This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Synthesizing Longitudinal Cutaneous Neurofibroma Imaging for Digital-Twin Burden Quantification

Christoph Sadée[1*], Alex Dils[1,2*], Krish Sangani[2], Lillian Rubino[1], Alexander Ferenchick[1,3], Varun Wadhwa[4], Anushka Poddar[5], Tobenna Onyemeh[6], Zimuzo Onah[7], Dhruv Sharma[4], Jackson Bae[8], Max Van Puyvelde[1], Carlos Romo[9], Melinda Jen[10], Nkiru Onodugo[11], Ayesha Akinkugbe[12,13], Qinmei Xu[1], Qingtao Kong[14], Rui Yang[15], Aisha Sokunbi[16], Shaoxiong Yao[17], Haomiao Huang[18], Ifeoma Okoye[6,11,19], Olivier Gevaert[1**], Kavita Sarin[20**]

Abstract. Rationale: Longitudinal tracking of cutaneous neurofibromas (cNFs) is needed to quantify progression, assess treatment response, and develop medical digital twins. Such applications require time-resolved imaging, which is difficult to acquire across years to decades of cNF development. In the absence of public longitudinal cNF datasets, synthetic imaging offers a practical alternative.

Methods: We integrated a physics-grounded morphoelastic finite-element growth model with a geometry-conditioned FLUX.2 Klein adapter to generate synthetic longitudinal images from real cNFs. The full dataset comprised 279 public cNF images with 42,305 annotated lesion contours; the adapter was trained on 229 images with 38,040 contours. In two blinded reader studies, four physicians identified de novo synthetic lesions or lesions grown from baseline. In addition, the effect of synthetic data on burden estimation was evaluated by comparing ConvNeXt models trained on real data alone and combined real–synthetic data.

Results: The pipeline generated lesion-growth sequences across diverse skin images. In reader study 1, mean synthetic-lesion recall was 17.1%, and 49.4% of clicks landed on synthetic lesions. In reader study 2, pooled grown-lesion recall was 47.5%, and 42.9% of clicks landed on grown lesions. Synthetic augmentation improved burden estimation on held-out real images, increasing R^2 from 0.775 to 0.849.

Conclusions: Physics-guided generative AI can create plausible time-resolved cNF images without longitudinal training data, supporting development of burden and treatment-response tools and medical digital twins.

[1]Division of Computational Medicine, Department of Medicine, Stanford University; [2]University of California, Berkeley; [3]University of Pennsylvania; [4]University of California, Davis; [5]University of Maryland, College Park; [6]University of Nigeria, Nsukka, Center for Excellence for Clinical Trials; [7]University of Nigeria Teaching Hospital Ituku-Ozalla; [8]University of California, Los Angeles; [9]Johns Hopkins School of Medicine; [10]Children's Hospital of Philadelphia, Dermatology Section; [11]University of Nigeria Teaching Hospital; [12]University of Lagos; [13]Lagos University Teaching Hospital; [14]Department of Dermatology, Jinling Hospital, Affiliated Hospital of the Medical School of Nanjing University, Nanjing, China; [15]Department of Dermatology, Second Affiliated Hospital of Nanjing Medical University, Nanjing, China; [16]University of Benin Teaching Hospital; [17]Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign; [18]Matter Venture Partners; [19]University of Nigeria, Nsukka; [20]Department of Dermatology, Stanford University School of Medicine.
*Co-first authors; **Co-senior authors.

Keywords: Mechanistic modeling · Generative AI · Medical digital twins · Cutaneous neurofibroma

1 Introduction

Cutaneous neurofibromas (cNFs), the most common tumor manifestation of neurofibromatosis type 1 (NF1), are benign nerve sheath tumors that typically emerge around puberty and increase in number and size throughout adulthood. Patients may develop anywhere from a few lesions to thousands, resulting in pain, disfigurement, and psychosocial morbidity [16, 12]. Despite this burden, cNF progression remains poorly characterized because lesions develop slowly over years to decades. Longitudinal imaging is therefore needed to understand how individual lesions evolve and to develop tools for measuring disease burden and treatment response. More broadly, such data are foundational to medical digital twins—patient-in-silico models that combine diverse data streams, remain synchronized with the patient, and model disease trajectories [19].

The imaging data required for these applications, however, remain largely unavailable. Public dermatology datasets that include cNFs, such as Fitzpatrick 17k [7], comprise two-dimensional photographs that lack the aligned three-dimensional surface geometry needed to characterize lesion protrusion, shape, and volumetric burden. None provide the repeated, time-resolved acquisitions necessary for modeling lesion growth. Three-dimensional imaging of cNFs is feasible using commercial systems such as Vectra [11], but available scans are either not public or include few cNF cases; existing datasets are cross-sectional and were not designed specifically to capture cNFs [20]. Prospective collection is also difficult because directly observing meaningful progression may require years or decades. Beyond model training, synthetic datasets can accelerate patient-facing AI development by allowing teams to build interfaces, data infrastructure, and end-to-end pipelines before scarce clinical data are available, reserving real patient data for calibration and validation. Mechanistic simulation and generative modeling may therefore provide a practical means of creating the longitudinal imaging needed to develop and evaluate tools.

Mechanistic models offer one route to simulating cNF progression. Biophysical growth models—ranging from reaction–diffusion formulations of tumor spread to continuum-mechanics models of tissue growth—have been used to simulate tumor evolution and incorporated into patient-specific, image-guided digital twins of tumor growth and treatment response [24, 14]. Applied to cNFs, a biomechanical growth model can simulate how a lesion enlarges and deforms the surrounding skin as a function of tissue properties, thereby recovering the three-dimensional shape of the growing lesion over time. However, such models do not reproduce the visual appearance of real lesions, including texture, color, surface detail, and the transition between the lesion and the surrounding skin. Mechanistic modeling can therefore supply plausible growth geometry but not the photorealistic appearance required for an image-based digital twin.

Generative AI provides a complementary capability. Generative models can now synthesize photorealistic images [18] and have been applied to dermatology dataset synthesis [21]. Because these models learn appearance directly from images, they can be trained on the two-dimensional cNF photographs that already exist and can generate realistic lesions conditioned on a specified location or geometry. However, without longitudinal training data, a generative model has no principled basis for determining how lesion geometry should change over time.

