F., Kabas, B., Dalmaz, O., Ozbey, M., & Çukur, T. (2024). *Self-consistent recursive diffusion bridge for medical image translation*. arXiv. <https://arxiv.org/abs/2405.06789>
 29. Yang, T., & Wang, L. (2022). Image translation-based unsupervised cross-modality domain adaptation for medical image segmentation. In *Proceedings of the MICCAI Workshop on Cross-Modality Domain Adaptation*.
 30. Armanious, K., et al. (2019). Unsupervised medical image translation using Cycle-MedGAN. In *Proceedings of the 27th European Signal Processing Conference (EUSIPCO 2019)* (pp. 1–5).