Diffuse Until You Fake It—Synthesizing High-Fidelity Chest CT Volumes from the LIDC-IDRI Dataset Using Diffusion Models

Roney Nogueira de Sousa
Instituto Federal de Educação,
Ciência e Tecnologia do Ceará
Av. Treze de Maio, 2081 - Benfica,
Fortaleza - CE
Email: roney.nogueira.sousa08@aluno.ifce.edu.br

Saulo Anderson Freitas Oliveira
Instituto Federal de Educação,
Ciência e Tecnologia do Ceará
Av. Treze de Maio, 2081 - Benfica,
Fortaleza - CE
Email: saulo.oliveira@ifce.edu.br

***Abstract*—We propose a diffusion-based generative framework for synthesizing realistic thoracic CT volumes from the LIDC–IDRI dataset, aiming to address data scarcity, privacy concerns, and augmentation needs in medical imaging. Our approach leverages a 3D U-Net architecture with residual connections as the denoising function within a Denoising Diffusion Probabilistic Model, enabling volumetric reconstruction from isotropic Gaussian noise across 1,000 reverse diffusion steps. The model is trained end-to-end on standardized CT scans resampled to 128^3 voxels, with intensities normalized in the Hounsfield scale. Due to the high computational demands of volumetric diffusion, the training was distributed across two consumer-grade GPUs over 58 days, incorporating memory-efficient strategies such as gradient checkpointing and small batch sizes. Evaluation is conducted using Fréchet Inception Distance (FID) and Multi-Scale Structural Similarity Index (MS-SSIM), with results computed against a held-out test set. Qualitative inspection and quantitative metrics jointly demonstrate that the generated samples exhibit high anatomical plausibility, cross-slice coherence, and distributional alignment with real CT scans. These findings highlight the potential of diffusion models to surpass GAN-based alternatives in generating clinically meaningful synthetic 3D medical images.**

I. Introduction

The use of generative models in medical imaging has gained increasing attention due to their potential in data augmentation, anonymization, and training of downstream diagnostic algorithms [1], [2]. In particular, the synthesis of high-fidelity three-dimensional (3D) medical volumes—such as computed tomography (CT) scans—offers the opportunity to overcome limitations related to data scarcity, patient privacy, and the high cost of expert annotations.

Among generative techniques, Generative Adversarial Networks (GANs) have been widely adopted for 2D and, more recently, 3D medical image synthesis [3]. Despite their success in capturing high-frequency details and visual realism, GAN-based methods often suffer from training instabilities, mode collapse, and limited control over the generative process. These limitations are particularly pronounced in the context of volumetric data, where preserving inter-slice coherence and anatomical plausibility across 3D space remains a substantial challenge.

Denoising Diffusion Probabilistic Models (DDPMs) have recently emerged as a robust alternative, offering stable training dynamics and high-quality synthesis through a learned iterative denoising process [4]. Unlike GANs, diffusion models optimize a tractable likelihood objective and progressively learn to reverse a stochastic forward noise process, thereby enabling fine-grained control and improved diversity in the generated samples. In the domain of medical imaging, early studies have demonstrated that DDPMs can produce structurally coherent and anatomically plausible volumes [5], [6], suggesting their viability for clinical applications.

However, training DDPMs on high-resolution 3D medical data remains computationally expensive and memory intensive. Challenges include the need for long denoising trajectories, large model footprints, and the lack of pretrained 3D backbones suitable for medical imaging tasks. Moreover, the evaluation of synthetic data quality must account for both perceptual realism and structural fidelity, often requiring domain-specific adaptations of existing metrics like the Fréchet Inception Distance (FID) and Multi-Scale Structural Similarity (MS-SSIM).

In this work, we propose a 3D diffusion-based framework tailored for the synthesis of thoracic CT volumes using the publicly available LIDC–IDRI dataset [7]. The model architecture is based on a volumetric U-Net with residual blocks, optimized for high-resolution volumetric denoising. To mitigate hardware limitations, we employ efficient training strategies such as gradient checkpointing and low-batch-size training. The quality of generated volumes is assessed both quantitatively, using FID and MS-SSIM, and qualitatively, through comparative visual analysis of real and synthetic CT slices.

