Data-Driven Neural-ODE Modeling for Breast Cancer Tumor Dynamics and Progression-Free Survival Predictions

Jinlin Xiang*, **Bozhao Qi**†, **Marc Cerou**‡, **Wei Zhao***, **Qi Tang**†
*450 Water St, Cambridge, MA, USA 02141
†55 Corporate Dr, Bridgewater, NJ, USA, 08807
‡Sanofi R&D, Chilly-Mazarin, France, 91380
Jinlin.Xiang@Sanofi.com, Bozhao.Qi@Sanofi.com, Marc.Cerou@Sanofi.com
Wei.Zhao@Sanofi.com, Qi.Tang@Sanofi.com

Abstract

Pharmacokinetic/Pharmacodynamic (PK/PD) modeling plays a pivotal role in novel drug development. Previous population-based PK/PD models encounter challenges when customized for individual patients. We aimed to investigate the feasibility of constructing a pharmacodynamic model for different phases of individual breast cancer pharmacodynamics, only leveraging limited data from early phases. To achieve that, we introduced an innovative approach, Data-driven Neural Ordinary Differential Equation (DN-ODE) modeling for multi-task, e.g., breast cancer tumor dynamics and progression-free survival predictions. To validate the DN-ODE approach, we conducted experiments with early-phase clinical trial data from the amcenestrant (an oral treatment for breast cancer) dataset (AMEERA 1-2) to predict pharmacodynamics in the later phase (AMEERA 3). Empirical investigations confirmed the efficacy of the DN-ODE, surpassing alternative PK/PD methodologies. Notably, we also introduced visualizations for each patient, demonstrating that the DN-ODE recognizes diverse tumor growth patterns (responded, progressed, and stable). Therefore, the DN-ODE model offers a promising tool for researchers and clinicians, enabling a comprehensive assessment of drug efficacy, identification of potential responders, and facilitation of trial design.

1 Introduction

Pharmacokinetic/Pharmacodynamic (PK/PD) modeling plays a critical role in pharmaceutical research and the development of new drugs [1]. These models simulate the interaction of drugs with the human body, optimize dosing regimens for individual patients, and assess drug efficacy and safety [2, 3]. We asked whether it is possible to develop a comprehensive breast tumor pharmacodynamic model for different phases of clinical trials using limited data (e.g., sparse medical records from the early phases) [4, 5]. Specifically, we aimed to predict individual pharmacodynamics for patients from different clinical trials, identify possible responders, and thereby assist researchers in designing the clinical trials.

While the typical population-based PK/PD approaches for modeling sparse clinical trial data have become firmly established, they still face challenges when it comes to estimating pharmacodynamic responses at the individual patient level [6, 7, 8]. Recent Artificial Intelligence for Science and healthcare applications have attracted significant interest, ranging from machine learning to deep learning and neural models [9, 10, 11, 12, 13]. Some studies compared various regression methodologies for pharmacokinetic modeling, including multiple learning regression, Support Vector Machine (SVM), random forest and online learning approach [14, 15, 16, 17]. Some studies focus on imbal-

DLDE-III Workshop in the 37th Conference on Neural Information Processing Systems (NeurIPS 2023).

anced covariates and adjust them for logistic regression [18, 19, 20]. Recent approaches unitized a Long Short-Term Memory (LSTM) model to simulate pharmacokinetic/pharmacodynamic data for hypothetical drugs [21, 22, 23]. However, these methods might not fully address individual pharmacodynamic modeling with limited clinical trial data. For example, Lu demonstrated that the LSTM model tends to persistently predict a drug reaction even after the patient stops taking drugs [24].

