

ORIGINAL RESEARCH

Open access

# Neural ODE-Based Tumor Dynamics Modeling for Pathologic Complete Response Prediction in Breast Cancer from Serial MRI

Alejandro Torres<sup>1\*</sup>, Miguel Fernandez<sup>1</sup>

## Abstract

Neoadjuvant chemotherapy (NAC) is standard for locally advanced breast cancer, with pathologic complete response (pCR) strongly predicting improved survival. However, only 30–40% of patients achieve pCR, while the rest undergo toxicity and delayed surgery without benefit. Current prediction methods rely on tumor volume at isolated time points or simple pre- and post-treatment comparisons, ignoring continuous tumor dynamics during therapy. Sparse and irregular MRI sampling further limits accurate modeling. We introduce a Neural Ordinary Differential Equation (Neural ODE) framework to model continuous tumor growth from sparse serial MRI during NAC. The model learns a time-continuous function describing tumor evolution and predicts individual response trajectories and final pCR status. The framework includes (1) MRI-based tumor segmentation, (2) construction of sparse longitudinal tumor volume series, (3) Neural ODE modeling of continuous dynamics via a neural network–parameterized derivative function, and (4) classification of the final latent state for pCR prediction. An optional module enables trajectory visualization and interpretability. This approach captures hidden continuous tumor behavior between scans, handles irregular sampling without imputation, and enables earlier response prediction. It is also computationally efficient using adjoint-based training and may reveal distinct growth patterns between responders and non-responders. Neural ODE-based modeling offers a more informative framework for predicting NAC response by capturing continuous tumor dynamics, with potential to improve pCR prediction over conventional volume-based methods.

**Keywords** Neural ordinary differential equations, Breast cancer, Neoadjuvant chemotherapy, Pathologic complete response, Serial MRI, Tumor growth modeling

\*Correspondence:

Alejandro Torres  
alejandro.torres@gmail.com

<sup>1</sup> Department of AI Healthcare Analytics, University of Chile, Santiago, Chile

## Introduction

Breast cancer remains the most commonly diagnosed malignancy among women worldwide, with neoadjuvant chemotherapy administered prior to surgical resection representing a standard treatment approach for locally advanced and high-risk early-stage disease [1, 2]. The primary goal of NAC is to downstage tumors, increase rates of breast-conserving surgery, and eradicate micrometastatic disease, with pathologic complete response serving as a powerful surrogate marker for

favorable long-term outcomes [3, 4]. Patients achieving pCR demonstrate significantly improved disease-free survival and overall survival compared to those with residual disease, making pCR a critical endpoint in both clinical practice and therapeutic trials [5, 6].

Despite the prognostic importance of pCR, only 30–40% of patients receiving standard NAC regimens achieve this favorable outcome, meaning that the majority undergo ineffective treatment characterized by toxicity, delayed surgery, and disease progression [7, 8]. The inability to

predict which patients will respond to NAC before treatment completion represents a significant clinical gap, as early identification of non-responders could enable treatment modification, switching to alternative regimens, or earlier surgical intervention [9, 10]. Current clinical practice relies on tumor volume changes measured from serial imaging—typically MRI at baseline, mid-treatment, and post-treatment—using criteria such as RECIST, but these approaches provide only coarse, discrete assessments that fail to capture the continuous dynamics of tumor response [11, 12].

The fundamental limitation of existing methods lies in their treatment of tumor evolution as a discrete process measured at isolated time points rather than a continuous dynamical system governed by underlying biological mechanisms [1, 13]. Serial MRI during NAC typically acquires images at 2-3 time points over 6-8 weeks, creating sparse and irregularly spaced data that conventional machine learning models struggle to handle without arbitrary interpolation or imputation [14, 15]. We propose a conceptual framework that addresses these limitations by employing Neural Ordinary Differential Equations to model continuous tumor growth dynamics from sparse serial MRI measurements, enabling more accurate prediction of pathologic complete response and potentially enabling earlier treatment guidance [1, 16].

## Background

### Neoadjuvant chemotherapy in breast cancer

Standard NAC regimens for breast cancer typically combine anthracycline-based and taxane-based agents administered over 4-6 cycles spanning approximately 12-18 weeks, with the specific regimen chosen based on tumor subtype, stage, and biomarker profile [2, 4]. Pathologic complete response is rigorously defined as ypT0/is ypN0—no residual invasive carcinoma in the breast and no tumor involvement in axillary lymph nodes—though some definitions permit isolated tumor cells, and pCR rates vary substantially by subtype from approximately 20-30% in hormone receptor-positive/HER2-negative tumors to 50-60% in triple-negative and HER2-positive disease [5, 6]. Clinical factors associated with pCR include younger age, higher tumor grade, elevated Ki-67 proliferation index, and specific genomic signatures, but these factors lack

sufficient predictive accuracy to guide individual treatment decisions [7, 9].

