# Variable deep learning training horizons reveal the temporal complexity of biological systems

Po-Hao Chiu[1*], Jacob I Evarts[2*], Patrick Feng[2*], Neda Bagheri[1,2§]

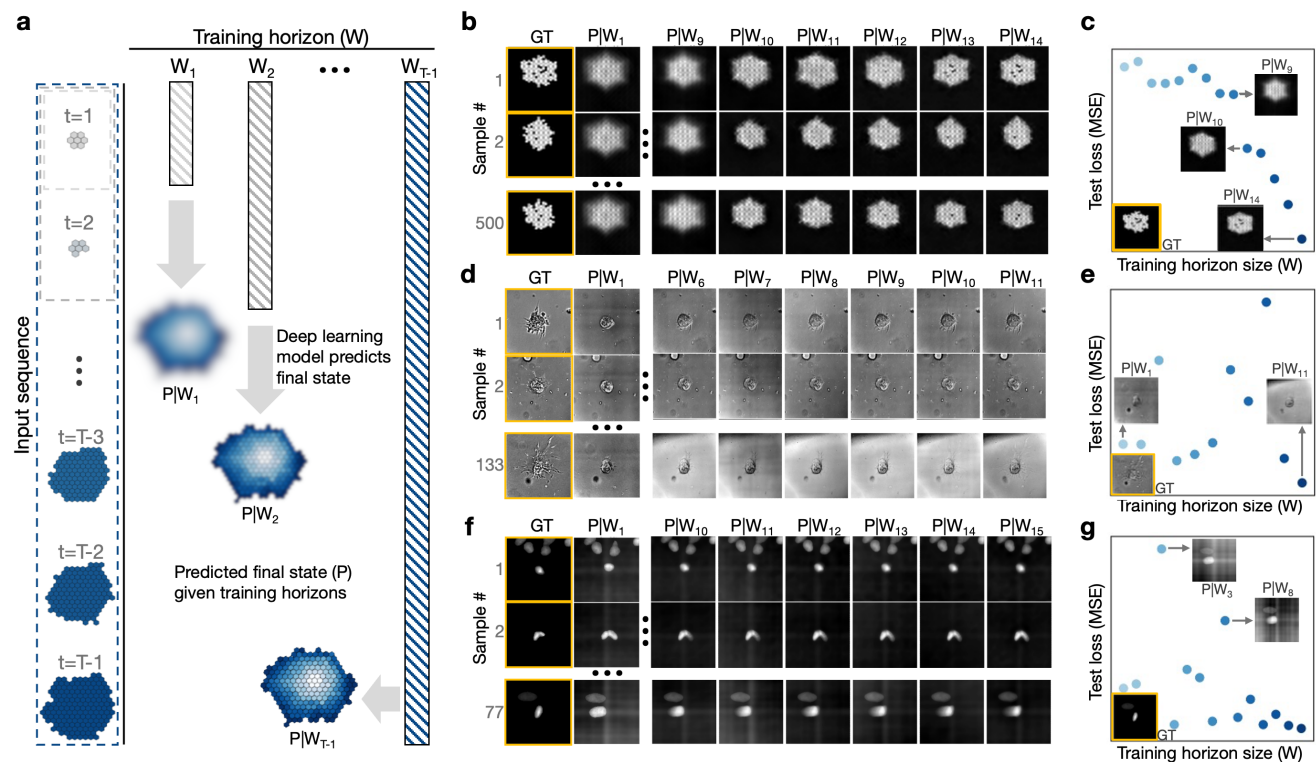[1]Chemical Engineering, University of Washington, Seattle, Washington, United States

[2]Biology, University of Washington, Seattle, Washington, United States

[§]To whom correspondence should be addressed: nbagheri@uw.edu

[*]These authors contributed equally.

## Abstract

The increasing quantity of time-series images presents new opportunities for extracting biological insights from data. Here, we introduce a deep learning framework with a variable input sequence length to predict cell and colony morphologies. We apply this framework to *in silico* and *in vitro* microscopy datasets, evaluating the impact of temporal data on performance. We find that while performance increases monotonically with increasing *in silico* training data, performance is varied in the *in vitro* case studies. The varying results reflect the intrinsic challenges stochastic, complex biological systems pose to data-driven modeling, and offer a new method through which we can identify biological transition points using temporal dynamics.