Here, we utilize Neural Ordinary Differential Equation (ODE) for irregularly sampled clinical trial data and directly predict multiple individual pharmacodynamics without necessitating a population-PD model [25, 26]. These pharmacodynamics could include attributes like tumor dynamics (estimation for tumor size) and progression-free survival (the percentage of patients in the group whose disease has not progressed) [27, 28, 29]. Specifically, the standard Neural ODE comprises three key components: the Encoder, the ODE solver, and the Decoder. To tackle the aforementioned challenges, the Neural ODE model is trained for all patients, enabling it to retain population-level features and preventing overfitting to individual patients. To predict individual pharmacodynamics, the Encoder encodes individual features as initial conditions and then feeds them to the ODE solver. Depending on individual features, the Neural ODE yields precise long-term pharmacodynamic estimations for diverse patients.

In a nutshell, we propose a novel Data-driven Neural Ordinary Differential Equation (DN-ODE) for multi-pharmacodynamic modeling, breast cancer tumor dynamics and progression-free survivals. In our approach, we use a shareable encoder for multi-task PD predictions. For each task, we only retrain the corresponding task ODE solver and decoder while leveraging the same shareable features. Our experiments show state-of-the-art performance on the clinical trial breast cancer dataset, outperforming alternative population-based PD methodologies.

2 Methodology

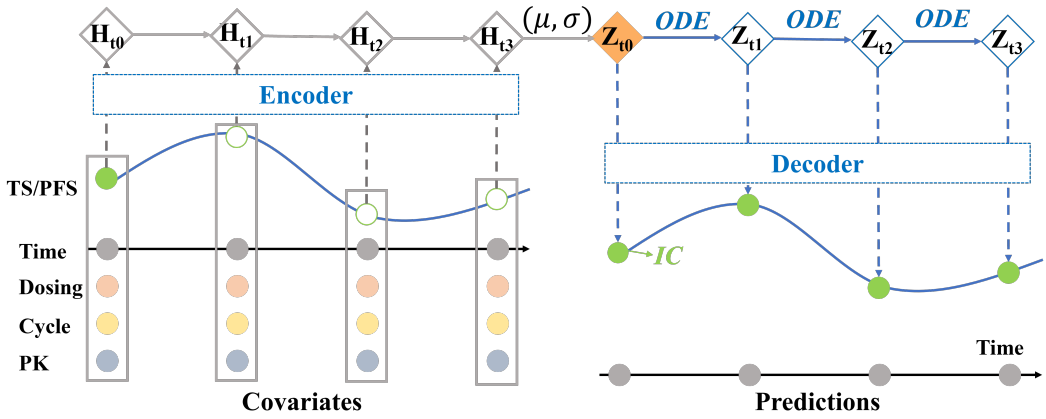

Figure 1: Computation pipeline of the DN-ODE model.

2.1 Data Preparation and Shared Encoder

The input covariates for the ODE model included time (gray dot), dosing (orange dot), cycle (yellow dot), PK estimations (light blue dot), and either a copied Tumor Size (TS) or Progression-Free Survival (PFS) label. To maintain consistency with previous work, we use the same dataset and obtain PK estimations from the previous PK model [28, 29], which consists of transit rate for absorption delay (Ktr), apparent volume in the central compartment (Vc/F), apparent volume in the peripheral compartment (Vp/F), apparent transit between the central and peripheral compartments (Q/F), and apparent clearance (CL/F). As for label inputs, we only retained the first-day values (represented by the Green dot in Figure 1) and set the values for the other days to 0 (represented by the Green circle in Figure 1). Time points without clinical records were filled with 0 or dropped [29].

Then, we feed the data into an encoder, which consists of two fully-connected layers and an RNN with 128 hidden units in each layer, following the neural-ODE official sample codes [25]. This encoder encoded the information into a 12-element array. Note that these numbers can be changed without affecting the results [24, 25].

Following the concept of an autoencoder in ODE works [25], the 12-element output (e.g., H_{t_3}) from

the Encoder defines the means and standard deviations of the latent state distributions. Specifically, the first 6 elements are employed to estimate the mean values (μ), and the last 6 elements are employed to estimate the standard deviation values (σ). Finally, Z_{t_0} (orange rhombus), which serves as the initial status for the ODE solver, is drawn from the Gaussian distribution derived from these mean and standard deviation estimates.