The clinical significance of pCR extends beyond immediate treatment outcomes, as patients achieving pCR following NAC have consistently demonstrated 5-year disease-free survival rates exceeding 85% compared to approximately 60% for those with residual disease, making pCR an accepted surrogate endpoint for regulatory approval of new neoadjuvant regimens [10, 11]. Conversely, patients with residual disease after NAC face elevated risks of recurrence and mortality, driving intense interest in developing methods to identify non-responders early enough to modify treatment during the therapeutic window [12, 13]. The inability to predict pCR before treatment completion represents a major barrier to personalized neoadjuvant therapy, as current decision-making relies on population-level averages rather than individual tumor biology [14, 15].

### Serial MRI for treatment monitoring

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is the preferred imaging modality for monitoring breast tumor response during NAC due to its superior soft tissue contrast, high spatial resolution, and ability to assess tumor vascularity through contrast uptake kinetics [5, 17]. Standard clinical protocols acquire MRI scans at three time points: baseline prior to treatment initiation (week 0), mid-treatment after 2-3 cycles (approximately weeks 3-6), and post-treatment following NAC completion (week 12-18), though the exact timing varies across institutions and clinical trials [18, 19]. Tumor volume is typically measured by manual or semi-automated segmentation of enhancing lesions on post-contrast sequences, with volumes calculated by summing voxel counts multiplied by voxel dimensions, achieving inter-reader reliability coefficients of 0.85-0.95 in experienced centers [20, 21].

The Response Evaluation Criteria in Solid Tumors (RECIST) guidelines define response categories based on percentage change in tumor diameter or volume, with partial response requiring at least 30% decrease in sum of diameters and progressive disease requiring at least 20% increase [11, 22]. However, RECIST and similar volume-based criteria were designed for assessing response at a single endpoint rather than modeling the continuous trajectory of tumor evolution during treatment, and they treat all volume changes as equally informative regardless of timing or growth dynamics [12, 23]. The sparse temporal

resolution of serial MRI—typically only 2-3 measurements over months of treatment—means that conventional methods cannot distinguish between different dynamic patterns that might reach the same final volume but reflect distinct biological responses, such as rapid initial shrinkage versus slow gradual decline [14, 24].

## Tumor growth models

Mathematical models of tumor growth have been developed for decades, with the exponential model representing the simplest formulation where tumor volume increases at a rate proportional to current volume, yielding the equation  $\frac{dV}{dt} = rV$  where  $r$  is the growth rate constant [1, 25]. While mathematically convenient, exponential growth is biologically unrealistic for solid tumors because it assumes no constraints on tumor expansion, ignoring limitations imposed by nutrient availability, vascular supply, and physical space [13, 26]. The logistic growth model addresses this limitation by introducing a carrying capacity  $K$  representing maximum tumor volume, with the governing equation  $\frac{dV}{dt} = rV(1 - \frac{V}{K})$ , producing sigmoidal growth that slows as volume approaches the carrying capacity [27, 28].

The Gompertz model provides an alternative formulation where the growth rate decays exponentially over time,  $\frac{dV}{dt} = rV e^{-at}$  which has been shown to better fit experimental tumor growth data than logistic models for many cancer types, particularly during early growth phases [15, 29]. Each of these classical models has identifiable parameters—growth rate, carrying capacity, and decay constants—that can be estimated from serial volume measurements, but the sparse data typical of clinical MRI (2-3 time points) makes reliable parameter estimation challenging or impossible [1, 16]. Neural ODEs offer a flexible alternative that learns the growth dynamics directly from data without assuming a specific parametric form, using a neural network to represent the derivative  $\frac{dV}{dt}$  as a function of volume and time, potentially discovering novel growth patterns that distinguish responding from non-responding tumors [1, 24].

**Table 1** clarifies why the manuscript’s central contribution is not merely a new classifier but a shift from discrete radiologic response assessment to continuous-time dynamical inference of treatment response.

**Table 1.** Analytical Comparison of Discrete Response Assessment versus Continuous-Time Neural ODE Tumor Dynamics Modeling

Analytical dimension	Conventional volume-change / RECIST-style approaches	Classical parametric growth models	Pro Neu fra
Representation of tumor evolution	Treats response as discrete change between isolated scans	Assumes tumor follows a pre-specified mathematical growth law	Treats as a contin dy pr
Temporal information use	Primarily baseline-to-mid or baseline-to-final difference	Uses serial timing but under rigid functional constraints	Uses a times c int between obs
Capacity to model heterogeneous response shapes	Limited; different trajectories may map to identical net volume change	Moderate, but constrained to exponential, logistic, or Gompertz-like behavior	High; r is lea data r pre-
Handling of irregular scan timing	Weak; often depends on fixed comparison windows or simplification	Possible in principle, but unstable with sparse clinical measurements	N contin mode patie obs t
Dependence on interpolation or imputation	Often implicit when comparing heterogeneous schedules	May require strong assumptions for sparse fitting	Does inte before
Biological flexibility	Low; response treated as	Moderate; embeds	High dya

