Medicine Without Boundaries: Generative AI for Translating Medical Data Across Modalities

Valerio Guarrasi[1,*], Francesco Di Feola[1,2], Giulio Iannello[1], Irene Iele[1], Linlin Shen[3], Massimiliano Mantegna[1], Daniele Molino[1], Elena Mulero Ayllón[1], Ludovica Pompilio[1], Aurora Rofena[1], Marco Salmè[1], Rosa Sicilia[1], Matteo Tortora[4] and Paolo Soda[1,2]

[1]*Unit of Artificial Intelligence and Computer Systems, Department of Engineering, Università Campus Bio-Medico di Roma, Rome, Italy*
[2]*Department of Diagnostics and Intervention, Radiation Physics, Biomedical Engineering, Umeå University, Umeå, Sweden*
[3]*Shenzhen University*
[4]*Department of Naval, Electrical, Electronics and Telecommunications Engineering, University of Genoa, Genoa, Italy*

Abstract

Generative artificial intelligence (AI) is transforming digital health by enabling synthetic data generation, enhancement, and multimodal integration. In this work, we investigate the role of generative models in clinical data translation, introducing contributions for intra-modality, inter-modality, and any-to-any translations. Intra-modality methods focus on improving data quality within a single modality, e.g., for example, through denoising, super-resolution, or harmonization. Inter-modality models enable translation between distinct modalities, e.g., generating PET from CT or CESM from mammography, supporting safer or more accessible diagnostics. Finally, any-to-any frameworks combine heterogeneous data types, including imaging, text, and signals, within shared latent spaces to enable multimodal synthesis and digital twin construction. We provide representative examples for each category, discuss their clinical relevance and methodological underpinnings, and outline the challenges that must be addressed to integrate generative AI into healthcare.

Keywords

Multimodal Learning, Image-to-image Translation, Digital Twin, Virtual Scanner

1. Introduction

Generative artificial intelligence (AI) is revolutionizing digital health by transforming the production, interpretation, and integration of clinical data [1, 2]. Models such as generative adversarial networks (GANs), diffusion models, and large language models (LLMs) can now synthesize realistic medical images, simulate disease progression, generate clinical reports, and align multimodal data with remarkable fluency [3]. These generative tools are advancing from isolated tasks to more integrated, data-driven reasoning in complex healthcare contexts [4, 5, 6].

A key development is generative clinical data translation, where models convert data from one form or modality to another [7, 8, 9, 10, 11]. We classify these tasks into three categories:

- *Intra-modality translation*: improving data within a single modality, e.g., denoising low-dose CT or synthesizing follow-up CTs.
- *Inter-modality translation*: generating one type of data from another, e.g., creating PET images from CT scans or histopathology from transcriptomics.
- *Any-to-any translation*: flexible conversion across multiple heterogeneous modalities, integrating images, text, signals, and structured data.

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
valerio.guarrasi@unicampus.it (V. Guarrasi); francesco.feola@umu.se (F. D. Feola); G.Iannello@unicampus.it (G. Iannello); irene.iele@unicampus.it (I. Iele); llshen@szu.edu.cn (L. Shen); M.Mantegna@unicampus.it (M. Mantegna); daniele.molino@unicampus.it (D. Molino); e.muleroayllon@unicampus.it (E. Mulero Ayllón); ludovica.pompilio@unicampus.it (L. Pompilio); aurora.Rofena@unicampus.it (A. Rofena); marco.salme@unicampus.it (M. Salmè); r.sicilia@unicampus.it (R. Sicilia); matteo.tortora@unige.it (M. Tortora); p.soda@unicampus.it (P. Soda)

These approaches enable diverse clinical applications. Intra-modality models standardize data quality across sites. Inter-modality models simulate missing or costly modalities [12, 13, 14], reducing diagnostic burden. Any-to-any systems support integrative tools like digital twins and clinical decision-support systems that mirror real-world complexity.

However, challenges remain. Generative outputs must be clinically realistic, interpretable, and trustworthy [15]. Risks like hallucinated features and spurious correlations threaten model reliability in clinical settings. Additionally, regulatory standards for generative models are still underdeveloped, hindering clinical implementation.

This paper systematically explores generative translation in medicine, organized around the three translation categories. For each, we detail illustrative use cases in Figure 1, discuss key methodological advances, and consider clinical impacts. We conclude by outlining open challenges, ethical issues, and future directions to ensure generative AI enriches, rather than merely automates, medical practice.

2. Intra-Modality Translation

Intra-modality translation involves generative tasks that refine data within a single imaging modality to improve quality, ensure standardization, and enable personalized clinical modeling. It addresses variability arising from different imaging protocols, scanner types, and reconstruction algorithms, thus enhancing data reliability for clinical use.

