From Baseline to Future CT: Large-Scale Diffusion Pretraining for Single-Scan IPF Progression Prediction

Niccolò McConnell[1], Yunjie Chen[1], Shanshan Wang[1], Daisuke Yamada[1], Ryoko Egashira[1], Adam Szmul[1], Shahab Aslani[1], John McCabe[1], Tianqi Yang[1], Helen McGregor[1], Pardeep Vasudev[1], Mehran Azimbagirad[1], Mark Jones[2], Bart Vanaudenaerde[3], Coline van Moorsel[4], Wim Wuyts[3], SUMMIT Consortium[1], Nesrin Mogulkoc[5], Paul Taylor[1], Sam M. Janes[1], Daniel C. Alexander[1], and Joseph Jacob[1]

[1] University College London, London, UK
niccolo.mcconnell.17@ucl.ac.uk
[2] University of Southampton, Southampton, UK
[3] KU Leuven, Leuven, Belgium
[4] St Antonius Ziekenhuis, Nieuwegein, The Netherlands
[5] Ege University Hospital, Izmir, Turkey

Abstract. Idiopathic pulmonary fibrosis (IPF) exhibits variable CT progression, from subtle interstitial changes to overt architectural distortion. Longitudinal imaging data is often sparse and confounded by non-rigid deformation, making accurate delineation of disease progression from single timepoint data a fundamental clinical imperative. We investigate single-scan IPF progression in thoracic CT by synthesising a future scan conditioned on the baseline scan, time horizon and patient covariates. To preserve fine lung structure under limited disease-specific longitudinal data, we decouple a transferable thoracic generative/longitudinal prior from disease-specific adaptation. We first learn a detail-preserving latent representation using patch-based autoencoding with overlap-average stitching. We then pretrain a covariate-conditioned latent diffusion model and longitudinal ControlNet on a large-scale thoracic CT screening dataset (>30,000 scans; >12,000 patients), and finally adapt the model to IPF progression synthesis with lung-focused patch-decoded supervision. Across internal IPF cohorts and a fully held-out external cohort, the screening-pretrained configuration achieves the strongest in-lung fidelity and lowest feature-space distance, and significantly outperforms training the progression adaptor without longitudinal pretraining. These results support thoracic diffusion pretraining as a reusable generative foundation for data-limited longitudinal disease modelling.

Keywords: Longitudinal CT synthesis · Idiopathic pulmonary fibrosis · Latent diffusion models · ControlNet · Thoracic foundation models.

1 Introduction

Idiopathic pulmonary fibrosis (IPF) is a lethal, progressive interstitial lung disease in which fibrotic remodelling and architectural distortion are largely irreversible once established [11]. CT features associated with IPF include peripheral reticulation and ground-glass opacities that precede later-stage features such as traction bronchiectasis and honeycombing, which together inform prognosis and treatment decisions [18]. Additionally, longitudinal CT in routine care is often sparse and irregular, with repeat imaging limited by radiation burden and variable clinical pathways across centres. Available follow-up data is further confounded by acquisition variability and non-rigid deformation from respiration and patient positioning [1]. A key unmet clinical challenge in IPF management centres on accurately predicting an individual's likely rate of disease progression. A model that synthesises future thoracic CTs from a single baseline scan could support trajectory analysis and progression-aware clinical management decisions.

Recent work has explored time-conditioned medical image synthesis and disease progression modelling with deep generative models [16,17,27]. Regarding IPF progression modelling, trajectory-centric approaches such as 4D-VQ-GAN [27] require multiple longitudinal scans for interpolation/trajectory completion. Latent diffusion with ControlNet-style conditioning [19,26,3,2] can produce high-fidelity conditioned generation, but is often limited to low spatial resolution due to computational cost – problematic for IPF where cues are subtle and spatially localised [12,10]. Furthermore, IPF progression is inherently volumetric, with fibrotic remodelling and traction bronchiectasis extending along branching airways and vessels. Unlike a recent diffusion model that explores slice-based (2.5D) synthesis [6], we perform full-volume synthesis in a high-resolution 3D latent space, which aims to preserve anatomical and volumetric consistency.