2.2 Ordinary Differential Equation (ODE) and Decoder

The tumor size or progression-free survivals are determined by the local initial state Z_{t_0} of each patient and a global set of latent dynamics shared across all patients, denoted by θ_f. For each patient, we feed the local initial state (Z_{t_0}) to the ODE solver (f) and produce Z ($Z = [Z_{t_1}, Z_{t_2}, Z_{t_3}, \ldots]$) for different times $t = [t_0, t_1, t_2, t_3, \ldots]$. We formally define this approach through a sampling procedure:

$$Z_{t_0} \sim p(Z_{t_0}), \tag{1}$$

$$Z_{t_1}, Z_{t_2}, Z_{t_3}, ... = \mathit{ODE}(Z_{t_0}, f, \theta_f, t_0, t_1, t_2, t_3, ...), \tag{2}$$

$$\mathit{each} \quad x_i \sim p(x|Z_{t_i}, \theta_x). \tag{3}$$

Specifically, the ODE function f is implemented as a four-layer fully connected network, with each layer having 16 hidden dimensions. At each time point $t_i, i \in [1, 2, 3, \ldots]$, we feed $Z_{t_(i-1)}$ and the corresponding time interval $t_i - t_{i-1}$ into the ODE function f. The output of f is Z_{t_i}, which is then circulated back into the next time point. Thus, we obtain a series of $Z = [Z_{t_1}, Z_{t_2}, Z_{t_3}, \ldots]$. Then, we feed Z into the decoder to produce a series of predictions for different times. When training the Progression model, we don't train the shared Encoder again, since features like PK have already been obtained.

3 Experiments and Results

3.1 Datasets and Metrics

In this experiment, we utilized clinical trial data on Tumor Size (TS) and Progression-Free Survival (PFS) [30, 31, 32]. Specifically, we used amcenestrant, an orally bioavailable selective Estrogen Receptor (ER) degrader developed for the treatment of ER+/HER2- advanced breast cancer. To train our model, we employed the phase 1/2, including AMEERA 1 (NCT03284957 [30]) and AMEERA 2 (NCT03816839 [31]) studies. These studies involved 75 patients who received amcenestrant monotherapy, with exposure to doses ranging from 20 to 600 mg daily. For inference, we utilized data from the AMEERA 3 clinical trial [32], which included 102 patients with measurable target lesions at baseline. We also present our dataset in a demographic summary table in Supplementary.

For evaluation, we employed both the RMSE loss and R^2 score as evaluation metrics. The RMSE loss provides an estimation of the difference between the predictions and ground truth, quantifying the accuracy of the model's predictions [33]. R^2 score goes a step further by considering whether the overall scale of the predictions aligns with the ground truth, providing insights into the goodness of fit of the model [34].

3.2 Experimental Results

Table 1: RMSE loss and R^2 on Train and Test Dataset

		Train		**Test**	
		Loss $\downarrow$	$R^2 \uparrow$	*Loss* $\downarrow$	$R^2 \uparrow$
Tumor Size	PK/PD [29]:	5.94	0.95	6.51 (fine-tuned)	0.93 (fine-tuned)
	DN-ODE (ours)	6.13	0.94	8.78	0.92
Progression-Free Survival	**DN-ODE** (ours)	0.20	0.96	0.21	0.95

Table 1 demonstrates the RMSE loss and R^2 scores corresponding to AMEERA 1-2 and AMEERA 3 clinical trial data. Although the previous PK/PD model yielded a more accurate prediction for tumor size, it necessitated additional fitting steps. Specifically, these steps encompassed recalculating the

distribution of population parameters as a prior distribution, followed by re-fitting it using individual data points. This procedure demanded supplemental information from the testing dataset. However, the Neural ODE approach only requires covariates for generating initial conditions, eliminating the need for additional fitting steps for the population-based model. Both tasks achieved R^2 scores surpassing 0.9, thereby highlighting the robust predictive capabilities of the Neural ODE approach. Across all training and testing scenarios, the ODE consistently outperforms the PK/PD methods in predicting progression-free survivals.