Here, we combine the two. We couple a mechanistic biomechanical growth model of a cNF lesion with a geometry-conditioned generative model that renders photorealistic lesions: the mechanistic model supplies the temporal trajectory and three-dimensional lesion geometry, while the generative model supplies the per-lesion appearance the mechanistic model lacks. We condition the generator on monocular depth and surface normals, obtained with Depth Pro [3] and DSINE [1], which then translates the simulated geometry into realistic skin. This makes it possible to directly visualize how cNFs evolve—or how they would change after treatment or surgical removal. We train the generative component on existing two-dimensional cNF photographs, assess lesion and growth realism in two blinded physician reader studies, and show that synthetic lesions improve downstream burden quantification. Together, these components generate the synthetic longitudinal imaging needed to build a medical digital twin.

2 Methods

To support the integrated pipeline, we collected 279 images containing cNFs from seven public sources and manually annotated 42,305 lesion contours (Table 1). We partitioned the data at the image level into training, validation, and test sets. Additionally, to model post-surgical appearance, we curated 11 public post-operative images and manually annotated scar regions.

Public source	**Images**	**Lesions**
Fitzpatrick 17k [7]	167	29,759
Dermatology Reasoning Dataset [5]	54	5,879
DermNet image collections [4]	40	5,669
PubMed-indexed case-report figures [15]	7	830
JDD Inclusive Derm Atlas [9]	6	79
SCIN [23]	3	30
ISIC Archive [8]	2	59
Total	**279**	**42,305**

Table 1. Public images and manually segmented lesion contours.

Below, we outline the synthetic-generation pipeline, which can be broken into three steps: (1) estimate depth and surface normals; (2) use the mechanistic model to propagate lesion growth and surrounding-skin deformation; and (3)

render each updated geometry as a new photorealistic skin image. Repeating steps 2 and 3 produces a time series from an initial lesion.

To capture lesion surface geometry, we estimated a metric depth map from each image using Depth Pro [3] and a unit surface-normal field using DSINE [1]. We then reconstructed a three-dimensional skin-surface mesh from the depth map; the resulting image, mask, depth, normal, surface, and layered-mesh modalities are summarized in Figure 1.

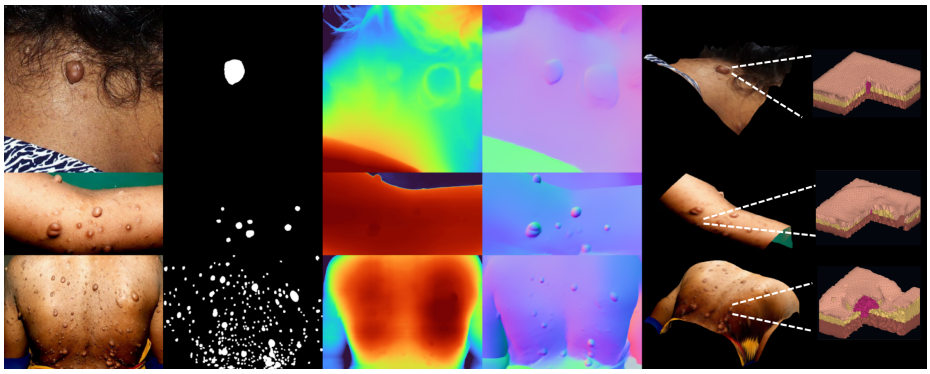

Fig. 1. Data modalities and geometric representations used in the pipeline. From left to right: input image, lesion mask, Depth Pro metric depth, DSINE surface normals, the reconstructed three-dimensional skin surface, and the layered tetrahedral mesh (dermis, subcutis, deep tissue, and embedded lesion) used by the mechanistic model.

The mechanistic component generated temporal disease trajectories by simulating single-lesion cNF growth in a layered three-dimensional skin mesh using a morphoelastic finite-growth model [17, 10]. The finite-element system solves for nodal displacements, equivalently the deformed positions of the tetrahedral-mesh vertices, while the deformation gradient $\mathbf{F}$ is computed element-wise from those nodal variables. The sequentially deformed surfaces were projected into image space to generate updated depth and surface-normal maps. Full formulation and solver details are provided in Appendices A.1–A.3.

To translate the simulated growth geometry into photorealistic skin images, we adapted FLUX.2 Klein [2], an inpainting model, to incorporate explicit surface geometry. In addition to the source RGB image, the model receives Depth Pro metric-depth and DSINE surface-normal maps that encode the simulated lesion and skin surface. This joint RGB-and-geometry conditioning preserves the color, texture, and surrounding context of the source image while rendering the lesion shape prescribed by the mechanistic simulation (Figure 2). Conditioning, training, and longitudinal-rendering details are provided in Appendices B.1–B.3.

To represent a treatment endpoint, we trained a separate adapter that replaces a selected lesion with the postoperative scar or pigmentary change expected after removal. This scar branch is invoked only for a simulated inter-

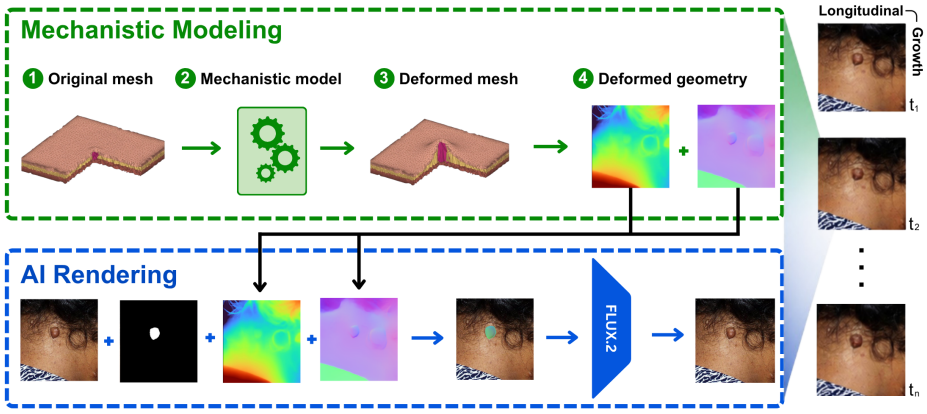

Fig. 2. Overview of the mechanistic–generative pipeline. A layered tissue mesh containing an embedded lesion (1) is passed through the morphoelastic growth model (2), producing a deformed mesh (3) whose surface is encoded as depth and surface-normal maps (4). These maps are composited into the conditioned model input, which renders the simulated geometry as photorealistic skin. Repeating this process across successive growth states produces a synthetic longitudinal image sequence.

vention, not during untreated growth; it uses a spatial mask without depth or surface-normal conditioning. Accordingly, the rightmost panel of each row in Figure 3 is a post-removal counterfactual rather than a fifth growth timepoint. Scar-training and implementation details are provided in Appendix B.4.