CT Harmonization CT image homogenization is crucial for minimizing variability from different reconstruction kernels. Recent work frames this as a style-transfer task using a Texture-Aware StarGAN [16], which performs unpaired, one-to-many translation between reconstruction styles. The model's multi-scale texture loss preserves fine anatomical detail across spatial and angular dimensions. An advanced variant, StarDiffusionGAN, combines diffusion models' synthesis fidelity with GANs' adversarial learning, enhancing robustness in heterogeneous clinical datasets.

Low-Dose CT Denoising Denoising LDCT images—where radiation dose reduction amplifies noise—requires preserving subtle structures. A new approach [17] integrates multi-scale texture descriptors via a differentiable self-attention mechanism to enhance detail preservation. To maintain performance under clinical distribution shifts, a Test-Time Adaptation (TTA) strategy uses a trainable reconstruction module that dynamically adapts feature transformations based on input characteristics, ensuring robustness to unseen data and enabling real-world deployment.

Virtual Treatment Generative models also enable virtual treatment simulations, such as predicting tumor evolution in NSCLC. Diffusion-based models conditioned on radiotherapy doses and patient-specific variables synthesize follow-up CT scans, simulating alternative treatment scenarios. Moving beyond static translation, these models learn temporal dynamics [18], offering more realistic disease progression simulations. Incorporating tumor segmentations from foundation models can further tailor these simulations [19], supporting digital twin frameworks for individualized treatment planning and longitudinal monitoring. Collectively, these intra-modality applications advance image quality, model generalizability, and personalized medicine [20], paving the way for integration into clinical workflows and transformative shifts in preprocessing, decision support, and simulation practices.

3. Inter-Modality Translation

Inter-modality translation generates clinical data in one modality from input in another, extending diagnostic capabilities and reducing reliance on invasive, costly, or radiation-intensive tests. This approach underpins *virtual scanners*, AI systems that simulate challenging-to-acquire modalities, enhancing diagnostic reach and workflow efficiency.

