

Conditional Latent Diffusion Model with Fourier-based Motion Modelling for Virtual Population Synthesis

Shaokun Lan^{1,2,3†}, Haoran Dou^{1,2,3†}, Jinghan Huang^{1,2,3}, Arezoo Zakeri^{1,2,4},
Fengming Lin^{1,2,3}, Zherui Zhou^{1,2,5}, Jinming Duan^{1,2,4}, and Alejandro F.
Frangi^{1,2,3,4,6*}

- ¹ Centre for Computational Imaging and Modelling in Medicine (CIMIM), University of Manchester, Manchester, UK
² Christabel Pankhurst Institute, University of Manchester, Manchester, UK
³ Department of Computer Science, University of Manchester, Manchester, UK
⁴ Division of Informatics, Imaging & Data Sciences, University of Manchester, Manchester, UK
⁵ Department of Electrical & Electronic Engineering, University of Manchester, Manchester, UK
⁶ NIHR Manchester Biomedical Research Centre, Manchester Academic Health Sciences Centre, University of Manchester, Manchester, UK
alejandro.frangi@manchester.ac.uk

Abstract. *In-silico* trials of medical devices require the generation of virtual populations of anatomies. In cardiovascular applications, virtual anatomy is typically represented as a 3D+t mesh sampled from a generative model. However, most existing mesh generators focus on static anatomy, while sequence models often lack explicit periodicity. To this end, we propose 4D F-MeshLDM, a conditional generative framework comprising a convolutional mesh VAE to encode meshes, a structural latent space that parameterises motion using a truncated Fourier series, and a diffusion prior that learns the latent distribution over Fourier coefficient tokens. By conditioning the diffusion process on clinical covariates via affine modulation, we enable controllable synthesis. Sampling tokens and performing inverse Fourier synthesis yield cycle-consistent latent trajectories, which can be decoded into 3D+t cardiac mesh sequences. Experiments on 5,000 UK Biobank subjects demonstrate that 4D F-MeshLDM outperforms state-of-the-art baselines in anatomical fidelity and achieves near-zero cycle closure error. Furthermore, the generated cohorts accurately preserve clinical functional indices, highlighting the potential of our framework for reliable *in-silico* cardiac trials.

Keywords: Latent Diffusion Models · Virtual Population Synthesis · Cardiac Motion · 4D Mesh Generation · Cycle Consistency

* Corresponding author: Alejandro F. Frangi.
Email: alejandro.frangi@manchester.ac.uk

† Shaokun Lan and Haoran Dou contributed equally to this work.

1 Introduction

As an alternative to clinical trials, *in-silico* trials use computational models to evaluate the safety and efficacy of medical devices [9, 13]. *In-silico* trials require a diverse virtual population that captures the anatomical and physiological variability of real patients, enabling device performance to be assessed across different cases. In cardiac applications, the virtual population can typically be represented as a generative model that produces realistic anatomy with plausible dynamic motion across the full cardiac cycle.

Statistical shape models and deep generators, like PCA and VAE-based mesh models, have been widely used to model cardiac anatomy [4–6, 11], but most methods generate static shapes. To address this issue, several deep generative approaches have been proposed for spatio-temporal synthesis of cardiac anatomy [16, 17], including implicit spatio-temporal fields [20], continuous-time generative dynamics [10], and disentangled population motion model [3]. However, these methods encourage periodicity implicitly through temporal conditioning, which can lead to drift and imperfect cycle closure, which are undesirable for device evaluation and simulation.