A further bottleneck relates to the limited size and heterogeneity of imaging data available in longitudinal IPF cohorts. Yet high-capacity 3D generative models are data-hungry and difficult to train robustly from scratch [7]. In contrast, large thoracic screening repositories comprise CT at scale [24], with broad anatomical and acquisition variability and repeat imaging that can support learning a transferable thoracic generative and longitudinal prior. We therefore learn this thoracic prior at scale and adapt it to IPF progression on smaller cohorts, transferring across differences in disease composition and acquisition context between pretraining and target data.

Contributions. We present a 3D latent generative framework for single-scan IPF progression prediction in thoracic CT: (i) large-scale thoracic screening pretraining that learns a transferable generative/longitudinal prior and improves fine-tuning on longitudinal IPF cohorts over training the ControlNet adaptor from scratch; (ii) a baseline-conditioned latent diffusion model with a longitudinal ControlNet to synthesise follow-up CT from one baseline scan with time-horizon and covariate inputs; and (iii) a tractable high-resolution design that preserves

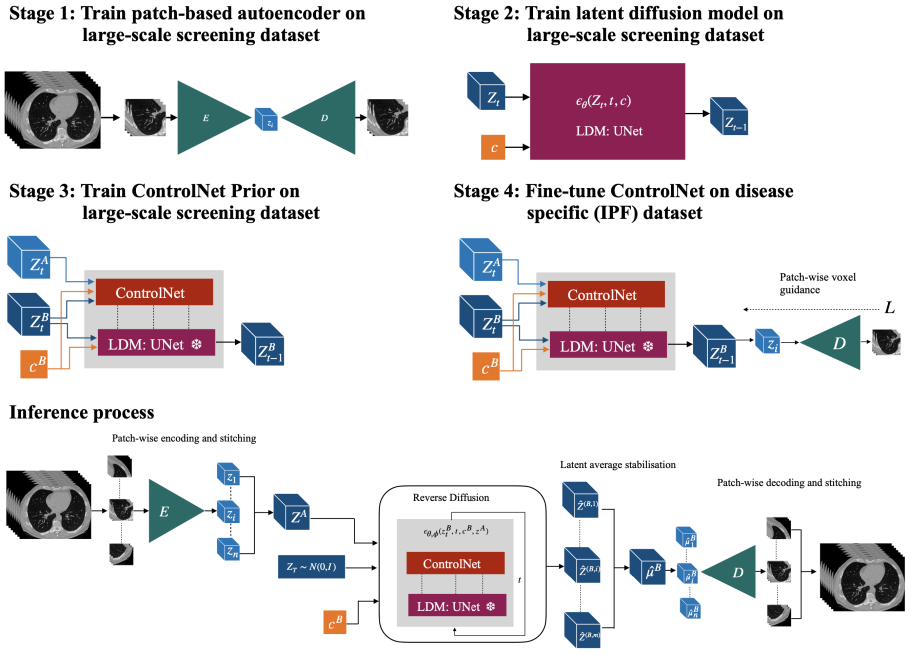

Fig. 1. Overview of the proposed training pipeline and inference process.

fine lung detail via lung-focused latent/decoded supervision and overlap-average stitching.

2 Method

We propose a four-stage pipeline for full-volume (256^3) future thoracic CT synthesis (Fig. 1) that preserves fine lung structure under compression and supports adaptation to a longitudinal IPF dataset. We decouple large-scale thoracic pretraining from IPF adaptation: (1) patch-based autoencoding; (2) covariate-conditioned latent diffusion pretraining [19] on a large-scale lung cancer screening dataset; (3) ControlNet pretraining [26] on longitudinal screening pairs; and (4) IPF-specific fine-tuning with lung-focused patch-decoded supervision. Inference uses overlap-average stitching and latent-average stabilisation.