**Figure 1. FNOs trained on varying temporal input horizons exhibit different performance outcomes:**

(**a**) The deep learning pipeline enables interrogation of the impact of the training window horizon on prediction accuracy. An ARCADE simulation sequence is shown with gradient colors that correspond with tumor age. A variable input window $W_i$ is passed through an FNO that is trained to predict the final timepoint of the time series, which we term the prediction horizon P. (**b**) FNO predictions based on increasing training horizons are shown for three of 500 samples of the in silico tumor simulation data. The FNO predicts images at the 15th time point based on an input window $W_i$ where $1 \leq i \leq 14$. The leftmost column shows the ground truth image highlighted in yellow, and each subsequent column shows the predicted image given a window length $W_i$. Each row corresponds to a unique sample. (**c**) The test loss (quantified by the mean squared error, MSE) of *in silico* simulation data is shown as a function of training window size. Results highlight nearly monotonic improvement of performance with increased training window horizon. The image in the bottom left corner is the ground truth of sample 1 in panel b (highlighted in yellow), and notable predictions are shown for select window horizons. (**d**) FNO predictions based on increasing training horizons are shown for three of 133 samples of the in vitro tumor spheroid data. The FNO predicts images at the 12th time point based on an input window $W_i$ where $1 \leq i \leq 11$. The leftmost column shows the ground truth image highlighted in yellow, and each subsequent column shows the

predicted image given a window length $W_i$. Each row corresponds to a unique sample. (**e**) The test loss (MSE) of *in vitro* spheroid images is shown as a function of training window size. Unlike the simulated tumor case study, there are considerable fluctuations in FNO performance even with an increase in window horizon and temporal information. The image in the bottom left corner is the ground truth of sample 133 in panel d (highlighted in yellow), and notable predictions are shown for select window horizons. (**f**) FNO predictions based on increasing training horizons are shown for three of 77 samples of the in vitro tumor cell death data. The FNO predicts images at the 16th time point based on an input window $W_i$ where $1 \leq i \leq 15$. The leftmost column shows the ground truth image highlighted in yellow, and each subsequent column shows the predicted image given a window length $W_i$. Each row corresponds to a unique sample. (**g**) The test loss (MSE) of individual cell death images is shown as a function of training window size. Results are as variable as the other in vitro dataset. However, the spikes in test loss seem to derive from noisy backgrounds which contrast with the pure black of the ground truth. Excluding those poorly performing models, the loss curve is notably more monotonic than the cultured spheroid one. The image in the bottom left corner is the ground truth (highlighted in yellow) and notable predictions are shown for select window horizons.

## Description

Biological imaging techniques have achieved a high degree of spatial and temporal resolution, situating them as an important data source alongside 'omics data for biological discovery (Bagheri et al., 2022; Kuhn Cuellar et al., 2022). This increase in high-resolution temporal imaging introduces new opportunities and challenges for analysis to extract insights from morphological dynamics. Machine learning shows promise in going beyond feature extraction towards learning the underlying rules governing a system (Soelistyo et al., 2022; Rotem et al., 2024). A better understanding of how temporal dynamics affect the performance of machine learning models can help uncover the timing of biological processes (Bao et al., 2025; Toulany et al., 2023) and outline the performance-cost tradeoff curve for generating high temporal resolution data (Cain et al., 2024; Aceituno et al., 2025). Here, we develop a deep learning model to forecast future morphological states of three biological systems from time-series images using input sequences of varying lengths to probe how much temporal information is needed to make accurate predictions about future states. The three systems include: (i) simulated tumors generated by an agent-based model, (ii) cultured tumor spheroids, and (iii) individual tumor cell death dynamics.