To evaluate realism, we carried out two blinded four-physician reader studies: in study 1 using only AI rendering, readers clicked lesions they believed were synthetic in images mixed with real lesions, whereas in study 2, readers clicked lesions they believed had been synthetically grown among unchanged real lesions. Study design, scoring definitions, and agreement analyses are detailed in Appendix C.3.

To test synthetic augmentation for cNF burden estimation, we trained matched pretrained ConvNeXt regressors [13] across three paired random seeds: the real-only arm used 60 real images, whereas the augmented arm used those images plus four simulated growth stages per image (240 synthetic images; 300 total). For each seed and arm, the checkpoint with the lowest error on 20 validation images was selected; the three selected models were ensembled and evaluated on 40 held-out real images. Additional details are provided in Appendix C.4.

3 Results

The coupled pipeline renders morphoelastic growth simulations as photorealistic longitudinal cNF images without requiring real time-series data. Conditioning on simulated depth and surface normals produces sequences in which lesions

emerge, enlarge, and dome outward while the surrounding skin is preserved. Each rendered timepoint therefore reflects the three-dimensional lesion geometry predicted by the mechanistic model. Across Fitzpatrick skin types, the renderings maintain realistic texture, surface relief, and skin-tone continuity (Figure 3).

Compared with the base FLUX.2 model, the lesion LoRA substantially reduced region and boundary errors while increasing PSNR and SSIM (Table 7); the scar LoRA produced smaller gains, most consistently reducing region and boundary MAE (Table 8).

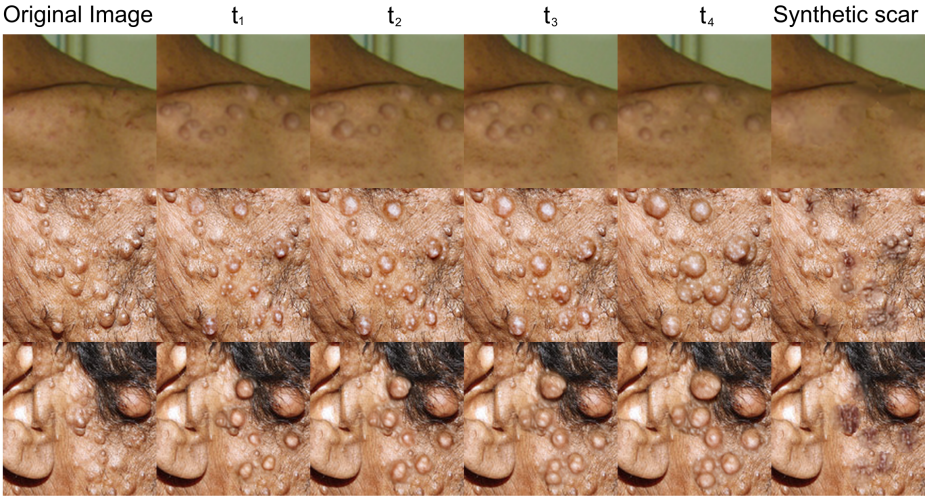

Fig. 3. Synthetic cNF growth and post-intervention appearance. From left to right: input image, four successive growth stages, and a post-removal scar.

Both reader studies indicate that synthetic and selectively grown cNFs were difficult to distinguish from surrounding real lesions. In the lesion-realism study, mean per-reader recall of synthetic lesions was 17.15%, and 49.44% of clicks landed on synthetic lesions. In the growth-discrimination study, readers collectively detected 75 of 158 unique grown lesions (47.47%), and 42.91% of clicks landed on grown lesions, indicating limited discrimination in both tasks (Table 2).

Synthetic augmentation improved image-level burden estimation on 40 held-out real images (Figure 4). The baseline model, trained on 60 real images, achieved $R^2 = 0.775$ and RMSE = 1.179; the augmented model, trained on those same 60 images plus 240 grown-lesion variants, achieved $\Delta R^2 = 0.074$ and 18% lower RMSE. These gains indicate that pipeline-generated synthetic data can improve cNF burden estimation on held-out real images.

Lesion-realism study					
Reader	**Clicks**	**On synthetic**	**On real**	**Precision**	**Target recall**
Reader 1	51	33	18	64.7%	28/417 (6.7%)
Reader 2	130	65	65	50.0%	58/417 (13.9%)
Reader 3	376	170	206	45.2%	161/417 (38.6%)
Reader 4	66	40	26	60.6%	39/417 (9.4%)
Pooled	**623**	**308**	**315**	**49.4%**	**185/417 (44.4%)**†
Growth-discrimination study					
Reader	**Clicks**	**On synthetic**	**On real**	**Precision**	**Target recall**
Reader 1	23	13	10	56.5%	13/158 (8.2%)
Reader 2	50	22	28	44.0%	22/158 (13.9%)
Reader 3	93	52	41	55.9%	52/158 (32.9%)
Reader 4	102	28	74	27.5%	28/158 (17.7%)
Pooled	**268**	**115**	**153**	**42.9%**	**75/158 (47.5%)**†

Table 2. Reader-study outcomes. The target was synthetic in the lesion-realism study and selectively grown in the growth-discrimination study. For the growth study, "On synthetic" denotes clicks on grown lesions and "On real" denotes every other click, scored as unchanged. Click counts and precision are click-level; †pooled target recall uses the union of unique targets detected across readers.

