



Data-Driven Discovery of Mathematical and Physical Relations in Oncology Data Using Human-Understandable Machine Learning

Daria Kurz^{1†}, Carlos Salort Sánchez^{2†} and Cristian Axenie^{3*†}

¹Interdisziplinäres Brustzentrum, Helios Klinikum München West, Akademisches Lehrkrankenhaus der Ludwig-Maximilians Universität München, Munich, Germany, ²TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany, ³Audi Konfuzius-Institut Ingolstadt Laboratory, Technische Hochschule Ingolstadt, Ingolstadt, Germany

OPEN ACCESS

Edited by:

George Bebis,
University of Nevada, Reno,
United States

Reviewed by:

Chang Chen,
University of Chicago, United States
Tin Nguyen,
University of Nevada, Reno,
United States
Sokratis Makrogiannis,
Delaware State University,
United States

*Correspondence:

Cristian Axenie
cristian.axenie@audi-konfuzius-
institut-ingolstadt.de

[†]These authors have contributed
equally to the work

Specialty section:

This article was submitted to
Medicine and Public Health,
a section of the journal
Frontiers in Artificial Intelligence

Received: 23 May 2021

Accepted: 08 October 2021

Published: 25 November 2021

Citation:

Kurz D, Sánchez CS and Axenie C
(2021) Data-Driven Discovery of
Mathematical and Physical Relations in
Oncology Data Using Human-
Understandable Machine Learning.
Front. Artif. Intell. 4:713690.
doi: 10.3389/frai.2021.713690

For decades, researchers have used the concepts of rate of change and differential equations to model and forecast neoplastic processes. This expressive mathematical apparatus brought significant insights in oncology by describing the unregulated proliferation and host interactions of cancer cells, as well as their response to treatments. Now, these theories have been given a new life and found new applications. With the advent of routine cancer genome sequencing and the resulting abundance of data, oncology now builds an “arsenal” of new modeling and analysis tools. Models describing the governing physical laws of tumor–host–drug interactions can be now challenged with biological data to make predictions about cancer progression. Our study joins the efforts of the mathematical and computational oncology community by introducing a novel machine learning system for data-driven discovery of mathematical and physical relations in oncology. The system utilizes computational mechanisms such as competition, cooperation, and adaptation in neural networks to simultaneously learn the statistics and the governing relations between multiple clinical data covariates. Targeting an easy adoption in clinical oncology, the solutions of our system reveal human-understandable properties and features hidden in the data. As our experiments demonstrate, our system can describe nonlinear conservation laws in cancer kinetics and growth curves, symmetries in tumor’s phenotypic staging transitions, the preoperative spatial tumor distribution, and up to the nonlinear intracellular and extracellular pharmacokinetics of neoadjuvant therapies. The primary goal of our work is to enhance or improve the mechanistic understanding of cancer dynamics by exploiting heterogeneous clinical data. We demonstrate through multiple instantiations that our system is extracting an accurate human-understandable representation of the underlying dynamics of physical interactions central to typical oncology problems. Our results and evaluation demonstrate that, using simple—yet powerful—computational mechanisms, such a machine learning system can support clinical decision-making. To this end, our system is a representative tool of the field of mathematical and computational oncology

and offers a bridge between the data, the modeler, the data scientist, and the practicing clinician.

Keywords: mathematical oncology, machine learning, mechanistic modeling, data-driven predictions, clinical data, decision support system

1 INTRODUCTION

The dynamics governing cancer initiation, development, and response to treatment are informed by quantitative measurements. These measurements carry details about the physics of the underlying processes, such as tumor growth, tumor–host cell encounters, and drug transport. Be it through mathematical modeling and patient-specific treatment trajectories—as in the excellent work of Werner et al. (2016)—through tumor’s mechanopathology—systematically described by Nia et al. (2020)—or through hybrid modeling frameworks of tumor development and treatment—identified by Chamseddine and Rejniak (2020)—capturing such processes from data can substantially improve predictions about cancer progression.

Machine learning algorithms are now leveraging automatic discovery of physics principles and governing mathematical relations for such improved predictions. Proof stands the proliferating body of such research—for representative results, see the works of Raissi (2018), Schaeffer (2017), Long et al. (2018), and Champion et al. (2019). However, the naive application of such algorithms is insufficient to infer physical laws underlying cancer progression. Simply positing a physical law or mathematical relation from data is useless without simultaneously proposing an accompanying ground truth to account for the inevitable mismatch between model and observations, as demonstrated in the work of de Silva et al. (2020).

Such a problem is even more important in clinical oncology where, in order to understand the links between the physics of cancer and signaling pathways in cancer biology, we need to describe the fundamental physical principles shared by most, if not all, tumors, as proposed by Nia et al. (2020). Here, mathematical models of the physical mechanisms and corresponding tumor physical hallmarks complement the heterogeneity of the experimental observations. Such a constellation is typically validated through *in vivo* and *in vitro* model systems where the simultaneous identification of both the structure and parameters of the dynamical system describing tumor–host interactions is performed (White et al., 2019).

Given the multidimensional nature of this system identification process, some concepts involved are nonintuitive and require deep and broad understanding of both the physical and biological aspects of cancer. To circumvent this, combining mechanistic modeling and machine learning is a promising approach with high potential for clinical translation. For instance, in a bottom-up approach, fusing cell-line tumor growth curve learning from heterogeneous data (i.e., caliper, imaging, microscopy) and unsupervised extraction of cytostatic pharmacokinetics, the study by Axenie and Kurz (2020a) introduced a novel pipeline for patient-tailored neoadjuvant

therapy planning. In another relevant study, Benzekry (2020) used machine learning to extract model parameters from high-dimensional baseline data (demographic, clinical, pathological molecular) and used mixed-effects theory to combine it with mechanistic models based on longitudinal data (e.g., tumor size measurements, pharmacokinetics, seric biomarkers, and circulating DNA) for treatment individualization.

Yet, despite the recent advances in mathematical and computational oncology, there are only a few systems trying to offer a human-understandable solution, or the steps to reach it—the most relevant are the studies by Jansen et al. (2020) and Lamy et al. (2019). But, such systems lack a rigorous and accessible description of the physical cancer traits assisting their clinical predictions. Our study advocates the improvement of mechanistic modeling with the help of machine learning. Our thesis goes beyond measurements-informed biophysical processes models, as described by Cristini et al. (2017), and toward human-understandable personalized disease evolution and therapy profiles learned from data, as foreseen by Kondylakis et al. (2020).

1.1 Study Focus

The purpose of this study is to introduce a system (and a framework) capable of learning human-understandable mathematical and physical relations from heterogeneous oncology data for patient-centered clinical decision support. To demonstrate the versatility of the system, we introduce multiple of its instantiations, in an end-to-end fashion (i.e., from cancer initiation to treatment outcome) for predictions based on available clinical datasets¹:

- **learning initiation patterns of preinvasive breast cancer** (i.e., ductal carcinoma *in situ* [DCIS]) from histopathology and morphology data available from the studies by Rodallec et al. (2019), Volk et al. (2011), Tan et al. (2015), and Mastri et al. (2019);
- **learning unperturbed tumor growth curves within and between cancer types** (i.e., breast, lung, leukemia) from imaging, microscopy, and caliper data available from the studies by Benzekry et al. (2019) and Simpson-Herren and Lloyd (1970);
- **extracting tumor phenotypic stage transitions** from three cell lines of breast cancer using imaging, immunohistochemistry, and histopathology data available from the studies by Rodallec et al. (2019), Volk et al. (2011), Tan et al. (2015), and Edgerton et al. (2011);

¹A copy of the used datasets along with the study generating them is included in the codebase associated with the manuscript.

- **simultaneously extracting the drug-perturbed tumor growth and drug pharmacokinetics** for neoadjuvant/adjuvant therapy sequencing using data available from the studies by Kuh et al. (2000), Volk et al. (2011), and Chen et al. (2014);
- **predicting tumor growth/recession** (i.e., estimating tumor volume after each chemotherapy cycle **under various chemotherapy regimens** administered to breast cancer patients, using real-world patient data available from the study by Yee et al. (2020) as well as cell lines studies from Rodallec et al. (2019), Volk et al. (2011), Tan et al. (2015), and Mastro et al. (2019).
- **determines cancer treatment planning**, for instance, adjuvant versus neoadjuvant chemotherapy (Sarapata and de Pillis, 2014).

In each of the instantiations, we use the same computational substrate (i.e., no specific task parametrization) and compare the performance of our system against state-of-the-art systems capable of extracting governing equations from heterogeneous oncology data from Cook et al. (2010), Mandal and Cichocki (2013), Weber and Wermter (2007), and Champion et al. (2019), respectively. The analysis focuses on (1) the accuracy of the systems in the learned mathematical and physical relations among various covariates, (2) the ability to embed more data and mechanistic models, and (3) the ability to provide a human-understandable solution and the processing steps to obtain that solution.

1.2 Study Motivation

In clinical practice, patient tumors are typically described across multiple dimensions from (1) high-dimensional heterogeneous data (e.g., demographic, clinical, pathological, molecular), and (2) longitudinal data (e.g., tumor size measurements, pharmacokinetics, immune screening, biomarkers), to (3) time-to-event data (e.g., progression-free or overall survival analysis), and, in the last years, (4) genetic sequencing that determine the genetic mutations driving their cancer. With this information, the clinical oncologist may tailor treatment to the patient's specific cancer.

But, despite the variety of such rich patient data available, tumor growth data, describing the dynamics of cancer development, from initiation to metastasis has some peculiarities. These features motivated the study and the approach proposed by our system. To summarize, tumor growth data:

- is typically **small**, with only a few data points measured, typically, at days-level resolution (Roland et al., 2009);
- is **unevenly sampled**, with irregular spacing among tumor size/volume observations (Volk et al., 2011);
- has **high variability** between and within tumor types (Benzekry et al., 2014) and type of treatment (Gaddy et al., 2017).
- is **heterogeneous** and sometimes **expensive or difficult to obtain** (e.g., biomarkers, functional magnetic resonance imaging (Abler et al., 2019), fluorescence imaging (Rodallec et al., 2019), flow cytometry, or calipers (Benzekry et al., 2019).

2 MATERIALS AND METHODS

In the current section, we introduce our system through the lens of practical examples of discovering mathematical and physical relations describing tumor–host–drug dynamics. We begin by introducing the basic computational framework as well as the various configurations in which the system can be used. The second part is dedicated to introducing relevant state-of-the-art approaches used in our comparative experimental evaluation.