	geometric size change only	simplified biological constraints	uncertainty, sensitivity, response, prediction
Individualization through covariates	Usually added only as parallel predictors	Limited unless explicitly embedded in parametric structure	Covariate initialization and/or feature
Usefulness for early prediction	Limited because reliable classification often requires later endpoint measurements	Limited by poor identifiability from few points	Stochastic prediction because trajectory can be expected to predict correlation
Interpretability of response pattern	Simple but shallow; percentage reduction is easy to read	Moderate through named parameters like growth rate or carrying capacity	Translating time growth
Main failure mode	Oversimplifies tumor biology and timing	Model misspecification when true dynamics depart from assumed form	Data late operation dependence for training

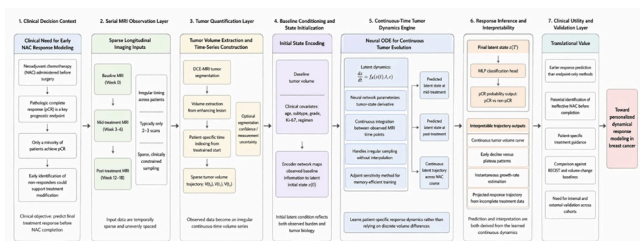
## Framework Overview

### High-level architecture

The proposed framework takes as input serial MRI scans acquired at baseline (week 0), mid-treatment (week 3-6), and post-treatment (week 12-18), processes these images through tumor segmentation to extract volume measurements at each time point, and then uses a Neural ODE to model the continuous trajectory of tumor volume from the sparse observations [1, 10]. The Neural ODE

component learns a parameterized function  $f_\theta$  that predicts the derivative of tumor volume with respect to time, enabling integration from any starting time to any target time to generate a complete growth curve that passes through or near the observed volume points [1, 13]. From this learned continuous trajectory, the framework extracts the final latent state at treatment completion and passes it through a multilayer perceptron classification head to output a probability of achieving pathologic complete response [5, 11].

Figure 1 shows the proposed framework, which conceptualizes pCR prediction as a continuous-time tumor dynamics problem in which sparse serial MRI measurements are transformed into a Neural ODE-derived latent trajectory that supports both response classification and trajectory-level interpretation.



**Figure 1.** Conceptual Architecture of Neural ODE-Based Continuous Tumor Dynamics Modeling for Pathologic Complete Response Prediction from Serial Breast MRI

The framework additionally incorporates clinical and pathological covariates—including patient age, tumor subtype (hormone receptor status, HER2 status), grade, and Ki-67 proliferation index—as initial condition information to modulate the learned dynamics for individual patients [7, 17]. By conditioning the Neural ODE on these covariates, the model can learn distinct growth dynamics for different tumor subtypes, potentially capturing known biological differences in chemosensitivity between triple-negative, HER2-positive, and hormone receptor-positive disease [18, 19]. The entire framework is trained end-to-end to simultaneously optimize both the accuracy of the reconstructed tumor volume trajectory and the classification of final response status [20, 21].

### Core assumptions

The framework rests on several core assumptions that must hold for valid application to clinical data, beginning with the assumption that tumor volume can be reliably and reproducibly measured from DCE-MRI using either manual

segmentation by expert radiologists or automated deep learning segmentation methods [22, 23]. A second key assumption is that at least two MRI time points are available per patient (baseline and either mid-treatment or post-treatment), as the Neural ODE requires at least two observations to constrain the learned dynamics, though three or more time points provide substantially better parameter identification [1, 24]. The framework also assumes that the underlying tumor growth dynamics during NAC can be approximated by a continuous ordinary differential equation, meaning that volume changes smoothly between observation times without abrupt jumps or discontinuities not captured by the measurement schedule [25, 26].

Additional assumptions concern the identifiability of model parameters from available data, including that the number of patients in the training cohort (typically  $N > 100-200$ ) is sufficient to learn the neural network parameters of  $f_{\theta}$  without overfitting, and that the time points across patients are sufficiently aligned to enable batch training while still accommodating individual variability in scan timing [27, 28]. The framework assumes that measurement errors in tumor volume extraction are independent and approximately normally distributed, which may not hold for very small tumors or those with irregular morphology where segmentation is challenging [5, 29]. Finally, the framework assumes that the relationship between tumor volume dynamics and pCR is stable across the clinical settings and patient populations to which the model is applied, requiring external validation before clinical deployment [15, 16].