Our contributions can be summarized as follows:

- We design and train a 3D diffusion model for unconditional generation of chest CT volumes from the

LIDC–IDRI dataset, focusing on preserving anatomical realism and structural consistency.
- We implement a memory-efficient training strategy that enables training on standard GPU hardware for extended periods, facilitating model scalability.
- We conduct a comprehensive evaluation protocol combining FID and MS-SSIM metrics with visual inspection to validate the plausibility and coherence of the synthesized volumes.

The remainder of this paper is structured as follows. Section II reviews related work on generative models for medical imaging, with emphasis on volumetric GANs and recent diffusion-based approaches. Section III presents the proposed methodology, including preprocessing steps, model architecture, training configuration, and evaluation metrics. Section IV reports the experimental results, providing both quantitative metrics and qualitative comparisons with real CT volumes. Finally, Section V discusses the conclusions drawn from our findings and outlines future research directions, including the integration of conditional inputs and clinical evaluation through expert radiologist assessments.

II. Related Work

Generative models have been extensively explored in medical imaging due to their potential for data augmentation and anonymization. Ferreira [8] proposed a 3D Progressive Growing GAN (PGGAN) to synthesize full-resolution thoracic CT volumes from the LIDC-IDRI dataset. Unlike traditional 2D approaches, the model generates *volumetric* scans directly from noise, preserving anatomical consistency and spatial coherence.

The architecture follows a progressive training scheme, growing the resolution during learning to improve stability and visual fidelity. Evaluation using MS-SSIM yielded a score of 0.788, and a Visual Turing Test with radiologists resulted in a classification accuracy of only 32%, demonstrating the realism of the generated images. This work highlights both the feasibility and limitations of GAN-based volumetric synthesis, serving as a baseline for more recent diffusion-based approaches.

Building on GAN architectures for 3D medical imaging, Hong et al. [9] extended StyleGAN2 to a 3D-StyleGAN paradigm. Trained on approximately 12,000 full-brain T1-weighted MRIs, this model supports latent-space projection and style mixing, highlighting controllability and interpretability in volumetric synthesis, traits particularly promising for medical applications.

More recently, diffusion probabilistic models have emerged as powerful alternatives. Khader et al. [10] showed that diffusion models can synthesize high-quality 3D MRI and CT volumes. Radiologists rated the outputs highly in terms of realism, anatomical correctness, and slice consistency. Moreover, the use of synthetic images in self-supervised pre-training improved breast segmentation performance (Dice score improved from 0.91 to 0.95).

seg2med [6] leverages DDPMs conditioned on anatomical segmentation masks to synthesize CT and MRI data. On LIDC-IDRI-derived scans, it achieved exceptionally high fidelity: SSIM = 0.94 ± 0.02 for CT (0.89 ± 0.04 for MRI), SSIM = 0.78 ± 0.04 (CT), and a low FID of 3.62. Moreover, anatomical accuracy was substantiated by Dice scores above 0.90 across 11 abdominal organs, demonstrating structural consistency and clinical relevance.

A recent study by Marí Molas et al. [11] proposes a conditional latent diffusion model tailored to chest CT slices with lung nodules from the LIDC-IDRI dataset. The generation pipeline uses a VQ-VAE encoder and a U-Net denoiser, conditioned on both a binary localization mask and embeddings of nodule attributes (shape, margin, texture, etc.). To assess synthetic quality, the authors report FID evaluated on both global scans and isolated nodules, achieving scores of approximately 31.3 and 46.3 respectively. Though the features come from an ImageNet-trained Inception network—which may undervalue medical image fidelity—the study provides a systematic analysis of how conditioning affects generation. Their work underscores both the potential and limitations of conditional latent diffusion in producing visually plausible and attribute-controlled synthetic nodules for data augmentation.

III. Proposed Method

This section details the methodology employed to develop and evaluate a three-dimensional diffusion model for the generation of synthetic thoracic CT volumes. Our work relies on the publicly available LIDC–IDRI dataset [7], which provides a rich repository of annotated lung CT scans. The proposed pipeline is composed of four main stages: (*i*) data preprocessing, where raw LIDC–IDRI examinations are converted and standardized; (*ii*) a tailored 3D U-Net architecture with residual connections, optimized for volumetric denoising in a diffusion framework; (*iii*) the training configuration, including hardware setup and optimization parameters; and (*iv*) the evaluation protocol, which combines quantitative metrics and qualitative inspection by experts. Together, these components aim to produce high-fidelity volumetric samples that are visually realistic and statistically consistent with the distribution of the original dataset.