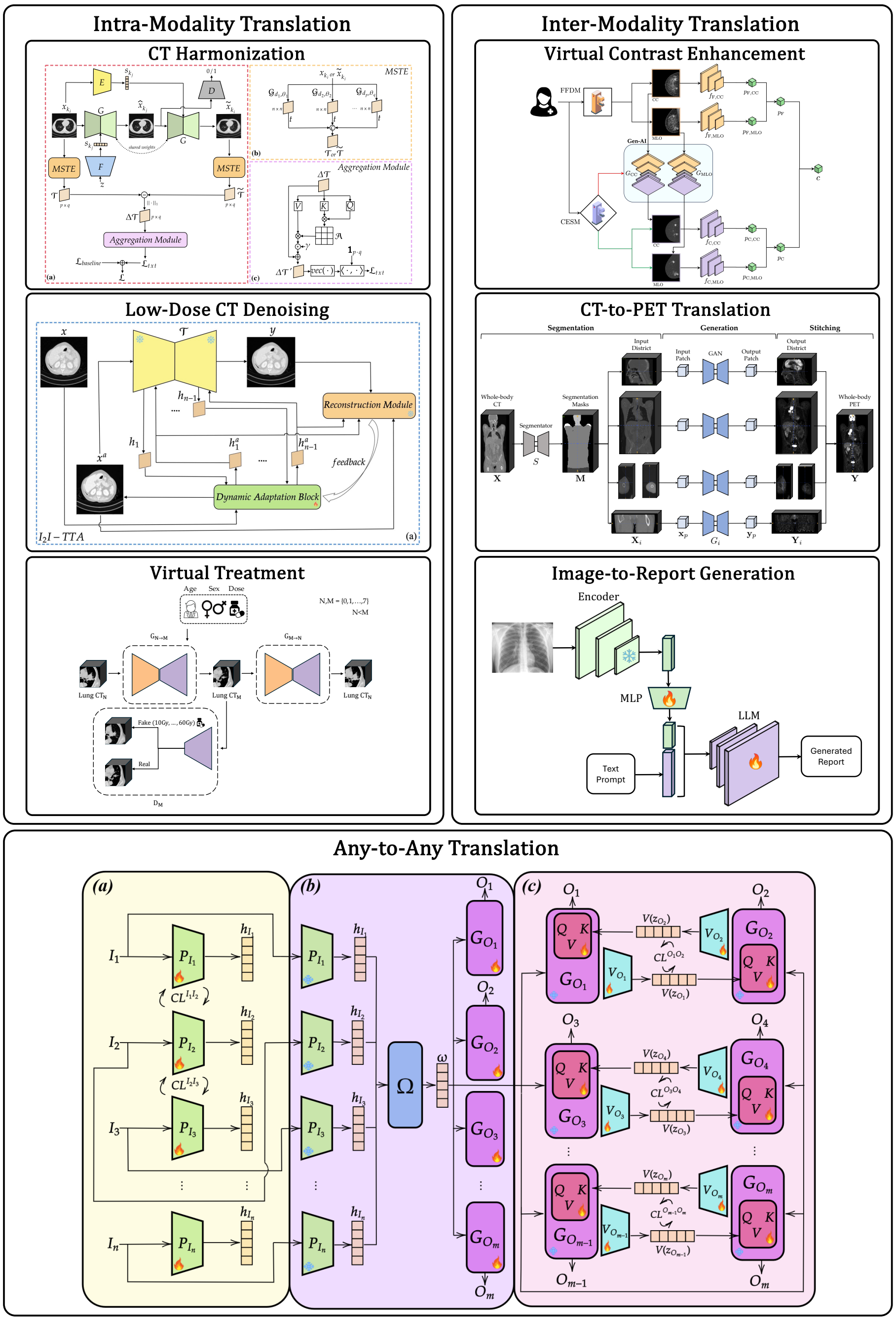

Figure 1: Overview of the presented research activities.

Virtual Contrast Enhancement Generative models can synthesize contrast-enhanced spectral mammography (CESM) images from standard full-field digital mammography (FFDM) [21, 22, 23]. Techniques like CycleGAN and Segmentation-aware CycleGAN (Seg-CycleGAN) incorporate segmentation priors and localized loss functions to enhance lesion visibility and maintain structural fidelity. These methods improve lesion localization and reduce false negatives, demonstrating promise for contrast-free breast cancer screening and diagnostics.

CT-to-PET Translation PET imaging, essential for metabolic insights in oncology, faces barriers from cost, radiation, and availability. Generative models can translate anatomical CT scans to synthetic PET images, providing non-invasive metabolic surrogates. Early methods used region-specific GANs [24], while newer strategies apply curriculum learning within unified translation networks. This twostep approach—initial global training followed by anatomy-informed curriculum learning—boosts performance across anatomically diverse body regions. Texture and intensity-based clustering further streamlines training by grouping similar districts, improving scalability and efficiency. These models excel in lesion visibility and tracer uptake simulation, making them vital for digital twin applications in oncology and personalized treatment planning.

Image-to-Report Generation Another inter-modality translation application is generating structured or narrative radiology reports directly from medical images. Vision-Language Models (VLMs), such as LLaVA, adapted for multilingual radiology tasks, have shown improvements in fluency and factual correctness when fine-tuned on instruction-based medical datasets [25]. Domain-specific fine-tuning outperforms general VLMs, highlighting the need for clinical and linguistic tailoring. These systems can reduce diagnostic bottlenecks and improve reporting access in underserved regions.

Inter-modality translation tasks highlight the potential of generative AI to expand diagnostics, enhance patient safety, and optimize workflows. Challenges remain in ensuring modality-specific accuracy, reducing hallucinations, and validating outputs in clinical trials. Retrieval Augmented Generation (RAG) is a promising approach, leveraging external knowledge to boost factual precision and mitigate hallucination risks. Despite RAG's benefits, broad clinical adoption will require continued progress in model architecture, domain adaptation, and multimodal supervision.

4. Any-to-Any Translation

Any-to-any translation represents the most versatile generative task, enabling flexible mapping across various modalities (e.g., medical images, clinical text, biosignals, structured data) using shared latent representations. This approach mirrors clinical practice's inherent multimodality, integrating heterogeneous data from imaging, EHRs, and reports.

Foundation Models for Multimodal Healthcare Large-scale generative frameworks like MedCoDi-M [26, 27] and models trained on datasets like MIMIC-CXR can jointly synthesize medical images and reports. These dual-output systems achieve high fidelity (e.g., FID, BLEU, ROUGE) and even surpass real data in disease classification, underscoring their value for synthetic data generation and clinical support.

MedCoDi-M uses contrastive learning to align modalities in a shared latent space, adapting to diverse tasks (e.g., report-from-image, image-from-text) through Multi-Prompt training. Its clinical validity, supported by expert evaluations, also addresses data democratization, privacy, and rare disease representation. Ongoing work extends these models from 2D images to 3D CT volumes, tackling challenges of memory, spatial continuity, and anatomical fidelity.