## Tumor Volume Extraction

### Segmentation from MRI

The first component of the framework extracts tumor volume measurements from serial DCE-MRI scans, requiring accurate segmentation of the enhancing tumor region from surrounding breast tissue, chest wall, and blood vessels [5, 17]. Manual segmentation by expert radiologists remains the reference standard, with typical protocols involving slice-by-slice contouring of tumor boundaries on post-contrast subtraction images, but this approach is time-consuming (requiring 15-30 minutes per scan) and subject to inter-observer variability [18, 20]. Automated deep learning segmentation methods, particularly U-Net and its variants including nnU-Net and attention U-Net, have demonstrated performance

approaching or matching human experts for breast tumor segmentation on DCE-MRI, with Dice similarity coefficients of 0.80-0.90 reported in validation studies [22, 23].

Quality control procedures are essential regardless of segmentation method, including visual inspection of segmentations to identify obvious errors such as inclusion of pectoral muscle or exclusion of tumor spiculations, and calculation of quality metrics like slice-wise volume consistency [24, 25]. For automated methods, cases with low segmentation confidence or unusual tumor morphology may require manual review and correction, particularly for non-mass enhancing lesions or tumors with ill-defined borders [26, 27]. The extracted volume measurements are recorded in cubic centimeters (cc) or milliliters, with typical breast tumor volumes at baseline ranging from 1-100 cc depending on stage and detection method [5, 28].

### Time series construction

Following segmentation, volume measurements are assembled into a time series for each patient, with time indexed from the baseline scan at week 0 and subsequent scans at their actual acquisition dates measured in days or weeks from baseline [5, 17]. The framework accommodates variable timing across patients—for example, mid-treatment scans obtained at week 3 for some patients and week 5 for others—by treating time as a continuous variable rather than requiring fixed intervals [1, 13]. Missing time points are handled naturally by the Neural ODE formulation, which does not require imputation or interpolation prior to modeling, though the presence of at least two time points per patient is necessary for parameter identification [1, 24].

The constructed time series includes not only volume measurements but also the timing of each scan relative to treatment initiation and the specific NAC regimen administered, as different chemotherapy combinations may induce different growth dynamics [14, 26]. For patients with more than three scans (e.g., additional early response assessment at week 2), all available time points are included to provide more constraints on the learned trajectory [1, 27]. The framework optionally incorporates uncertainty estimates for each volume measurement based on segmentation confidence or inter-rater variability, allowing the model to downweight noisy observations during training [28, 29].

# Neural ODE for Tumor Dynamics

## Neural ODE formulation

The core of the framework is a Neural Ordinary Differential Equation that models the continuous evolution of tumor state over time, formulated as  $\frac{dz}{dt} = f_{\theta}(z(t), t, c)$ , where  $z(t)$  is the latent state vector at time  $t$ ,  $f_{\theta}$  is a neural network with parameters  $\theta$ , and  $c$  represents conditioning clinical covariates including tumor subtype and patient age [1, 2]. Unlike classical parametric growth models (exponential, logistic, Gompertz) that impose a fixed functional form, the neural network learns the growth dynamics directly from training data, potentially discovering patterns that differentiate responders from non-responders [1, 13]. The latent state  $z(t)$  may be multidimensional, encoding not only tumor volume but also unobserved biological variables such as proliferation rate, drug sensitivity, or immune infiltration that influence treatment response [8, 14].

The Neural ODE integrates from an initial time  $t_0$  to any target time  $t_1$  using a numerical ODE solver, typically a Runge-Kutta method or adaptive solver such as Dormand-

$$z(t_1) = z(t_0) + \int_{t_0}^{t_1} f_{\theta}(z(t), t, c) dt$$

Prince [1, 15]. The framework uses the adjoint sensitivity method for backpropagation, which computes gradients with respect to  $\theta$  by solving a second augmented ODE backward in time, achieving constant memory cost independent of the number of solver steps [1, 16]. This memory efficiency is critical for clinical applications where long treatment durations (12-18 weeks) may require many integration steps, and for large training cohorts of hundreds or thousands of patients [5, 17].

## Initial condition encoding

The initial latent state  $z(0)$  is encoded from the baseline MRI tumor volume and available clinical covariates using a small encoder neural network that maps the observed volume at week 0 ( $V_0$ ) and covariate vector  $c$  to a latent representation [1, 18]. This encoding allows the framework to account for the fact that different tumors with identical baseline volumes may have different growth potentials and chemosensitivity based on their underlying biology, reflected in their initial latent state [5, 19]. The encoder can optionally incorporate additional baseline features such as tumor morphology descriptors (spiculation, margin

characteristics) or radiomic texture features extracted from the baseline MRI [20, 21].

For patients with only baseline and post-treatment scans (no mid-treatment), the initial state is still encoded from baseline data, and the Neural ODE integrates forward to the post-treatment time point, with the reconstruction loss comparing predicted and observed final volume [1, 22]. The framework can also be extended to bidirectional dynamics by integrating backward from a later time point to earlier times, though this requires careful handling of temporal causality for prediction tasks [1, 23].