2.1 System Basics

Physical interactions of cancer cells with their environment (e.g., local tissue, immune cells, drugs) determine the physical characteristics of tumors through distinct and interconnected mechanisms. For instance, cellular proliferation and its inherent abnormal growth patterns lead to increased solid stress (Nia et al., 2016). Subsequently, cell contraction and cellular matrix deposition modify the architecture of the surrounding tissue, which can additionally react to drugs (Griffon-Etienne et al., 1999) modulating the stiffness (Rouvière et al., 2017) and interstitial fluid pressure (Nathanson and Nelson, 1994). But such physical characteristics also interact among each other initiating complex dynamics, as demonstrated in Nia et al. (2020).

Our system can capture such complex dynamics through a network-based paradigm for modeling, computation, and prediction. It can extract the mathematical description of the interactions exhibited by multiple entities (e.g., tumor, host cells, cytostatic drugs) for producing informed predictions. For guiding the reader, we present a simple, biologically grounded example in **Figure 1**.

In this example, our system learns simultaneously the power-law tumor growth under immune escape (Benzekry et al., 2014) and the nonlinear potentiation–inhibition model of natural killer (NK) cells–tumor interactions (Ben-Shmuel et al., 2020), while exhibiting the known overlapping cytotoxic T lymphocytes (CTLs)–NK cell mutual linear regulation pattern (Uzhachenko and Shanker, 2019). As shown in **Figure 1**, our system offers the means to learn the mathematical relations governing the physical tumor–immune interactions, without supervision, from available clinical data (**Figure 1**—input data relations and learned and decoded relations). Furthermore, the system can infer unavailable (i.e., expensive to measure) physical quantities (i.e., after learning/training) in order to make predictions on the effects of modifying

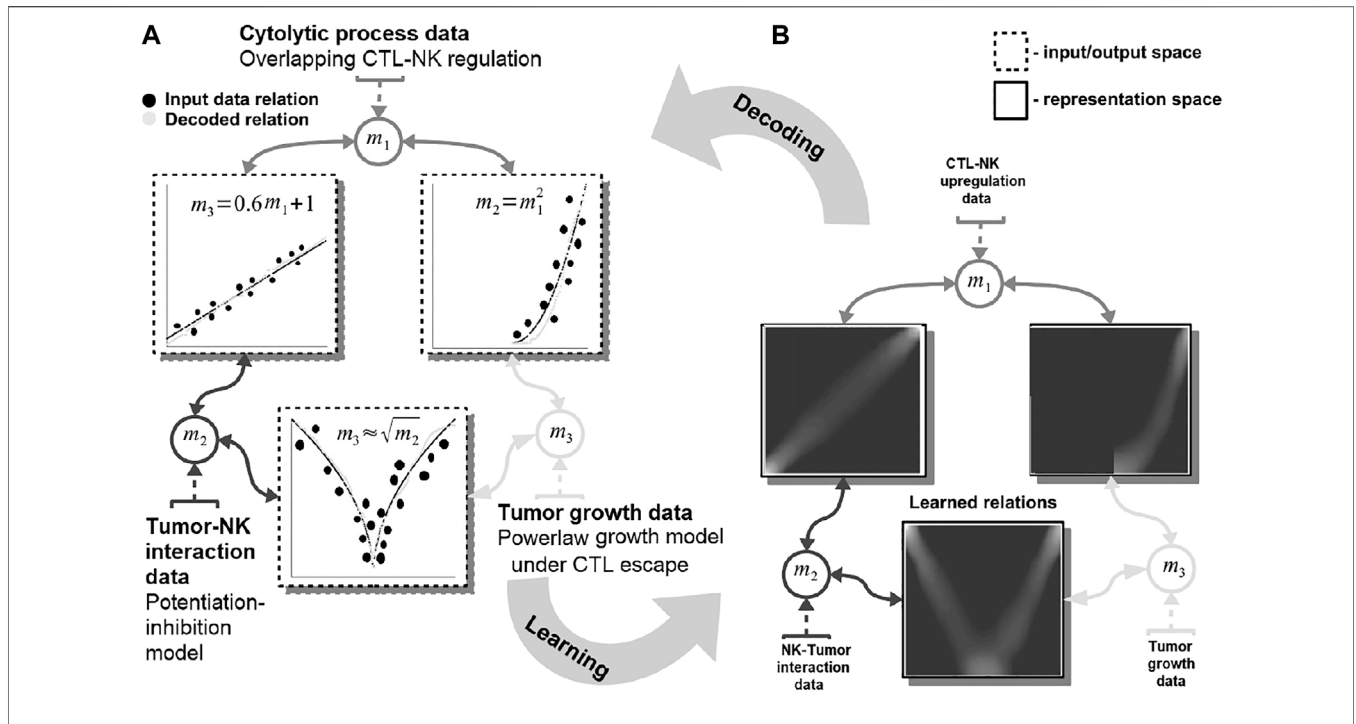


FIGURE 1 | Basic functionality of the proposed machine learning system. Data are fed in the system through the representations maps, m_i which encode each quantity in a distributed (array-like) representation. The system dynamics brings all available quantities into agreement and learns the underlying mathematical relations among them [see representation space—(A)]. The relations resemble the mathematical model of the interactions: power-law tumor growth under immune escape, nonlinear potentiation–inhibition tumor–immune interaction, and linear regulation pattern among immune system cells. The learned mathematical relations are then compared with the data (i.e., the ground truth) and the mechanistic model output [see input/output space—(B)]. Note: to simplify the visualization data points are depicted as clusters (i.e., the size of a cluster reflects the number of points concentrated in a region).

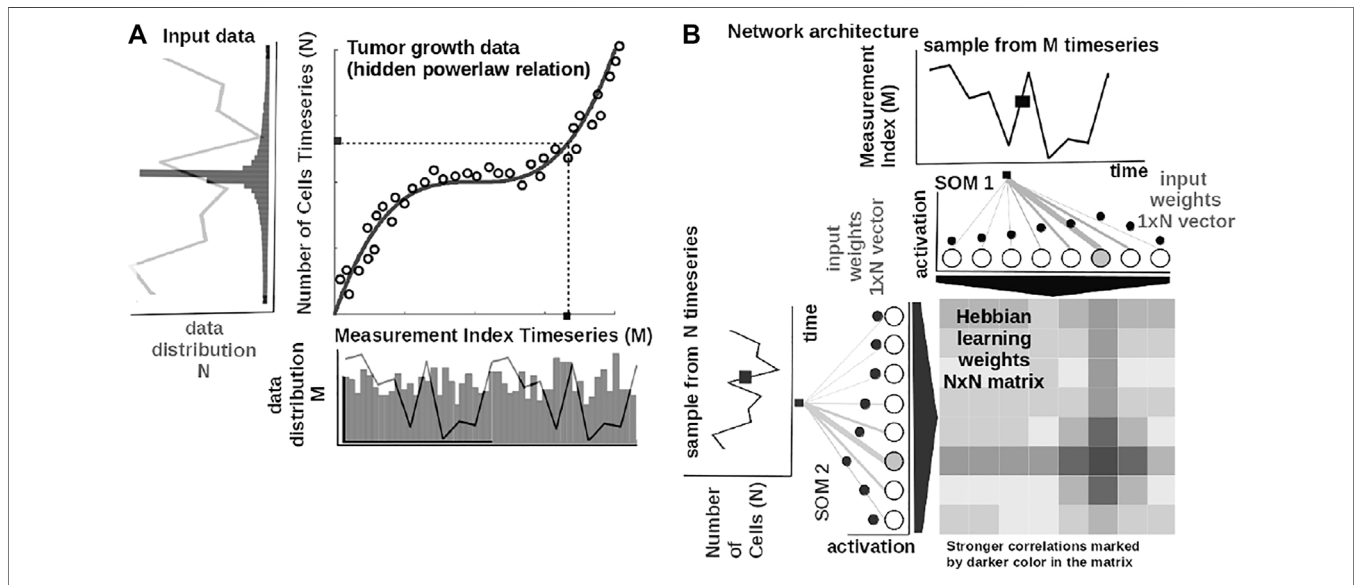


FIGURE 2 | Basic functionality of the system. (A) Tumor growth data following a nonlinear mathematical relation and its distribution—relation is hidden in the time series (i.e., number of cells vs. measurement index). Data from Comen et al. (2016). (B) Basic architecture of our system: one-dimensional (array) SOM networks with N neurons encoding the time series (i.e., number of cells vs. measurement index), and an $N \times N$ Hebbian connection matrix (coupling the two SOMs) that will encode the mathematical relation after training.

the pattern of interactions among the tumor and the immune system. For instance, by not feeding the system with the innate immune response (i.e., the NK cells dynamics), the system infers, based on the CTL–NK cell interaction pattern and the tumor growth pattern, a plausible tumor–NK cell mathematical relation in agreement with observations (**Figure 1B**, squared root nonlinearity).

Basically, our system acts as constraint satisfaction network converging to a global consensus given local (i.e., the impact of the measured data) and global dynamics of the physics governing the interactions (see the clear patterns depicting the mathematical models of interaction in **Figure 1**). The networked structure allows the system to easily interconnect multiple data quantities measuring different biological components (Markowetz and Troyanskaya, 2007) or a different granularity of representation of the underlying interaction physics (Cornish and Markowetz, 2014).

2.2 Computational Substrate

The core element of our study is an unsupervised machine learning system based on Self-Organizing Maps (SOMs) Kohonen (1982) and Hebbian learning (HL) Chen et al. (2008). The two components are used in concert to represent and extract the underlying relations among correlated data. In order to introduce the computational steps followed by our system, we provide a simple example in **Figure 2**. Here, we feed the system with data from a cubic growth law (third power-law) describing the effect of drug dose density over 150 weeks of adjuvant chemotherapy in breast cancer (data from Comen et al., 2016). The two data sources (i.e., the cancer cell number and the irregular measurement index over the weeks) follow a cubic dependency (cmp. **Figure 2A**). Before being presented the data, our system has no prior information about the data distribution and its generating process (or model). The system learns the underlying (i.e., hidden) mathematical relation directly from the pairs of input data without supervision.

The input SOMs (i.e., one-dimensional [1-D] lattice networks with N neurons) extract the probability distribution of the incoming data, depicted in **Figure 2A**, and encode samples in a distributed activity pattern, as shown in **Figure 2B**. This activity pattern is generated such that the closest preferred value of a SOM neuron to the input will be strongly activated and will decay, proportional with distance, for neighboring units. This process is fundamentally benefiting from the quantization capability of SOM. The tasks we solve in this work have low dimensionality, basically allowing a 1-D SOM to provide well-behaved distributed representations. 1-D SOMs are proven mathematically to converge and handling boundary effects. For higher-dimensional data, our system can be coupled with a reduction technique (i.e., principal component analysis, t-Distributed Stochastic Neighbor Embedding) to reduce data to 1-D time series, without a large penalty in complexity. In addition, this process is extended with a dimension corresponding to the latent representation of network resource allocation (i.e., number of neurons allocated to represent the input data space). After

learning, the SOMs specialize to represent a certain (preferred) value in the input data space and learn its probability distribution, by updating its tuning curves shape.