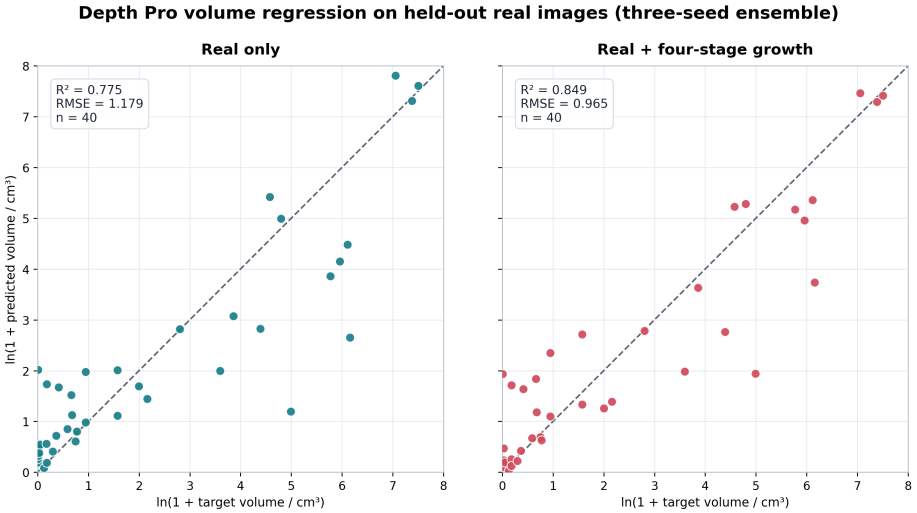

Fig. 4. Depth-derived volume regression on 40 held-out real images (three-seed ensemble). Predicted versus target log-volume for the real-only model (left) and the model trained with real plus four-stage synthetic growth images (right); the dashed line marks perfect agreement. Augmentation increases R^2 from 0.775 to 0.849 and reduces RMSE from 1.179 to 0.965.

4 Discussion

We show that photorealistic, time-resolved cNF imaging can be generated without collecting longitudinal data. By driving a geometry-conditioned generative model with a mechanistic growth simulation, we produce sequences in which individual lesions emerge and enlarge with realistic appearance, sidestepping the years-to-decades acquisition that direct longitudinal imaging would demand.

This capability is especially valuable for a rare disease such as NF1, where large, prospectively collected, time-resolved cohorts are impractical to assemble. Synthetic data offer a route to populate the appearance and progression space that real datasets leave sparse, and our burden experiment shows the practical payoff: augmenting a volume-regression model with synthetic images improved image-level burden estimation. Generated progression sequences could, in turn, support candidate longitudinal models and the development of digital twins, with real longitudinal data reserved for calibration and validation.

More broadly, the work illustrates how mechanistic modeling and generative AI complement one another: the physics model enforces plausible three-dimensional growth, while the generative model renders detail the physics cannot. This coupling is a concrete step toward generating the imaging that a patient-in-silico formulation within digital twins would require [19], and more generally toward making simulated disease state visually explicit.

Several limitations remain. The mechanistic model is a simplification—it grows a single lesion at a time under isotropic morphoelastic growth and does not capture lesion initiation, interaction, or the biological drivers of cNF development. Appearance is conditioned on monocular depth and normals, which may encode lighting or estimator bias rather than true surface shape. Furthermore, a true medical digital twin would require extending this initial patient-in-silico formulation through continuous updating and extensive verification, validation, and uncertainty quantification (VVUQ).

5 Conclusion

Quantifying cNF burden and tracking change over time is central to managing NF1, yet the longitudinal imaging needed to develop and validate such measurements is impractical to collect on years-to-decades disease timescales. We show that this imaging can instead be synthesized: a mechanistic growth model supplies the temporal, three-dimensional evolution of a lesion, and a geometry-conditioned generative model renders it as photorealistic skin. Across two blinded reader studies, physicians showed limited ability to identify synthetic or selectively grown lesions, and synthetic augmentation improved automated burden quantification, supplying the imaging a cNF digital twin needs to measure and monitor disease burden over time without requiring prospectively acquired longitudinal cohorts.

References

1. Bae, G., Davison, A.J.: Rethinking inductive biases for surface normal estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9535–9545 (2024). https://doi.org/10.1109/CVPR52733.2024.00911
2. Black Forest Labs: FLUX.2 [klein] 4B Base model card. `https://huggingface.co/black-forest-labs/FLUX.2-klein-base-4B` (2026), accessed 2026-07-20
3. Bochkovskii, A., Delaunoy, A., Germain, H., Santos, M., Zhou, Y., Richter, S.R., Koltun, V.: Depth pro: Sharp monocular metric depth in less than a second. In: The Thirteenth International Conference on Learning Representations (2025), `https://openreview.net/forum?id=aueXfY0Clv`
4. DermNet: Dermnet image library. `https://dermnetnz.org/images` (2026), accessed 2026-07-20
5. Ferreira Dias, D.: Dermatology reasoning dataset: Structured chain-of-thought annotations across five public sources. Hugging Face Datasets, `https://huggingface.co/datasets/danielfdias98/derm-reasoning` (2026), accessed 2026-07-20
6. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychological Bulletin **76**(5), 378–382 (1971). https://doi.org/10.1037/h0031619
7. Groh, M., Harris, C., Soenksen, L., Lau, F., Han, R., Kim, A., Koochek, A., Badri, O.: Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 1820–1828 (2021). https://doi.org/10.1109/CVPRW53098.2021.00201
8. International Skin Imaging Collaboration: ISIC archive. `https://www.isic-archive.com/` (2026), accessed 2026-07-20
9. Journal of Drugs in Dermatology: The full spectrum of dermatology: A diverse and inclusive dermatology atlas. `https://jddonline.com/project-atlas/` (2026), accessed 2026-07-20
10. Kuhl, E.: Growing matter: A review of growth in living systems. Journal of the Mechanical Behavior of Biomedical Materials **29**, 529–543 (2014). https://doi.org/10.1016/j.jmbbm.2013.10.009
11. Lau, J.C.L., Fleming, J., Good, M., Lim, A., Saunderson, R.B., Phan, T.A., Schlub, T., Siow, S.F., Lacson, N., Romo, C., Blakely, J., Bergqvist, C., Berman, Y.D.: Comparing 3d imaging devices for the measurement of cutaneous neurofibromas in patients with neurofibromatosis type 1. Skin Research and Technology **30**(9), e70020 (2024). https://doi.org/10.1111/srt.70020
12. Lin, M.J., Yao, H., Vera, K., Patel, E., Johnson, M., Caroline, P., Ramos, J., Mehta, J., Hu, X., Blakeley, J.O., Romo, C.G., Sarin, K.Y.: Cutaneous neurofibromas and quality of life in adults with neurofibromatosis type 1. JAMA Dermatology **160**(10), 1091–1098 (2024). https://doi.org/10.1001/jamadermatol.2024.2912
13. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986 (2022). https://doi.org/10.1109/CVPR52688.2022.01167
14. Lorenzo, G., Ahmed, S.R., Hormuth, D.A., Vaughn, B., Kalpathy-Cramer, J., Solorio, L., Yankeelov, T.E., Gomez, H.: Patient-specific, mechanistic models of tumor growth incorporating artificial intelligence and big data. Annual Review of Biomedical Engineering **26**, 529–560 (2024). https://doi.org/10.1146/annurev-bioeng-081623-025834