3.3 Visualizations

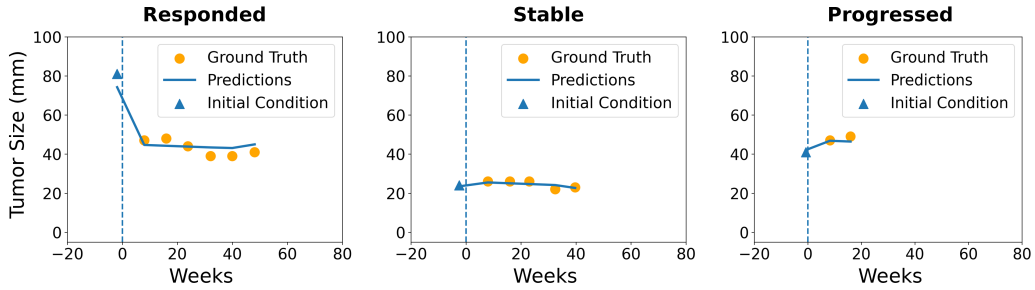

Figure 2: Tumor size predictions over time.

Figure 2 illustrates the consistent performance of the Neural ODE model for Tumor size predictions. The neural model distinguishes three distinct tumor growth patterns in both AMEERA 1-2 and AMEERA 3: responded, stable, and progressed. For each patient during two consecutive visits, if a reduction of over 30% in tumor size occurs, we classify it as a responder. Conversely, if an increase of more than 20% in tumor size occurs, we categorize it as progression. Cases that do not meet either of these criteria are designated as stable diseases. For both training and testing, we only provide the initial tumor size and corresponding date as the initial condition (navy blue triangle) to the ODE model, but ODE generates accurate predictions that closely align with the ground truth. We provide a case-by-case comparison of PK/PD and ODE modeling in Supplementary.

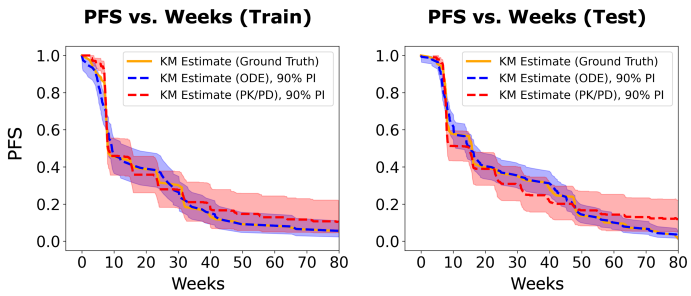

Figure 3: Progression-free survivals over time.

Figure 3 illustrates the population-level progression-free survival across weeks using the Kaplan-Meier (KM) estimation [35]. Progression-free survival is defined as the time from randomization to progression as per RECIST 1.1 criteria: 1) At least 20% and 5 mm from the lowest observed tumor size of target lesions; 2) Progression from non-target lesions or appearance of new lesions (event if death) [36]. The solid orange line represents the KM estimate derived from the original datasets, while the dashed blue line shows our predicted progression-free survival with a 90% prediction interval (blue shaded area) and the dashed red line shows PK/PD predicted progression-free survival with a 90% prediction interval (red).Since the PK/PD model requires an additional fitting step for the test dataset, we used the calibrated model [28, 29] for KM predictions on the test dataset. Remarkably, our predictions closely align with the ground truth, demonstrating a good fit for the observed data.

4 Conclusion

We proposed an innovative approach, the Data-driven Neural Ordinary Differential Equation (DN-ODE) model, to capture the dynamics of breast cancer tumor development and predict progression-free survivals. Our experiments demonstrated DN-ODE's ability to identify various tumor growth patterns and potential responders, offering valuable insights to doctors and informing the design of future clinical trials.

References

[1] Lewis B Sheiner, Stuart Beal, Barr Rosenberg, and Vinay V Marathe. Forecasting individual pharmacokinetics. *Clinical Pharmacology & Therapeutics*, 26(3):294–305, 1979.

