Genentech
A Member of the Roche Group

Explainable Deep Learning for Tumor Dynamic Modeling and Overall Survival Prediction

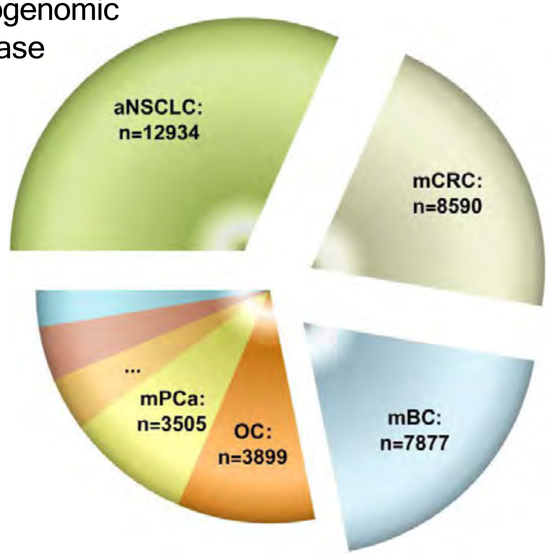
James Lu, Genentech

IQ Workshop on *Machine Intelligence for Quantitative Modeling in Drug Discovery & Development Applications*
15-16 September 2022

Oncology Data Challenge in the Digital Age

- Current trends in technology & digitization generate growth across:
  - the number of patients
  - the dimensionality of longitudinal measurements
  - the multimodality of data

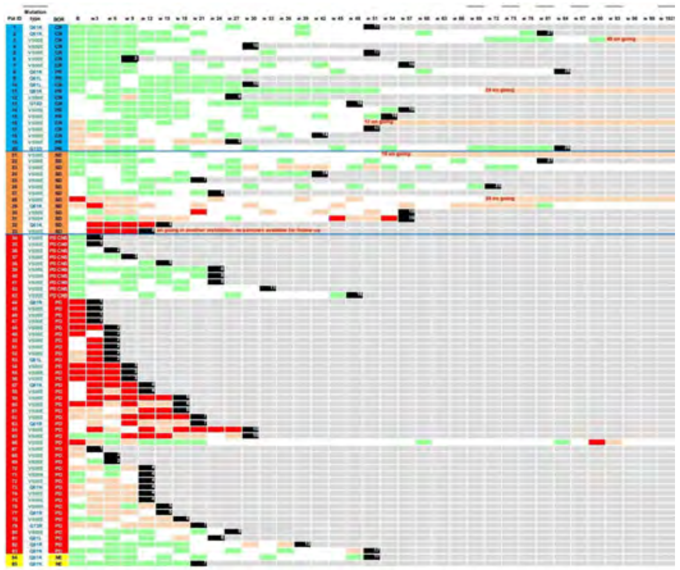
eg, Flatiron Clinicogenomic Database



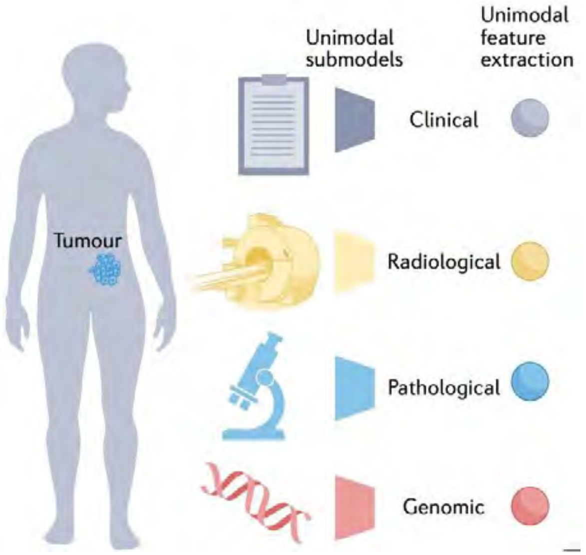
> 40K patients

Liu *et al*, Nature Med (2022)

eg, ctDNA



Seremet *et al*, J Transl. Med. (2019)



Boehm *et al*, Nature Reviews Cancer (2021)

Genentech
A Member of the Roche Group

A Brief History of Evolution in Tumor Dynamics for OS Prediction

From parametric/Cox to ML models

	2018	2019	2020	2021
	Clinical Trials: Immunotherapy. Clinical Cancer Research. **A Model of Overall Survival Predicts Treatment Outcomes with Atezolizumab versus Chemotherapy in Non-Small Cell Lung Cancer Based on Early Tumor Kinetics**. Laurent Claret[1], Jin Y. Jin[2], Charles Ferté[3], Helen Winter[2], Sandhya Girish[2], Mark Stroh[2], Pei He[4], Marcus Ballinger[5], Alan Sandler[5], Amita Joshi[2], Achim Rittmeyer[6], David Gandara[7], Jean-Charles Soria[3], and René Bruno[1]	**Machine-Learning and Stochastic Tumor Growth Models for Predicting Outcomes in Patients With Advanced Non–Small-Cell Lung Cancer**. Kien Wei Siah, SM[1]; Sean Khozin, MD, MPH[2]; Chi Heem Wong, SM[1]; and Andrew W. Lo, PhD[1,3]	**Machine Learning and Mechanistic Modeling for Prediction of Metastatic Relapse in Early-Stage Breast Cancer**. Chiara Nicolò, PhD[1,2]; Cynthia Périer, PhD[1,2]; Melanie Prague, PhD[3,4]; Carine Bellera, PhD[4,5]; Gaëtan MacGrogan, MD[6,7]; Olivier Saut, PhD[1,2]; and Sébastien Benzekry, PhD[1,2]	**Application of Machine Learning for Tumor Growth Inhibition – Overall Survival Modeling Platform**. Phyllis Chan[1,*,†], Xiaofei Zhou[1,2,4,†], Nina Wang[1], Qi Liu[1], René Bruno[3] and Jin Y. Jin[1]
Tumor Model	$\mathrm{TS}(t)=\begin{cases}\mathrm{TS}_0\cdot[\exp(\mathrm{KG}\cdot(t-\mathrm{tlag}))-1] & \text{if } t<\mathrm{tlag}\\ \mathrm{TS}_0\cdot[\exp(-\mathrm{KS}\cdot(t-\mathrm{tlag}))+\exp(\mathrm{KG}\cdot(t-\mathrm{tlag}))-1] & \text{if } t^3\geq\mathrm{tlag}\end{cases}$	$dY(t)=(\mu+\beta^T x+\sigma\xi(t))Y(t)\ dt,\quad Y(t),\sigma>0;\ \beta,x\in\mathbb{R}^m$ $dY(t)=(\mu+\beta^T x)Y(t)\ dt+\sigma Y(t)\ dW(t)$	$V_p^i(t)=e^{\frac{a^i}{b^i}\left(1-e^{-b^i t}\right)}$	$\mathrm{TS}(t)=\begin{cases}\mathrm{TS}_0\cdot[\exp(\mathrm{KG}\cdot(t-\mathrm{tlag}))-1] & \text{if } t<\mathrm{tlag}\\ \mathrm{TS}_0\cdot[\exp(-\mathrm{KS}\cdot(t-\mathrm{tlag}))+\exp(\mathrm{KG}\cdot(t-\mathrm{tlag}))-1] & \text{if } t^3\geq\mathrm{tlag}\end{cases}$
Num. of patients	n=1003	n=7805	n=642	n=668
Num. of var.	~20 categorical+continuous	~50 categorical+continuous	~20 categorical+continuous	~30 categorical+continuous
Link to Survival	Parametric/Cox	Cox, Random Survival Forest, Multi-layer Perceptron	Random Survival Forest	Lasso, Random Forest, Boosting, Kernel Machine