2.1 Patch-based latent representation for thoracic CT

We train a 3D convolutional autoencoder (E, D) on the large-scale screening dataset to map $x \in \mathbb{R}^{1\times 256^3}$ to $z = E(x) \in \mathbb{R}^{C\times 64^3}$ (C=4; 4× downsampling). We avoid stronger compression because latents below 64^3 attenuated high-frequency lung structure (airways, vessels, peripheral fibrosis patterns) and

reduced parenchymal detail in senior board-certified radiologist review. The autoencoder uses a lightweight residual architecture with group normalisation and is trained to reconstruct the input CT scan (Sec. 3.1). To preserve subtle local parenchymal texture while keeping training tractable, we optimise (E, D) on lung-biased 96^3 patches rather than full 256^3 volumes. Patch-wise training was used for memory efficiency while maintaining local high-frequency supervision. Full-volume 256^3 synthesis is obtained at inference by overlap-average stitching. We use

$$\mathcal{L}_{\mathrm{AE}} = \lambda_{\ell_1}\mathcal{L}_{\ell_1} + \lambda_{\mathrm{ssim}}\mathcal{L}_{\mathrm{SSIM}} + \lambda_{\nabla}\mathcal{L}_{\nabla} + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}} + \lambda_{\mathrm{GAN}}\mathcal{L}_{\mathrm{GAN}}, \tag{1}$$

where $\mathcal{L}_{\ell_1}$, $\mathcal{L}_{\mathrm{SSIM}}$, and $\mathcal{L}_{\nabla}$ enforce voxel/edge fidelity, $\mathcal{L}_{\mathrm{KL}}$ regularises the posterior, and $\mathcal{L}_{\mathrm{GAN}}$ is a 3D PatchGAN [5] loss on decoded patches for high-frequency realism. The discriminator is introduced after a short warm-up with gradual weight ramping.

For full-volume diffusion, each 256^3 scan is encoded with sliding-window 96^3 crops (25% overlap) and stitched by overlap-average accumulation using posterior means. Lung masks are downsampled to latent resolution for later lung-focused weighting.

2.2 Large-scale latent diffusion pretraining

We pretrain a latent diffusion model on full-volume latents $z \in \mathbb{R}^{4\times 64^3}$ from the large-scale CT screening dataset to learn a generic thoracic prior spanning anatomical and acquisition variability. The denoiser ϵ_θ is conditioned on covariates $\mathbf{c} = \{age, sex, height, bmi, pack\_years, lung\_vol\}$ [18,22], with a global latent scaling factor s for stability ($\tilde{z} = s\,z$). Using a DDPM schedule [4] (T=1000), we train a 3D U-Net denoiser with

$$\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{z,\epsilon,t}\Big[\epsilon - \epsilon_\theta(\tilde{z}_t, t, \mathbf{c})_2^2\Big], \tag{2}$$

with lung-focused spatial weighting on the latent grid (from downsampled lung masks).

2.3 ControlNet for progression modelling and disease transfer

Given a baseline–follow-up latent pair (z_0, z_1) with interval Δt, we diffuse the follow-up to obtain $\tilde{z}_t$ and condition a ControlNet branch on

$$\mathbf{u} = \mathrm{concat}(z_0, \Delta t \cdot \mathbf{1}) \in \mathbb{R}^{(C+1)\times 64^3}, \tag{3}$$

where Δt is broadcast as a constant time map. ControlNet predicts residual features

$$\{\Delta\mathbf{h}^{(\ell)}\}, \Delta\mathbf{h}^{(\mathrm{mid})} = \mathrm{ControlNet}_\phi(\tilde{z}_t, t, \mathbf{c}, \mathbf{u}), \tag{4}$$

which are injected into the frozen diffusion U-Net during noise prediction,

$$\hat{\epsilon} = \epsilon_\theta(\tilde{z}_t, t, \mathbf{c}; \{\Delta\mathbf{h}^{(\ell)}\}, \Delta\mathbf{h}^{(\mathrm{mid})}), \tag{5}$$

optimising ϕ with θ fixed to preserve the screening-pretrained thoracic prior.