[2] Hartmut Derendorf and Bernd Meibohm. Modeling of pharmacokinetic/pharmacodynamic (pk/pd) relationships: concepts and perspectives. *Pharmaceutical research*, 16:176–185, 1999.

[3] Ene I Ette and Paul J Williams. Population pharmacokinetics i: background, concepts, and models. *Annals of Pharmacotherapy*, 38(10):1702–1706, 2004.

[4] Otto Metzger, Christina Herold, Coralie Poncet, Heidi De Swert, Jose Casas-Martin, Ann Partridge, Samia Guita, Lisa Carey, Eva Schumacher, Theodora Goulioti, et al. Abstract ot1-04-01: Ameera-6: Phase 3 study of adjuvant amcenestrant versus tamoxifen for patients with hormone receptor-positive early breast cancer, who have discontinued adjuvant aromatase inhibitor therapy due to treatment-related toxicity. *Cancer Research*, 83(5_Supplement):OT1–04, 2023.

[5] Yuanzhou Wei, Dan Zhang, Meiyan Gao, Yuanhao Tian, Ya He, Bolin Huang, and Changyang Zheng. Breast cancer prediction based on machine learning. *Journal of Software Engineering and Applications*, 16(8):348–360, 2023.

[6] Jogarao VS Gobburu and Patrick J Marroum. Utilisation of pharmacokinetic-pharmacodynamic modelling and simulation in regulatory decision-making. *Clinical pharmacokinetics*, 40:883–892, 2001.

[7] Diane R Mould and Richard Neil Upton. Basic concepts in population modeling, simulation, and model-based drug development—part 2: introduction to pharmacokinetic modeling methods. *CPT: pharmacometrics & systems pharmacology*, 2(4):1–14, 2013.

[8] PH Van Der Graaf. Introduction to population pharmacokinetic/pharmacodynamic analysis with nonlinear mixed effects models. *CPT: pharmacometrics & systems pharmacology*, 3(12):e153, 2014.

[9] Danton S Char, Nigam H Shah, and David Magnus. Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, 378(11):981, 2018.

[10] Jinlin Xiang, Shane Colburn, Arka Majumdar, and Eli Shlizerman. Knowledge distillation circumvents nonlinearity for optical convolutional neural networks. *Applied Optics*, 61(9):2173–2183, 2022.

[11] Benjamin Ribba, Dominic Stefan Bräm, Paul Gabriel Baverel, and Richard Wilson Peck. Model enhanced reinforcement learning to enable precision dosing: A theoretical case study with dosing of propofol. *CPT: Pharmacometrics & Systems Pharmacology*, 11(11):1497–1510, 2022.

[12] Yifan Yang, Chang Liu, and Zheng Zhang. Particle-based online bayesian sampling. *arXiv preprint arXiv:2302.14796*, 2023.

[13] Chenyu You, Jinlin Xiang, Kun Su, Xiaoran Zhang, Siyuan Dong, John Onofrey, Lawrence Staib, and James S Duncan. Incremental learning meets transfer learning: Application to multi-site prostate mri segmentation. In *International Workshop on Distributed, Collaborative, and Federated Learning*, pages 3–16. Springer, 2022.

[14] Jie Tang, Rong Liu, Yue-Li Zhang, Mou-Ze Liu, Yong-Fang Hu, Ming-Jie Shao, Li-Jun Zhu, Hua-Wen Xin, Gui-Wen Feng, Wen-Jun Shang, et al. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Scientific reports*, 7(1):42192, 2017.

[15] Thorsten Joachims. Making large-scale svm learning practical. Technical report, Technical report, 1998.

[16] Yang Zheng, Jinlin Xiang, Kun Su, and Eli Shlizerman. Bi-maml: Balanced incremental approach for meta learning. *arXiv preprint arXiv:2006.07412*, 2020.