15. National Library of Medicine: Pubmed. `https://pubmed.ncbi.nlm.nih.gov/` (2026), accessed 2026-07-20
16. Ortonne, N., Wolkenstein, P., Blakeley, J.O., Korf, B., Plotkin, S.R., Riccardi, V.M., Miller, D.C., Huson, S., Peltonen, J., Rosenberg, A., Carroll, S.L., Verma, S.K., Mautner, V., Upadhyaya, M., Stemmer-Rachamimov, A.: Cutaneous neurofibromas: Current clinical and pathologic issues. Neurology **91**(2 Suppl 1), S5–S13 (2018). https://doi.org/10.1212/WNL.0000000000005792
17. Rodriguez, E.K., Hoger, A., McCulloch, A.D.: Stress-dependent finite growth in soft elastic tissues. Journal of Biomechanics **27**(4), 455–467 (1994). https://doi.org/10.1016/0021-9290(94)90021-3
18. 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). https://doi.org/10.1109/CVPR52688.2022.01042
19. Sadée, C., Testa, S., Barba, T., Hartmann, K., Schuessler, M., Thieme, A., Church, G.M., Okoye, I., Hernandez-Boussard, T., Hood, L., Shmulevich, I., Kuhl, E., Gevaert, O.: Medical digital twins: Enabling precision medicine and medical artificial intelligence. The Lancet Digital Health **7**(7), 100864 (2025). https://doi.org/10.1016/j.landig.2025.02.004
20. Saint, A., Ahmed, E., Shabayek, A.E.R., Cherenkova, K., Gusev, G., Aouada, D., Ottersten, B.: 3DBodyTex: Textured 3d body dataset. In: 2018 International Conference on 3D Vision. pp. 495–504 (2018). https://doi.org/10.1109/3DV.2018.00063
21. Sinha, A., Kawahara, J., Pakzad, A., Abhishek, K., Ruthven, M., Ghorbel, E., Kacem, A., Aouada, D., Hamarneh, G.: DermSynth3D: Synthesis of in-the-wild annotated dermatology images. Medical Image Analysis **95**, 103145 (2024). https://doi.org/10.1016/j.media.2024.103145
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
23. Ward, A., Li, J., Wang, J., Lakshminarasimhan, S., Carrick, A., Campana, B., Hartford, J., Sreenivasaiah, P.K., Tiyasirisokchai, T., Virmani, S., Wong, R., Matias, Y., Corrado, G.S., Webster, D.R., Smith, M.A., Siegel, D., Lin, S., Ko, J., Karthikesalingam, A., Semturs, C., Rao, P.: Creating an empirical dermatology dataset through crowdsourcing with web search advertisements. JAMA Network Open **7**(11), e2446615 (2024). https://doi.org/10.1001/jamanetworkopen.2024.46615
24. Wu, C., Lorenzo, G., Hormuth, D.A., Lima, E.A.B.F., Slavkova, K.P., DiCarlo, J.C., Virostko, J., Phillips, C.M., Patt, D., Chung, C., Yankeelov, T.E.: Integrating mechanism-based modeling with biomedical imaging to build practical digital twins for clinical oncology. Biophysics Reviews **3**(2), 021304 (2022). https://doi.org/10.1063/5.0086789

A Mechanistic Model

A.1 Layered domain and discretization

The computational domain comprised four mutually exclusive compartments: dermis, subcutis, deep tissue, and the embedded cNF lesion. The dermal layer was a 1.5-mm inward offset from the relief-bearing skin surface. The deep-tissue interface was placed at 40% of the block's vertical range from the fixed base, with subcutis between that interface and the dermis. The lesion head was embedded beneath the surface and connected through a stalk to the dermal–subcutaneous region. The conforming mesh contained 18,733 nodes and 80,266 first-order tetrahedra. Coordinates were converted from millimeters to meters for the finite-element solution and back to millimeters for reporting.

Let $\Omega_0 \subset \mathbb{R}^3$ denote this reference mesh. The primary unknown was the displacement field $\mathbf{u} : \Omega_0 \to \mathbb{R}^3$, discretized at the tetrahedral-mesh vertices. The deformation map, deformed position, and deformation gradient were

$$\boldsymbol{\varphi}(\mathbf{X}) = \mathbf{X} + \mathbf{u}(\mathbf{X}), \qquad \mathbf{x} = \boldsymbol{\varphi}(\mathbf{X}), \qquad \mathbf{F} = \nabla_X \boldsymbol{\varphi} = \mathbf{I} + \nabla_X \mathbf{u}. \tag{1}$$

Thus, the finite-element system solved for nodal displacements (equivalently, deformed vertex positions); $\mathbf{F}$ was evaluated element-wise from those nodal variables.

A.2 Growth kinematics and constitutive response

The total deformation gradient was decomposed multiplicatively as

$$\mathbf{F} = \mathbf{F}_e \mathbf{F}_g, \qquad \mathbf{F}_e = \mathbf{F}\mathbf{F}_g^{-1}, \tag{2}$$

so that growth altered the local stress-free configuration and the elastic part carried stress [17, 10]. With lesion and dermal indicators χ_t and χ_s, respectively, the implemented growth tensor was

$$\mathbf{F}_g = \chi_t \theta \mathbf{I} + \chi_s \mathbf{F}_{g,s} + (1 - \chi_t - \chi_s)\mathbf{I}. \tag{3}$$

Lesion growth was isotropic; therefore its prescribed stress-free volume multiplier was θ^3. Subcutis and deep tissue had identity growth.