Stages 3 and 4 use the same ControlNet architecture and frozen diffusion backbone: Stage 3 pretrains a generic longitudinal adaptor on screening pairs, and Stage 4 fine-tunes it on IPF. During adaptation, in addition to $\mathcal{L}_{\text{diff}}$, we use latent change-consistency regularisers $\mathcal{L}_{\Delta} = (\hat{z}_1 - z_0) - (z_1 - z_0)_1$ and $\mathcal{L}_{\nabla} = \nabla\hat{z}_1 - \nabla z_{11}$, where $\hat{z}_1$ is reconstructed from $\hat{\epsilon}$ via the scheduler.

To improve local parenchymal detail while keeping 256^3 supervision tractable, we add patch-decoded voxel-level losses during adaptation. We sample K latent patches (default K=3, p=24), decode them with the frozen decoder, and optimise voxel ℓ_1, voxel-gradient, and blurred ℓ_1 (Gaussian-smoothed) losses. The blurred term emphasises coarse-scale anatomical agreement, reducing sensitivity to small residual baseline–follow-up misregistration. Sampling is lung-biased to focus supervision on clinically relevant parenchyma.

We optimise

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \lambda_{\Delta}\mathcal{L}_{\Delta} + \lambda_{\nabla}\mathcal{L}_{\nabla} + \lambda_{\text{vox}}\mathcal{L}_{\text{vox}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}} + \lambda_{\text{blur}}\mathcal{L}_{\text{blur}} + \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}}. \quad (6)$$

For *Proposed*, $\lambda_{\text{GAN}} = 0$; the adversarial term is enabled only in *With GAN*.

2.4 Inference and latent-average stabilisation

Given baseline scan x_0, we obtain z_0 via sliding-window encoding and overlap-average stitching, then sample a follow-up latent with denoising diffusion implicit model (DDIM) [23] using exponential moving average (EMA) weights [14]. Let $z_T \sim \mathcal{N}(0, I)$ and denote the conditional reverse process by

$$\hat{z}_1 = \mathcal{I}_{\phi}(z_T, z_0, \Delta t, \mathbf{c}). \quad (7)$$

To reduce sampling variance, we average m independent latent samples [17],

$$\mu_1 \approx \frac{1}{m}\sum_{i=1}^{m} \mathcal{I}_{\phi}(z_T^{(i)}, z_0, \Delta t, \mathbf{c}), \quad (8)$$

then decode μ_1 with patch-wise overlap-average decoding.

Implementation details. Training used AdamW on NVIDIA A6000 GPUs with LR/batch/epochs for AE/LDM/pretrain-CN/IPF-CN of 10^{-4}/6/120, $3{\times}10^{-5}$/16/100, $6{\times}10^{-5}$/4/50, and $4{\times}10^{-6}$/4/30; AE and LDM/CN channels were $(64, 128, 256)$ and $(128, 256, 512)$, respectively, with two ResBlocks, DDIM=300, and m=5. Code is available at `https://github.com/niccolo246/baseline-to-future-ct`.

3 Experiments

3.1 Data, preprocessing, and evaluation

We evaluate on longitudinal IPF CT from the Netherlands, Turkey, Belgium, and the UK, excluding acute exacerbation, severe infection, and post-transplant

scans. The Netherlands/Turkey/Belgium data are used for training/internal testing and the UK cohort for external testing; pretraining and IPF adaptation use disjoint sources. Foundational pretraining uses the large-scale thoracic CT screening dataset SUMMIT (31,110 scans; 12,443 patients; repeat scans up to 3 years) to learn a broad thoracic generative/longitudinal prior across diverse comorbidities and progression patterns. IPF adaptation/evaluation uses only the IPF dataset: internal training includes 629 patients (1,286 scans; 986 pairs), internal testing 158 patients (316 scans; 220 pairs), and external testing 62 patients (70 scans; 17 pairs), providing a site-shift evaluation.