[17] Yifan Yang, Alec Koppel, and Zheng Zhang. A gradient-based approach for online robust deep neural network training with noisy labels. *arXiv preprint arXiv:2306.05046*, 2023.

[18] Jody D Ciolino, Renée H Martin, Wenle Zhao, Edward C Jauch, Michael D Hill, and Yuko Y Palesch. Covariate imbalance and adjustment for logistic regression analysis of clinical trial data. *Journal of biopharmaceutical statistics*, 23(6):1383–1402, 2013.

[19] Jinlin Xiang and Eli Shlizerman. Tkil: tangent kernel approach for class balanced incremental learning. *arXiv preprint arXiv:2206.08492*, 2022.

[20] Jinlin Xiang and Eli Shlizerman. Tkil: Tangent kernel optimization for class balanced incremental learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3529–3539, 2023.

[21] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.

[22] Pingping Song, Yuhan Dong, and Kai Zhang. Gcd-pkaug: A gradient consistency discriminator-based augmentation method for pharmacokinetics time courses. In *International Conference on Neural Information Processing*, pages 3–14. Springer, 2022.

[23] Xiangyu Liu, Chao Liu, Ruihao Huang, Hao Zhu, Qi Liu, Sunanda Mitra, and Yaning Wang. Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *International journal of clinical pharmacology and therapeutics*, 59(2):138, 2021.

[24] James Lu, Kaiwen Deng, Xinyuan Zhang, Gengbo Liu, and Yuanfang Guan. Neural-ode for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *Iscience*, 24(7), 2021.

[25] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[26] Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam M Oberman. How to train your neural ode. *arXiv preprint arXiv:2002.02798*, 2, 2020.

[27] Kristine R Broglio and Donald A Berry. Detecting an overall survival benefit that is derived from progression-free survival. *JNCI: Journal of the National Cancer Institute*, 101(23):1642–1649, 2009.

[28] Marc Cerou, Hoai-Thu Thai, Laure Deyme, Sophie Fliscounakis-Huynh, Sylvaine Cartot-Cotton, and Christine Veyrat-Follet. Semi-mechanistic pharmacokinetic-pharmacodynamic modeling of tumor size dynamics in advanced breast cancer patients treated with single-agent amcenestrant. *PAGE*, 2022.

[29] Marc Cerou, Hoai-Thu Thai, Laure Deyme, Sylvaine Cartot-Cotton, and Christine Veyrat-Follet. Semi mechanistic joint modeling of tumor dynamics and pfs in advanced breast cancer: leveraging data from early amcenestrant phase i-ii trials. *PAGE*, 2023.

[30] Aditya Bardia, Sarat Chandarlapaty, Hannah M Linden, Gary A Ulaner, Alice Gosselin, Sylvaine Cartot-Cotton, Patrick Cohen, Séverine Doroumian, Gautier Paux, Marina Celanovic, et al. Ameera-1 phase 1/2 study of amcenestrant, sar439859, in postmenopausal women with er-positive/her2-negative advanced breast cancer. *Nature Communications*, 13(1):4116, 2022.

[31] Haruru Kotani, Kenji Tamura, Toru Mukohara, Kan Yonemori, Yumiko Kawabata, Xavier Nicolas, Tomoyuki Tanaka, and Hiroji Iwata. O5-2 ameera-2: Phase 1 study of oral serd amcenestrant (sar439859) in japanese women with er+/her2-advanced breast cancer. *Annals of Oncology*, 33:S470, 2022.

[32] SM Tolaney, A Chan, K Petrakova, S Delaloge, M Campone, H Iwata, P Peddi, PA Kaufman, E De Kermadec, Q Liu, et al. 212mo ameera-3, a phase ii study of amcenestrant (amc) versus endocrine treatment of physician's choice (tpc) in patients (pts) with endocrine-resistant er+/her2- advanced breast cancer (abc). *Annals of Oncology*, 33:S634–S635, 2022.

[33] Cort J Willmott. On the validation of models. *Physical geography*, 2(2):184–194, 1981.