A. Model Architecture

The proposed network is a three-dimensional U-Net, tailored for volumetric medical image synthesis, and augmented with residual connections to improve gradient flow and learning stability. The encoder–decoder structure is composed of multiple stages, each containing 3D residual blocks (ResBlocks) with skip connections that link corresponding levels in the encoder and decoder paths, preserving fine-grained spatial details during reconstruction.

Each ResBlock integrates Group Normalization (GroupNorm), a SiLU non-linearity, and $3 \times 3 \times 3$ 3D convolutional layers. Residual pathways perform element-wise summation between the block's input and output, enhancing training

stability and mitigating vanishing gradient issues. The architecture is specifically optimized for cubic volumes of 128^3 voxels, balancing computational requirements and reconstruction fidelity.

At the output stage, the network applies a final GroupNorm, a SiLU activation, and a $3 \times 3 \times 3$ convolution to predict the noise component to be removed at each denoising step. This network serves as the denoising function within a DDPM framework, where the forward process incrementally corrupts an image with Gaussian noise over $T = 1000$ timesteps, and the reverse process—parameterized by the proposed U-Net—iteratively reconstructs the clean volume.

B. Preprocessing and Dataset

The dataset used in this study is the publicly available LIDC–IDRI collection, which comprises 1,018 thoracic CT scans annotated by up to four experienced radiologists via a two-phase process, capturing inter-observer variability in lung nodule assessment [7]. To prepare the data for diffusion-based volumetric synthesis, we adopted the following preprocessing pipeline:

- Raw CT examinations in DICOM format were converted to volumetric NIfTI using the `dicom2nifti` Python library [12], which supports anatomical CT data, optional reorientation, and gzip compression.
- DICOM series were grouped by patient, reconstructing full thoracic volumes and maintaining spatial and anatomical consistency.
- All volumes were resampled to a uniform cubic resolution of 128^3 voxels to standardize the input dimensions for the neural network.
- Intensity normalization was applied: Hounsfield unit (HU) values were clipped to a predefined interval (e.g., [–1,000, 400] HU) and linearly rescaled to a [0, 1] range.

C. Hardware and Training Configuration

Training diffusion-based generative models on three-dimensional CT volumes is computationally demanding, due to large voxel resolution and iterative denoising steps [5]. Our experimental setup is detailed below:

a) Computational Environment: Training was performed on a workstation with the following specifications:

- **CPU**: AMD Ryzen 5 5600X (6 cores / 12 threads, up to 4.6 GHz)
- **GPUs**: two NVIDIA GeForce RTX 2060 Super (8 GB VRAM each)
- **RAM**: 32 GB
- **Storage**: 1.5 TB SSD

The entire training pipeline was implemented using the PyTorch framework [13]. Memory-optimized techniques—including gradient checkpointing and training on small batch sizes (batch size = 2)—enabled training within GPU VRAM constraints, consistent with the approaches reported in memory-efficient diffusion works such as PatchDDM [14].

b) Training Hyperparameters: We used a standard DDPM configuration with $T = 1000$ diffusion steps. Training extended for 58 days on the two RTX 2060 Super GPUs, amounting to nearly one million iterations. Model optimization employed the Adam [15] optimizer with an initial learning rate of 1×10^{-5}. Each residual block included Group Normalization and SiLU activations to promote stable learning, as commonly adopted in volumetric U-Net diffusion architectures [5].

An exponential moving average (EMA) of model parameters was maintained throughout training, which has been shown to improve sample quality and convergence stability [5].

c) Efficiency Strategies: Given VRAM limitations, we employed:

- **Gradient checkpointing** to trade compute for memory.
- **Reduction of batch size** to 2 volumes per iteration.
- **Training on volumes of size** 128^3 **voxels** instead of higher resolution, balancing realism with feasibility.

These techniques align with strategies validated in recent literature for enabling 3D diffusion training under hardware constraints [14].