Ethics statement. All scans were de-identified and analysed under institutional ethics and data-governance approvals at the contributing centres; only de-identified imaging and non-identifying covariates were used for model development.

Pair construction, preprocessing, and evaluation frame. We form baseline–follow-up pairs (x_0, x_1) with $0 \leq \Delta t \leq 3$ years. Lung masks are obtained with TotalSegmentator [25]. All scans are resampled to isotropic 1.35 mm spacing, placed on a fixed 256^3 grid, centred by the lung-mask centroid, and intensity-clipped to $[-1200, 800]$ and normalised to $[0, 1]$. Follow-up scans are non-rigidly registered to baseline using Elastix [9].

Metrics. Unless stated otherwise, metrics are computed *in-lung* using the follow-up lung mask. We report SSIM, MSE, and PSNR on normalised volumes; perceptual distance ("ViT feat.") as embedding MSE from an open-source volumetric lung CT foundation encoder [13]; and absolute airway volume error (mL). For the primary comparison (*Proposed* vs *Scratch*), we use paired one-sided Wilcoxon signed-rank tests across pairs with metric-specific alternatives and Holm correction.

3.2 Comparison models

We perform IPF-adaptation ablations while keeping the screening-pretrained autoencoder and diffusion backbone fixed. *Diff. only* optimises only the latent diffusion noise-prediction loss $\mathcal{L}_{\text{diff}}$, removing all auxiliary latent and patch-decoded losses. *No Blur* removes only $\mathcal{L}_{\text{blur}}$, *With GAN* uses a non-zero adversarial loss, and *Scratch* trains the ControlNet directly on IPF without longitudinal screening-pair pretraining. As related single-scan IPF follow-up synthesis methods differ in input assumptions (e.g., text-based lung-function conditioning) and generation regimes (e.g., slice-based vs full-volume) [6], we compare against two baselines in the same 64^3 latent-space and conditioning setting: (i) a deterministic 3D U-Net [20] regressor mapping $(z_0, \Delta t, \mathbf{c}) \to \hat{z}_1$, with Δt as a constant channel and FiLM conditioning [15]; and (ii) a latent conditional VAE [8,21] modelling $p(z_1 \mid z_0, \Delta t, \mathbf{c})$ with a learned conditional prior/posterior, matched to our latent resolution and scaling.

Table 1. Internal test set results (mean ± std). Higher is better for SSIM and PSNR; lower is better for all other metrics. MSE and ViT distance are reported in units of $\times 10^{-2}$. *AE ceiling* denotes deterministic reconstruction of the ground-truth follow-up via the frozen autoencoder.

Method	SSIM↑	MSE↓	PSNR↑	ViT dist.↓	Airway err. (mL)↓
Ours + ablations					
Proposed	**0.947 ± 0.019**	**1.21 ± 0.69**	**31.01 ± 2.14**	**1.89 ± 1.87**	**18.27 ± 15.95**
With GAN	0.941 ± 0.020	2.14 ± 0.94	27.29 ± 1.86	6.14 ± 3.56	20.27 ± 16.57
No Blur	0.940 ± 0.020	2.30 ± 0.93	26.91 ± 1.75	7.51 ± 4.38	27.48 ± 22.80
Diff. only	0.933 ± 0.021	3.22 ± 1.42	25.53 ± 1.99	7.94 ± 4.12	20.65 ± 16.88
Scratch	0.916 ± 0.024	3.54 ± 1.16	24.93 ± 1.44	6.40 ± 3.28	48.32 ± 29.54
Comparison methods					
Latent-UNet	0.930 ± 0.021	2.60 ± 0.40	28.02 ± 1.55	4.76 ± 0.49	66.36 ± 29.26
Latent-CVAE	0.936 ± 0.020	1.92 ± 0.45	30.95 ± 1.94	2.39 ± 0.42	61.20 ± 28.67
AE ceiling	0.981 ± 0.008	0.234 ± 0.135	37.12 ± 2.58	0.107 ± 0.060	–