[34] Alessandro Di Bucchianico. Coefficient of determination (r 2). *Encyclopedia of statistics in quality and reliability*, 2008.

[35] Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore. Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, 1(4):274, 2010.

[36] Colleen M Costelloe, Hubert H Chuang, John E Madewell, and Naoto T Ueno. Cancer response criteria and bone metastases: Recist 1.1, mda and percist. *Journal of Cancer*, 1:80, 2010.

Supplementary Material

Table 2: Demographic summary of the data.

		Training dataset			Testing dataset
		AM-1 Part A, $N = 21$	**AM-1 Part B**, $N = 46$	**AM-2**, $N = 8$	**AM-3**, $N = 102$
Ages (years)	Mean (Min∼Max)	59 (40∼86)	64 (37∼88)	66 (48∼76)	58 (29∼84)
Weight (kg)	Mean (Min∼Max)	71 (49∼105)	73 (41∼116)	52 (43∼74)	72 (41∼109)
Height (cm)	Mean (Min∼Max)	163 (153∼176)	162 (145∼176)	156 (147∼167)	161 (148∼175)
BMI	Mean (Min∼Max)	27 (16∼38)	28 (17∼41)	21 (19.5∼26.8)	27.5 (16.8∼43.6)

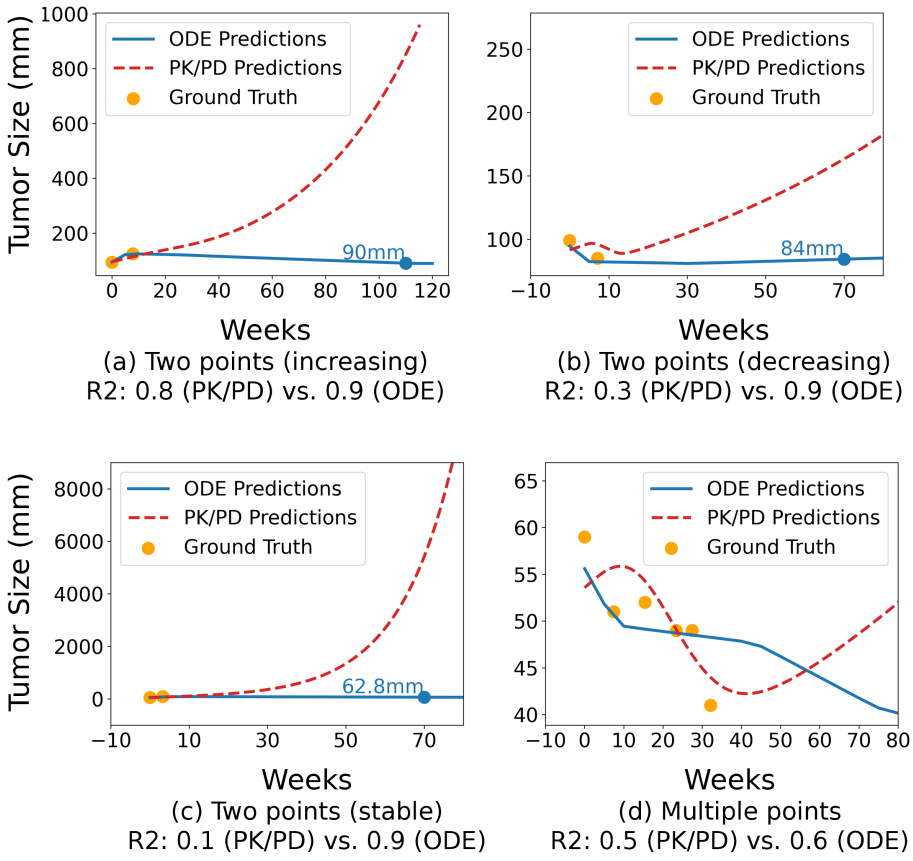

Figure 4: Ablation Study: PK/PD modeling vs. DN-ODE modeling.