In this work, we propose 4D F-MeshLDM, a conditional latent diffusion model for full-cycle cardiac mesh generation that enforces cycle periodicity and consistency by construction. The cardiac motion of each subject is encoded as a compact set of Fourier coefficient tokens via a truncated Fourier series in the latent space of a mesh VAE, and a transformer-based diffusion prior is trained to model the distribution of these tokens conditioned on clinical covariates. The main contributions are as follows:

- We propose the first cardiac motion generative model that guarantees mathematically exact cycle consistency by construction. By parameterising latent trajectories as truncated Fourier series, we eliminate endpoint drift and closure errors, strictly enforcing periodicity without relying on learned recurrent dynamics [16] or soft temporal conditioning [3, 10, 20].
- To the best of our knowledge, we introduce the first conditional latent diffusion model for 3D+t mesh sequences. We model the complex joint distribution of motion and anatomy in a compact spectral domain via a Transformer-based diffusion prior operating on Fourier coefficient tokens, enabling scalable and controllable synthesis.
- Experiments demonstrate that 4D F-MeshLDM significantly outperforms state-of-the-art baselines [3, 10, 16]. We achieve nearly zero cycle closure error and superior anatomical fidelity, validating the model’s reliability for *in-silico* trials.

2 Methodology

Our framework aims to generate anatomically plausible and cycle-consistent 3D+t cardiac mesh sequences conditioned on clinical covariates. As illustrated

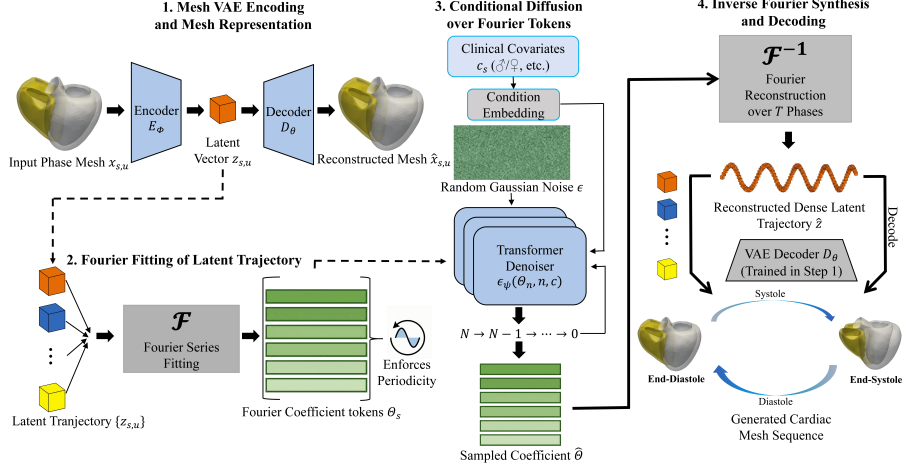


Fig. 1. Overview of the 4D F-MeshLDM pipeline. Dashed arrows denote training-time encoding and fitting, while solid arrows denote the generative pathway.

in Fig. 1, our framework decouples the spatial representation of cardiac anatomy from the modeling of its periodic motion trajectory. A mesh VAE encodes the mesh from each phase to a latent code. Then, cardiac motion is represented by fitting a truncated Fourier series to the latent trajectory, producing a set of coefficient tokens. A conditional diffusion prior models the token distribution, followed by inverse Fourier synthesis with VAE decoding, which generates a full cardiac cycle.

2.1 Anatomical Latent Representation

We train a mesh variational autoencoder based on CoMA [19] to learn compact spatial features. For a mesh $x_{s,u}$ of a given subject s at frame u , the encoder E_ϕ maps geometry to a latent distribution $q_\phi(z_{s,u} | x_{s,u})$. The decoder D_θ reconstructs the mesh $\hat{x}_{s,u}$ using Chebyshev spectral convolutions and hierarchical pooling. We optimize the standard VAE objective with a weight β :

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q_\phi} [\|x_{s,u} - \hat{x}_{s,u}\|_2^2] + \beta \text{KL}(q_\phi(z_{s,u} | x_{s,u}) \| \mathcal{N}(0, I)). \quad (1)$$

After training, the encoder mean $z_{s,u} \in \mathbb{R}^D$ serves as the deterministic representation for motion modelling.