4 Results

4.1 Internal test set

Table 1 summarises internal in-lung/anatomical fidelity. *Proposed* performs best overall in image fidelity and perceptual alignment, with competitive airway fidelity. The dominant effect is longitudinal pretraining: *Scratch* degrades all metrics, indicating that cohort-scale IPF data alone is insufficient for stable high-capacity 3D diffusion progression modelling. The primary comparison (*Proposed* vs *Scratch*) was significant for all pre-specified fidelity/perceptual metrics (SSIM, MSE, PSNR, ViT feature distance; paired one-sided Wilcoxon signed-rank tests; all Holm-corrected $p < 0.001$). Ablations that remove auxiliary supervision (*Diff. only*), remove coarse blurred supervision (*No Blur*), or add adversarial refinement (*With GAN*) generally worsen global similarity and perceptual distance. Among comparison methods, latent cVAE outperforms the deterministic latent U-Net. The high *AE ceiling* suggests the remaining gap is dominated by progression modelling rather than decoding.

Qualitative comparisons (Fig. 2) mirror the quantitative trends: *Proposed* better preserves high-frequency parenchymal/airway detail and captures local fibrotic progression patterns (e.g., traction bronchiectasis), whereas comparison models tend to over-smooth changes and miss local architectural distortion.

4.2 External test set

Table 2 shows the same overall pattern on the held-out external site. *Proposed* remains strongest in global fidelity and perceptual alignment, while *Scratch* again degrades markedly across metrics. The primary comparison (*Proposed* vs *Scratch*) remains significant for the main fidelity/perceptual metrics (SSIM, MSE, PSNR, and ViT feature distance; paired one-sided Wilcoxon signed-rank tests; all Holm-corrected $p < 0.05$).

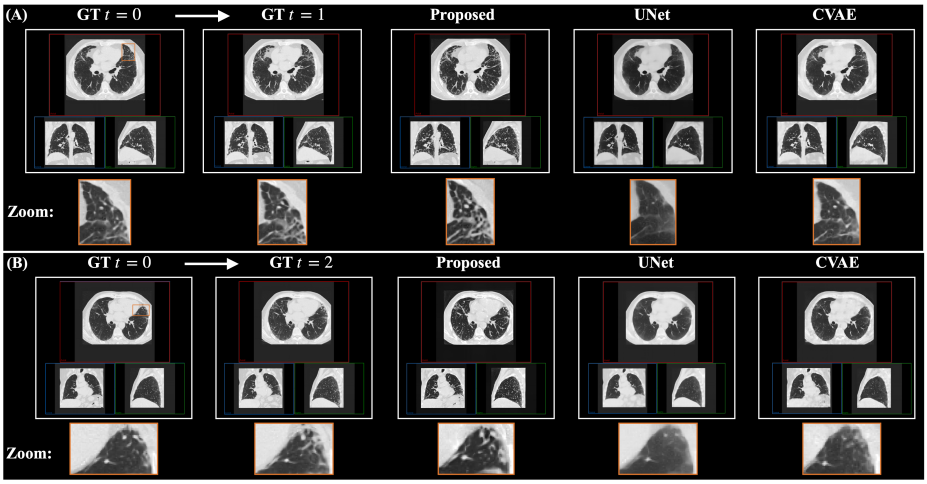

Fig. 2. Qualitative longitudinal thoracic CT synthesis. Two representative patient cases (A–B) are shown with baseline (GT $t = 0$), ground-truth follow-up (GT $t = 1$ or $t = 2$ years), and generated follow-up scans from *Proposed*, latent U-Net, and latent cVAE. Axial/coronal/sagittal views and zoomed regions highlight local parenchymal progression patterns.