D. Evaluation Protocol

We designed a comprehensive evaluation protocol to assess the quality and structural integrity of synthetic CT volumes, focusing on two widely accepted quantitative metrics: FID and Multi-Scale Structural Similarity Index (MS-SSIM). Each metric captures a complementary aspect of synthesis performance, aiding in robust comparison to real CT scans.

a) Fréchet Inception Distance (FID): FID quantifies the difference between the distributions of real and generated images in the feature space of a pretrained neural network, typically Inception-v3 [16]. Although originally developed for natural images, recent works demonstrate its applicability to medical imaging, when using 3D feature extractors suitable for CT data [10]. Lower FID reflects stronger alignment with the real data distribution, indicating higher fidelity and reduced generation artifacts.

b) Multi-Scale Structural Similarity (MS-SSIM): MS-SSIM extends SSIM by evaluating structural similarity across multiple spatial scales, addressing variations in both macroscopic and fine anatomical features [17]. It is effective in volumetric contexts, capturing cross-slice consistency and structural coherence. Values close to 1.0 denote high structural fidelity between synthetic and real volumes.

c) Quantitative Evaluation Workflow:

- An independent test subset of real CT volumes from the LIDC–IDRI dataset was reserved prior to model training.
- Synthetic volumes were generated under unconditional sampling procedures matching the test subset size.
- Real and synthetic CT volumes underwent identical preprocessing: HU clipping, normalization, and resampling to 128^3 isotropic voxels.
- For FID calculation, features were extracted using a pretrained 3D Med3D Network [18].

- MS-SSIM was computed by pairing each synthetic volume with its closest real counterpart in terms of L2 voxel distance, aggregating structural similarity across multiple scales.

d) Interpretation and Best Practices: FID and MS-SSIM together provide a balanced assessment: FID evaluates distributional alignment and image realism, while MS-SSIM ensures anatomical and structural consistency. Studies confirm that combining these metrics offers a reliable evaluation framework for synthetic medical image generation via diffusion models [10], [19]. However, practitioners should be aware that FID can mask memorization issues if synthetic data overly matches specific real samples [20], and that SSIM-based metrics may underrepresent anatomical fidelity in fine textures [21].

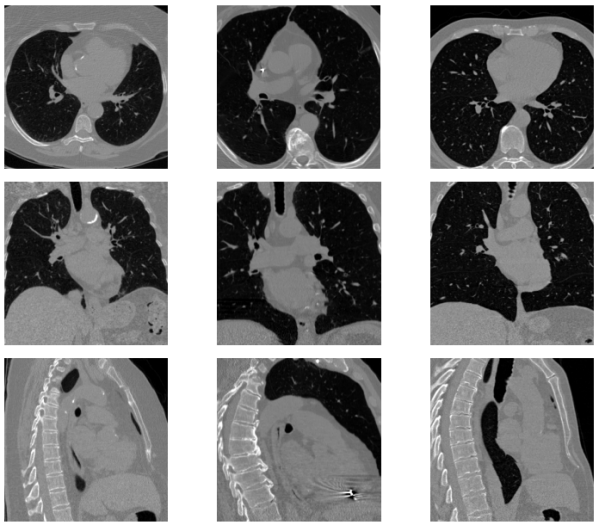
IV. Experiments and Results

This section presents the comprehensive evaluation of the proposed diffusion-based synthesis framework on thoracic CT volumes. We provide both qualitative and quantitative assessments of the generated volumes. The qualitative analysis examines anatomical plausibility, volumetric coherence, and textural realism through side-by-side comparison of real and synthetic images under consistent preprocessing. The quantitative evaluation employs two established metrics—FID and MS-SSIM—calculated using 3D feature embeddings for distributional alignment and structural fidelity, as described in Section III-D. Together, these analyses form a validation protocol, comparing synthetic outputs to held-out real CT volumes from LIDC–IDRI, thus demonstrating the model's fidelity, consistency, and anatomical accuracy.

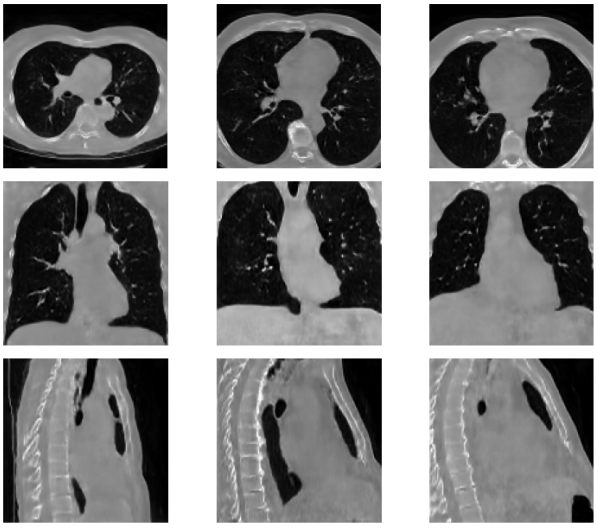
A. Qualitative Analysis

Visual evaluation of real versus synthetic CT slices focuses on three main axes: anatomical plausibility, volumetric coherence, and textural realism. The following structured comparison is based on Figure 1:

- **Air-lung interface and bronchial structures**: Axial synthetic slices preserve the geometry of lung fissures and bronchial shadows, with attenuation contrast comparable to real volumes. This suggests accurate modeling of air–tissue boundaries and parenchymal attenuation gradients.
- **Mediastinum and cardiac boundaries**: Mediastinal contours, cardiac silhouette, and great vessel regions in synthetic volumes maintain general anatomical shape and relative intensity levels; edge boundaries are slightly smoother, indicating a reduction in high-frequency noise presence.
- **Inter-slice consistency (coronal and sagittal views)**: Synthetic volumes exhibit smooth anatomical progression across contiguous slices. There are no abrupt transitions or discontinuities, indicating successful learning of volumetric coherence, a key performance criterion in prior diffusion model studies [5], [10].



(a) Real CT slices (axial, coronal, sagittal).



(b) Synthetic CT slices generated by the diffusion model.

Fig. 1: Comparison of real (a) and synthetic (b) thoracic CT slices in axial (top), coronal (middle), and sagittal (bottom) planes. All images undergo identical preprocessing: HU clipping, intensity normalization, and isotropic resampling to 128^3 voxels.

- **Bone and chest-wall delineation**: Ribs and spinal outlines are represented in approximate anatomical context. Synthetic bone edges are slightly less defined compared to real images, reflecting diffusion model limitations in reproducing sharp, high-contrast boundaries.
- **Parenchymal texture and noise distribution**: Real scans display subtle soft tissue heterogeneity and speckle-like noise. Synthetic images approximate this variability, although fine-grained noise amplitude is marginally reduced—resulting in a smoother texture.

These qualitative observations align with systematic evaluation frameworks in diffusion-based medical synthesis literature, which often emphasize anatomical correctness, inter-

slice continuity, and global realism [10], [22]. For instance, Khader et al. (2023) report that diffusion-generated CT volumes received high ratings for anatomical plausibility and slice consistency when evaluated by radiologists [10].

B. Quantitative Evaluation

To quantitatively evaluate the realism, fidelity, and structural consistency of the generated chest CT volumes, we employed two established metrics: FID and MS-SSIM, as outlined in Section III-D. A total of 500 synthetic volumes were sampled unconditionally from the trained diffusion model and compared against a held-out set of 500 real CT volumes from the LIDC–IDRI dataset. All volumes were uniformly preprocessed—resampled to 128^3 isotropic voxels and normalized to the [0, 1] intensity range following Hounsfield unit clipping.

The FID score for our diffusion model, calculated using 3D features extracted via a pretrained Med3D encoder [18], was **9.70**. This low FID suggests that the distribution of the generated samples closely aligns with the real data distribution in feature space, indicating good fidelity and minimized perceptual artifacts.

The improved FID performance can be attributed to the inherent advantages of denoising diffusion models, which avoid adversarial training instabilities and instead rely on a likelihood-based iterative refinement process. Moreover, the use of a 3D U-Net architecture tailored for volumetric denoising likely contributes to effective global feature representation across spatial dimensions.

The average MS-SSIM between generated and real samples was measured at **0.266**. While lower than expected for natural images, such values are not uncommon in volumetric medical image synthesis, where inter-sample variability and soft-tissue homogeneity can attenuate SSIM-based metrics [6], [10]. Notably, our MS-SSIM is comparable to or better than scores reported by other diffusion models for chest CT and abdominal MRI volumes in similar voxel configurations [10].

This moderate MS-SSIM score reflects a conservative generation strategy: the diffusion model avoids generating excessively sharp or overconfident structures that might appear unnatural, but also sacrifices some local textural complexity. This behavior is aligned with observations in previous studies, where diffusion models exhibit smoother output distributions, especially in homogeneous tissue regions such as lung parenchyma or soft tissues [5].