Tangential dermal adaptation was represented using a through-thickness direction $\mathbf{N}$ obtained from a Laplace field between the top and bottom surfaces:

$$\mathbf{F}_{g,s} = \theta_s(\mathbf{I} - \mathbf{N} \otimes \mathbf{N}) + \mathbf{N} \otimes \mathbf{N}. \tag{4}$$

After each converged increment, the elastic area stretch $a_e = J_e \|\mathbf{F}_e^{-T}\mathbf{N}\|$ updated the local dermal multiplier,

$$\theta_s^{k+1} = \mathrm{clip}\left[\theta_s^k + 0.2\max(a_e - 1.05, 0), 1, 1.6\right]. \tag{5}$$

All compartments used a compressible Neo-Hookean response. With $I_e = \mathrm{tr}(\mathbf{F}_e^T \mathbf{F}_e)$, $J_e = \det \mathbf{F}_e$, $J_g = \det \mathbf{F}_g$, $\mu = E/[2(1+\nu)]$, and $\lambda = E\nu/[(1+\nu)(1-2\nu)]$, the energy per reference volume was

$$\psi = J_g \left[\frac{\mu}{2}(I_e - 3) - \mu \ln J_e + \frac{\lambda}{2}(\ln J_e)^2 \right]. \tag{6}$$

Table 3 reports both the prescribed material constants and the derived Lamé and bulk moduli. These homogeneous parameters were used to generate geometry and were not fitted to an individual patient.

Compartment	E (Pa)	ν	μ (Pa)	λ (Pa)	K (Pa)
Dermis / skin	5.0×10^5	0.40	178,571.4	714,285.7	833,333.3
Subcutis	2.0×10^3	0.40	714.3	2,857.1	3,333.3
Deep tissue	1.0×10^6	0.40	357,142.9	1,428,571.4	1,666,666.7
cNF lesion	5.0×10^3	0.40	1,785.7	7,142.9	8,333.3

Table 3. Material settings. The bulk modulus is $K = \lambda + 2\mu/3 = E/[3(1-2\nu)]$.

A.3 Boundary conditions, finite-element solution, and image propagation

For a prescribed growth state $\mathbf{F}_g$, quasi-static mechanical equilibrium was written in the reference configuration as

$$-\nabla_X \cdot \mathbf{P}(\mathbf{u}) = \mathbf{B} \quad \text{in } \Omega_0, \qquad \mathbf{P} = \frac{\partial \psi}{\partial \mathbf{F}}, \tag{7}$$

where $\mathbf{P}$ is the first Piola–Kirchhoff stress and $\mathbf{B}$ is the body-force density. The bottom surface was fixed in all displacement components. Each side wall had zero normal displacement but could slide tangentially. The top surface was traction-free. A constant body force of 9810 $\mathrm{N/m^3}$ acted in direction $[0, -1, 0]$. With test function $\mathbf{v}$, each growth increment satisfied

$$\mathcal{R}(\mathbf{u}; \mathbf{v}) = \int_{\Omega_0} \nabla \mathbf{v} : \mathbf{P} \, dX - \int_{\Omega_0} \mathbf{v} \cdot \mathbf{B} \, dX = 0. \tag{8}$$

The system used first-order vector Lagrange elements, Newton iteration, and a direct MUMPS factorization. Each increment allowed 60 Newton iterations and up to six retries with a halved increment after restoring the last converged state. The lesion stretch advanced from $\theta = 1.00$ to 1.35 over 30 increments, producing 31 states.

Deformed lesion volume was recomputed directly from the state-specific tetrahedra,

$$V_t = \sum_{K \in \Omega_t} \frac{1}{6} \left| \det[\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \mathbf{x}_4 - \mathbf{x}_1] \right|. \tag{9}$$

Lesion volume and maximum surface displacement increased monotonically. No state contained an element with $J_e \leq 0$; the minimum elastic Jacobian over the trajectory was 0.122693 (Table 4).

State	θ	Volume (mm^3)	Max. shift (mm)	min J_e	Folded
0	1.000	2775.8	0.000	–	0
5	1.058	3042.0	0.431	0.642	0
10	1.117	3337.6	0.872	0.413	0
15	1.175	3667.4	1.366	0.277	0
20	1.233	4034.7	1.902	0.203	0
25	1.292	4442.6	2.473	0.161	0
30	1.350	4893.9	3.071	0.123	0

Table 4. Selected states from the 31-state morphoelastic trajectory.

For each converged state, the solved displacement field updated every surface and lesion vertex using the deformation map defined above. Because the lesion and surrounding skin were carried by the same conforming mesh, the deformation preserved their spatial correspondence. The deformed surface was then projected into image space to generate matched depth and surface-normal maps; these geometry maps conditioned the photorealistic renderer described in Appendix B.3. Repeating the solve, projection, and rendering over all 31 states produced a label-consistent longitudinal sequence from one real starting image rather than generating unrelated tumors.

B FLUX Image Generation

B.1 Geometry conditioning

The lesion adapter used 229 training images with 38,040 annotated contours; validation and test contained 25 images each. Depth Pro produced a metric distance map D. Let q_1 and q_{99} be the first and 99th percentiles over finite positive pixels. The generator received normalized inverse depth

$$\widetilde{D}_{\mathrm{inv}}(p) = \mathrm{clip}\left(\frac{q_{99} - D(p)}{\max(q_{99} - q_1, 10^{-6})}, 0, 1\right). \tag{10}$$

DSINE normals were resized bilinearly and renormalized with a 10^{-6} norm floor. The x and y components were mapped from $[-1, 1]$ to $[0, 1]$.

All lesion contours in an image were combined by union. At native resolution (H, W), the union was dilated by an elliptical element of radius

$$r = \max(2, \mathrm{round}[0.006 \min(H, W)]). \tag{11}$$

RGB, inverse depth, normals, and masks were letterboxed to 512×512 with preserved aspect ratio. RGB used area interpolation, continuous geometry used

bilinear interpolation, and masks used nearest-neighbor interpolation. If I is RGB and M_e is the dilated edit support, the three-channel condition was

$$C(p) = \begin{cases} I(p), & M_e(p) = 0, \\ \left[\widetilde{D}_{\mathrm{inv}}(p), \dfrac{N_x(p)+1}{2}, \dfrac{N_y(p)+1}{2}\right], & M_e(p) = 1. \end{cases} \tag{12}$$

The training target was the unmodified image. The instruction asked the model to reconstruct clinical skin inside each geometry-coded region while following the supplied depth and normals and preserving surrounding RGB.

B.2 LoRA for FLUX lesion

FLUX.2 Klein Base 4B was frozen and loaded with 4-bit NF4 quantization with double quantization and bfloat16 dequantization/compute. Rank-32 LoRA matrices with scale 32 yielded 33,423,360 trainable parameters. Adapters targeted the query, key, value, and output projections in each of five joint transformer blocks and the fused query–key–value/MLP and output projections in each of 20 single transformer blocks. Text encoders, latent encoder/decoder, and all base-transformer weights remained frozen.