Practically, given an input value $s^p(k)$ from one time series at time step k , the network follows the processing stages in **Figure 3**. For each i th neuron in the p th input SOM, with preferred value $w_{in,i}^p$ and tuning curve size $\xi_i^p(k)$, the generated neural activation is given by

$$a_i^p(k) = \frac{1}{\sqrt{2\pi}\xi_i^p(k)} e^{-\frac{(s^p(k)-w_{in,i}^p)^2}{2\xi_i^p(k)^2}}. \quad (1)$$

The most active (i.e., competition winning) neuron of the p th population, $b^p(k)$, is the one that has the highest activation given the time series data point at time k

$$b^p(k) = \underset{i}{\operatorname{argmax}} a_i^p(k). \quad (2)$$

The competition for highest activation (in representing the input) in the SOM is followed by a cooperation process that captures the input space distribution. More precisely, given the winning neuron, $b^p(k)$, the cooperation kernel,

$$h_{b,i}^p(k) = e^{-\frac{\|r_i - r_b\|^2}{2\sigma(k)^2}}, \quad (3)$$

allows neighboring neurons in the network (i.e., found at position r_i in the network) to precisely represent the input data point given their location in the neighborhood $\sigma(k)$ of the winning neuron. The topological neighborhood width $\sigma(k)$ decays in time, to avoid artifacts (e.g., twists) in the SOM. The kernel in **Eq. 3** is chosen such that adjacent neurons in the network specialize on adjacent areas in the input space, by “pulling” the input weights (i.e., preferred values) of the neurons closer to the input data point,

$$\Delta w_{in,i}^p(k) = \alpha(k)h_{b,i}^p(k)(s^p(k) - w_{in,i}^p(k)). \quad (4)$$

This process updates the tuning curves width ξ_i^p given the spatial location of the neuron in the network, the distance to the input data point, the cooperation kernel size, and a decaying learning rate $\alpha(k)$,

$$\Delta \xi_i^p(k) = \alpha(k)h_{b,i}^p(k)((s^p(k) - w_{in,i}^p(k))^2 - \xi_i^p(k)^2). \quad (5)$$

To illustrate these mechanisms, we consider the learned tuning curves shapes for five neurons in the input SOMs (i.e., neurons 1, 6, 13, 40, 45) encoding the breast cancer cubic tumor growth law, depicted in **Figure 4**. We observe that higher input probability distributions are represented by dense and sharp tuning curves (e.g., neuron 1, 6, 13 in SOM1), whereas lower or uniform probability distributions are represented by more sparse and wide-tuning curves (e.g., neuron 40, 45 in SOM1).

This way, the system optimally allocates neurons such that a higher amount of neurons represent areas in the input space, which need a finer resolution, and a lower amount for more coarsely represented input space areas. Neurons in the two SOMs are then linked by a fully (all-to-all) connected matrix of synaptic

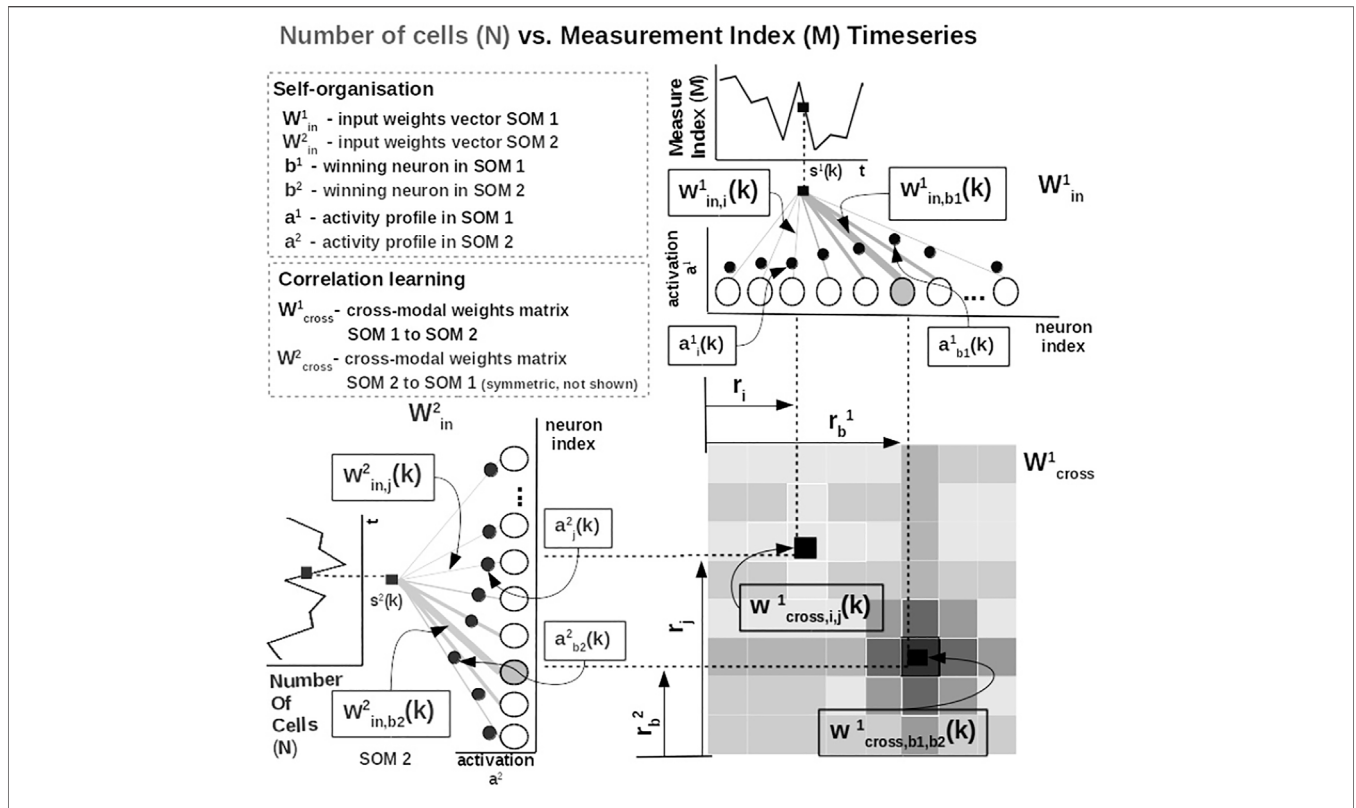


FIGURE 3 | Detailed computational steps of our system, instantiated for tumor growth learning given the observed number of cells and the measurement index data from Comen et al. (2016).

connections, where the weights are computed using HL. The connections between uncorrelated (or weakly correlated) neurons in each SOM (i.e., w_{cross}) are suppressed (i.e., darker color), whereas correlated neuron connections are enhanced (i.e., brighter color), as depicted in **Figure 3**. Each connection weight $w_{cross,i,j}^p$ between neurons i, j in the input SOMs is updated with an HL rule as follows:

$$\Delta w_{cross,i,j}^p(k) = \eta(k)(a_i^p(k) - \bar{a}_i^p(k))(a_j^q(k) - \bar{a}_j^q(k)), \quad (6)$$

where

$$\bar{a}_i^p(k) = (1 - \beta(k))\bar{a}_i^p(k-1) + \beta(k)a_i^p(k), \quad (7)$$

is an exponential decay (i.e., momentum), and $\eta(k)$, $\beta(k)$ are monotonic (inverse-time) decaying functions. HL ensures a weight increase for correlated activation patterns and a weight decrease for anticorrelated activation patterns. The Hebbian weight matrix encodes the coactivation patterns between the input SOMs, as shown in **Figure 2B**, and, eventually, the learned mathematical relation given the data, as shown in **Figure 4**. Such a representation, as shown in **Figure 4**, demonstrates the human-understandable output of our system that employs powerful, yet simple and transparent, processing principles, as depicted in **Figure 3**.

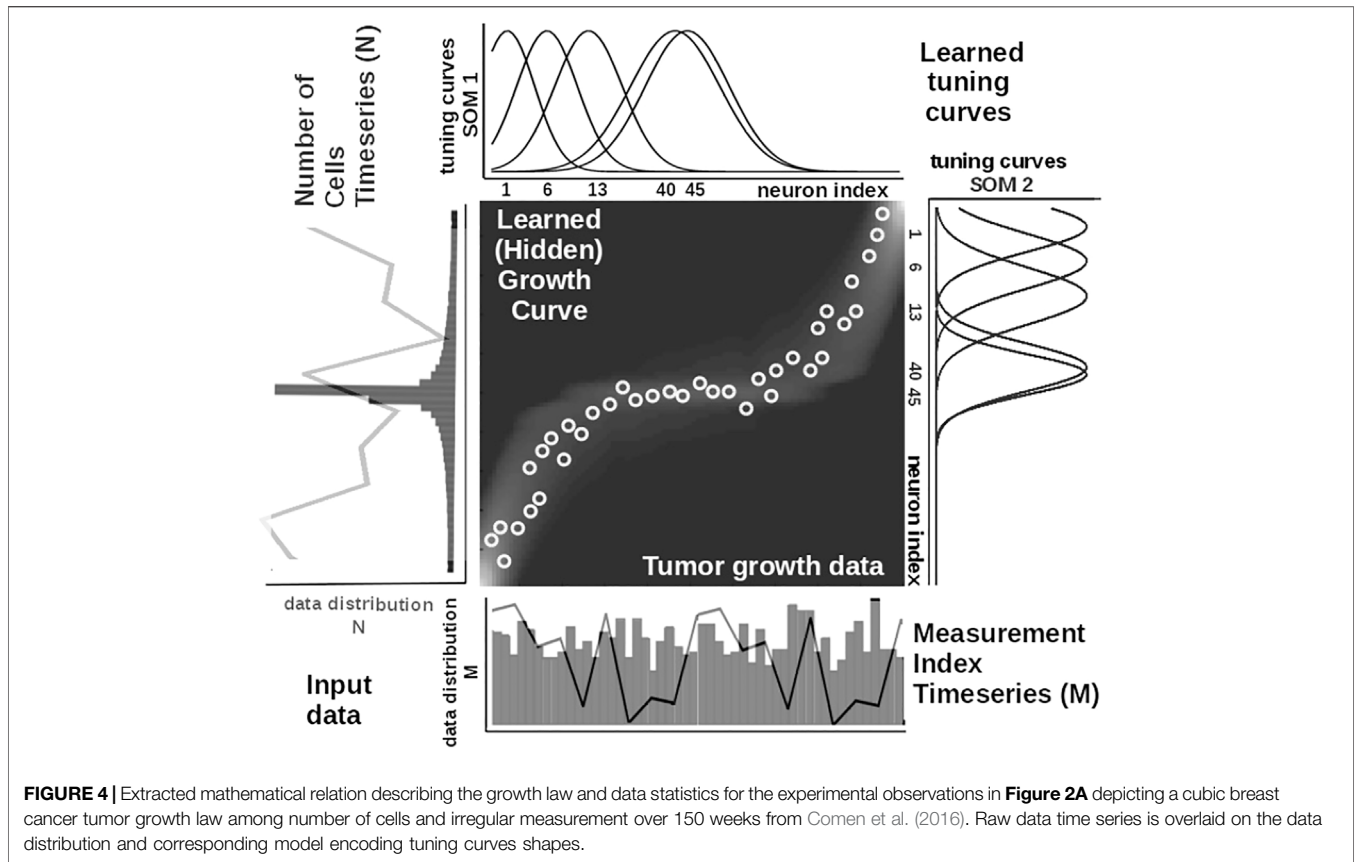
Input SOM self-organization and Hebbian correlation learning operate at the same time in order to refine both the input data representation and the extracted