…

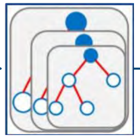
Genentech
A Member of the Roche Group

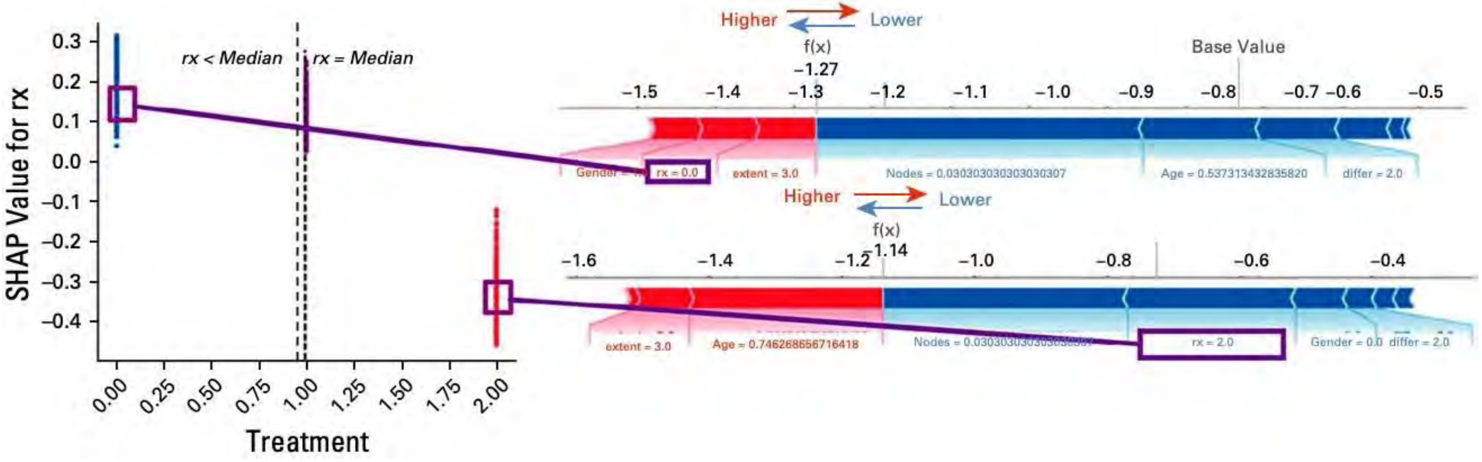
Explaining Nonlinear ML Models for Survival Predictions


ARTIFICIAL INTELLIGENCE

Computing the Hazard Ratios Associated With Explanatory Variables Using Machine Learning Models of Survival Data

Sameeer Sundrani[1,2] and James Lu, PhD[1]

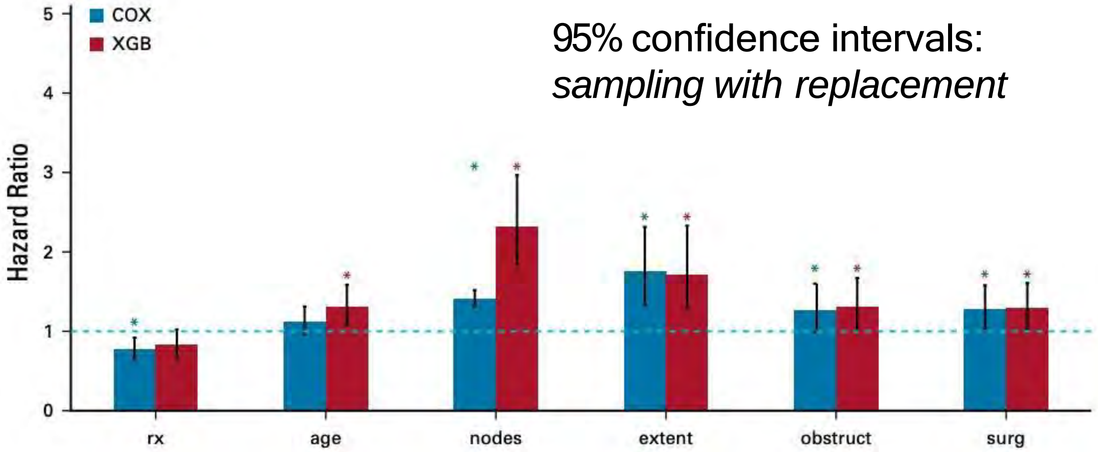
$x_i \rightarrow$  $\rightarrow f(x_i)$: $\quad h_i(t) = \exp\big(f(x_i)\big) \times h_0(t),$



Shapley Additive Values (SHAP) as a unifying way to both explain variable contribution (Φ) to ML model prediction and quantify contribution to the hazard function:

$$f(x_i) = \Phi_0 + \sum_{j=1}^{P} \Phi_j(f, x_i),$$

$$h_i(t) = \underbrace{\exp\big(\Phi_1(f, x_i)\big)}_{\text{var. 1}} \times \underbrace{\exp\big(\Phi_2(f, x_i)\big)}_{\text{var. 2}} \times \cdots \times \exp(\Phi_0)) \times h_0(t),$$



95% confidence intervals:
sampling with replacement

Lundberg *et al*, Nature Mach Intell (2020)

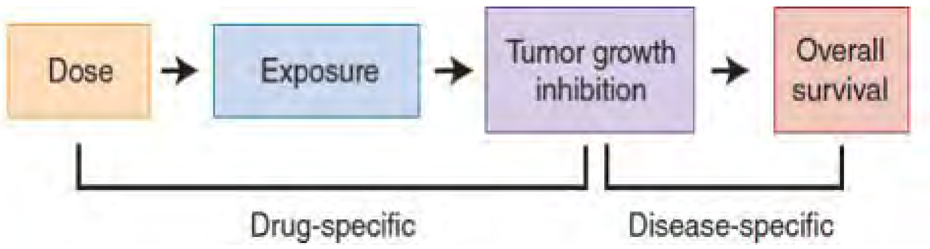
Towards Next Generation Oncology Disease Modeling

Established TGI-OS

e.g., $$\mathrm{TS}(t) = \begin{cases} \mathrm{TS}_0 \cdot [\exp(\mathrm{KG} \cdot (t{-}\mathrm{tlag})){-}1] & \text{if } t < \mathrm{tlag} \\ \mathrm{TS}_0 \cdot [\exp(-\mathrm{KS} \cdot (\mathrm{t}{-}\mathrm{tlag})) + \exp(\mathrm{KG} \cdot (\mathrm{t}{-}\mathrm{tlag})) - 1] & \text{if } t^3 \geq \mathrm{tlag} \end{cases}$$