The lesion model received the following text prompt:

> *Clinical close-up image of human skin with one raised cutaneous neurofibroma inside the masked area, realistic textures and skin bumps, natural skin color, realistic lighting, medically plausible, seamless blending with the surrounding skin.*

Setting	Lesion-adapter value
Training records	229 condition/target pairs
Resolution	512×512
Optimizer updates	300
Batch / accumulation	1 / 1
Optimizer	AdamW with 8-bit optimizer states
Learning rate	10^{-4}
Schedule	Constant; 20 warm-up updates
LoRA rank / scale	32 / 32
Frozen-base precision	4-bit NF4; bfloat16 compute
Training guidance value	1
Arithmetic	bfloat16 mixed precision; TF32 enabled
Memory controls	Gradient checkpointing; cached latents
Data-loading workers	8
Checkpoint interval	100 updates; three retained
Random seed	20260714

Table 5. Geometry-conditioned lesion-adapter training settings.

B.3 Rendering longitudinal growth

The surface used for mechanics was first converted to a lesion-centered STL, with the lesion placed at its center. Let d_ℓ denote the baseline lesion diameter. The image-derived skin surface was retained within a circle of radius $3d_\ell$. A square outer boundary was placed one additional lesion diameter beyond this circle, and the resulting domain extended 5 mm below the skin surface. For each mechanical state, the deformed surface vertices were interpolated onto a lesion-centered 256×256 grid. Baseline relief was subtracted and negative differences were clipped to zero. With matched source distance D_0 in meters and simulated relief h_k in millimeters, the state distance was

$$D_k = D_0 - 10^{-3} h_k. \tag{13}$$

State normals combined the source normal slope with centered finite differences of h_k and were renormalized to unit length. The edit support was an ellipse with baseline radii 60 and 75 pixels, scaled by $(V_k/V_0)^{1/3}$ and dilated with a 9×9 ellipse. Rendering used 30 sampling steps, guidance 4.0, a fixed seed, and one prompt instructing the model to enlarge the same lesion without satellites, rings, or halos. A nine-pixel interior feather blended the edit.

For both lesion and scar synthesis, the reported image was hard-composited,

$$\widehat{I}(p) = M(p)G(p) + [1 - M(p)]I(p), \tag{14}$$

where G is the raw generator output and M the binary edit support. Source pixels outside the support were therefore preserved exactly.

B.4 Scar FLUX model

The scar data contained 359 annotated regions. Training used 340 annotations, each converted into two deterministic context crops, for 680 training records; validation used 13 annotations and test used six. Each crop and its union mask was resized to 512×512. The mask was dilated by five pixels with an elliptical 11×11 element. Pixels inside the support were set to magenta $[255, 0, 255]$, while pixels outside retained source RGB. No depth or normal channels were supplied. The instruction asked the model to replace each magenta region with a postoperative cNF-removal scar while preserving the surrounding skin and clinical context.

The scar adapter used the same frozen FLUX.2 base, LoRA rank and scale, quantization, optimizer family, learning rate, batch size, precision, and random seed as the lesion adapter. It was trained independently for 600 updates and was conditioned only by the spatial mask.

At inference, the generated region was composited with Eq. 14. The scar editor models postoperative appearance; it does not simulate biological wound healing or forecast a patient-specific scar trajectory.

Setting	Scar-adapter value
Training records	680 crops from 340 masks
Resolution	512×512
Optimizer updates	600
Batch / accumulation	1 / 1
Optimizer	AdamW with 8-bit optimizer states
Learning rate	10^{-4}
Schedule	Constant; 30 warm-up updates
LoRA rank / scale	32 / 32
Frozen-base precision	4-bit NF4; bfloat16 compute
Training guidance value	1
Arithmetic	bfloat16 mixed precision; TF32 enabled
Memory controls	Gradient checkpointing; cached latents
Data-loading workers	8
Checkpoint interval	200 updates; three retained
Random seed	20260714

Table 6. Mask-conditioned scar-adapter training settings.

C Experiments

C.1 Base FLUX.2 versus lesion LoRA

The lesion comparison used all 25 held-out images and 1,816 test contours. Base and adapted arms received identical geometry conditions, prompt, target, support, seed, 50 sampling steps, and guidance 4.0; only LoRA activation differed. Images were scaled to $[0, 1]$. For edit region $R = \{p : M(p) = 1\}$,

$$\mathrm{MAE}_R = \frac{1}{3|R|} \sum_{p \in R} \sum_{c=1}^{3} |X_c(p) - Y_c(p)|, \tag{15}$$

$$\mathrm{MSE}_R = \frac{1}{3|R|} \sum_{p \in R} \sum_{c=1}^{3} [X_c(p) - Y_c(p)]^2. \tag{16}$$

RMSE was $\sqrt{\mathrm{MSE}_R}$, PSNR was $-10 \log_{10}[\max(\mathrm{MSE}_R, 10^{-12})]$, and SSIM used data range one [22]. Boundary MAE used the 7×7 morphological gradient of the support. Ten thousand image-level bootstrap resamples formed the paired 95% intervals in Table 7.

Metric	Foundation	Lesion LoRA	Gain [95% CI]	LoRA better
Region MAE	0.2525 ± 0.1301	0.0629 ± 0.0264	0.1897 [0.1399, 0.2449]	24/25
Region RMSE	0.3060 ± 0.1359	0.0857 ± 0.0358	0.2203 [0.1662, 0.2787]	24/25
Region PSNR (dB)	11.07 ± 3.77	22.15 ± 4.03	11.08 [8.76, 13.41]	24/25
Region SSIM	0.2042 ± 0.1394	0.5534 ± 0.1315	0.3492 [0.2816, 0.4146]	24/25
Boundary MAE	0.1053 ± 0.0568	0.0212 ± 0.0090	0.0841 [0.0633, 0.1070]	25/25

Table 7. Matched FLUX.2 foundation-versus-LoRA lesion results. Values are image-level mean ± sample standard deviation; positive gain favors LoRA.