## Trajectory prediction

Given the initial latent state  $z(0)$  and the learned dynamics  $f_{\theta}$ , the ODE solver generates a continuous trajectory  $z(t)$  for  $t \in [0, T]$  where  $T$  is the post-treatment time point, from which tumor volume  $v(t)$  is extracted as a linear projection of the latent state or as a component of  $z(t)$  [1, 24]. The predicted volume at each observed time point (mid-treatment  $t_{mid}$ , post-treatment  $t_{post}$ ) is compared to the extracted MRI volumes  $V_{mid}$  and  $V_{post}$  to compute a reconstruction loss, encouraging the learned dynamics to match the observed data [1, 25]. The continuous nature of the trajectory means the framework can predict volumes at any arbitrary time, including unobserved time points such as week 2 or week 10, enabling dense monitoring of tumor evolution from sparse measurements [5, 26].

## Response Prediction

### pCR classification

The final latent state at treatment completion  $z(T)$  serves as the input to a classification head that predicts the probability of pathologic complete response, implemented as a multilayer perceptron with one or two hidden layers followed by a sigmoid activation function producing an output in the [0,1] range [1, 5]. This classification head is trained jointly with the Neural ODE dynamics, allowing gradients from the pCR prediction task to influence the learned representation of tumor growth dynamics [6, 7]. By integrating the prediction task into the same optimization objective as trajectory reconstruction, the framework learns dynamics that are not only faithful to observed volumes but also discriminative of eventual response status [10, 11].

The binary classification output corresponds to the two clinically relevant categories: pCR (ypT0/is ypN0) versus

non-pCR (any residual invasive disease), though the framework could be extended to ordinal or multi-class outcomes such as residual cancer burden categories [3, 4]. For patients with incomplete treatment courses where post-treatment MRI is unavailable, the framework can predict pCR probability from mid-treatment data alone by integrating the Neural ODE only to the available time point and applying a modified classification head trained for early prediction [12, 13]. The classifier produces both a binary prediction (pCR vs non-pCR) and a confidence score, the latter being valuable for clinical decision-making where uncertain cases may warrant additional imaging or biopsy [14, 15].

## Early response prediction

A key clinical advantage of the framework is its ability to predict final pCR status from mid-treatment data alone, enabling potential treatment modification before NAC completion [16, 17]. Using only baseline and mid-treatment MRI scans (typically weeks 0 and 3-6), the Neural ODE integrates forward to the projected post-treatment time point T (e.g., week 18) based on the learned dynamics from the training cohort, producing an extrapolated trajectory and a predicted final latent state  $z(T_{\text{projected}})$  [1, 18]. The accuracy of this early prediction depends on how well the learned dynamics generalize across patients and whether mid-treatment volume changes reliably signal eventual response, a relationship that the framework learns directly from training data [19, 20].

Early prediction could support several clinical interventions: switching non-responders to alternative chemotherapy regimens, adding targeted therapies, or proceeding directly to surgery without completing ineffective cycles [9, 10]. The framework can also be applied iteratively as additional time points become available, updating predictions as more data accrues and potentially increasing confidence before committing to treatment changes [21, 22]. Simulation studies using retrospective data would be needed to establish the optimal timing for early prediction and the confidence thresholds that justify different clinical actions [5, 23].

## Interpretability

### Learned growth dynamics

The Neural ODE framework offers inherent interpretability through visualization of learned tumor growth trajectories,

enabling clinicians to examine how predicted volume evolves over time for individual patients and compare patterns between responders and non-responders [1, 24]. By extracting the predicted continuous volume curve  $v(t)$  for each patient, the framework reveals temporal features such as the rate of initial decline, presence of plateau phases, or late regrowth that may distinguish response phenotypes not apparent from discrete volume measurements [5, 25]. Responders typically demonstrate rapid early volume reduction within the first 2-4 weeks followed by continued decline, while non-responders may show slow reduction, stable disease, or early progression, though the specific patterns are learned from data rather than prescribed [14, 26].

Beyond trajectory visualization, the learned neural network  $f_{\theta}$  can be analyzed to extract biologically meaningful parameters such as the instantaneous growth rate at any time point, calculated as  $(1/v) * dv/dt$  for volume-proportional dynamics [1, 27]. Comparing growth rate trajectories between response groups may reveal critical windows where divergence occurs, potentially identifying the optimal timing for response assessment [15, 28]. The framework can also generate counterfactual predictions—for example, estimating what a patient's tumor volume would have been without treatment by integrating the dynamics learned from an untreated control cohort, providing a personalized estimate of treatment effect [5, 29].