2.2 Cardiac Motion Modelling via Fourier Fitting

Our framework models cardiac motion in the frequency domain to separate the spatial trajectory of the motion from its cardiac cycle duration. For a subject s with a sequence length of T_s frames indexed by $u = 0, \dots, T_s - 1$, we define

the normalised phase at frame u as $\tau_u = \frac{u}{T_s-1} \in [0, 1]$. The latent trajectory is approximated by a truncated Fourier series of order K :

$$z_{s,u} \approx a_{s,0} + \sum_{k=1}^{K-1} (a_{s,k} \cos(2\pi k \tau_u) + b_{s,k} \sin(2\pi k \tau_u)), \quad (2)$$

where $a_{s,k}, b_{s,k} \in \mathbb{R}^D$ are subject-specific coefficients. Crucially, this phase definition implies $\tau_0 = 0$ and $\tau_{T_s-1} = 1$. Since the Fourier basis is 1-periodic, this formulation guarantees mathematically exact cycle closure, ensuring $z_{s,0} \equiv z_{s,T_s-1}$ by construction.

We can express Eq. (2) linearly. We define a basis vector $A_u \in \mathbb{R}^{1 \times B}$ ($B = 2K - 1$) containing the harmonic terms:

$$A_u = [1, \cos(2\pi \tau_u), \sin(2\pi \tau_u), \dots, \cos(2\pi(K-1)\tau_u), \sin(2\pi(K-1)\tau_u)]. \quad (3)$$

The coefficients are stacked into a matrix $\Theta_s \in \mathbb{R}^{B \times D}$, where rows correspond to the terms $\{a_{s,0}, a_{s,1}, b_{s,1}, \dots\}$. The trajectory is thus given by $z_{s,u} \approx A_u \Theta_s$.

To obtain the ground-truth coefficients Θ_s for training, we stack the observed frame-wise latents $\{z_{s,u}\}_{u=0}^{T_s-1}$ into a trajectory matrix $Z_s \in \mathbb{R}^{T_s \times D}$. We then solve the following least-squares equation:

$$\Theta_s = \arg \min_{\Theta} \|A_s \Theta - Z_s\|_F^2, \quad (4)$$

where $A_s \in \mathbb{R}^{T_s \times B}$ is the subject-specific basis matrix constructed using the subject's cycle length T_s . Crucially, while A_s depends on the temporal sampling T_s , the resulting coefficients Θ_s represent the time-invariant motion feature of the subject. In the following discussion, we omit the subscript s and denote Θ as the token set for brevity.

2.3 Conditional Diffusion on Fourier Tokens

We treat the coefficient matrix Θ as a sequence of B tokens and model their joint distribution $p(\Theta \mid c)$ using a Denoising Diffusion Probabilistic Model (DDPM) [7]. The forward process diffuses the coefficients into Gaussian noise over N steps. The reverse generative process is learned by a denoiser network ϵ_ψ , implemented as a Transformer.

The Transformer treats the B frequency components as a sequence of tokens. To preserve frequency information, we augment each token with a learnable Fourier-index embedding. Clinical covariates c are embedded and injected into the network via Adaptive Layer Normalisation (AdaLN) [14], allowing the model to generate motion patterns specific to the target population. The objective is to predict the noise ϵ added at step n :

$$\mathcal{L}_{\text{Diff}} = \mathbb{E}_{\Theta_0, \epsilon, n, c} [\|\epsilon - \epsilon_\psi(\Theta_n, n, c)\|_2^2]. \quad (5)$$

2.4 Generative Inference

The generative inference process is detailed below. First, a coefficient matrix $\hat{\Theta} \sim p_\psi(\Theta | c)$ is sampled from the conditional diffusion prior on covariate c . This determines the time-invariant geometric motion pattern of the heart. Second, we dynamically adjust the cycle duration to reflect the physiological relationship between demographics and heart rate. We model the duration of the target cycle T' as a function of covariate c with $T' = f(c)$, where $f(\cdot)$ is a mapping derived from the statistics of the training population. Unlike existing methods that require additional data processing or ad-hoc temporal resampling to alter sequence lengths, our method inherently supports generating populations with variable cardiac cycle durations. Then, based on this subject-specific duration T' , a dynamic sampling basis matrix $\mathbf{A}' \in \mathbb{R}^{T' \times B}$ can be constructed using the normalised phases $\tau' = \{0, \frac{1}{T'-1}, \frac{2}{T'-1} \dots, 1\}$. The dense latent trajectory is then reconstructed by combining the generated morphology with the subject-specific temporal basis:

$$\hat{Z} = \mathbf{A}'\hat{\Theta}. \quad (6)$$

Finally, the full mesh sequence is obtained by decoding each frame: $\hat{x}_u = D_\theta(\hat{z}_u)$. Our work yields a strictly periodic mesh sequence that is consistent with both the morphological and temporal characteristics of the target population defined by c .

3 Experimental Results

3.1 Dataset

We conducted evaluation experiments on a cohort of 5,000 time-resolved biven-tricular (BiV) surface meshes reconstructed from cine CMR available from the UK Biobank (UKBB). For perfectly fair comparisons with baselines which are constrained to fixed temporal lengths, we specifically focused on subjects where all $T_s = T = 50$ in this work. For these subjects, $T = 50$ covers the full cardiac cycle starting from end-diastole (ED) at $u = 0$, with $u = 49$ being the last frame returning to ED. All meshes share a common BiV template topology and consistent vertex ordering across subjects and cardiac phases after reconstruction, enabling point-to-point geometric evaluation. Each mesh was reconstructed by [21]. The cohort was split into 4,000/500/500 for training, validation, and testing, respectively. Gender and age were selected as the covariates in this work. We synthesised a virtual cohort with the same number of subjects in the test set to ensure a fair comparison.

3.2 Implementation Details

All experiments were implemented in PyTorch and trained on 1 NVIDIA H200 GPU with a total memory of 141 GB. We trained a mesh VAE with latent dimension $D = 64$ using AdamW with learning rate 5×10^{-4} and KL weight

$\beta = 10^{-5}$ for 400 epochs. To model cardiac dynamics, we fitted a truncated Fourier series with a frequency $K = 13$, which achieved the best performance amongst all K values in the ablation study. A tokenized Transformer DDPM (width 256, depth 8, 8 heads) using 500 diffusion steps was trained with learning rate 3×10^{-4} for 1,000 epochs.

3.3 Baselines and Evaluation Metrics

The 4D F-MeshLDM was compared with three state-of-the-art cardiac motion generation baselines: (1) CHeart [16], a spatio-temporal mesh generator based on recurrent neural networks; (2) 4D CardioSynth [3], a VAE-based synthesis method with spatio-temporal disentanglement; and (3) CardiacFlow [10], a continuous-time generative dynamics model.

We evaluated the distribution quality of anatomy generation using specificity [2], measured as the nearest-neighbour distance from generated samples to the real population, and coverage@5 [12], measured as the fraction of real samples that have at least one generated sample within the 5-nearest-neighbourhood. To evaluate the spatio-temporal fidelity of subject-specific reconstructions, we calculated the Sequence Root Mean Square (RMS) error, which is defined as the root mean square of the Euclidean distances between corresponding vertices of the generated and ground-truth meshes, averaged across all time frames within the sequence. Moreover, following prior work on spatiotemporal regularisation in cardiac motion analysis [1, 18], we quantified temporal smoothness by the mean of the second-order finite differences of the BiV volume trajectory, denoted as volume smoothness, and vertex trajectories, denoted as mesh smoothness. Cycle closure errors were evaluated by volume cycle consistency and mesh cycle consistency, which represent start-to-end discrepancies. To assess clinical relevance, we derived standard ventricular functional indices from the full-cycle left ventricle (LV) and right ventricle (RV) volume trajectories, including end-diastolic volume (EDV), end-systolic volume (ESV), ejection fraction (EF), which are routinely reported in cardiovascular magnetic resonance (CMR) examinations [8, 15]. We computed the mean absolute error (MAE) of these functional indices between the ground truth sequence and the generated sample paired by subject covariates.