mathematical relation. This is visible in the encoding and the decoding functions where the input activations a are projected through the input weights w_{in} (Eq. 1) to the Hebbian matrix and then decoded through the w_{cross} correlation weights (Eq. 8).

In order to recover the real-world value from the network, we use a decoding mechanism based on (self-learned) bounds of the input data space. The input data space bounds are obtained as minimum and maximum of a cost function of the distance between the current preferred value of the winning neuron (i.e., the value in the input which is closest [in Euclidian space] to the weight vector of the neuron) and the input data point in the SOM (i.e., using Brent's optimization Brent, 2013). Depending on the position of the winning neuron in the SOM, the decoded/recovered value $y(t)$ from the SOM neurons weights is computed as follows:

$$y(t) = \begin{cases} w_{in,i}^p + d_i^p & \text{if } i \geq \frac{N}{2} \\ w_{in,i}^p - d_i^p & \text{if } i < \frac{N}{2} \end{cases}$$

where $d_i^p = \sqrt{2\xi_i^p(k)^2 \log(\sqrt{2\pi} a_i^p(k) \xi_i^p(k)^2)}$ for the winning neuron with index i in the SOM, a preferred value $w_{in,i}^p$ and $\xi_i^k(k)$ tuning curve size and $\Delta a_i^p(k) = w_{cross,i,j}^p(k)a_j^q(k)$. The activation $a_i^p(k)$ is computed by projecting one data point through SOM q and subsequently through the Hebbian matrix



to compute the paired activity (i.e., at the other SOM p , Eq. (8)) describing the other data quantity.

$$\Delta a_i^p(k) = w_{cross,i,j}^p(k) a_j^q(k) \quad (8)$$

where $w_{cross,i,j}^p(k) = rot90(w_{cross,i,j}^q(k))$ and $rot90$ is a clockwise rotation. The processes described in the previous equations denote the actual inference process following the training phase (i.e., classifying new data). Basically, after applying the input time series and finding the winner in the input SOM population, the decoding decision is based on the position of the winner. Two bounds (i.e., left and right) are defined with respect to the winner's position such that the recovered value is obtained by running Brent's algorithm between the preferred values of the neurons with indices given by the bounds. The method is guaranteed to converge to global minima (of the cost function), and it is immune to boundary effects, if winners are placed at the extremes of the SOM population. A thorough analysis of the learned relations in the Hebbian matrix demonstrated that because of the asymmetric neighborhood function in the input SOMs, the activity saturated at the edges of the latent representation space. Interestingly, this was also visible in the coactivation pattern, such that the higher activity values characterize the bounds of the Hebbian representation toward the edges. When decoding the activity pattern from the Hebbian matrix, we were able to recover a relatively good probability distribution shape. This, interesting and useful,

behavior emphasizes the joint effect that the SOM distributed representation boundary effects and the Hebbian temporal coactivation have upon the data. The resulting distributions have a convex profile, concentrating a large number of samples toward the edges of the histogram with a large variance, whereas precisely decoded areas follow a relatively uniform distribution. We noticed that the decoder treated equally (i.e., accuracy of decoding) linear relations with strong boundary conditions and symmetric nonlinear relations without boundary conditions. The decoding step is a fundamental aspect contributing to the human-understandable output of our system. This demonstrates that simple operations, such as competition and cooperation in neural networks, can exploit the statistics of clinical data and provide a human-understandable representation of the governing mathematical relations behind tumor growth processes.

2.3 Comparable Systems

In this section, we briefly introduce four state-of-the-art approaches that we comparatively evaluated against our system. Ranging from statistical methods, to machine learning, and up to deep learning (DL), the selected systems were designed to extract governing equations from the data.

Cook et al. The system of *Cook et al. (2010)* uses a combination of simple computational mechanisms, like winner-take-all (WTA) circuits, HL, and homeostatic activity regulation, to extract mathematical relations among different data sources.

Real-world values presented to the network are encoded in population code representations. This approach is similar to our approach in terms of the sparse representation used to encode data. The difference resides in the fact that in our model the input population (i.e., SOM network) connectivity is learned. Using this capability, our model is capable of learning the input data bounds and distribution directly from the input data, without any prior information or fixed connectivity. Furthermore, in this system, the dynamics between each population encoded input is performed through plastic Hebbian connections. Starting from a random connectivity pattern, the matrix finally encoded the functional relation between the variables that it connects. The Hebbian linkage used between populations is the correlation detection mechanism used also in our model, although in our formulation we adjusted the learning rule to accommodate both the increase and decrease of the connection weights.

Weber and Wermter. Using a different neurally inspired substrate, the system of Weber and Wermter (2007) combines competition and cooperation in a self-organizing network of processing units to extract coordinate transformations. More precisely, the model uses simple, biologically motivated operations, in which coactivated units from population-coded representations self-organize after learning a topological map. This basically assumes solving the reference frame transformation between the inputs (mapping function). Similar to our model, the proposed approach extends the SOM network by using sigma-pi units (i.e., weighted sum of products). The connection weight between this type of processing units implements a logical AND relation. The algorithm produces invariant representations and a topographic map representation.

Mandal and Cichocki. Going away from biological inspiration, the system of Mandal and Cichocki (2013) used a type of nonlinear canonical correlation analysis (CCA), namely, alpha-beta divergence correlation analysis (ABCA). The ABCA system extracts relations between sets of multidimensional random variables. The core idea of the system is to first determine linear combinations of two random variables (called canonical variables/variants) such that the correlation between the canonical variables is the highest among all such linear combinations. As traditional CCA is only able to extract linear relations between two sets of multidimensional random variable, the proposed model comes as an extension to extract nonlinear relations, with the requirement that relations are expressed as smooth functions and can have a moderate amount of additive random noise on the mapping. The model employs a probabilistic method based on nonlinear correlation analysis using a more flexible metric (i.e., divergence/distance) than typical CCA.

Champion et al. As DL is becoming a routine tool for data discovery, as shown in the recent work of Champion et al. (2019), Raissi (2018), Schaeffer (2017), and de Silva et al. (2020), we also consider a DL system (inspired from Champion et al., 2019) and evaluate it along the other methods. To apply this prediction method to tumor growth, we need to formulate the setup as a time series prediction problem. At any given point, we have the dates and values of previous observations. Using these two features, we

can implement DL architectures that predict the size of the tumor at a future step. Recurrent neural networks (RNNs) are the archetypal DL architectures for time series prediction. The principal characteristic of RNN, compared with simpler DL architectures, is that they iterate over the values that have been observed, obtaining valuable information from it, like the rate at which the objective variable grows, and use that information to improve prediction accuracy. The main drawback of using DL in the medical field is the need of DL models to be presented with large amounts of data. We address this problem by augmenting the data. We use support vector machines (SVMs) for augmenting data, to obtain expected tumor development with normal noise generates realistic measurements. This approach presents the expected average development of a tumor.

3 EXPERIMENTAL SETUP AND RESULTS

In order to evaluate our data-driven approach to learn mathematical and physical relations from heterogeneous oncology data, we introduce the five instantiations and their data briefly introduced in the Study Focus section.

3.1 Datasets

In our experiments, we used publicly available tumor growth, pharmacokinetics, and chemotherapy regimens datasets (Table 1), with *in vitro* or *in vivo* clinical tumor volume measurements, for breast cancer (datasets 1, 2, 5, 6, 7) and other cancers (e.g., lung, leukemia—datasets 3 and 4, respectively). This choice is to probe and demonstrate transfer capabilities of the system to tumor growth patterns induced by different cancer types. The choice of the dataset for each of the experiments was determined by the actual task we wanted to demonstrate. For instance, for demonstrating the capability to predict preinvasive cancer volume, we used the DCIS dataset. For the between-cancer predictions, we used four (i.e., two breast and two nonbreast) out of the whole seven datasets, whereas for the within-cancer-type analysis, we only looked at the breast cancer growth prediction (i.e., four datasets). For the *in vivo* experiment, we only considered the I-SPY2 trial data.

It is important to note that tumor cancer types are staged based on the size and spread of tumors, basically their volume. However, because leukemia occurs in the developing blood cells in the bone marrow, its staging is different from solid tumors. In order to emphasize the versatility of the evaluated systems, for the leukemia datasets, we used experiments that monitored human leukemic cell engraftment over time by monitoring tumor volume in scaffolds (Antonelli et al., 2016). For the pharmacokinetics experiments (i.e., mainly focused on taxanes family for experiments on MCF-7 breast cancer cell line from Tan et al. (2015)), we used the data from Kuh et al. (2000) describing intracellular and extracellular concentrations of Paclitaxel during uptake. The datasets and the code for all the systems used in our evaluation are available on GitLab².

²Experimental codebase: <https://gitlab.com/akii-microlab/math-comp-ml>.

TABLE 1 | Description of the datasets used in the experiments.