Chosen tumor model

Parametric statistical model



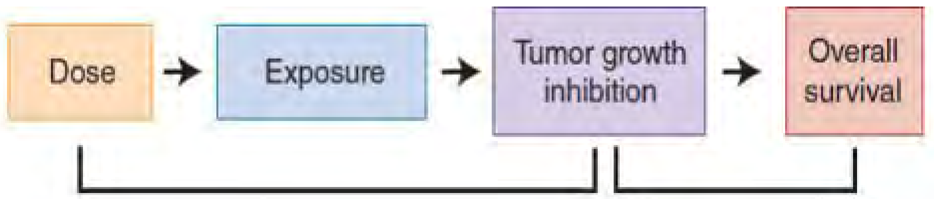
Chan *et al*, Prediction of overall survival in patients across solid tumors following atezolizumab treatments: A tumor growth inhibition-overall survival modeling framework, CPT:PSP (2021)

Augmentation with Machine Intelligence

High Content Data

Data-driven, interpretable Neural-ODE

Nonlinear ML model

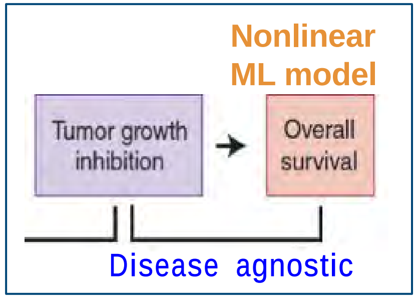


Disease agnostic

Potential benefits:

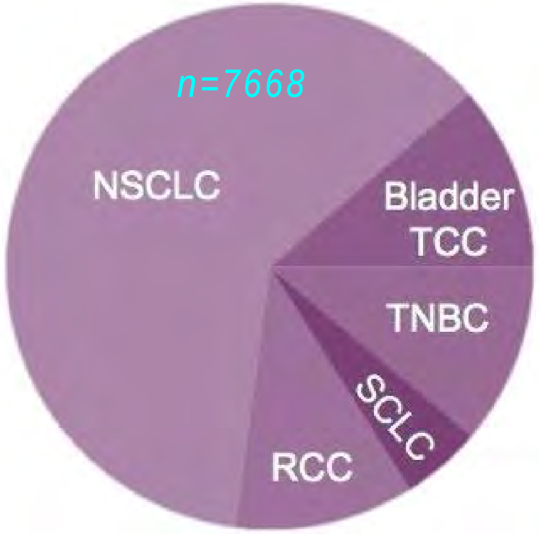
- *Improved predictivity (at the patient and/or trial level)*
- *Cross-molecule learning*
- *Disease understanding & extrapolation*

Pan-Indication Machine Learning Model for TGI-OS

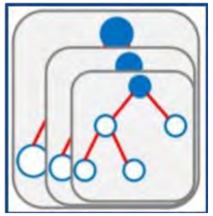


A single ML model able to predict across different solid tumors

Patient Data Across Solid Tumors



Indication-Independent ML model

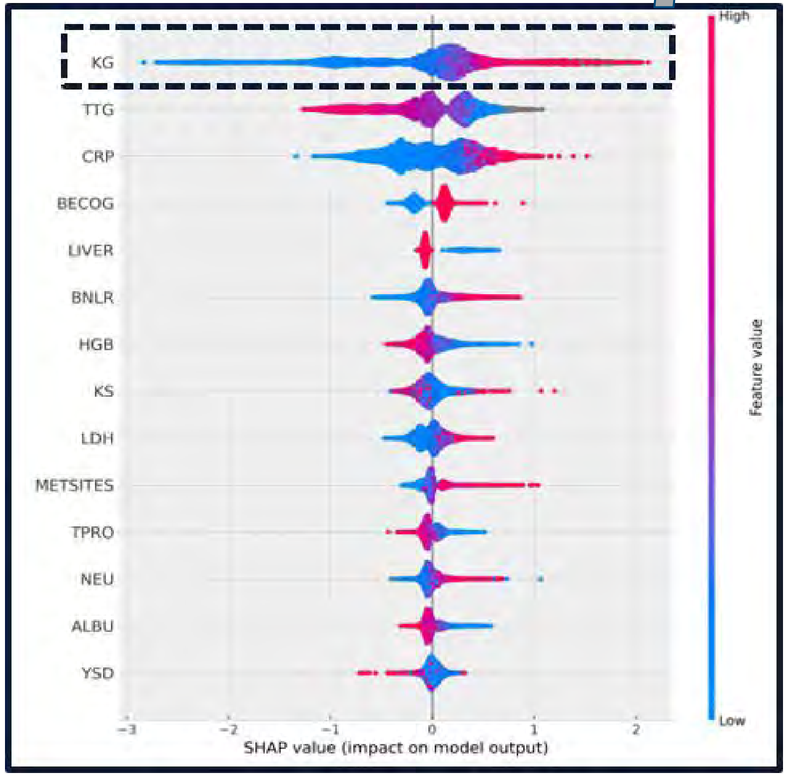


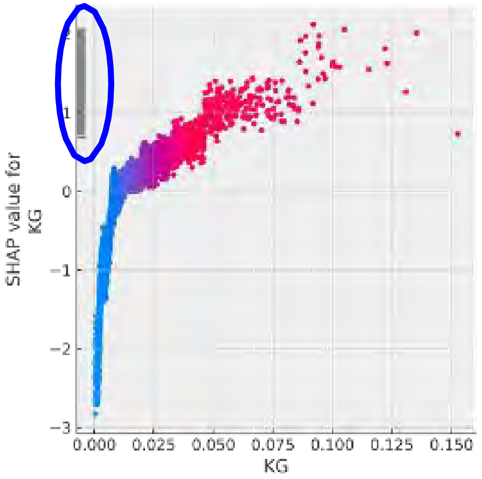
Advantages:

- Better predictivity (c-index≅0.8)
- Handles *TGI non-evaluable*
- Better extrapolation to novel tumor types

Model Explanation

$$TS(t) = \begin{cases} TS_0 \cdot [\exp(KG \cdot (t - tlag)) - 1] & \text{if } t < tlag \\ TS_0 \cdot [\exp(-KS \cdot (t - tlag)) + \exp(KG \cdot (t - tlag)) - 1] & \text{if } t^3 \geq tlag \end{cases}$$