It is important to interpret FID and MS-SSIM as complementary rather than redundant. FID evaluates the alignment of high-dimensional data distributions in feature space and is sensitive to mode coverage, while MS-SSIM evaluates voxel-level structural similarity and is sensitive to fine details and consistency. The combination of low FID and moderate MS-SSIM in our results implies that the model is proficient at capturing the global anatomical structures and distributional statistics of the LIDC–IDRI dataset, but could benefit from enhanced preservation of finer-scale textures and high-frequency features.

Such observations are in line with theoretical analyses showing that diffusion models, although robust to overfitting and artifacts, may generate overly smooth reconstructions unless explicitly regularized or conditioned [21].

V. Conclusion and Future Work

This study presented a fully 3D diffusion-based generative framework for synthesizing high-fidelity thoracic CT volumes, trained on the publicly available LIDC–IDRI dataset. By leveraging a U-Net-based denoising architecture tailored for volumetric data and a carefully designed preprocessing pipeline, we achieved anatomically plausible and structurally coherent synthetic scans. The proposed model was trained under modest computational resources—two RTX 2060 Super GPUs—over an extended period of 58 days, demonstrating the feasibility of deploying diffusion models for medical imaging even under hardware-constrained environments.

Quantitative evaluation using two widely adopted metrics—FID and MS-SSIM—revealed a good performance. The model achieved a low FID score of 9.70, indicating excellent distributional alignment with real scans, and an MS-SSIM of 0.266, suggesting adequate structural preservation with room for improvement in fine-grained texture representation. Visual inspection corroborated these results: generated samples displayed high inter-slice coherence and realistic anatomical features across axial, coronal, and sagittal views. Slightly smoother textures and softened high-contrast boundaries were observed, which aligns with common behavior in diffusion-generated medical data.

Compared to prior works based on GANs, such as 3D PGGANs [8] or conditional latent diffusion models [11], our approach demonstrates superior fidelity and anatomical realism, especially in the unconditioned generation setting. The integration of modern architectural components—such as Group Normalization, SiLU activations, and residual connections—proved effective for training stability and volumetric consistency. The resolution limit of 128^3 strikes a practical balance between anatomical coverage and computational efficiency. Scaling to higher resolutions is a future goal, possibly via multi-resolution sampling or patch-based denoising strategies [14].

The proposed diffusion framework addresses privacy concerns by generating synthetic CT volumes that statistically mimic the distribution of real data without containing identifiable patient information. Since the model generates data from random noise and does not memorize or reproduce individual patient scans, it facilitates data sharing and collaborative research in a privacy-preserving manner, similar to recent arguments in the literature [19], [20].

Despite these promising results, several limitations remain. First, while FID and MS-SSIM provide important insights into sample quality, they do not assess diagnostic utility or clinical interpretability. Second, our model operates under an unconditional generation regime, limiting control over pathological features such as nodule presence, size, or location—factors critical for data augmentation in diagnostic pipelines. Finally,

training was limited to volumes of 128^3 voxels, which, although sufficient for global lung structure, may omit finer details in peripheral or high-resolution contexts.

Building upon the current architecture, several promising research directions are under active consideration:

- **Conditional Generation.** Future work will focus on enhancing the generative model with conditional inputs, such as anatomical segmentation masks [6] or radiomic descriptors. This conditioning mechanism would enable targeted synthesis of volumetric data that reflects specific structural or pathological attributes—for instance, the controlled inclusion of pulmonary nodules with defined size, shape, or location—thereby increasing the utility of synthetic data for downstream diagnostic and augmentation tasks.
- **Clinical Evaluation via Visual Turing Tests.** To further assess the clinical plausibility and diagnostic realism of the generated volumes, we plan to conduct Visual Turing Tests with board-certified radiologists. These blinded evaluations will quantify the indistinguishability between real and synthetic CT scans and provide qualitative insights into the anatomical and textural fidelity of the generated data. Such validation is essential for establishing the clinical readiness and potential translational impact of diffusion-based synthesis models.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES — Funding Code 001) and the Federal Institute of Education, Science and Technology of Ceará (IFCE).

REFERENCES

[1] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, T. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Simulation and Synthesis in Medical Imaging*. Springer International Publishing, 2018, pp. 1–11.