Experimental dataset setup				
Dataset	Cancer type	Data type	Data points	Data freq.
1	Breast ¹ (MDA-MB-231 cell line)	Fluorescence imaging	7	2x/week
2	Breast ² (MDA-MB-435 cell line)	Digital caliper	14	2x/week
3	Lung ³	Caliper	10	7x/week
4	Leukemia ⁴	Microscopy	23	7x/week
5	Breast ⁵ (MCF-7 cell line)	Microscopic imaging	8	1x/week
6	Breast ⁶ (LM2-4LUC + cell line)	Digital caliper	10	3x/week
7	Breast ⁷ (stage 2/3 cancers)	Functional magnetic resonance imaging	5	1x/week
8	Breast ⁸ (ductal carcinoma <i>in situ</i>)	Histopathology	5	1x/week

¹Dataset from the study by Rodallec et al. (2019)

²Dataset from the study by Volk et al. (2011)

³Dataset from the study by Benzekry et al. (2019)

⁴Dataset from the study by Simpson-Herren and Lloyd (1970)

⁵Dataset from the study by Tan et al. (2015)

⁶Dataset from the study by Mastri et al. (2019)

⁷Dataset from the study by Yee et al. (2020)

⁸Dataset from the study by Edgerton et al. (2011)

3.2 Procedures

In order to train the different approaches we considered in our study, basically the datasets were preprocessed to represent two-dimensional dynamics, namely, tumor growth or drug concentration evolution and irregular time evolution, respectively. Each of the two time series was directly encoded in neural distributed neural populations for the work by Cook et al., Weber et al., and our approach, whereas the approaches of Mandal et al. and Champion et al. fused the time series in a single input vector. For the training part, the work by Mandal et al. used alternating conditional expectation algorithm to calculate optimal transformations by fast boxcar averaging the rank-ordered data, whereas the Champion et al. approach used backpropagation. The neurally inspired approaches in Cook et al., Weber et al., and our system used HL, Sigma-Pi (Sum-Product) learning, and a combination of competition and cooperation for correlation learning, respectively. Finally, for inference, we used the systems resulting from the training phase (without modification) for one pass (forward pass) of unseen data through the system (i.e., basically accounting to a series of linear algebra operations).

Our system in all of our experiments, data depicting tumor growth, pharmacokinetics, and chemotherapy regimens are fed to our system, which encodes each time series in the SOMs and learns the underlying relations in the Hebbian matrix. The SOMs are responsible for bringing the time series in the same latent representation space where they can interact (i.e., through their internal correlation). Throughout the experiments, each of the SOM has $N = 100$ neurons, the Hebbian connection matrix has size $N \times N$, and parametrization is done as follows: $\alpha = [0.01, 0.1]$ decaying, $\eta = 0.9$, $\sigma = \frac{N}{2}$ decaying following an inverse time law. The training procedure of our system follows the next steps:

- normalize the input dataset;
- set up condition to reach relaxed state (i.e., no more fluctuations $[\Delta\epsilon]$ in the SOM neural activation and Hebbian matrix);

- for each new data item, go through the pairs of neural populations (i.e., SOMs) and compute activation;
- for cross-connection among SOMs compute the Hebbian matrix entries;
- after convergence (i.e., reached $\Delta\epsilon$), the system comprises the learned relation encoded in the matrix;

The testing procedure of our system follows the next steps:

- decode the encoded relation from the Hebbian matrix;
- denormalize data to match the original input space;
- compare with ground truth.

An important aspect is that for our system, after convergence (i.e., reaching an $\Delta\epsilon$ of changes in weights), the content of the Hebbian matrix is decoded. This amounts to a process in which the (now) static layout of values in the matrix actually depicts the underlying function $y = f(x)$. Our system is basically updating the weights and shapes of the tuning curves (i.e., preferred values) of the SOMs and the cross-SOM Hebbian weights in the training process. After training, for inference and testing, the decoded function (i.e., using Brent's derivative-free optimization method) accounts for a typical regression neural network for which cross-validation is applied. More precisely, we ran a fourfold cross-validation for each dataset.

Cook et al. For the neural network system proposed by Cook et al. (2010), in all our experiments, we used neural populations with 200 neurons each, a 0.001 WTA settling threshold, 0.005 scaling factor in homeostatic activity regulation, 0.4 amplitude target for homeostatic activity regulation, and 250 training epochs. More details and the reference codebase are available on GitLab.

Weber et al. For the neural network system proposed by Weber and Wermter (2007), in all our experiments, we used a network with 15 neurons, 0.001 learning rate, 200,000 training epochs, and unit normalization factor. The fully parametrized codebase is available, along the other systems reference implementations, on GitLab.

Mandal et al. For the CCA-based system proposed by Mandal and Cichocki (2013), in all our comparative experiments, we used a sample size of 100, replication factor 10, 0.5 divergence factor, 1,000 variable permutations, and 1.06 bandwidth for Gaussian kernel density estimate. The full codebase is provided, along the other systems reference implementations, on GitLab.

Champion et al. For our DL implementation, we used the system of Champion et al. (2019) as a reference. We then modified the structure to accommodate the peculiarities of the clinical data. In all the experiments, the DL system contained hidden layers of size 128 neurons, trained for 100 epochs, with a mini-batch size of 1, and 50% augmentation percentage. The full codebase is provided, along the other systems reference implementations, on GitLab. Another important implementation aspect is that we use a combination of SVM and DL approaches. While SVM can work with a limited amount of data, DL models tend to perform worse when big data are not available. Therefore, we test multiple approaches to artificially augment the training data:

- **DL with no augmentation, DL.** We train the model directly from the data without further transformations.
- **DL with SVM augmentation, DL + SVM.** We used the SVM model trained beforehand to enhance the data. We set a number of observations that we want to enhance and generate random timestamps we use for prediction using SVM. Then we add those artificial values as new observations for training.
- **DL with SVM augmentation and random noise, DL + SVM + noise.** We follow the same process as in SVM augmentation, but before adding the predictions to the training pool, we add normal noise.

For the SVM we use one input feature, the days passed, and one output feature, the size of the tumor. For DL, we use the gated recurrent units (GRUs; Chung et al. (2014)) as building blocks to design a structure inspired by the work of Champion et al. (2019). The architecture consists on one GRU layer, one ReLU activation, a fully connected layer, and another ReLU activation. We designed a simple architecture to better suit the model to the scarce availability of data inspired by the study by Berg and Nyström (2019). As the DL model is a recurrent model, our input data consist of all data available from a certain patient up to a point. Both models normalize the data (both days and tumor size) by dividing by the maximum value observed. For consistency across methods, we run a fourfold cross-validation for each dataset (except dataset 0, which has only two samples; therefore, we run a twofold cross-validation). We present the average results over the cross-validation. The complete parametrization and implementation are available on GitLab.

3.3 Results

As previously mentioned, we evaluate the systems on a series of instantiations depicting various decision support tasks relevant for clinical use. All of the five models were evaluated through multiple metrics (Table 2) on each of the four cell line datasets. In order to evaluate the distribution of the measurement error as a

TABLE 2 | Evaluation metrics for data-driven relation learning systems. We consider N —number of measurements, σ —standard deviation of data, p —number of parameters of the model.

Metric	Equation
SSE	$\sum_{i=1}^N \left(\frac{y^i - y_m^i}{\sigma} \right)^2$
RMSE	$\sqrt{\frac{SSE}{N-p}}$
sMAPE	$\frac{1}{N} \sum_{i=1}^N \left(2 \frac{ y^i - y_m^i }{(y^i + y_m^i)} \right)$
AIC	$N \ln \left(\frac{SSE}{N} \right) + 2p$
BIC	$N \ln \left(\frac{SSE}{N} \right) + \ln(N)p$

function of the measured volumes of the tumors, the work of Benzekry et al. (2014) recommended the following model for the standard deviation of the error σ_i at each measurement time point i ,

$$\sigma_i = \begin{cases} \sigma (y_m^i)^\alpha, & \text{if } y_m^i \geq y^i \\ \sigma (y^i)^\alpha, & \text{if } y_m^i < y^i \end{cases}$$

This model shows that when overestimating ($y_m \geq y$), the measurement error α is subproportional, and when underestimating ($y_m < y$), the obtained error is the same as the measured data points. In our experiments, we consider $\alpha = 0.84$ and $\sigma = 0.21$ as a good trade-off of error penalty and enhancement. We use this measurement error formulation to calculate the typical performance indices (i.e., sum of squared errors [SSE], root mean squared error [RMSE], symmetric mean absolute percentage error [sMAPE]) and goodness-of-fit and parsimony (i.e., Akaike information criterion [AIC] and Bayesian information criterion [BIC]), as shown in Table 2.

3.3.1 Learning Growth Patterns of Preinvasive Breast Cancer

Analyzing tumor infiltration patterns, clinicians can evaluate the evolution of neoplastic processes, for instance, from DCIS to breast cancer. Such an analysis can provide very important benefits, in early detection, in order to (1) increase patient survival, (2) decrease the likelihood for multiple surgeries, and (3) determine the choice of adjuvant versus neoadjuvant chemotherapy. For a full analysis and in-depth discussion of our system’s capabilities for such a task, refer to Axenie and Kurz (2020b). For this task, we assessed the capability of the evaluated systems to learn the dependency between histopathologic and morphological data. We fed the systems with DCIS data from Edgerton et al. (2011), namely, time series of nutrient diffusion penetration length within the breast tissue (L), ratio of cell apoptosis to proliferation rates (A), and radius of the breast tumor (R). The study by Edgerton et al. (2011) postulated that the value of R depends on A and L following a “master equation” Eq. 9

$$A = 3 \frac{L}{R} \left(\frac{1}{\tanh\left(\frac{R}{L}\right)} - \frac{L}{R} \right) \tag{9}$$

whose predictions are consistent with nearly 80% of *in situ* tumors identified by mammographic screenings. For this

TABLE 3 | Evaluation of the data-driven relation learning systems.

Dataset/system	Evaluation metrics		
	SSE	RMSE	sMAPE
Breast (DCIS), Edgerton et al. (2011)			
Cook et al.	56.321	0.4867	0.5901
Weber et al.	59.879	0.5099	0.6512
Mandal et al.	62.346	0.5617	0.6800
Champion et al.	58.645	0.4721	0.6054
Our system	54.216	0.4656	0.5734

initial evaluation of the data-driven mathematical relations learning systems, we consider three typical performance metrics (i.e., SSE, RMSE, and sMAPE, respectively) against the experimental data (i.e., ground truth and Eq. 9):

As one can see in Table 3, our system overcomes the other approaches on predicting the nonlinear dependency between radius of the breast tumor (R) given the nutrient diffusion penetration length within the breast tissue (L) and ratio of cell apoptosis to proliferation rates (A) from real *in vivo* histopathologic and morphological data.

3.3.2 Learning Unperturbed Tumor Growth Curves Within and Between Cancer Types

In the second task, we evaluated the systems on learning unperturbed (i.e., growth without treatment) tumor growth curves. The choice of different cancer types (i.e., two breast cell lines, lung, and leukemia) is to probe and demonstrate between- and within-tumor-type prediction versatility.