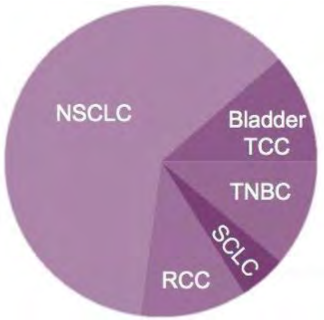


ML inference of hazard rate for *TGI non-evaluable* patients

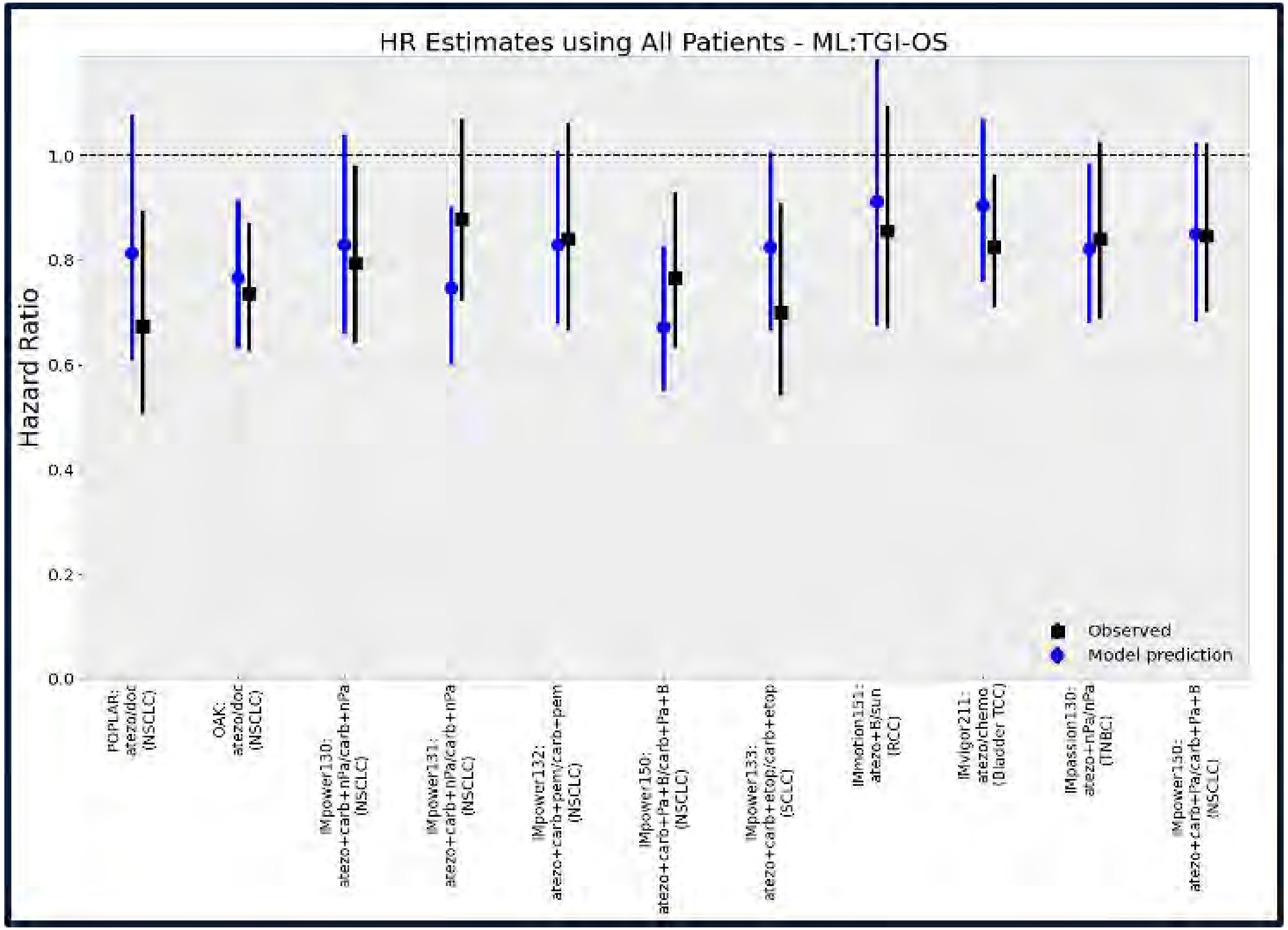
Duda *et al*, ASCPT (2021)

Pan-Indication Machine Learning Model Predicts Hazard Ratio

- Predicted hazard ratios (HRs) in test sets across 11 arms of 10 clinical trials over 5 solid tumor cancer types in agreement with observed HRs



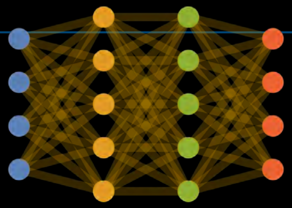
- Opportunities in tumor dynamic modeling:
  - Enable prediction from earlier tumor measurements
  - Bias in tumor size predictions
  - Use of multimodal data



The Merging of Deep Learning with Dynamical Systems

Deep Learning

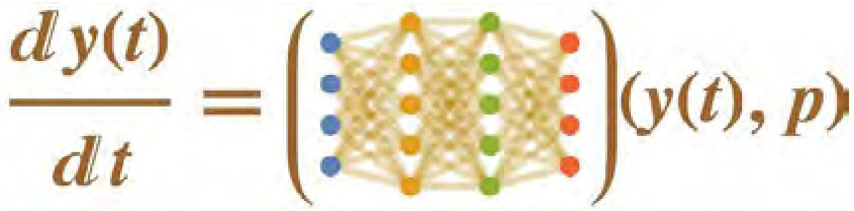
- Excels in approximating high dimensional/nonlinear functions
- Learn to improve model as data increases

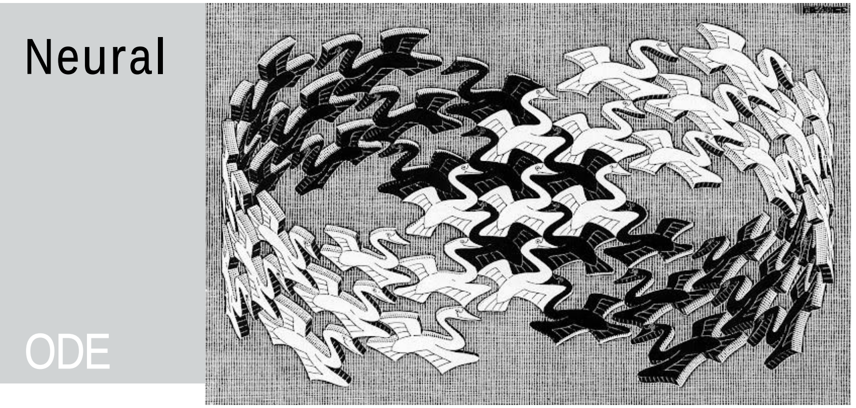


Dynamical Systems

- Enables the assumption of a time-invariant system
- Enables the abstraction of longitudinal data to low, fixed dimensional metrics, *p*

$$y'(t) = f(y(t), p)$$





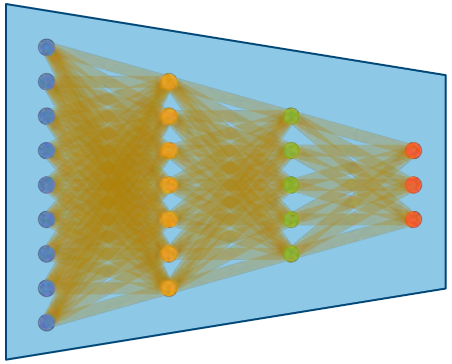
Tumor Dynamics Neural-ODE (TDNODE): an autonomous dynamical system that learns from tumor dynamic data and enables extraction of metrics that can predict patient survival

The Architecture of Tumor Dynamics Neural-ODE enhances Interpretability

Encoder-Decoder Architecture

<u>Data</u>

Longitudinal measurements of Sum-of-Longest Diameters (SLD)