Fourier neural operators (FNOs) are a unique neural network architecture developed for learning dynamics of systems governed by smooth, continuous processes, and have found success in emulating scientific partial differential equations (PDEs) (Kovachki et al., 2021). Although FNOs have shown strong performance in forecasting time-series across physical systems (Kurth et al., 2023; Li et al., 2023; Long et al., 2024), their sensitivity to temporal context, particularly with biological image data, remains an open question that impacts if and how one might characterize time-dependent biological processes. In this work, we train FNOs to predict a future state of the system (the final timepoint of the dataset), which we term the prediction horizon (P), from a preceding image sequence of variable length, $W_i$, the training horizon, where i refers to the number of time points in the training window (Fig. 1a). We hypothesized that prediction performance would increase as a function of the training window.

As expected, FNO performance increases monotonically with training window size in the *in silico* case study (Fig. 1b, c). Models trained on the simulated tumor dataset accurately predict general emergent properties, such as tumor size and location; however, more biologically relevant characteristics only appear when the model is trained with greater values of $W_i$. Specifically, for $W_i \leq 6$, there is no qualitative or quantitative (measured by test loss) improvement. When $W_{6 < i \leq 9}$, accurate tumor borders begin to appear. Between $W_9$ and $W_{10}$, the test loss decreases significantly and a clear accurate shape for the tumor emerges. In cases where $W_i > 10$, the predicted tumor has clear borders and the shape becomes increasingly precise with increasing $W_i$.

When the pipeline is applied to the tumor spheroid dataset, a different trend emerges. The models trained on the *in vitro* spheroids exhibit highly variant accuracy (Fig. 1d, e). Qualitatively, predictions with $W_{i > 6}$ capture finer boundary details and steadily decrease in test loss as the window size increases. However, the prediction with $W_9$ deviates from this trend, exhibiting an unusually high loss. We believe this variation in performance is due to the relatively short time-scale of the image dataset and the lack of substantial morphological changes over that period. Despite variations in the loss function, a qualitative assessment of performance points to consistent improvement with increasing $W_i$, suggesting that pixel-wise loss may not be a robust metric due to non-biological variations in the background of the predicted images. Nevertheless, the FNO seems to be learning increasingly more biological features and morphological characteristics of *in vitro* tumors with increasing $W_i$.

Having assessed performance in two distinct tumor spheroid datasets, we evaluated the impact of the training window on a cancer cell death time-lapse dataset with a finer temporal resolution. This system features a distinct biological event (cell death) that is characterized by a rapid condensing of fluoresced materials. Quantitatively, the test loss (Fig. 1f, g), exhibits non-monotonicity similar to the tumor spheroid dataset. Despite fluctuations in model performance, the FNO appears to improve as the training horizon increases. The primary inconsistencies in the loss curve qualitatively align with poor prediction of the image background, rather than in the prediction of biologically relevant dynamics. Qualitatively, results

indicate that prediction accuracy for biologically relevant dynamics improves with increasing $W_i$, as the FNO captures more accurate cell morphologies. This trend demonstrates that learning temporal dynamics is highly dependent on biological variance and technical noise, and that qualitative validation should be considered in concert with quantitative metrics.

In this paper, we introduce a variable training horizon pipeline using FNOs to investigate the complexity of temporal trends of several biological systems using imaging data. Our results reveal a contrast between simulated and real-world biological dynamics. The simulated tumor growth model, governed by smooth dynamics, showed a clear monotonic improvement in prediction accuracy as more temporal information was provided. However, in both the tumor spheroid and tumor cell death dynamics datasets, increasing the temporal information shows variant, non-monotonic performance landscapes. Factors that could limit the performance of data-driven models include ineffective sampling of critical transitions, biological "switches", and timescales where dynamics can be obscured by biological stochasticity and/or technical noise. These findings align with recent work suggesting that the loss landscape for auto-regressive prediction in dynamic systems can be surprisingly uneven (Aceituno et al., 2025), which we find to be exacerbated by the nonlinear dynamics inherent in biological processes. However, models trained on denoised microscopy images obtained by image processing methods (e.g., cell segmentation) produce smoother performance landscapes despite occasional outliers. Our results highlight both a challenge and an opportunity: while data-driven methods struggle with the unpredictability of biological systems, the model failure points can be used to identify key biological transitions.