3.4 Experimental Results

Quantitative Analysis of the Generation Quality: As shown in Table 1, we conducted unconditional and conditional cardiac cycle generation to comprehensively demonstrate the superiority of our method. In the unconditional setting, 4D F-MeshLDM_{uncond} achieved the best results across all metrics and anatomies. For BiV, Coverage@5 achieved 71.2%, while specificity improved from 2.41 ± 0.53 mm to 2.32 ± 0.36 mm. The largest fidelity gain was observed for RV specificity, which decreased from 2.89 ± 0.56 mm to 2.41 ± 0.34 mm. In the conditional generation, 4D F-MeshLDM_{cond} achieved the best specificity and Coverage@5 for all anatomies. For BiV, Coverage@5 increased to 98.60%, compared with 92.79% for CHeart. 4D F-MeshLDM_{cond} also achieved the best BiV specificity

with 2.76 ± 0.52 mm, improving over CHeart (2.80 ± 0.51 mm) and CardiacFlow (2.95 ± 0.56 mm). 4D F-MeshLDM_{cond} consistently achieved the lowest sequence reconstruction errors across all anatomical structures. Specifically, for BiV, our method reduced the Seq. RMSE to 7.54 ± 2.75 mm, substantially outperforming both CHeart (8.19 ± 2.97 mm) and CardiacFlow (9.07 ± 3.32 mm). These results demonstrate that our framework not only preserves the diversity and fidelity of the population distribution but also more accurately captures the precise, subject-specific dynamic motion trajectories of individual patients.

Temporal Consistency: To evaluate the cycle consistency, we adapted 4D CardioSynth to conditional model and generated conditional BiV mesh sequences for each method. Table 2 shows that 4D F-MeshLDM provides substantially stronger temporal cycle consistency than prior methods. Volume cycle consistency was reduced to 4.506×10^{-6} mL² and mesh cycle consistency was reduced to $(5.564 \pm 0.309) \times 10^{-3}$ mm. Because of the Fourier parameterisation, our 4D F-MeshLDM actually yields a theoretical zero cycle closure error. The non-zero values are solely due to the normal cardiac cycle drift from the data, which is nearly-zero for populations without severe pathologies, whereas baselines suffer from structural drift. A lower second-order difference values of smoothness indicate smoother temporal evolution, but excessive smoothness may not be consistent with the true systolic-diastolic dynamics of the cardiac cycle. 4D F-MeshLDM achieves the lowest mesh smoothness and near-zero cycle closure error, indicating temporally coherent motion without endpoint drift. Overall, the near-zero cycle closure errors in both volume and mesh metrics and the best mesh smoothness strongly support the effectiveness of the periodic latent trajectory. Regarding inference time, marginally slower than CHeart (0.0919s), CardiacFlow (0.0941s), and 4D CardioSynth (0.0945s), our 4D F-MeshLDM still achieves rapid generation, producing an entire 3D+t sequence in just 0.1630s.

Clinical relevance: Table 2 also demonstrates the results of clinical relevance evaluation. Since $EF = 1 - ESV/EDV$ which shows that correlated EDV and ESV errors can partially cancel in the ratio, the lower EF MAE indicates better preservation of relative pump function of 4D F-MeshLDM, while the absolute EDV/ESV errors of the proposed model remain comparable to the baselines rather than uniformly superior.

Ablation Study: We conducted the ablation study on the impact of the Fourier parameterisation by varying the frequency K . The vFID was computed as a Fréchet distance between the distributions of real and generated volume trajectories [10]. Fig. 2 visualises the mean $\pm$ std volume curves for generated sequences against real data. Removing Fourier fitting substantially degraded trajectory realism with vFID = 51.06, and the generated data exhibited significant bias in both LV and RV volumes over the cycle. The Fourier fitting consistently improved vFID and aligned the systolic-diastolic dynamics with the real distribution. The best performance was achieved at $K = 13$ with the lowest vFID.

Table 1. Quantitative Analysis of cardiac cycle generation. Specificity and Seq. RMSE are shown as mean $\pm$ std.