5 Discussion

We presented a practical framework for 3D IPF progression modelling that decouples a transferable thoracic prior (learned at scale) from disease-specific adaptation (learned on limited cohorts). Across internal and external evaluations, the screening-pretrained configuration achieved the strongest in-lung fidelity and lowest feature-space distance, with consistent trends on the held-out external cohort. It also significantly outperformed *Scratch* on the primary fidelity/perceptual metrics (paired Wilcoxon signed-rank tests, Holm-corrected), supporting longitudinal pretraining as the key factor in robust high-capacity progression modelling. This comparison is conservative because all ablations retain the screening-pretrained autoencoder and diffusion backbone and differ only in the progression adaptor. Limitations include reliance on in-lung similarity rather than IPF-specific endpoints or formal radiologist scoring, an illustrative rather than definitive qualitative analysis, and a modest external cohort (17 pairs). Direct comparison with Jiang et al. [6] remains confounded by conditioning, dataset, and 2.5D versus 3D generation differences. Overall, these results support thoracic diffusion pretraining as a reusable generative foundation for longitudinal lung CT modelling and motivate extension to other lung diseases.

Acknowledgments. This work was funded in whole or in part by the Wellcome Trust [227835/Z/23/Z], the NIHR UCLH Biomedical Research Centre, and the UKRI Centre for Doctoral Training in AI-enabled Healthcare Systems [EP/S021612/1]. The SUMMIT Study is funded by GRAIL LLC through a research grant awarded to S.M.J.

Table 2. External test set results (mean ± std). Higher is better for SSIM and PSNR; lower is better for all other metrics. MSE and ViT distance are reported in units of $\times 10^{-2}$. *AE ceiling* denotes deterministic reconstruction of the ground-truth follow-up via the frozen autoencoder (upper bound under the chosen latent compression).

Method	SSIM↑	MSE↓	PSNR↑	ViT dist.↓	Airway err. (mL)↓
Ours + ablations					
Proposed	**0.947 ± 0.013**	**1.05 ± 0.45**	**30.21 ± 2.25**	**2.25 ± 1.19**	17.33 ± 14.51
With GAN	0.940 ± 0.011	1.93 ± 0.65	27.31 ± 1.38	5.17 ± 2.17	**12.40 ± 10.93**
No Blur	0.939 ± 0.011	2.13 ± 0.67	26.83 ± 1.18	6.15 ± 2.33	19.48 ± 15.01
Diff. only	0.934 ± 0.011	2.77 ± 1.02	25.78 ± 1.53	6.82 ± 2.42	26.91 ± 18.95
Scratch	0.915 ± 0.010	3.19 ± 0.92	25.06 ± 1.17	5.48 ± 2.14	34.67 ± 22.40
Comparison methods					
Latent-UNet	0.928 ± 0.008	2.25 ± 0.40	27.10 ± 1.67	3.48 ± 0.52	58.38 ± 18.69
Latent-CVAE	0.935 ± 0.009	1.59 ± 0.38	29.87 ± 1.85	4.36 ± 0.19	53.60 ± 19.91
AE ceiling	0.979 ± 0.007	0.290 ± 0.170	36.54 ± 3.80	0.0866 ± 0.0306	–

Disclosure of Interests. J.J. reports consultancy, advisory board, lecture fees, and/or grant funding from Boehringer Ingelheim, F. Hoffmann-La Roche, GlaxoSmithKline, NHSX, Takeda, Wellcome Trust, Microsoft Research, and Gilead Sciences, and is named on UK patent applications 2113765.8 and GB2211487.0. S.M.J. reports advisory board fees from Bard1 Lifescience, grant income from GRAIL Inc., lecture fees from Chiesi and AstraZeneca, unpaid advisory board membership for GRAIL, and a spouse employed by AstraZeneca. All other authors declare no competing interests relevant to the content of this article.