## Methods

*In silico* **tumor microenvironment simulations.** Data was generated using ARCADE, an agent-based model (ABM) characterizing tumor growth in a heterogeneous and dynamic vascular microenvironment. We simulated 500 tumors under varying microenvironmental conditions over 15 days, capturing a snapshot of the *in silico* 2D-slice of tissue every 24 hours. The simulations used for this study are similar to those described in (Yu and Bagheri, 2020).

*In vitro* **cancer spheroid microscopy images.** Time-lapse images of tumor spheroids monitor the effects of interface stiffness on the invasiveness of tumor development (Thi Kim Ngan Ngo, 2022). This dataset contains time-lapse microscopy of 133 tumor spheroids cultured under varying extracellular matrix stiffness and surface topographies. We trained models on snapshots of these tumors taken every 2 hours over a period of 24 hours.

*In vitro* **cell death microscopy images**. A LNCaP cell line was treated with doxorubicin, a cell death–inducing compound, and time-lapse images monitored the impact of this treatment (Vicar et al., 2020). The dataset we used spans 24 hours with a frame rate 1 frame per 3 minutes, and a spatial resolution of 1.59 px/µm. We employed StarDist (Schmidt et al., 2018), a deep learning-based cell segmentation model, to identify individual cell locations. We tracked individual cell's location across time frames through a Kalman tracker. The timing of the cell death event was determined from the fluorescence intensity over time frames where a sharp increase indicates DNA condensation during cell death. We used 1 to 15 time frames prior to the cell death event for the window horizon input and predicted the system morphology 5 frames after the cell death.

**Spatiotemporal FNO model architecture**. The FNO model was developed with a 3D architecture—2 spatial dimensions and 1 temporal dimension—to effectively capture the spatiotemporal dynamics of tumor emergence. We implemented the FNO using the neuraloperator library (Kovachki et al., 2021; Kossaifi et al., 2024), which uses the PyTorch library (Paszke et al., 2019). Hyperparameters were defined as: modes=(24, 24, 8), layers=8, width=32. Datasets were partitioned into model training (80%), validation (10%), and test (10%) sets. All images were center-cropped and down-sampled to a fixed spatial resolution of $128 \times 128$. Model training minimized the H1 loss using the Adam optimizer (Kingma and Ba, 2017).

**Data and code availability.** All datasets and code are open and accessible. The deep learning model source code is available on Zenodo at 10.5281/zenodo.17478675. The ARCADE ABM v2.4 source code (used to generate the synthetic tumor images) is available on Zenodo at 10.5281/zenodo.10622155. The tumor spheroid and cell death datasets were previously published (Thi Kim Ngan Ngo, 2022; Vicar et al., 2020).

## References

Aceituno PV, Miller JW, Marti N, Farag Y, Boussange V. 2025. Temporal horizons in forecasting: a performance-learnability trade-off. DOI: 10.48550/arXiv.2506.03889

Angione C, Silverman E, Yaneske E. 2022. Using machine learning as a surrogate model for agent-based simulations. PLOS ONE. 17: e0263150. DOI: 10.1371/journal.pone.0263150

Bagheri N, Carpenter AE, Lundberg E, Plant AL, Horwitz R. 2022. The new era of quantitative cell imaging—challenges and opportunities. Molecular Cell. 82: 241.

Bao H, Zhang S, Yu Z, Xu H. 2025. Deep Learning-Based High-Resolution Time Inference for Deciphering Dynamic Gene Regulation from Fixed Embryos. DOI: 10.1101/2025.03.17.643618

Cain JY, Evarts JI, Yu JS, Bagheri N. 2024. Incorporating temporal information during feature engineering bolsters emulation of spatio-temporal emergence. Bioinformatics. 40: btae131. DOI: 10.1093/bioinformatics/btae131

Fadikar A, Higdon D, Chen J, Lewis B, Venkatramanan S, Marathe M. 2018. Calibrating a Stochastic, Agent-Based Model Using Quantile-Based Emulation. SIAM/ASA Journal on Uncertainty Quantification. 6: 1685. DOI: 10.1137/17M1161233

Kingma DP, Ba J. 2017. Adam: A Method for Stochastic Optimization. DOI: 10.48550/arXiv.1412.6980