These innovations position any-to-any models as intelligent intermediaries that integrate, synthesize, and reason across modalities—key to digital twins, adaptive decision support, and equitable healthcare [28, 29, 30, 31, 32, 33]. Future deployment will hinge on rigorous clinical validation and clear regulatory frameworks to ensure safe, impactful use.

5. Conclusion

Generative translation in healthcare spans from intra-modality (noise reduction, harmonization) to inter-modality (contrast-enhanced or functional imaging) and fully multimodal any-to-any frameworks. Each contributes uniquely to data quality, multimodal reasoning, and personalized decision support.

Realizing this potential requires:

1. **Research:** Focus on robust, generalizable models, standardized evaluations, and multi-institutional datasets.

2. **Clinical Collaboration:** Involve clinicians in model co-design and validation to ensure relevance and safety in real-world workflows.
3. **Regulatory Clarity:** Define standards for synthetic data quality, model documentation (e.g., datasheets), and human oversight in high-risk applications [34, 35, 36, 37].

Acknowledgments

Irene Iele, Massimiliano Mantegna, Daniele Molino, Ludovica Pompilio, Aurora Rofena are Ph.D. students enrolled in the National Ph.D. in Artificial Intelligence, course on Health and Life Sciences, organized by Università Campus Bio-Medico di Roma. We acknowledge financial support from: i) PNRR MUR project PE0000013-FAIR; ii) PRIN 2022 MUR 20228MZFAA-AIDA (CUP C53D23003620008); iii) Unione europea- Next Generation EU, Missione 4 Componente C2 - Investimento 1.1 "Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN)" CUP di progetto (CUP C53D23009280001); iv) FCS MISE (CUP B89J23000580005); v) MAECI (grant n. CN23GR09); vi) PNRR M6/C2 project PNRR-MCNT2-2023-12377755; vii) University Campus Bio-Medico di Roma under the program "University Strategic Projects" within the project "AI-powered Digital Twin for next-generation lung cancEr cAre (IDEA)"; viii) Kempe Foundation project JCSMK24-0094.

Declaration on Generative AI

During the preparation of this work, the authors did not use any AI tool.

References

[1] C. Adornato, et al., Virtual scanner: leveraging resilient generative ai for radiological imaging in the era of medical digital twins, in: 2024 Ital-IA Intelligenza Artificiale-Thematic Workshops, Ital-IA 2024, Naples, Italy, May 29-30, 2024, CEUR-WS, 2024, pp. 60–65.
[2] F. Aksu, et al., Towards AI-driven next generation personalized healthcare and well-being, in: 2024 Ital-IA Intelligenza Artificiale-Thematic Workshops, Ital-IA 2024, Naples, Italy, May 29-30, 2024, CEUR-WS, 2024, pp. 360–365.
[3] V. Guarrasi, et al., Building an ai-enabled metaverse for intelligent healthcare: opportunities and challenges, in: CEUR WORKSHOP PROCEEDINGS, volume 3486, 2023, pp. 134–139.
[4] L. Furia, et al., Exploring early stress detection from multimodal time series with deep reinforcement learning, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2023, pp. 1917–1920.
[5] L. Nibid, et al., Deep pathomics: A new image-based tool for predicting response to treatment in stage III non-small cell lung cancer, Plos one 18 (2023) e0294259.
[6] M. Tortora, et al., RadioPathomics: multimodal learning in non-small cell lung cancer for adaptive radiotherapy, IEEE Access 11 (2023) 47563–47578.
[7] A. Francesconi, et al., Class balancing diversity multimodal ensemble for Alzheimer's disease diagnosis and early detection, Computerized Medical Imaging and Graphics 123 (2025) 102529.
[8] V. Guarrasi, et al., A systematic review of intermediate fusion in multimodal deep learning for biomedical applications, Image and Vision Computing (2025) 105509.
[9] G. Di Teodoro, et al., A graph neural network-based model with Out-of-Distribution Robustness for enhancing Antiretroviral Therapy Outcome Prediction for HIV-1, Computerized Medical Imaging and Graphics 120 (2025) 102484.
[10] V. Guarrasi, et al., Timing Is Everything: Finding the Optimal Fusion Points in Multimodal Medical Imaging, arXiv preprint arXiv:2505.02467 (2025).
[11] C. M. Caruso, et al., A multimodal ensemble driven by multiobjective optimisation to predict overall survival in non-small-cell lung cancer, Journal of Imaging 8 (2022) 298.