Method	Specificity (mm, $\downarrow$)			Coverage@5 (% , $\uparrow$)			Seq. RMSE (mm, $\downarrow$)		
	LV	RV	BiV	LV	RV	BiV	LV	RV	BiV
<i>Unconditional Generation</i>									
4D CardioSynth	1.94 $\pm$ 0.33	2.89 $\pm$ 0.56	2.41 $\pm$ 0.53	60.2	66.5	64.4	-	-	-
4D F-MeshLDM _{uncond}	1.89$\pm$0.32	2.41$\pm$0.37	2.32$\pm$0.36	63.5	72.8	71.2	-	-	-
<i>Conditional Generation</i>									
CHeart	2.32 $\pm$ 0.49	2.92 $\pm$ 0.56	2.80 $\pm$ 0.51	91.40	93.02	92.79	7.67 $\pm$ 2.79	8.63 $\pm$ 3.23	8.19 $\pm$ 2.97
CardiacFlow	2.40 $\pm$ 0.56	3.08 $\pm$ 0.59	2.95 $\pm$ 0.56	96.05	96.44	96.98	8.51 $\pm$ 3.13	9.58 $\pm$ 3.59	9.07 $\pm$ 3.32
4D F-MeshLDM _{cond}	2.27$\pm$0.51	2.89$\pm$0.34	2.76$\pm$0.52	98.14	99.07	98.60	6.82$\pm$2.41	7.97$\pm$2.91	7.54$\pm$2.75

Table 2. Quantitative analysis of temporal consistency and functional indices. (Mesh metrics: mean $\pm$ std. Volume/Mesh cycle consistency: MSE/RMSE between first and last frame meshes.)

Method	Temporal metrics ($\downarrow$)				LV MAE ($\downarrow$)			RV MAE ($\downarrow$)		
	Vol. smooth.	Mesh smooth.	Vol. Consistency (mL ²)	Mesh Consistency (mm)	EDV (mL)	ESV (mL)	EF	EDV (mL)	ESV (mL)	EF
CHeart	0.548	0.058 $\pm$ 0.026	14.496	1.106 $\pm$ 0.396	15.21	13.10	0.062	9.21	5.10	0.058
CardiacFlow	0.229	0.046 $\pm$ 0.010	0.133	0.143 $\pm$ 0.019	23.09	21.23	0.073	12.57	7.04	0.062
4D CardioSynth	0.140	0.037 $\pm$ 0.006	4.352	0.861 $\pm$ 0.216	14.25	13.17	0.080	8.72	4.95	0.055
4D F-MeshLDM	0.361	0.029$\pm$0.017	4.506$\times 10^{-6}$	(5.564$\pm$0.309)$\times 10^{-3}$	18.86	16.00	0.049	9.91	5.76	0.047

This result indicates that Fourier fitting plays a key role in realistic cardiac cycle generation.

Visual Assessment: Fig. 3 visualises a BiV mesh sequence generated by 4D F-MeshLDM conditioned on age 70 and female sex, which exhibits a coherent contraction-relaxation pattern across the cycle. The surfaces preserve anatomically plausible global morphology throughout the motion and the sequence demonstrates cycle consistency.

4 Conclusion

This paper introduced 4D F-MeshLDM, a conditional latent diffusion model for full-cycle 3D+t cardiac mesh synthesis with exact periodicity. A mesh VAE encodes phase anatomy, while a truncated Fourier representation enforces cycle-consistent latent trajectories whose coefficient tokens are modelled by a conditional Transformer DDPM. On 5,000 UK Biobank subjects, our method demonstrated high generation fidelity and achieved near-zero cycle closure error. Future work will investigate the performance of this work on populations with diverse cardiac cycle durations. We will also explore relaxing the fixed-topology limitation to better capture severe pathologies. Conditioning will be further extended to more diverse clinical covariates for downstream digital-twin and in-silico trials pipelines.