## Training Considerations

### Loss function

The total loss function combines three components: trajectory reconstruction loss, classification loss, and optional regularization terms, with hyperparameters  $\lambda_{recon}$ ,  $\lambda_{class}$ , and  $\lambda_{reg}$  controlling their relative contributions [1, 5]. The reconstruction loss is typically mean squared error between predicted and observed tumor volumes at each available time point, averaged across all patients and time points: 
$$L_{recon} = \left(\frac{1}{N}\right) \sum \sum ||v_{(t_j)} - \hat{v}_{(t_j)}||^2$$
 [1, 6]. The classification loss is binary cross-entropy between predicted pCR probabilities and ground-truth labels: 
$$L_{class} = - [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$
 [7, 8].

Regularization terms may include weight decay on neural network parameters to prevent overfitting, especially when training cohorts are modest in size ( $N < 200$ ), and ODE-specific regularization such as penalizing rapid changes in  $f_\theta$  to encourage smooth dynamics [1, 14]. For the adjoint sensitivity training method, no additional memory regularization is required as the constant-memory property already provides efficiency [1, 15]. The loss is minimized using stochastic gradient descent or Adam optimizer, with mini-batches of patients and gradients computed through the ODE solve using the adjoint method [16, 17].

## Handling sparse data

A fundamental advantage of Neural ODEs is their natural handling of irregularly sampled data, as the ODE solver can integrate between any time points without requiring interpolation or fixed time grids [1, 18]. Each patient can have a different number of MRI time points (2, 3, or more) and different acquisition timings (e.g., mid-treatment at week 3 vs week 5), and the framework processes them all within the same training loop by computing reconstruction loss only at observed times [1, 19]. This flexibility is particularly valuable for clinical data where scan schedules vary due to patient logistics, protocol changes, or image quality issues requiring repeat scans [20, 21].

For patients with only two time points (baseline and post-treatment), the framework still learns meaningful dynamics because the reconstruction loss constrains the integrated trajectory to match both observed volumes, though parameter identifiability is weaker than with three or more time points [1, 22]. The adjoint method enables memory-efficient training even with long integration windows (0 to 18 weeks) and many solver steps, as memory usage does not scale with integration duration [1, 23]. Data augmentation strategies such as time-point dropout (randomly withholding some observed time points during training) can improve robustness to sparse data at test time [24, 25].

## Evaluation Strategy

### Prediction metrics

The primary evaluation metric for pCR prediction is the area under the receiver operating characteristic curve (AUROC), which measures the model's ability to discriminate between responders and non-responders across all classification thresholds and is insensitive to class imbalance [5, 11]. Secondary metrics include

accuracy (proportion correct predictions), sensitivity (true positive rate for pCR detection), specificity (true negative rate), positive predictive value, and negative predictive value, all of which should be reported with confidence intervals [6, 7]. For the trajectory reconstruction task, root mean squared error (RMSE) between predicted and observed tumor volumes at mid-treatment and post-treatment time points quantifies how faithfully the Neural ODE captures observed dynamics [1, 12].

Comparison to baseline methods is essential to establish the framework's value, with appropriate comparators including: (1) RECIST-based volume change (e.g., percent reduction from baseline to mid-treatment), (2) logistic regression using baseline volume and clinical factors, (3) standard deep learning classifiers applied to volume time series without ODE structure, and (4) classical parametric growth models (exponential, logistic, Gompertz) with parameter estimation [3, 13]. Statistical comparisons of AUROC between models should use DeLong's test for paired samples, with  $p < 0.05$  indicating significant improvement [14, 15].

### Validation protocols

Internal validation using k-fold cross-validation (typically 5 or 10 folds) is the minimum standard, ensuring that reported performance reflects generalization to unseen patients from the same population [16, 17]. Patients should be split at the patient level, not the scan level, to avoid data leakage where scans from the same patient appear in both training and validation sets [18, 19]. External validation on an independent cohort from a different institution or clinical trial is critical before clinical deployment, as performance often degrades when applied to populations with different case mixes, MRI protocols, or treatment regimens [5, 20].

**Table 2** translates the manuscript's conceptual proposal into a clinical-methodological agenda by showing which assumptions are load-bearing, how they can fail, and what validation strategy is needed before translational use.

**Table 2.** Assumption–Risk–Validation Matrix for Translating Neural ODE-Based pCR Prediction from Conceptual Framework to Clinical Study Design

Framework assumption or design commitment	Why it is necessary in this manuscript	Principal threat if violated	Observed manifestation of failure