C.2 Base FLUX.2 versus scar LoRA

The scar comparison used six held-out masks. Base and adapted arms used the same mask condition, prompt, target, support, seed, 28 sampling steps, guidance 4.0, and hard-compositing rule. Only scar-LoRA activation differed. Metrics followed the lesion comparison and are reported in Table 8.

Metric	Foundation	Scar LoRA	Paired gain	LoRA better
Region MAE	0.1455 ± 0.1090	0.1274 ± 0.0757	0.0181	5/6
Region RMSE	0.1707 ± 0.1187	0.1685 ± 0.0888	0.0022	4/6
Region PSNR (dB)	17.36 ± 6.67	16.99 ± 6.22	-0.37	4/6
Region SSIM	0.5107 ± 0.1893	0.5254 ± 0.2011	0.0147	3/6
Boundary MAE	0.0514 ± 0.0326	0.0268 ± 0.0192	0.0246	5/6

Table 8. Matched FLUX.2 foundation-versus-LoRA scar results. Values are mask-level mean ± sample standard deviation; positive gain favors LoRA.

C.3 Physician reader studies

Four physicians independently viewed the same 20-image lesion-realism set without masks, geometry maps, model labels, or answer-key overlays. They clicked every lesion believed to be synthetic. A click inside a synthetic component was assigned to that component; repeated clicks did not create additional unique detections. Clicks on real lesions and background were retained as false selections. Click precision was the fraction of all clicks landing on synthetic components, per-reader recall was the fraction of 417 synthetic lesions uniquely detected, and pooled recall used the union of detections across readers. These definitions yield 49.44% click precision, 17.15% mean per-reader recall, and 44.36% pooled union recall. Fleiss' κ was computed over the binary per-lesion reader decisions [6].

The growth-discrimination study used 20 images containing 505 annotated lesions, including 158 selectively grown targets and 347 unchanged lesions. Four physicians were instructed to click every lesion believed to have been selectively grown and were not shown the original image, mask, target count, or answer key. They placed 268 clicks: 115 on grown lesions, while the remaining 153 clicks were scored as unchanged. The per-reader unchanged counts were 10, 28, 41, and 74. Per-reader grown-lesion recall was 8.23%, 13.92%, 32.91%, and 17.72% (mean 18.20%), with click precision of 56.52%, 44.00%, 55.91%, and 27.45%,

respectively; pooled click precision was 42.91%. After deduplicating detections within readers and taking the union across readers, 75 of the 158 grown lesions were detected, yielding pooled grown-lesion recall of $75/158 = 47.47\%$.

C.4 Depth-volume regression

The downstream cohort contained 120 real images divided into 60 training, 20 validation, and 40 held-out test images. Four geometry states were generated for each training image using growth factors $g \in \{1.12, 1.26, 1.42, 1.60\}$, producing 240 synthetic training records. The real-only arm therefore used 60 records and the augmented arm used the same 60 real records plus 240 synthetic records.

For each lesion pixel, a local skin plane $\mathbf{n} \cdot \mathbf{P} + d = 0$ gave ray–plane distance $B_{\text{raw}} = -d/(\mathbf{n} \cdot \mathbf{r})$, where $\mathbf{r} = [(u - c_x)/f_x, (v - c_y)/f_y, 1]^T$. With observed Depth Pro distance D, relief was bounded by the projected lesion extent,

$$L = \max\left(\frac{w_{\text{box}} \operatorname{median}(D)}{f_x}, \frac{h_{\text{box}} \operatorname{median}(D)}{f_y} \right), \tag{17}$$

$$\delta = \min\{\max(B_{\text{raw}} - D, 0), L\}, \qquad B_{\text{eff}} = D + \delta. \tag{18}$$

The pixel-frustum contribution and image target were

$$v(p) = \frac{B_{\text{eff}}(p)^3 - D(p)^3}{3 f_x f_y}, \qquad V = 10^6 \sum_p v(p) \text{ cm}^3. \tag{19}$$

At lesion overlap, the maximum lesion-specific contribution was counted once.

The regressor input was the numeric metric-depth map only. Log depth was standardized from the 60 real training maps using mean -0.5306794 and sample standard deviation 0.7287829, clipped at five standard deviations, resized with preserved aspect ratio, and centered on a 512×512 canvas. The only stochastic spatial transform was horizontal reflection with probability 0.5.

The backbone was ConvNeXt-Tiny with ImageNet-1K V1 initialization [13]. Its input kernel was collapsed from three channels to one, and its classifier was replaced by a single Linear$(768, 1)$ scalar head. The standardized target was

$$y = \frac{\ln(1 + V/\text{cm}^3) - 2.1898819}{2.4698925}, \tag{20}$$

with both constants estimated from the 60 real training targets.

Both arms used batch size 16, 600 optimizer updates, and 9,600 record exposures. AdamW used learning rate 10^{-4} for the backbone and 10^{-3} for the scalar head, weight decay 10^{-4}, and global gradient-norm clipping at one. A 30-update linear warm-up preceded cosine decay. The loss was SmoothL1 with $\beta = 0.5$. Training used bfloat16 autocasting and eight data workers. Paired seeds 1301, 1302, and 1303 began from identical model states within each seed, and the checkpoint with minimum validation RMSE was selected.

Seed	Step	Arm	Log RMSE	Raw R^2	Raw RMSE	MAE	ρ
1301	30	Real only	1.3728	0.1516	379.50	112.45	0.814
1301	90	Augmented	0.9989	0.7449	208.11	81.89	0.874
1302	300	Real only	1.1618	0.8321	168.83	76.58	0.860
1302	450	Augmented	1.0353	−0.0769	427.55	148.98	0.879
1303	210	Real only	1.1451	0.6583	240.84	92.45	0.867
1303	30	Augmented	1.0867	0.6076	258.09	94.78	0.851

Table 9. Individual-run results on the 40 held-out real images. Log RMSE uses $\ln(1 + V/\mathrm{cm}^3)$; raw RMSE and MAE are in cm^3; ρ is Spearman correlation.

The three-seed log-space ensemble increased R^2 from 0.774635 to 0.848812 and reduced RMSE from 1.178689 to 0.965416, corresponding to $\Delta R^2 = 0.074177$ and an 18.095% relative RMSE reduction. A paired hierarchical bootstrap with 20,000 resamples gave median $\Delta R^2 = 0.068962$ (95% interval $[-0.009649, 0.190921]$) and median augmented-minus-real RMSE -0.197585 (95% interval $[-0.435208, 0.026161]$).