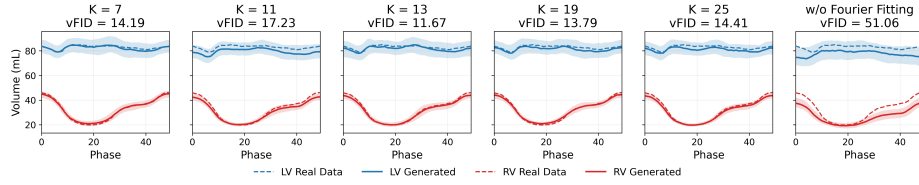


Fig. 2. Ablation study on the Fourier latent trajectory model with the different truncation frequency K on LV/RV volume trajectories.

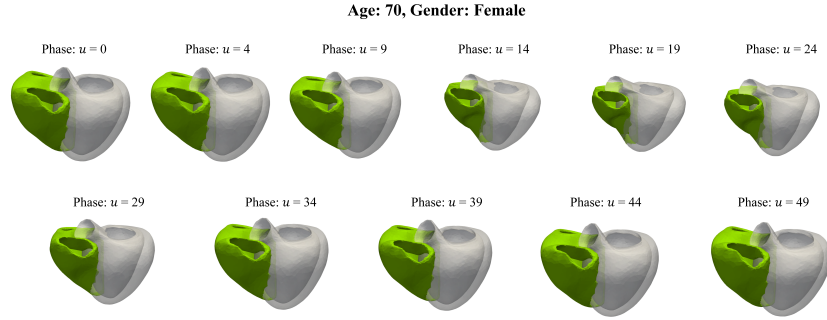


Fig. 3. Visual qualitative example of a generated cardiac cycle from 4D F-MeshLDM.

Acknowledgements. AFF acknowledges support from the Royal Academy of Engineering under the RAEng Chair in Emerging Technologies (INSILEX CiET1919/19), ERC Advanced Grant–UKRI Frontier Research Guarantee (INSILICO EP/Y030494/1), the UK Centre of Excellence on in-silico Regulatory Science and Innovation (UK CEiRSI) (10139527), the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Centre (BRC) (NIHR203308), the BHF Manchester Centre of Research Excellence (RE/24/30017), and the CRUK RadNet Manchester (C1994/A28701).

Disclosure of Interests. AF is Co-founder and shareholder of OculomeX Health Ltd, adsilico Ltd, and Synaptive Consulting Ltd.

References

1. Chen, H., Gao, J., Chen, Z., Gao, C., Huo, S., Jiang, M., Pu, J., Hu, C.: Improve myocardial strain estimation based on deformable groupwise registration with a locally low-rank dissimilarity metric. *BMC Medical Imaging* **24**(1), 330 (2024)
2. Davies, R.H., Twining, C.J., Cootes, T.F., Taylor, C.J.: Building 3-d statistical shape models by direct optimization. *IEEE Transactions on Medical Imaging* **29**(4), 961–981 (2009)
3. Dou, H., Huang, J., Zakeri, A., Zhou, Z., Mu, T., Duan, J., Frangi, A.: 4d cardiosynth: Synthesising dynamic virtual heart populations through spatiotemporal