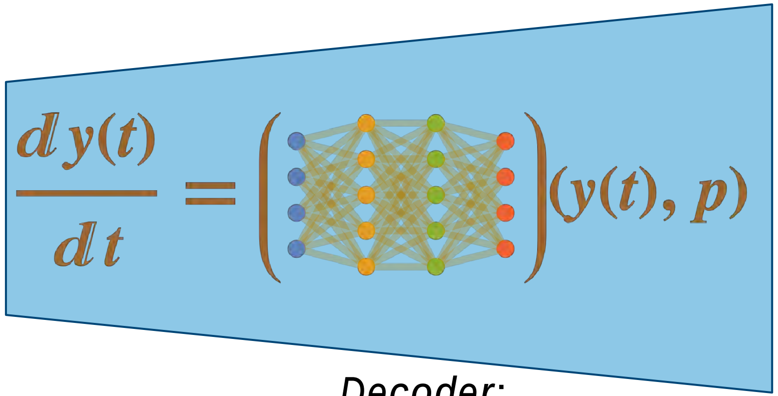


Encoder: patient-specific data abstraction

p

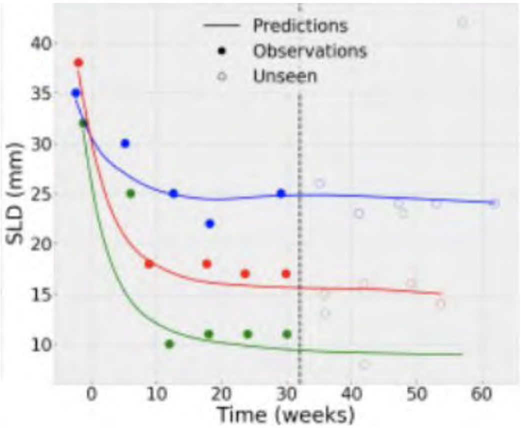
TDNODE metrics

Overall Survival?



Decoder: predictions using a <u>single</u> system of ODEs based on metrics *p*

<u>Prediction</u>

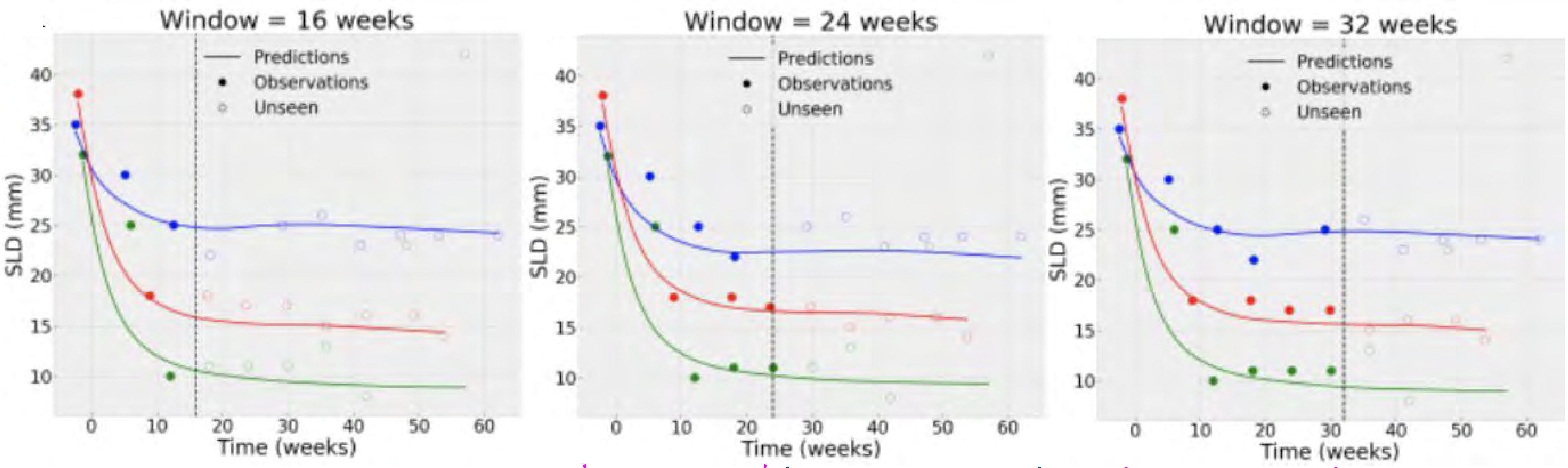


Benefits:

- Abstraction of patient data into a low-dimensional latent representation, *p*
- Representation of observed dynamics with a time-invariant dynamical system
- Interpretation of tumor metrics *p* and utilization in survival prediction

Data Augmentation Enables Robust Tumor Size Predictions

- Data augmentation can enrich existing data to improve accuracy & robustness of TDNODE
- Illustration of performing data augmentation on patient data using a 3 set of observation windows:



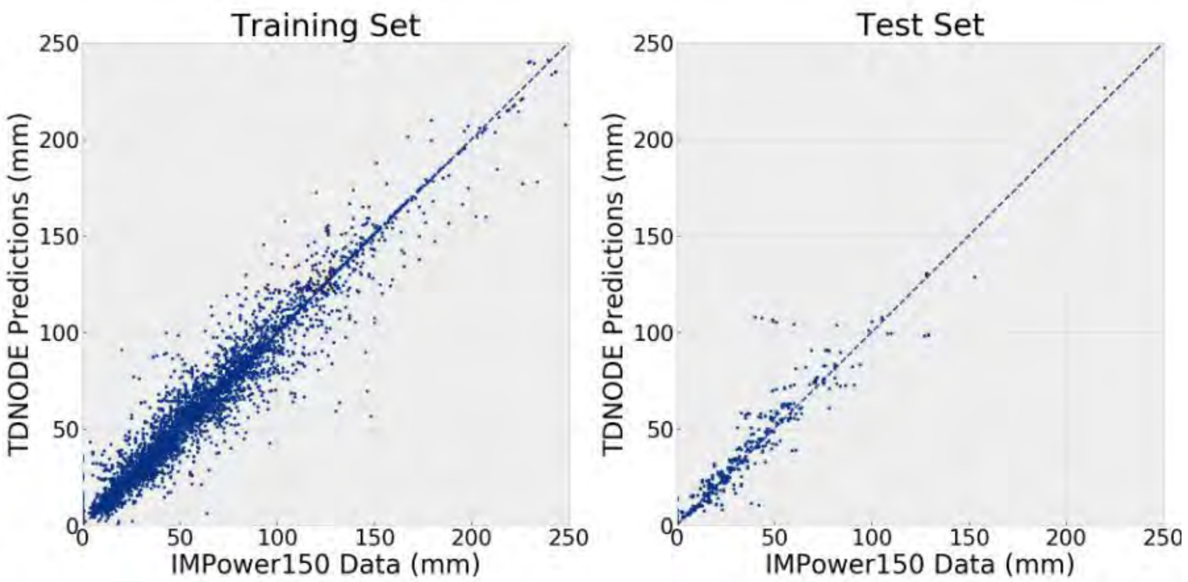
Data fed into the encoder

Prediction to be optimized

Data fed into the encoder

Data fed into the encoder

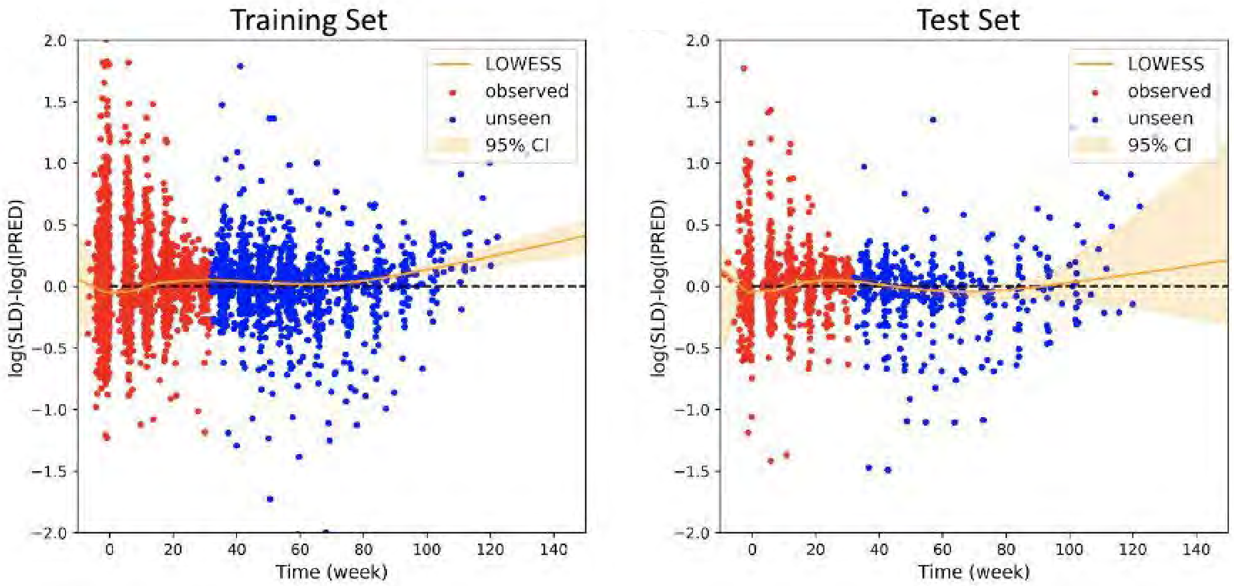
TDNODE provides Accurate Tumor Predictions with Minimal Bias



Treatment ARM	#Predictions, $t > t_N$	RMSE (median±STD)	R2 Score (median±STD)
Arm 1: Atezo.+Carb.+Pac.	208	10.4±1.40	0.828±0.05
Arm 2: Atezo.+Carb.+Pac.+Bev.	214	7.80±0.62	0.925±0.01
Arm 3: Carb.+Pac.+Bev.	79	5.70±0.59	0.961±0.02
All Treatment arms	**501**	**8.72±0.77**	**0.900±0.02**

Several formulations of tumor dynamic models are known to show biased predictions when data is truncated.

By the formulation of loss function and data augmentation, TDNODE provides minimal bias in predicting future tumor values:



From Tumor Data to OS Predictions

Pharmacometrics TGI-OS Modeling

Longitudinal measurements of
Sum-of-Longest Diameters (SLD)

chosen tumor model

$$TS(t) = \begin{cases} TS_0 \cdot [\exp(KG \cdot (t - tlag)) - 1] & \text{if } t < tlag \\ TS_0 \cdot [\exp(-KS \cdot (t - tlag)) + \exp(KG \cdot (t - tlag)) - 1] & \text{if } t^3 \geq tlag \end{cases}$$

population approach

Select TGI metrics (eg, KG)

trial-and-error

Parametric Survival Model

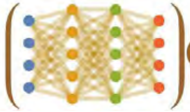
variance-covariance matrix

OS curves & Hazard Ratio (median & 95% PI)

TDNODE-OS.ML

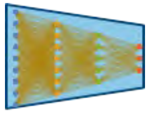
Longitudinal measurements of
Sum-of-Longest Diameters (SLD)

$$\frac{dy(t)}{dt} = \Big(\quad \Big)(y(t), p)$$

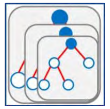


supervised ML

Encoder Output



supervised ML



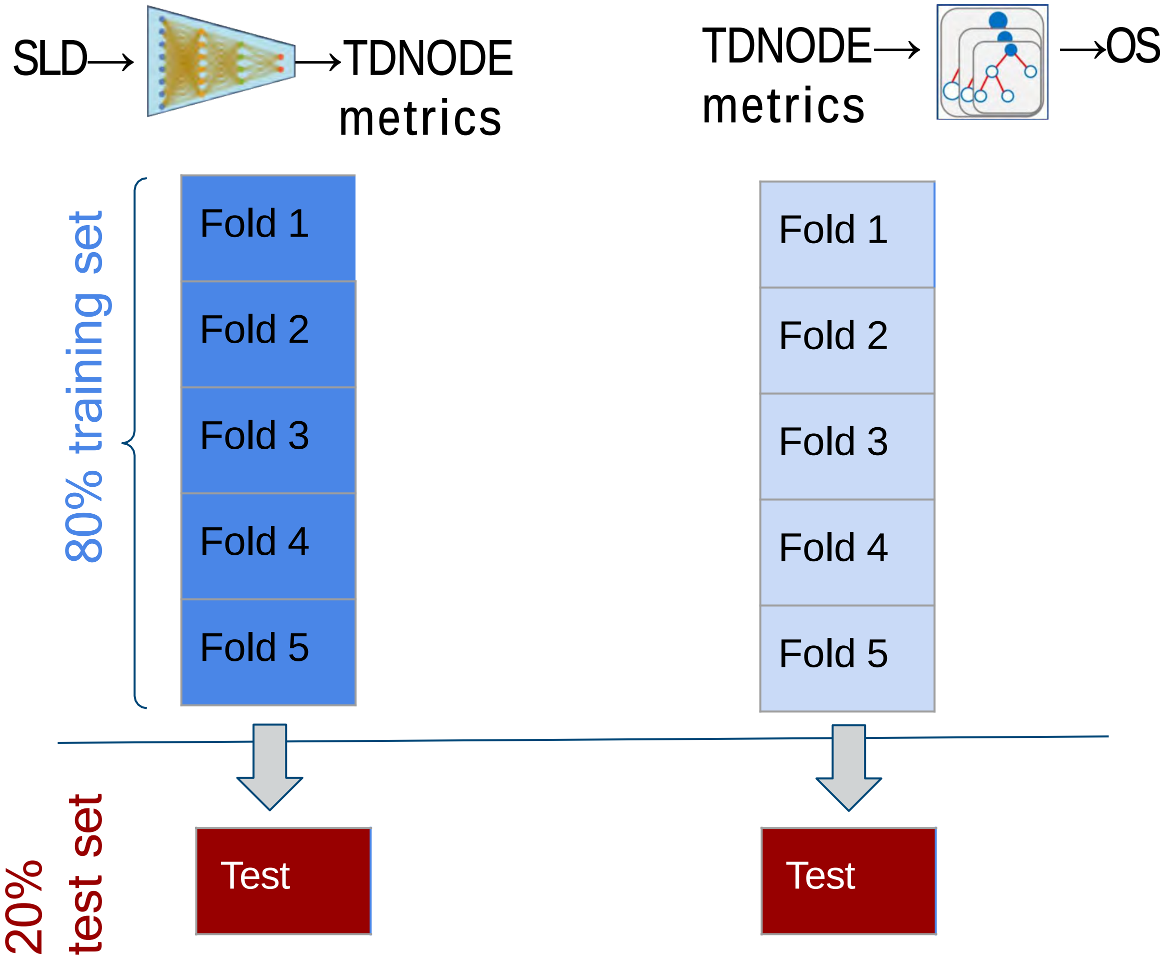
sampling with replacement

OS curves & Hazard Ratio (median & 95% PI)