References

1. Elicker, B.M., Kallianos, K.G., Henry, T.S.: The role of high-resolution computed tomography in the follow-up of diffuse lung disease: Number 2 in the series "radiology" edited by nicola sverzellati and sujal desai. European Respiratory Review **26**(144) (2017)
2. Fang, Z., et al.: Conditional diffusion model for x-ray segmentation data generation. Journal of Artificial Intelligence Practice **7**(1), 7–10 (2024)
3. 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: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4430–4441. IEEE (2025)
4. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
6. Jiang, C., Xing, X., Nan, Y., Fang, Y., Zhang, S., Walsh, S., Yang, G., Shen, D.: A lung structure and function information-guided residual diffusion model for predicting idiopathic pulmonary fibrosis progression. Medical image analysis **103**, 103604 (2025)

7. Karras, T., Aittala, M., Lehtinen, J., Hellsten, J., Aila, T., Laine, S.: Analyzing and improving the training dynamics of diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 24174–24184 (2024)
8. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
9. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging **29**(1), 196–205 (2009)
10. Lee, E., Seo, J.B., Lee, H.J., Chae, E.J., Lee, S.M., Oh, S.Y., Kim, N.: Quantitative assessment of global and regional air trappings using non-rigid registration and regional specific volume change of inspiratory/expiratory ct scans: studies on healthy volunteers and asthmatics. Korean Journal of Radiology **16**(3), 632 (2015)
11. Martinez, F.J., Collard, H.R., Pardo, A., Raghu, G., Richeldi, L., Selman, M., Swigris, J.J., Taniguchi, H., Wells, A.U.: Idiopathic pulmonary fibrosis. Nature reviews Disease primers **3**(1), 1–19 (2017)
12. Martini, K., Jungblut, L., Sartoretti, T., Langhart, S., Yalynska, T., Nemeth, B., Frauenfelder, T., Euler, A.: Impact of radiation dose on the detection of interstitial lung changes and image quality in low-dose chest ct–assessment in multiple dose levels from a single patient scan. European Journal of Radiology **166**, 110981 (2023)
13. McConnell, N., Vasudev, P., Yamada, D., Cheng, D., Azimbagirad, M., McCabe, J., Aslani, S., Shahin, A.H., Zhou, Y., et al.: A computationally frugal, open-source chest ct foundation model for thoracic disease detection in lung cancer screening programmes. Communications Medicine (2026)
14. Morales-Brotons, D., Vogels, T., Hendrikx, H.: Exponential moving average of weights in deep learning: Dynamics and benefits. arXiv preprint arXiv:2411.18704 (2024)
15. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
16. Pombo, G., Gray, R., Cardoso, M.J., Ourselin, S., Rees, G., Ashburner, J., Nachev, P.: Equitable modelling of brain imaging by counterfactual augmentation with morphologically constrained 3d deep generative models. Medical Image Analysis **84**, 102723 (2023)
17. Puglisi, L., Alexander, D.C., Ravì, D., Initiative, A.D.N., et al.: Brain latent progression: Individual-based spatiotemporal disease progression on 3d brain mris via latent diffusion. Medical image analysis p. 103734 (2025)
18. Raghu, G., Remy-Jardin, M., Myers, J.L., Richeldi, L., Ryerson, C.J., Lederer, D.J., Behr, J., Cottin, V., Danoff, S.K., Morell, F., et al.: Diagnosis of idiopathic pulmonary fibrosis. an official ats/ers/jrs/alat clinical practice guideline. American journal of respiratory and critical care medicine **198**(5), e44–e68 (2018)
19. 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. pp. 10684–10695 (2022)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Advances in neural information processing systems **28** (2015)

22. Song, H., Sun, D., Ban, C., Liu, Y., Zhu, M., Ye, Q., Yan, W., Ren, Y., Dai, H.: Independent clinical factors relevant to prognosis of patients with idiopathic pulmonary fibrosis. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research **25**, 4193 (2019)
23. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
24. Team, N.L.S.T.R.: The national lung screening trial: overview and study design. Radiology **258**(1), 243–253 (2011)
25. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
26. 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. pp. 3836–3847 (2023)
27. Zhao, A., Xu, M., Shahin, A.H., Wuyts, W., Jones, M.G., Jacob, J., Alexander, D.C.: 4d-vq-gan: A world model for synthesizing medical scans at any time point for personalized disease progression modeling of idiopathic pulmonary fibrosis. In: Medical Imaging with Deep Learning (2025)