Our system provides overall better accuracy between- and within-tumor-type growth curve prediction, as shown in Table 4

TABLE 4 | Evaluation of the data-driven relation learning systems on tumor growth curve extraction.

Dataset/system	Evaluation metrics (smaller value is better)					
	SSE	RMSE	sMAPE	AIC	BIC	Rank3
Breast4 cancer						
Cook et al.	7,009.6	37.4423	1.7088	52.3639	52.2557	2
Weber et al.	8,004.9	44.7350	1.7088	55.2933	55.1310	5
Mandal et al.	7,971.8	39.9294	1.7088	53.2643	53.1561	4
Champion et al.	6,639.1	40.7403	1.4855	53.9837	53.8215	3
Our system	119.3	4.1285	0.0768	19.8508	19.8508	1
Breast5 cancer						
Cook et al.	0.2936	0.1713	0.1437	-40.5269	-39.5571	4
Weber et al.	0.2315	0.1604	0.1437	-41.3780	-39.9233	2
Mandal et al.	0.3175	0.1782	0.1437	-39.5853	-38.6155	5
Champion et al.	0.2699	0.1732	0.1512	-39.5351	-38.0804	3
Our system	0.0977	0.0902	0.0763	-57.7261	-57.7261	1
Breast6 cancer						
Cook et al.	3.0007	0.7071	1.0606	50.1322	51.2887	2
Weber et al.	3.2942	0.8116	1.6626	56.4133	55.1915	5
Mandal et al.	3.1908	0.7292	1.3506	53.2643	52.5421	4
Champion et al.	3.4772	0.8339	1.1288	53.9837	53.7775	3
Our system	0.7668	0.3096	0.2615	19.3208	19.1298	1
Breast7 cancer						
Cook et al.	45.6031	2.3875	1.2216	-40.0084	-39.9975	4
Weber et al.	56.0738	2.8302	1.8346	-41.2345	-39.1234	2
Mandal et al.	53.2428	2.5797	1.4816	-39.5853	-37.1260	5
Champion et al.	54.7189	2.7958	1.5086	-39.1234	-38.0664	3
Our system	0.2008	0.1417	0.0364	-57.1221	-57.6112	1
Lung cancer						
Cook et al.	44.5261	2.2243	1.5684	19.3800	20.1758	2
Weber et al.	54.1147	2.6008	1.5684	23.5253	24.7190	5
Mandal et al.	53.2475	2.4324	1.5684	21.3476	22.1434	4
Champion et al.	50.6671	2.5166	1.5361	22.8012	23.9949	3
Our system	3.6903	0.5792	0.2121	-12.0140	-12.0140	1
Leukemia						
Cook et al.	223.7271	3.2640	1.6368	56.3235	58.5944	2
Weber et al.	273.6770	3.6992	1.6368	62.9585	66.3649	5
Mandal et al.	259.9277	3.5182	1.6368	59.7729	62.0439	4
Champion et al.	248.5784	3.5255	1.6001	60.7461	64.1526	3
Our system	35.2541	1.2381	0.3232	9.8230	9.8230	1

Notes: 3—Calculated as best in 3/5 metrics; 4—MDA-MB-231 cell line; 5—MDA-MB-435 cell line; 6—MCF-7, T47D cell line; 7—LM2-4LUC + cell line.

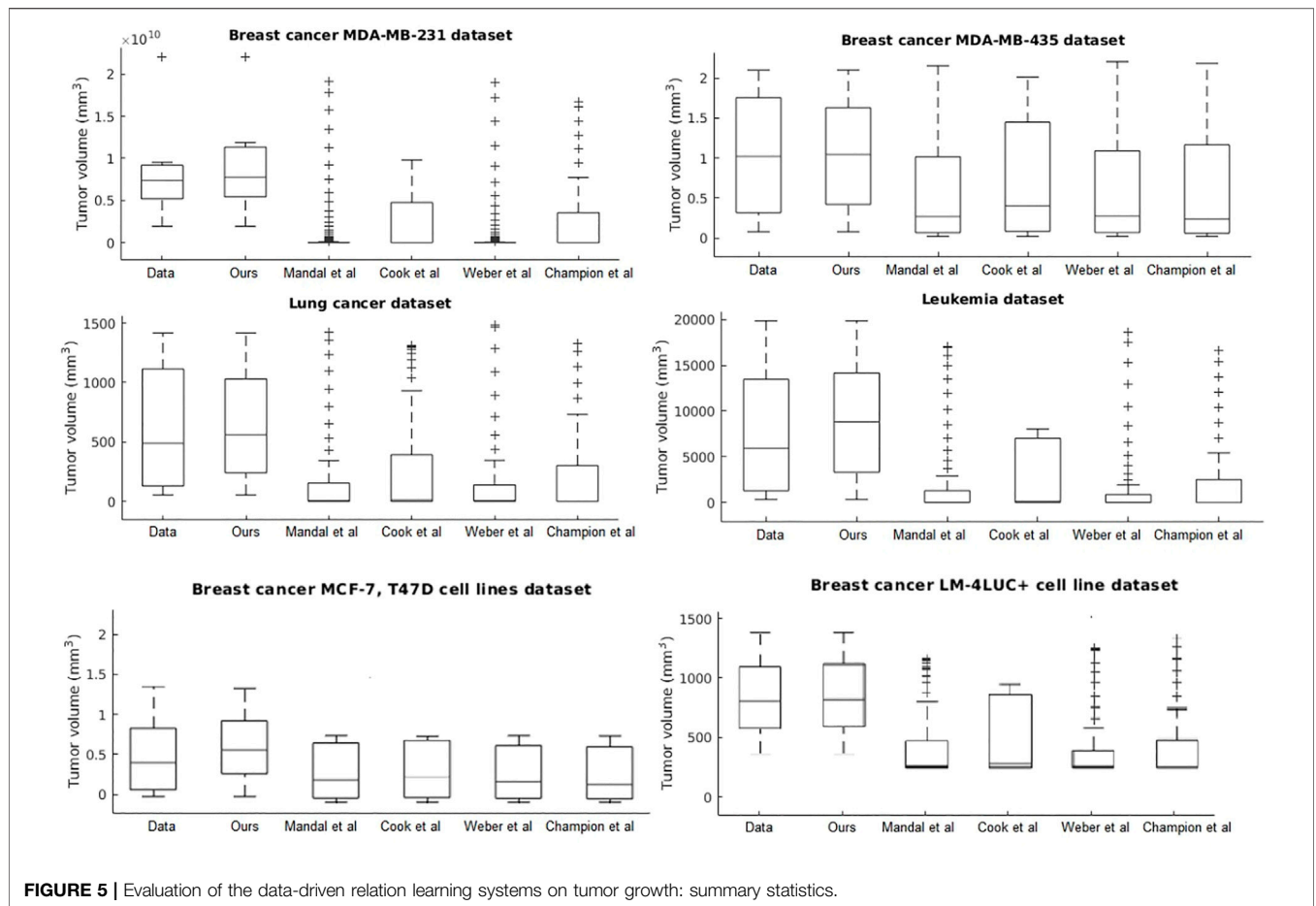


FIGURE 5 | Evaluation of the data-driven relation learning systems on tumor growth: summary statistics.

and the summary statistics (depicted in **Figure 5**). The superior performance is given by the fact that our system can overcome the other approaches when facing incomplete biological descriptions, the diversity of tumor types, and the small size of the data. Interested readers can refer to Axenie and Kurz (2021) for a deeper performance analysis of our system.

3.3.3 Extracting Tumor Phenotypic Stage Transitions

The next evaluation task looks at learning the mathematical relations describing the phenotypic transitions of tumors in breast cancer. For this experiment, we considered the study of 17 breast cancer patients in the study by Edgerton et al. (2011). Typically, in the breast cancer phenotypic state space, quiescent cancer cells (Q) can become proliferative (P) or apoptotic (A). In addition, nonnecrotic cells become hypoxic if the oxygen supply drops below a threshold value. But, hypoxic cells can recover to their previous state or become necrotic, as shown by Macklin et al. (2012).

In this instantiation, we focus on a simplified three-state phenotypic model (i.e., containing P, Q, A states). The transitions among tumor states are stochastic events generated by Poisson processes. Each of the data-driven relation learning systems is fed with time series of raw immunohistochemistry and morphometric data for each of the 17 tumor cases (Edgerton et al., 2011; **Supplementary Tables S1, S2**) as follows: cell cycle

time τ_P , cell apoptosis time τ_A , proliferation index PI , and apoptosis index AI . Given this time series input, each system needs to infer the mathematical relations for α_P , the mean quiescent-to-proliferation (Q-P) transition rate, and α_A , the quiescent-to-apoptosis (Q-A) transition rate, respectively (**Figure 6**). Their analytical form state transition is given by:

$$\alpha_P = \frac{\frac{1}{\tau_P} (PI + PI^2) - \frac{1}{\tau_A} AIPI}{1 - AI - PI}, \alpha_A = \frac{\frac{1}{\tau_A} (AI - AI^2) + \frac{1}{\tau_P} AIPI}{1 - AI - PI} \tag{10}$$

Q-A and Q-P state transitions of cancer cells are depicted in **Figure 6**, where we also present the relation that our system learned. Both in **Figure 6** and **Table 5**, we can see that our system is able to recover the correct underlying mathematical function with respect to ground truth (clinically extracted and modeled **Eq. 10** from the study by Macklin et al. (2012)).