- disentanglement. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12 (2025)
4. Dou, H., Ravikumar, N., Frangi, A.F.: A conditional flow variational autoencoder for controllable synthesis of virtual populations of anatomy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 143–152. Springer (2023)
 5. Dou, H., Virtanen, S., Ravikumar, N., Frangi, A.F.: A generative shape compositional framework to synthesize populations of virtual chimeras. *IEEE Transactions on Neural Networks and Learning Systems* **36**(3), 4750–4764 (2024)
 6. Gooya, A., Davatzikos, C., Frangi, A.F.: A bayesian approach to sparse model selection in statistical shape models. *SIAM Journal on Imaging Sciences* **8**(2), 858–887 (2015)
 7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: *Advances in Neural Information Processing Systems*. vol. 33, pp. 6840–6851 (2020)
 8. Hundley, W.G., Bluemke, D.A., Bogaert, J., Flamm, S.D., Fontana, M., Friedrich, M.G., Grosse-Wortmann, L., Karamitsos, T.D., Kramer, C.M., Kwong, R.Y., et al.: Society for cardiovascular magnetic resonance (scmr) guidelines for reporting cardiovascular magnetic resonance examinations. *Journal of Cardiovascular Magnetic Resonance* **24**(1), 29 (2022)
 9. Konduri, P.R., Marquering, H.A., Van Bavel, E.E., Hoekstra, A., Majoie, C.B., Investigators, I.: In-silico trials for treatment of acute ischemic stroke. *Frontiers in Neurology* **11**, 558125 (2020)
 10. Ma, Q., Meng, Q., Qiao, M., Matthews, P., O’Regan, D., Bai, W.: Cardiacflow: 3d+t four-chamber cardiac shape completion and generation via flow matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 89–99 (2025)
 11. Mozyrska, J., Beetz, M., Melas-Kyriazi, L., Grau, V., Banerjee, A., Bueno-Orovio, A.: 3d cardiac anatomy generation using mesh latent diffusion models. *arXiv preprint arXiv:2508.14122* (2025)
 12. Naeem, M.F., Oh, S.J., Uh, Y., Choi, Y., Yoo, J.: Reliable fidelity and diversity metrics for generative models. In: International Conference on Machine Learning. pp. 7176–7185 (2020)
 13. Pappalardo, F., Russo, G., Tshinanu, F.M., Viceconti, M.: In silico clinical trials: concepts and early adoptions. *Briefings in bioinformatics* **20**(5), 1699–1708 (2019)
 14. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4195–4205 (2023)
 15. Puntmann, V.O., Valbuena, S., Hinojar, R., Petersen, S.E., Greenwood, J.P., Kramer, C.M., Kwong, R.Y., McCann, G.P., Berry, C., Nagel, E., et al.: Society for cardiovascular magnetic resonance (scmr) expert consensus for cmr imaging end-points in clinical research: part i-analytical validation and clinical qualification. *Journal of Cardiovascular Magnetic Resonance* **20**(1), 67 (2018)
 16. Qiao, M., Wang, S., Qiu, H., De Marvao, A., O’Regan, D., Rueckert, D., Bai, W.: Cheart: A conditional spatio-temporal generative model for cardiac anatomy. *IEEE Transactions on Medical Imaging* **43**(3), 1259–1269 (2023)
 17. Qiao, M., McGurk, K.A., Wang, S., Matthews, P.M., O’Regan, D.P., Bai, W.: A personalized time-resolved 3d mesh generative model for unveiling normal heart dynamics. *Nature Machine Intelligence* **7**(5), 800–811 (2025)
 18. Qiao, M., Zheng, J., Zhang, W., Ma, Q., Li, L., Kainz, B., O’Regan, D.P., Matthews, P.M., Niederer, S., Bai, W.: Mesh4d: A motion-aware multi-view variational autoencoder for 3d+ t mesh reconstruction. In: International Conference

- on Medical Image Computing and Computer-Assisted Intervention. pp. 343–353 (2025)
19. Ranjan, A., Bolkart, T., Sanyal, S., Black, M.: Generating 3d faces using convolutional mesh autoencoders. In: Proceedings of the European Conference on Computer Vision. pp. 704–720 (2018)
 20. Sørensen, K., Diez, P., Margeta, J., El Youssef, Y., Pham, M., Pedersen, J., Kühl, T., De Backer, O., Kofoed, K., Camara, O., Paulsen, R.: Spatio-temporal neural distance fields for conditional generative modeling of the heart. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 422–432 (2024)
 21. Xia, Y., Chen, X., Ravikumar, N., Kelly, C., Attar, R., Aung, N., Neubauer, S., Petersen, S.E., Frangi, A.F.: Automatic 3d+ t four-chamber cmr quantification of the uk biobank: integrating imaging and non-imaging data priors at scale. *Medical Image Analysis* **80**, 102498 (2022)