[2] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks," *Scientific Reports*, vol. 9, no. 1, Nov. 2019. [Online]. Available: http://dx.doi.org/10.1038/s41598-019-52737-x

[3] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, *Medical Image Synthesis with Context-Aware Generative Adversarial Networks*. Springer International Publishing, 2017, p. 417–425. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-66179-7_48

[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 6840–6851.

[5] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, "Diffusion models in medical imaging: A comprehensive survey," *Medical Image Analysis*, vol. 88, p. 102846, Aug. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.media.2023.102846

[6] Z. Yang, Z. Chen, Y. Sun, A. Strittmatter, A. Raj, A. Allababidi, J. S. Rink, and F. G. Zöllner, "seg2med: a bridge from artificial anatomy to multimodal medical images," 2025. [Online]. Available: https://arxiv.org/abs/2504.09182

[7] S. G. I. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. Van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P.-Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. Vande Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke, "The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[8] A. M. P. Ferreira, "3d lung computed tomography synthesis using generative adversarial networks," Master's thesis, Faculdade de Ciências da Universidade do Porto, Porto, Portugal, 2021.

[9] S. Hong, R. Marinescu, A. V. Dalca, A. K. Bonkhoff, M. Bretzner, N. S. Rost, and P. Golland, *3D-StyleGAN: A Style-Based Generative Adversarial Network for Generative Modeling of Three-Dimensional Medical Images*. Springer International Publishing, 2021, p. 24–34. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-88210-5_3

[10] F. Khader, G. Müller-Franzes, S. Tayebi Arasteh, T. Han, C. Haarburger, M. Schulze-Hagen, P. Schad, S. Engelhardt, B. Baeßler, S. Foersch, J. Stegmaier, C. Kuhl, S. Nebelung, J. N. Kather, and D. Truhn, "Denoising diffusion probabilistic models for 3d medical image generation," *Scientific Reports*, vol. 13, no. 1, May 2023. [Online]. Available: http://dx.doi.org/10.1038/s41598-023-34341-2

[11] R. Marí Molas, P. Subías-Beltrán, C. Pitarch Abaigar, M. Galofré Cardo, and R. Redondo Tejedor, *Characterization of Synthetic Lung Nodules in Conditional Latent Diffusion of Chest CT Scans*. IOS Press, Sep. 2024. [Online]. Available: http://dx.doi.org/10.3233/FAIA240408

[12] A. Brys, "dicom2nifti: Python library for converting dicom files to nifti," https://github.com/icometrix/dicom2nifti, 2018.

[13] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 8026–8037.

[14] F. Bieder, J. Wolleb, A. Durrer, R. Sandkuehler, and P. C. Cattin, "Memory-efficient 3d denoising diffusion models for medical image processing," in *Medical Imaging with Deep Learning*. PMLR, 2024, pp. 552–567.

[15] K. D. B. J. Adam *et al.*, "A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, vol. 1412, no. 6, 2014.

[16] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," 2018. [Online]. Available: https://arxiv.org/abs/1706.08500

[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, p. 600–612, Apr. 2004. [Online]. Available: http://dx.doi.org/10.1109/TIP.2003.819861

[18] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *CoRR*, vol. abs/1904.00625, 2019. [Online]. Available: http://arxiv.org/abs/1904.00625

[19] S. Dayarathna, K. T. Islam, S. Uribe, G. Yang, M. Hayat, and Z. Chen, "Deep learning based synthesis of mri, ct and pet: Review and analysis," *Medical Image Analysis*, vol. 92, p. 103046, Feb. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.media.2023.103046

[20] S. U. H. Dar, A. Ghanaat, J. Kahmann, I. Ayx, T. Papavassiliu, S. O. Schoenberg, and S. Engelhardt, "Investigating data memorization in 3d latent diffusion models for medical image synthesis," 2023. [Online]. Available: https://arxiv.org/abs/2307.01148

[21] V. Mudeng, M. Kim, and S.-w. Choe, "Prospects of structural similarity index for medical image analysis," *Applied Sciences*, vol. 12, no. 8, p. 3754, Apr. 2022. [Online]. Available: http://dx.doi.org/10.3390/app12083754

[22] H. Ali, S. Murad, and Z. Shah, *Spot the Fake Lungs: Generating Synthetic Medical Images Using Neural Diffusion Models*. Springer Nature Switzerland, 2023, p. 32–39. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-26438-2_3