3.3.4 Simultaneously Extracting Drug-Perturbed Tumor Growth and Drug Pharmacokinetics

Chemotherapy use in the neoadjuvant and adjuvant settings generally provides the same long-term outcome (de Wiel et al., 2017). But what is the best choice for a particular patient? This question points at those quantifiable patient-specific factors (e.g.,

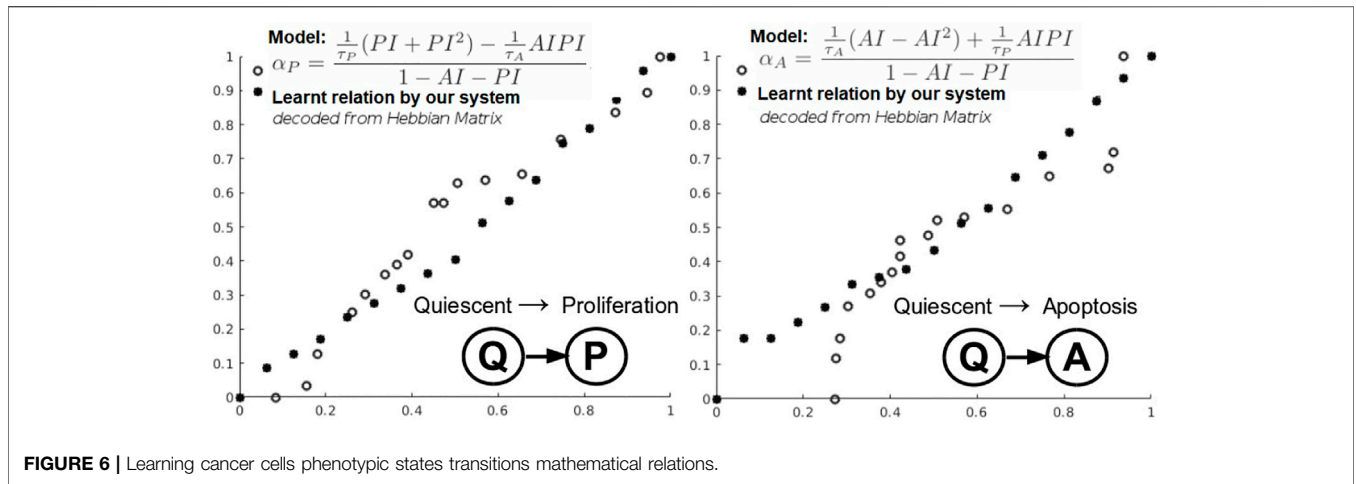


FIGURE 6 | Learning cancer cells phenotypic states transitions mathematical relations.

TABLE 5 | Evaluation of the data-driven relation learning systems for extracting phenotypic transitions relations.

State transition/system	Evaluation metrics		
	SSE	RMSE	sMAPE
Quiescent(Q) to proliferation(P) transition relation			
Cook et al.	0.820	0.240	0.190
Weber et al.	0.865	0.294	0.196
Mandal et al.	0.904	0.320	0.214
Champion et al.	0.845	0.274	0.189
Our system	0.750	0.210	0.172
Quiescent(Q) to apoptosis(A) transition relation			
Cook et al.	0.421	0.162	0.140
Weber et al.	0.484	0.178	0.154
Mandal et al.	0.490	0.182	0.151
Champion et al.	0.441	0.166	0.147
Our system	0.398	0.153	0.131

Note that none of the evaluated system had prior knowledge of the data distribution or biological assumptions. To have a more detailed overview on the capabilities of our system to capture phenotypic dynamics, refer to Axenie and Kurz (2020c).

tumor growth curve under chemotherapy, drug pharmacokinetics) that influence the sequencing of chemotherapy and surgery in a therapy plan. A large variety of breast cancer tumor growth patterns used in cancer treatments planning were identified experimentally and clinically and modeled over the years (Gerlee, 2013). In addition, progress in pharmacokinetic modeling allowed clinicians to investigate the effect of covariates in drug administration, as shown in the work by Zaheed et al. (2019). Considering breast cancer, paclitaxel is a typical drug choice with broad use in monotherapy as well as immune-combined therapies (Stage et al., 2018).

In the current section, we present the experimental results of all the evaluated systems and consider (1) accuracy in learning the chemotherapy-perturbed tumor growth model and (2) accuracy in learning the pharmacokinetics of the chemotoxic drug (i.e., paclitaxel) dose. For the tumor growth curve extraction, we considered four cell lines of breast cancer (i.e., MDA-MB-231, MDA-MB-435, MCF-7, LM2-LUC + cell lines; Table1). The evaluation results of the systems in the perturbed tumor

growth scenario are provided in Figure 7. Note that our system learns the temporal relationships among the quantities fed to the two sides of the system (Figure 3), which can, subsequently, be used to infer one (unavailable) quantity based on the one available. For instance, if the system had learned the change in volume at irregular time points, given a next time point, the system will recover the most plausible volume value—basically accounting for a one-step-ahead prediction. For a longer prediction horizon, one can recurrently apply this process for new predictions and so on.

Table 6 presents the results using SVM and the different versions of DL. We can see that usually vanilla DL outperforms SVM. DL is a more complex model, as well as uses more input data, so this result is expected. Once we add the augmentation from SVM, the model has a comparable performance to SVM. Our theory is that this is caused by DL learning to imitate SVM instead of real data. Once we add noise to the augmentation, the data become more realistic and usually yield improvements in performance.

For the pharmacokinetics learning experiments, we used the data from the computational model of intracellular pharmacokinetics of paclitaxel of Kuh et al. (2000) describing the kinetics of paclitaxel uptake, binding, and efflux from cancer cells in both intracellular and extracellular contexts.

As one can see in Figure 8A, the intracellular concentration kinetics of paclitaxel is highly nonlinear. Our system is able to extract the underlying function describing the data without any assumption about the data and other prior information, opposite to the model from (Kuh et al., 2000). Interestingly, our system captured a relevant effect consistent with multiple paclitaxel studies (Stage et al., 2018), namely, that the intracellular concentration increased with time and approached plateau levels, with the longest time to reach plateau levels at the lowest extracellular concentration—as shown in Figure 8.

Analyzing the extracellular concentration in Figure 8B, we can see that our system extracted the trend and the individual variation of drug concentration after the administration of the drug (i.e., in the first 6 h) and learned an accurate fit without any prior or other biological assumptions. Interestingly, our system captured the fact that the intracellular drug concentration

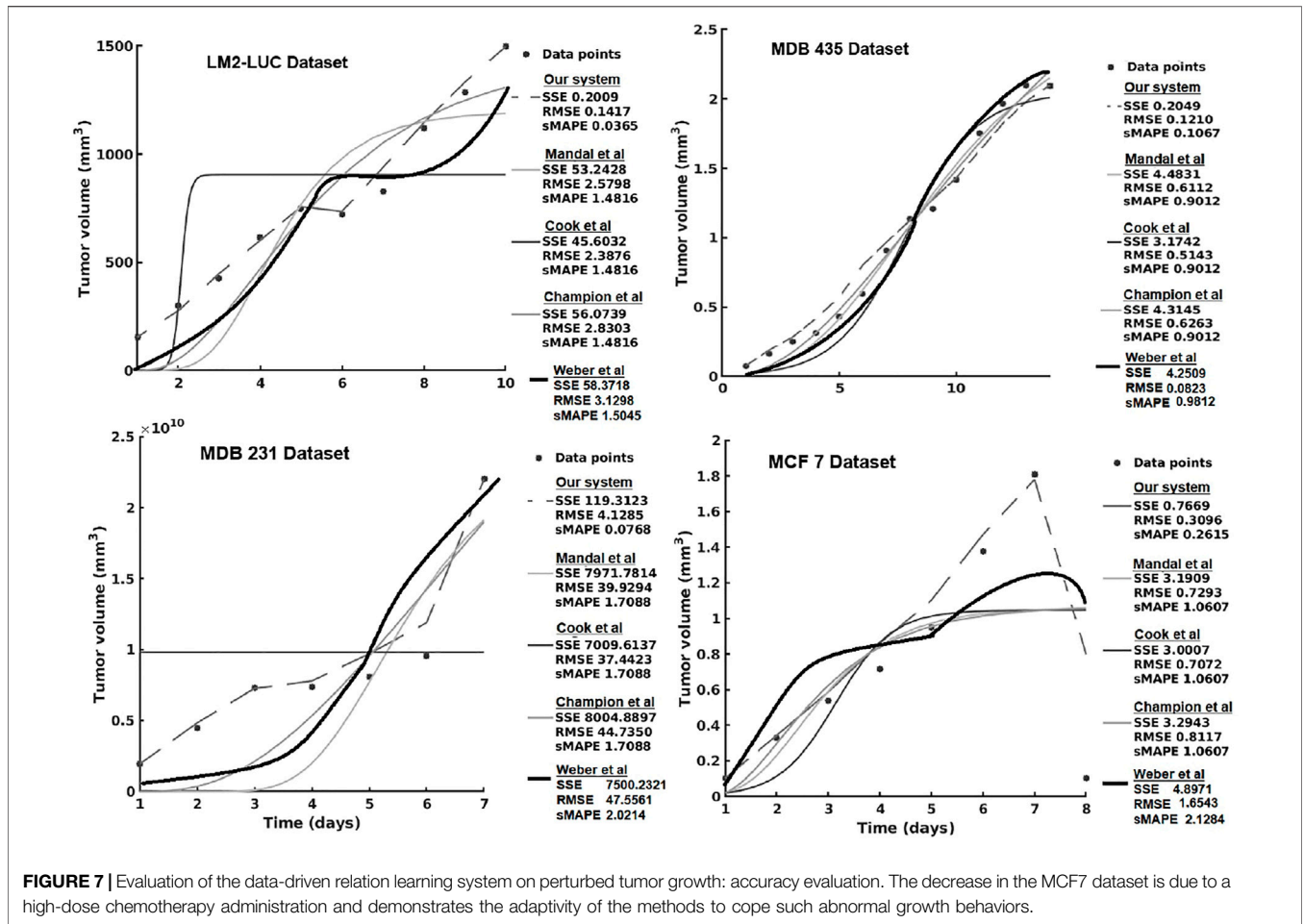


FIGURE 7 | Evaluation of the data-driven relation learning system on perturbed tumor growth: accuracy evaluation. The decrease in the MCF7 dataset is due to a high-dose chemotherapy administration and demonstrates the adaptivity of the methods to cope such abnormal growth behaviors.

TABLE 6 | Description of the DL approach inspired by Champion et al. (2019).

Evaluation of the deep learning approach					
Dataset	Cancer (cell line)	RMSE ^{SVM}	RMSE ^{DL}	RMSE ^{DL+SVM}	RMSE ^{DL+SVM+noise}
1	Breast (MDA-MB-231)	1.8424	1.5382	1.7544	1.7088
2	Breast (MDA-MB-435)	1.0977	1.5990	0.9584	0.9012
3	Breast (MCF-7)	1.4112	1.7295	1.3632	1.0607
4	Breast (LM2-4LUC+)	1.8945	1.8345	1.7620	1.4816

Note that for all of the evaluation datasets, the best performing DL approach [i.e., inspired by Champion et al. (2019)] is the combined DL—SVM—noise configuration.

increased linearly with extracellular concentration decrease, as shown in **Figure 8**.

The overall evaluation of pharmacokinetics learning is given in **Table 7**.

In this series of experiments, all of the systems learned that changes in cell number were represented by changes in volume, which (1) increased with time at low initial total extracellular drug concentrations due to continued cell proliferation and (2) decreased with time at high initial total extracellular drug concentrations due to the antiproliferative and/or cytotoxic drug effects, as reported by Kuh et al. (2000). In order to assess the impact the predictions have on therapy sequencing (i.e., neoadjuvant vs. adjuvant chemotherapy), refer to Axenie and Kurz (2020a).