Chan *et al*, CPT:PSP (2021)

TDNODE Metrics Can Accurately Predict Survival at the Individual Level

- While TDNODE metrics have only been trained on tumor dynamic (SLD) data, they can be used to predict patient OS: *demonstrates improved predictive performance compared to TGI metrics*

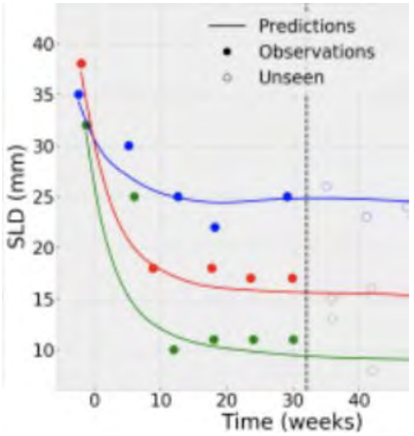


Model	Inputs	Input size	C-index via 5-fold CV	C-index on test set
TGI-OS.ML	KG, KS, TTG	3-dim	0.72±0.01	0.68
TDNODE-OS.ML	*p*	6-dim	**0.84±0.02**	**0.83**

TDNODE Metrics Predict Survival Curve & Hazard Ratios

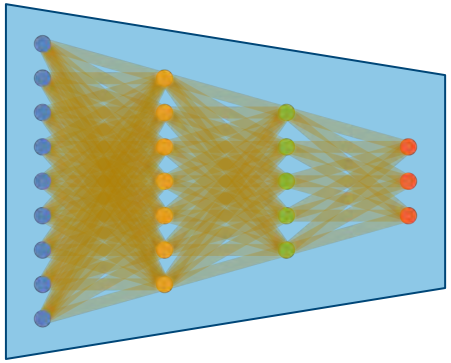
Model Predictions on n=216 (unseen) Test Patients

bootstrapped 500 times



Longitudinal tumor data

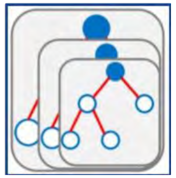
from Test Patients



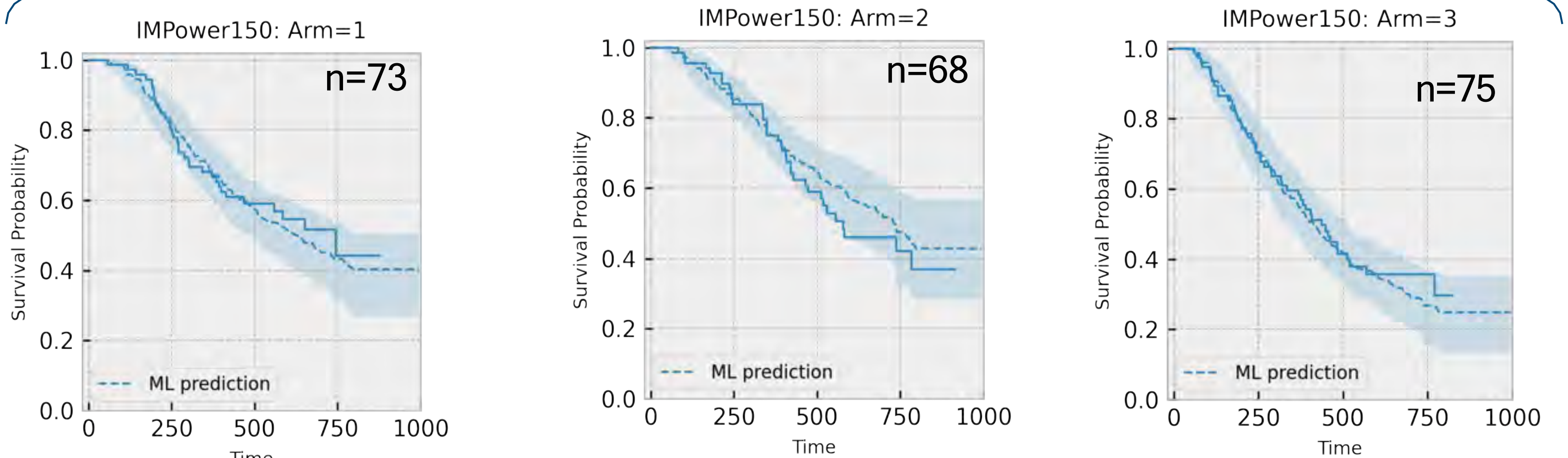
Encoder

TDNODE metrics

p



XGBoost:
OS predictor



Arm-1-vs-3

- Observed data:
  - HR=0.64 [0.41, 1.00]
- TDNODE-OS.ML:
  - HR=0.62 [0.47, 0.82]

Arm-2-vs-3

- Observed data:
  - HR=0.66 [0.42, 1.02]
- ML model (test set):
  - HR=0.54 [0.40, 0.72]