DN-ODE VS. Traditional PK/PD modeling

To comprehensively compare the DN-ODE with traditional PK/PD methods, we conducted an exhaustive analysis encompassing different scenarios, as shown in Figure 4. These cases can be broadly classified into two main scenarios: 1) involving two-point ground truth (a-c) and 2) involving multiple-point ground truth (d). We emerge irregular data points (only two points) due to patients' early dropout. Within these cases, we can further distinguish between three distinct situations: increasing tumor size (tumor growth), decreasing tumor size (tumor regression), and maintaining stable tumor size (minor changes).

Implementation Details

The Neural ODE structure consists of three components: the encoder, the Ordinary Differential Equation (ODE), and the decoder. The encoder includes two Fully-Connected layers and a Recurrent Neural Network (RNN). We feed in time, Initial Condition (IC), dosing information, cycle, and estimated PK (Given by [29]). To follow previous works [23, 24], we set the hidden layer size to 128 and the output length to 12, using the ReLU activation function. The ODE includes four fully connected layers, with an input length of 6 (sampled from the 12 available). Each hidden layer has 128 hidden units, and the last fully connected layer produces a 6-length vector. Then, the decoder concatenates this 6-length vector with time and IC values on the first day, resulting in a 16-length vector. The decoder includes two fully connected layers with an input length of 16, 32 hidden units, and a single output as the prediction.

During training, we selected a batch size of 1 to accommodate the different length of the clinical trial data (e.g., different patients has different numbers of CT records). To approximate a batch size of $n = 4$, we employed loss backpropagation for every 4 samples. The hyperparameters included an initial learning rate of 0.001 with the Adam optimizer, along with L2 regularization set to 0.1 and an RMSE loss function. For weight initialization, we utilized a normal distribution $N(0, 0.001)$, and the bias was set to 0. The training process ran for 80 epochs, with the learning rate decreasing by a factor of 10 every 20 epochs. Additionally, for the ODE part, we set the ODE solver tolerance to 0.0001.

Interpretative Visualizations for DN-ODE

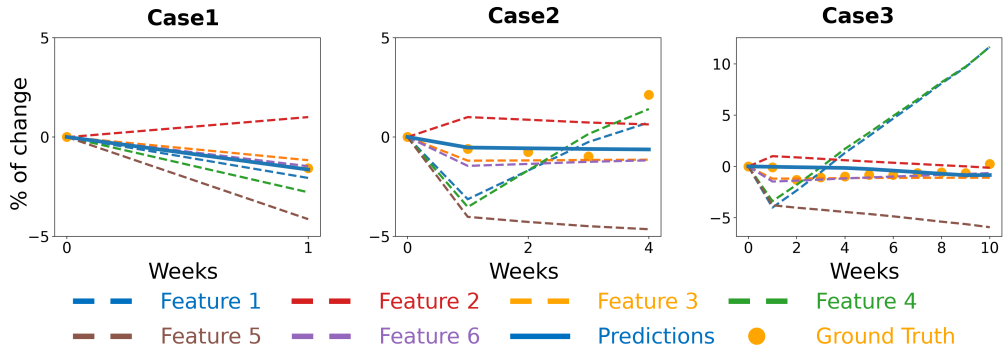

Figure 5: Interpretative Visualizations for DN-ODE: features (all dashed lines) vs. predictions (solid blue line), ground truth (orange dots)

To investigate why the ODE yields robust predictions for breast tumor dynamics, we provide visualizations in Figure 5. We selected three individual patient cases and plotted the percentage change of features versus time, the percentage change of predictions versus time, and the percentage change of ground truth versus time. To clearly show the linear relationships between features and predictions, we magnified the prediction and ground truth values by a factor of 5.

These visualizations reveal that certain features exhibit strong linear correlations with our final predictions. For instance, in Case 1, both the blue and green features, as well as the brown features, trend downwards when the tumor size decreases. Similarly, in Case 2, the blue and green features mirror the dynamics of the ground truth—decreasing as the tumor size decreases and increasing as it grows. By relying on such features, our ODE solver and decoder are able to generate accurate predictions.