Tumor volume can be measured reproducibly from serial DCE-MRI	The model depends on volume trajectories as its primary observed signal	Segmentation noise may be mistaken for biological dynamics	Implant trajectories oscillate unstably, prediction poor calibration	The learned latent state is predictive of pCR rather than only reconstructive of volume	The framework aims to link dynamics to outcome, not just fit curves	Good reconstruction may coexist with weak response discrimination	Low AUROC despite plausible trajectories
At least two clinically meaningful time points are available per patient	The dynamics require observed temporal anchors	Underconstrained trajectories become weakly identifiable	Good fit but extrapolation generalization and uncertainty	Mid-treatment dynamics are informative enough for early prediction	The clinical promise rests on actionable pre-completion inference	Early predictions may be overconfident or clinically premature	Strong timing performance but weak 3-month performance
Actual scan timing carries biological information and should be retained	The framework's advantage depends on modeling irregular time directly	Temporal normalization may erase clinically informative response timing	Reduce overfitting over baseline endpoint model	Performance gains over conventional methods are clinically meaningful	Conceptual novelty must translate into comparative utility	Improvement may be statistically trivial or clinically irrelevant	Marginal AUROC without sensitivity-specific calibration
Tumor response during NAC is reasonably approximable by a smooth ODE	Neural ODEs assume continuous latent evolution between observations	Abrupt biological shifts may not be represented well by smooth dynamics	System underfitting sudden progression treatment switch	Learned trajectories are interpretable enough for clinical trust	Adoption depends on more than raw predictive accuracy	Black-box dynamics may undermine clinician confidence	Correlation prediction with understanding temporal ratios
Clinical covariates meaningfully modulate dynamics	Patient heterogeneity is central to individualized prediction	Covariates may add noise or encode site-specific confounding	Apparent subgroup performance differential failure extent	Model validity is stable across sites, scanners, and regimens	Clinical deployment requires robustness beyond a single dataset	Domain shift may alter both imaging-derived volumes and response patterns	Calibration drift, subgroup-specific collapse dependent error
Training cohort is large and diverse enough to estimate $\theta$	Flexible dynamics require sufficient sample support	Overfitting to institution-specific patterns or subtype imbalance	Inflated AUROC marginally extended degradation	Sensitivity analyses should examine how prediction performance varies with the number and timing of available MRI time points, for example comparing models trained using: (A) baseline only, (B) baseline + mid-treatment, (C) baseline + post-treatment, and (D) all three time points [1, 21]. Additional sensitivity analyses should assess robustness to segmentation errors by adding synthetic			

noise to extracted volumes, and to timing variability by jittering scan time stamps [22, 23]. Subgroup analyses by tumor subtype (triple-negative, HER2-positive, hormone receptor-positive/HER2-negative) are essential, as predictive performance may differ substantially across biologically distinct patient populations [4, 24].

## Conclusion

We have presented a conceptual framework that leverages Neural Ordinary Differential Equations to model continuous tumor growth dynamics from sparse serial MRI measurements for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. By treating tumor evolution as a continuous dynamical system rather than discrete volume measurements, the framework captures the underlying biological processes that distinguish responders from non-responders, potentially enabling earlier and more accurate prediction than conventional volume-based approaches. The framework naturally accommodates irregularly sampled clinical data, integrates clinical covariates, and provides interpretable visualizations of learned growth trajectories that could support clinical decision-making.

The key advantages of this approach include: continuous modeling of tumor dynamics between sparse observation points, graceful handling of variable scan timing and missing time points without imputation, end-to-end training that simultaneously optimizes trajectory reconstruction and response prediction, memory-efficient training via the adjoint method enabling scaling to large cohorts, and inherent interpretability through trajectory visualization and growth rate extraction. These properties address fundamental limitations of current methods that treat MRI time points as independent or rely on simple volume change calculations, ignoring the rich dynamical

information contained in how tumors evolve during treatment.

Implementation of this framework on existing breast cancer NAC datasets with serial MRI—such as the I-SPY2 trial or ACRIN 6657 trial—would establish whether Neural ODE-based growth modeling improves pCR prediction compared to conventional methods. Future extensions could incorporate additional data modalities including dynamic contrast-enhanced pharmacokinetic parameters, diffusion-weighted MRI, or liquid biopsy biomarkers into the latent state representation. As computational oncology moves toward personalized treatment planning, frameworks that respect the continuous, dynamical nature of tumor response to therapy will become increasingly essential for precision medicine.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 05 Oct 2023 Revised: 12 Dec 2023 Accepted: 03 Jan 2024