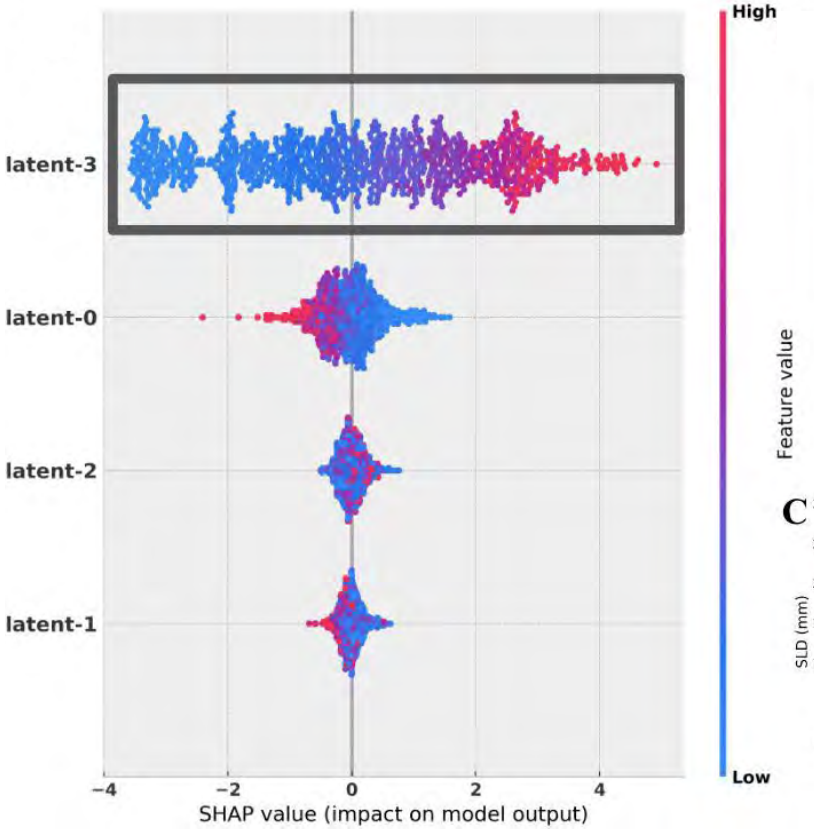
Genentech
A Member of the Roche Group

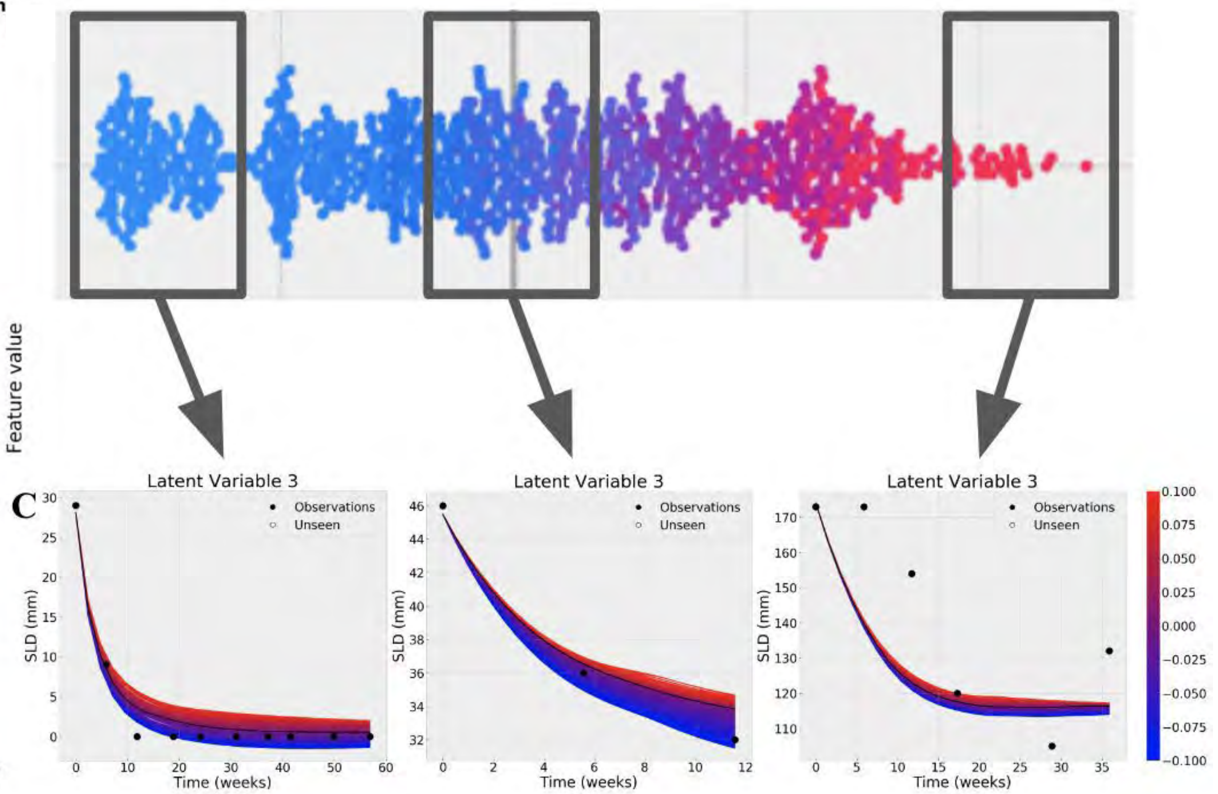
Interpreting ML Survival Model & **Influential** TDNODE Metrics

SHAP analysis: dependence of OS on p

Dependence of TDNODE model on parameter p

$$\frac{dy(t)}{dt} = (\text{NN})(y(t), p)$$

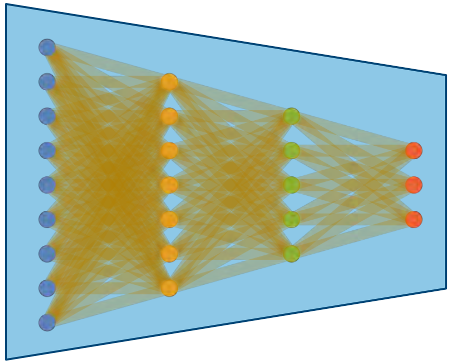




Tumor Dynamics Neural-ODE as the Foundation for Incorporating Multimodal Data

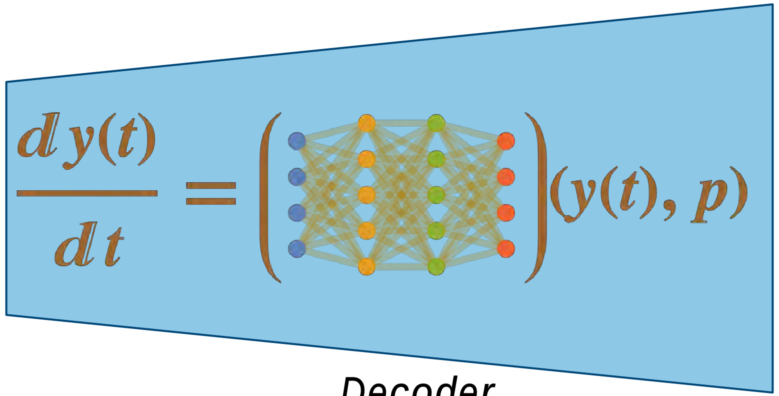
Data

SLD
+
ctDNA
+



Encoder

p

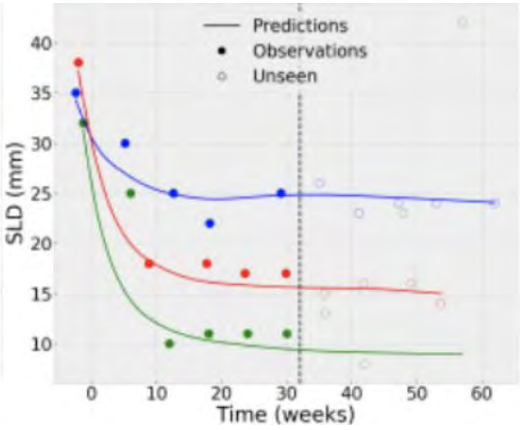


Decoder

Latent Patient Representation

Overall Survival

Prediction



Conclusion

- The increasing needs for highly predictive Oncology Disease Progression Models using longitudinal, multimodal data calls for advanced AI algorithms
- Tumor Dynamics Neural-ODE (TDNODE) provides an explainable DL approach that mirrors the workflow of PMx TGI-OS paradigm
- TDNODE provides a principled foundation for incorporating multimodal with longitudinal data

Acknowledgement

Genentech

- Mark Laurie, Omid Bazgir, Rashed Harun, Phyllis Chan, Rene Bruno, Kenta Yoshida, Ben Wu, Jin Jin, Chunze Li, Amita Joshi

Genentech

A Member of the Roche Group