Kossaifi J, Kovachki N, Li Z, Pitt D, Liu Schiaffini M, Duruisseaux V, et al., Anandkumar A. 2024. A Library for Learning Neural Operators. DOI: 10.48550/ARXIV.2412.10354

Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, Stuart A, Anandkumar A. 2021. Neural Operator: Learning Maps Between Function Spaces. DOI: 10.48550/ARXIV.2108.08481

Kuhn Cuellar L, Friedrich A, Gabernet G, De La Garza L, Fillinger S, Seyboldt A, et al., Nahnsen S. 2022. A data management infrastructure for the integration of imaging and omics data in life sciences. BMC Bioinformatics. 23: 61. DOI: 10.1186/s12859-022-04584-3

Kurth T, Subramanian S, Harrington P, Pathak J, Mardani M, Hall D, et al., Anandkumar A. 2023. FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. DOI: 10.1145/3592979.3593412

Lai G, Chang WC, Yang Y, Liu H. 2017. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. DOI: 10.48550/ARXIV.1703.07015

Lan Y, Shin SY, Nguyen LK. 2025. From shallow to deep: The evolution of machine learning and mechanistic model integration in cancer research. Current Opinion in Systems Biology. 40: 100541. DOI: 10.1016/j.coisb.2025.100541

Li B, Wang H, Feng S, Yang X, Lin Y. 2023. Solving Seismic Wave Equations on Variable Velocity Models With Fourier Neural Operator. IEEE Transactions on Geoscience and Remote Sensing. 61: 1. DOI: 10.1109/TGRS.2023.3333663

Long R, Xie H, Lian D. 2024. Time Series Forecasting with Multi-scale Decomposition and Fourier Neural Operators. 2024 7th International Conference on Computer Information Science and Application Technology (CISAT): 952. DOI: 10.1109/CISAT62382.2024.10695364

Mao BH, Nguyen Thi KM, Tang MJ, Kamm RD, Tu TY. 2023. The interface stiffness and topographic feature dictate interfacial invasiveness of cancer spheroids. Biofabrication. 15: 015023. DOI: 10.1088/1758-5090/acaa00

Murray EJ, Marshall BDL, Buchanan AL. 2021. Emulating Target Trials to Improve Causal Inference From Agent-Based Models. American Journal of Epidemiology. 190: 1652. DOI: 10.1093/aje/kwab040

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al., Chintala S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. DOI: 10.48550/ARXIV.1912.01703

Rotem O, Schwartz T, Maor R, Tauber Y, Shapiro MT, Meseguer M, et al., Zaritsky A. 2024. Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization. Nature Communications. 15: 7390. DOI: 10.1038/s41467-024-51136-9

Schmidt U, Weigert M, Broaddus C, Myers G. 2018. Cell Detection with Star-convex Polygons.

Soelistyo CJ, Vallardi G, Charras G, Lowe AR. 2022. Learning biophysical determinants of cell fate with deep neural networks. Nature Machine Intelligence. 4: 636. DOI: 10.1038/s42256-022-00503-6

Thi Kim Ngan Ngo. 2022. The interface stiffness and topographic feature dictate interfacial invasiveness of cancer spheroids. DOI: 10.17632/4V625RP3CJ.1

Toulany N, Morales Navarrete H, Capek D, Grathwohl J, Unalan M, Muller P. 2023. Uncovering developmental time and tempo using deep learning. Nature Methods. 20: 2000. DOI: 10.1038/s41592-023-02083-8

Vicar T, Raudenska M, Gumulec J, Balvan J. 2020. The Quantitative-Phase Dynamics of Apoptosis and Lytic Cell Death. Scientific Reports. 10: 1566. DOI: 10.1038/s41598-020-58474-w

Yu JS, Bagheri N. 2020. Agent-Based Models Predict Emergent Behavior of Heterogeneous Cell Populations in Dynamic Microenvironments. Frontiers in Bioengineering and Biotechnology. 8: 249. DOI: 10.3389/fbioe.2020.00249

**Conflicts of Interest:** The authors declare that there are no conflicts of interest present.