[12] C. M. Caruso, et al., A deep learning approach for overall survival prediction in lung cancer with missing values, Computer Methods and Programs in Biomedicine 254 (2024) 108308.
[13] C. M. Caruso, et al., MARIA: A multimodal transformer model for incomplete healthcare data, Computers in Biology and Medicine 196 (2025). doi:10.1016/j.compbiomed.2025.110843.
[14] C. M. Caruso, et al., Not Another Imputation Method: A Transformer-based Model for Missing Values in Tabular Datasets, arXiv preprint arXiv:2407.11540 (2024).
[15] M. Salmè, et al., Beyond the Generative Learning Trilemma: Generative Model Assessment in Data Scarcity Domains, arXiv preprint arXiv:2504.10555 (2025).
[16] F. Di Feola, et al., Texture-Aware StarGAN for CT data harmonisation, arXiv preprint arXiv:2503.15058 (2025).
[17] F. Di Feola, et al., Multi-scale texture loss for CT denoising with GANs, AI Open (2025).
[18] E. M. Ayllón, et al., Can Foundation Models Really Segment Tumors? A Benchmarking Odyssey in Lung CT Imaging, in: 2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2025, pp. 375–380.
[19] E. Cordelli, et al., Time-window SIQR analysis of COVID-19 outbreak and containment measures in Italy, in: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2020, pp. 277–282.
[20] M. Mantegna, et al., Benchmarking GAN-Based vs Classical Data Augmentation on Biomedical Images, in: International Conference on Pattern Recognition, Springer, 2024, pp. 92–104.
[21] A. Rofena, et al., A deep learning approach for virtual contrast enhancement in Contrast Enhanced Spectral Mammography, Computerized Medical Imaging and Graphics 116 (2024) 102398.
[22] A. Rofena, et al., Augmented Intelligence for Multimodal Virtual Biopsy in Breast Cancer Using Generative Artificial Intelligence, arXiv preprint arXiv:2501.19176 (2025).
[23] A. Rofena, et al., Lesion-Aware Generative Artificial Intelligence for Virtual Contrast-Enhanced Mammography in Breast Cancer, arXiv preprint arXiv:2505.03018 (2025).
[24] V. Guarrasi, et al., Whole-Body Image-to-Image Translation for a Virtual Scanner in a Healthcare Digital Twin, in: 2025 IEEE 38th International Symposium on Computer-Based Medical Systems, 2025, pp. 528–534.
[25] M. Salmè, et al., Evaluating Vision Language Model Adaptations for Radiology Report Generation in Low-Resource Languages, arXiv preprint arXiv:2505.01096 (2025).
[26] D. Molino, et al., Any-to-Any Vision-Language Model for Multimodal X-ray Imaging and Radiological Report Generation, 2025. URL: https://arxiv.org/abs/2505.01091. arXiv:2505.01091.
[27] D. Molino, et al., MedCoDi-M: A Multi-Prompt Foundation Model for Multimodal Medical Data Generation, 2025. URL: https://arxiv.org/abs/2501.04614. arXiv:2501.04614.
[28] K. Mogensen, et al., An optimized ensemble search approach for classification of higher-level gait disorder using brain magnetic resonance images, Computers in Biology and Medicine 184 (2025) 109457.
[29] F. Ruffini, et al., Multi-Dataset Multi-Task Learning for COVID-19 Prognosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 251–261.
[30] V. Guarrasi, et al., Multi-objective optimization determines when, which and how to fuse deep networks: An application to predict COVID-19 outcomes, Computers in Biology and Medicine 154 (2023) 106625.
[31] V. Guarrasi, et al., Optimized fusion of CNNs to diagnose pulmonary diseases on chest X-Rays, in: International Conference on Image Analysis and Processing, Springer, 2022, pp. 197–209.
[32] L. Tronchin, et al., Beyond a Single Mode: GAN Ensembles for Diverse Medical Data Generation, arXiv preprint arXiv:2503.24258 (2025).
[33] V. Guarrasi, et al., Pareto optimization of deep networks for COVID-19 diagnosis from chest X-rays, Pattern recognition 121 (2022) 108242.
[34] A. N. Caragliano, et al., Doctor-in-the-Loop: An explainable, multi-view deep learning framework for predicting pathological response in non-small cell lung cancer, Image and Vision Computing (2025) 105630.

[35] A. N. Caragliano, et al., Multimodal Doctor-in-the-Loop: A Clinically-Guided Explainable Framework for Predicting Pathological Response in Non-Small Cell Lung Cancer, arXiv preprint arXiv:2505.01390 (2025).

[36] V. Guarrasi, et al., Multimodal explainability via latent shift applied to COVID-19 stratification, Pattern Recognition 156 (2024) 110825.

[37] E. Cordelli, et al., Making AI trustworthy in multimodal and healthcare scenarios, in: CEUR Workshop Proceedings, volume 3486, 2023, pp. 353–358.