Published online: 20 July 2024

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Chen RT, Rubanova Y, Bettencourt J, Duvenaud DK. Neural ordinary differential equations. *Adv Neural Inf Process Syst*. 2018;31:6571-83.
- Rubanova Y, Chen RT, Duvenaud DK. Latent ordinary differential equations for irregularly-sampled time series. *Adv Neural Inf Process Syst*. 2019;32:5321-31.
- Rackauckas C, Ma Y, Martensen J, Warner C, Zubov K, Supekar R, et al. Universal differential equations for scientific machine learning. *arXiv [Preprint]*. 2020:arXiv:2001.04385.
- Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*. 2019;378:686-707.
- Zhu Z, Albadawy E, Saha A, Zhang J, Harowicz MR, Mazurowski MA. Deep learning for identifying radiogenomic associations in breast cancer. *Comput Biol Med*. 2019;109:85-90.
- Braman N, Adoui ME, Vulchi M, Turk P, Etesami M, Fu P, et al. Deep learning-based prediction of response to HER2-targeted neoadjuvant chemotherapy from pre-treatment dynamic breast MRI: a multi-institutional validation study. *arXiv [Preprint]*. 2020:arXiv:2001.08570.
- Ha R, Chin C, Karcich J, Liu MZ, Chang P, Mutasa S, et al. Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset. *J Digit Imaging*. 2019;32(5):693-701.
- Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. *Breast Cancer Res Treat*. 2019;173(2):455-63.
- Peng Y, Cheng Z, Gong C, Zheng C, Zhang X, Wu Z, et al. Pretreatment DCE-MRI-based deep learning outperforms radiomics analysis in predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. *Front Oncol*. 2022;12:846775.
- Joo S, Ko ES, Kwon S, Jeon E, Jung H, Kim JY, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Sci Rep*. 2021;11(1):18800.
- Qu YH, Zhu HT, Cao K, Li XT, Ye M, Sun YS. Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning method. *Thorac Cancer*. 2020;11(3):651-8.
- Jin C, Yu H, Ke J, Ding P, Yi Y, Jiang X, et al. Predicting treatment response from longitudinal images using multi-task deep learning. *Nat Commun*. 2021;12(1):1851.
- Li F, Yang Y, Wei Y, He P, Chen J, Zheng Z, et al. Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. *J Transl Med*. 2021;19(1):348.
- Dammu H, Ren T, Duong TQ. Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. *PLoS One*. 2023;18(1):e0280148.
- Panthi B, Mohamed RM, Adrada BE, Boge M, Candelaria RP, Chen H, et al. Longitudinal dynamic contrast-enhanced MRI radiomic models for early prediction of response to neoadjuvant systemic therapy in triple-negative breast cancer. *Front Oncol*. 2023;13:1264259.
- Mohamed RM, Panthi B, Adrada BE, Boge M, Candelaria RP, Chen H, et al. Multiparametric MRI-based radiomic models for early prediction of response to neoadjuvant systemic therapy in triple-negative breast cancer. *Sci Rep*. 2024;14(1):16073.
- Zhou Z, Adrada BE, Candelaria RP, Elshafeey NA, Boge M, Mohamed RM, et al. Prediction of pathologic complete response to neoadjuvant systemic therapy in triple negative breast cancer using deep learning on multiparametric MRI. *Sci Rep*. 2023;13(1):1171.
- Zeng H, Qiu S, Zhuang S, Wei X, Wu J, Zhang R, et al. Deep learning-based predictive model for pathological complete response to neoadjuvant chemotherapy in breast cancer from biopsy pathological images: a multicenter study. *Front Physiol*. 2024;15:1279982.
- Guo J, Chen B, Cao H, Dai Q, Qin L, Zhang J, et al. Cross-modal deep learning model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. *NPJ Precis Oncol*. 2024;8(1):189.
- Carriero A, Groenhoff L, Vologina E, Basile P, Albera M. Deep learning in breast cancer imaging: state of the art and recent advancements in early 2024. *Diagnostics (Basel)*. 2024;14(8):848.

Wong C, Fu Y, Li M, Mu S, Chu X, Fu J, et al. MRI-based artificial intelligence in rectal cancer. *J Magn Reson Imaging*. 2023;57(1):45-56.

Ravichandran K, Braman N, Janowczyk A, Madabhushi A. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. In: *Medical Imaging 2018: Computer-Aided Diagnosis*. Bellingham (WA): SPIE; 2018. p. 79-88.

Raissi M, Karniadakis GE. Hidden physics models: machine learning of nonlinear partial differential equations. *J Comput Phys*. 2018;357:125-41.

Ansari AF, Heng A, Lim A, Soh H. Neural continuous-discrete state space models for irregularly-sampled time series. In: *International Conference on Machine Learning*. PMLR; 2023. p. 926-51.

Bilic A, Chen C. BC-MRI-SEG: a breast cancer MRI tumor segmentation benchmark. In: *2024 IEEE 12th International*

*Conference on Healthcare Informatics (ICHI)*. IEEE; 2024. p. 674-8.

Kim J, Park H. Radiomics-guided multimodal self-attention network for predicting pathological complete response in breast MRI. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2024. p. 1-5.

Sharma D, Purushotham S, Reddy CK. MedFuseNet: an attention-based multimodal deep learning model for visual question answering in the medical domain. *Sci Rep*. 2021;11(1):19826.

Zhao Q, Liu Z, Adeli E, Pohl KM. Longitudinal self-supervised learning. *Med Image Anal*. 2021;71:102051.

Vieira BH, Liem F, Dadi K, Engemann DA, Gramfort A, Bellec P, et al. Predicting future cognitive decline from non-brain and multimodal brain imaging data in healthy and pathological aging. *Neurobiol Aging*. 2022;118:55-65.