3.3.5 Predicting Tumor Growth/Recession Under Chemotherapy

In the last series of experiments, we used real patient data from the I-SPY 1 TRIAL: ACRIN 6657 (Yee et al., 2020). Data for the 136 patients treated for breast cancer in the I-SPY-1 clinical trial were obtained from the cancer imaging archive³ and the Breast Imaging Research Program at UCSF. The time series data contained only the largest tumor volume from magnetic resonance imaging measured before therapy, 1 to 3 days after therapy, between therapy cycles, and before surgery,

³<https://wiki.cancerimagingarchive.net/display/Public/ISPY1>.

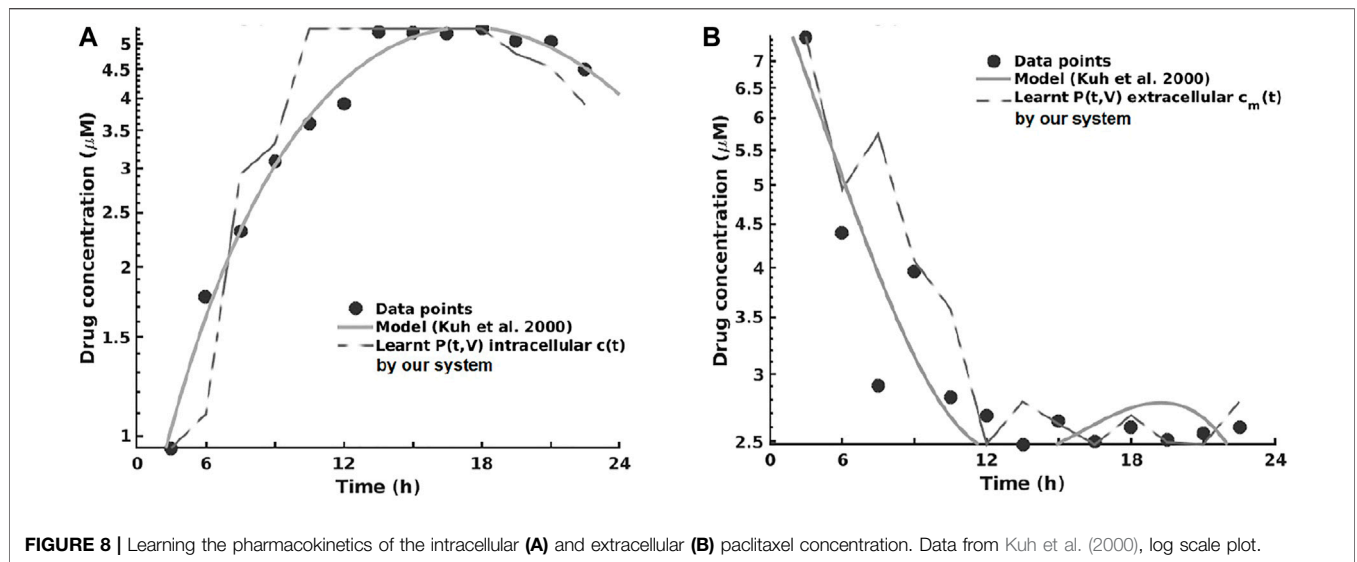


TABLE 7 | Evaluation of the data-driven relation learning systems for pharmacokinetics extraction.

Pharmacokinetics data/system	Evaluation metrics		
	SSE	RMSE	sMAPE
Intracellular paclitaxel			
Cook et al.	0.6234	0.2812	0.1487
Weber et al.	0.7212	0.3689	0.1794
Mandal et al.	0.6743	0.3046	0.1602
Champion et al.	0.6539	0.2607	0.1500
Our system	0.5960	0.2141	0.1403
Extracellular paclitaxel			
Cook et al.	0.5676	0.2341	0.1213
Weber et al.	0.6674	0.2891	0.1289
Mandal et al.	0.6128	0.2974	0.1366
Champion et al.	0.5790	0.2633	0.1156
Our system	0.5484	0.2054	0.1068

respectively. To summarize, the properties of the dataset are depicted in **Figure 9**.

As we can observe in **Table 8**, our system learns a superior fit to the tumor growth data, with respect to the other systems, despite the limited number of samples (i.e., 7 data points for MDA-MD-231 cell line dataset and up to 14 data points for MDA-MD-435 cell line dataset). It is important to note that when analyzing tumor growth functions and response under chemotherapy, we faced the high variability among patients given by the typical constellation of hormone receptor indicators (i.e., HR and HER2neu, which covered the full spectrum of positive and negative values) for positive and negative prognoses. All data-driven learning systems capture such aspects to some extent. Our system learns a superior fit overall the three metrics, capturing the intrinsic impact chemotherapy has upon the tumor growth function, despite the limited number of samples (i.e., 4 data points of the dataset overall evaluation dataset of 20% of patients). An extended evaluation of our system on a broader set of datasets for therapy outcome prediction is given by Kurz and Axenie (2020).

4 DISCUSSION

We complement the quantitative evaluation in the previous section with an analysis of the most important features of all the systems capable to extract mathematical relations in the aforementioned clinical oncology tasks. As the performance evaluation was done in the previous section, we will now focus on other specific comparison terms relevant for the adoption of such systems in clinical practice.

One initial aspect is the design and functionality. Using either distributed representations (Cook et al., 2010; Weber and Wermter, 2007; Champion et al., 2019) or compact mathematical forms Mandal and Cichocki (2013), all methods encoded the input variables in a new representation to facilitate computation. At this level, using neural network dynamics (Cook et al., 2010; Weber and Wermter, 2007) or pure mathematical multivariate optimization (Mandal and Cichocki, 2013; Champion et al., 2019), the solution was obtained through iterative processes that converged to consistent representations of the data. Our system employs a lightweight learning mechanism, offering a transparent processing scheme and human-understandable representation of the learned relations as shown in **Figure 4**. Besides the capability to extract the correlation among the two features, the system can simultaneously extract the shape of the distribution of the feature spaces. This is an important feature when working with rather limited medical data samples.

A second aspect refers to the amount of prior information embedded by the designer in the system. It is typical that, depending on the instantiation, a new set of parameters is needed, making the models less flexible. Although less intuitive, the pure mathematical approaches (Mandal and Cichocki, 2013) (i.e., using CCA) need less tuning effort due to the fact that their parameters are the result of an optimization procedure. On the other side, the neural network approaches (Cook et al., 2010; Weber and Wermter, 2007; Champion et al., 2019) need a more judicious parameter tuning, as their dynamics

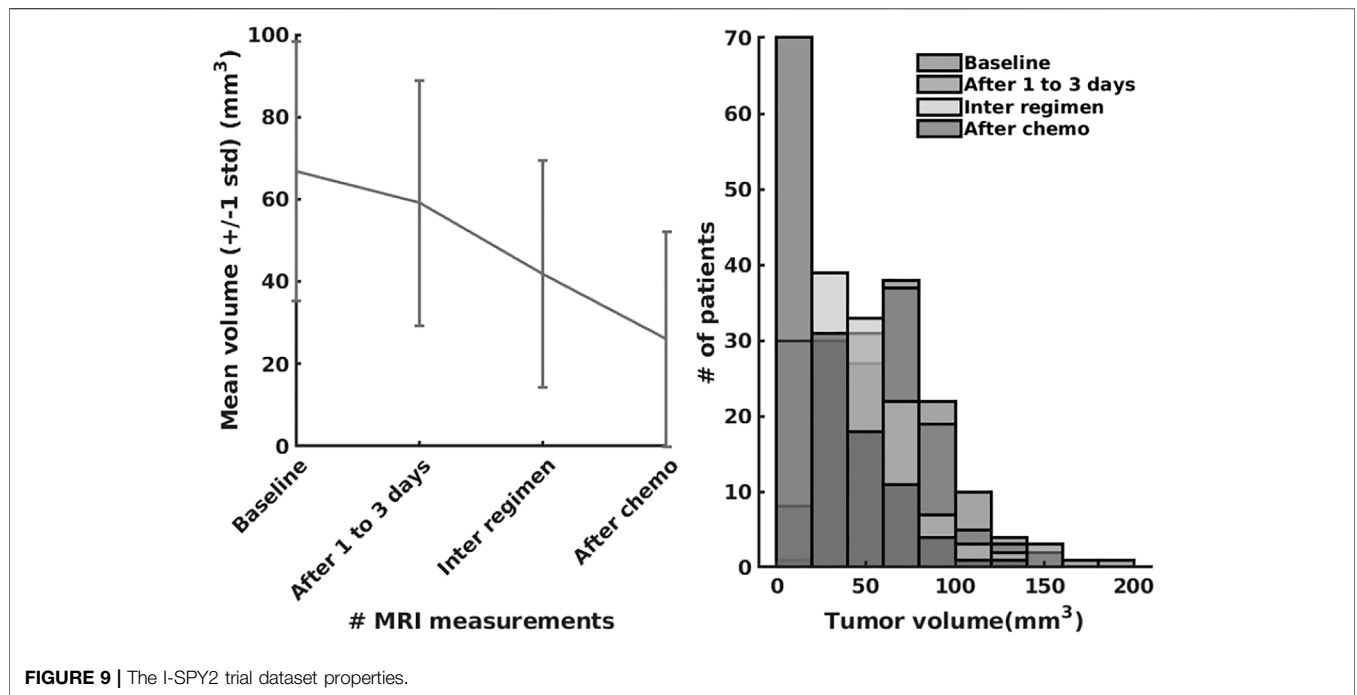


FIGURE 9 | The I-SPY2 trial dataset properties.

TABLE 8 | Evaluation of the data-driven relation learning systems on real patient breast cancer data.

Dataset/model	Evaluation metrics		
	SSE	RMSE	sMAPE
I-SPY2 trial Yee et al. (2020)			
Cook et al.	1.3735	1.1439	0.1133
Weber et al.	1.7543	1.2005	0.2539
Mandal et al.	2.963	1.0963	0.7834
Champion et al.	2.0747	1.04100	0.1073
Our system	0.8650	0.4650	0.0389

are more sensitive and can reach either instability (e.g., recurrent networks) or local minima. Except parametrization, prior information about inputs is generally needed when instantiating the system for a certain scenario. Sensory value bounds and probability distributions must be explicitly encoded in the models through explicit distribution of the input space across neurons in the studies by Cook et al. (2010) and Weber and Wermter (2007), linear coefficients in vector combinations (Mandal and Cichocki, 2013), or standardization routines of input variables (Champion et al., 2019). Our system exploits only the available data to simultaneously extract the data distribution and the underlying mathematical relation governing tumor growth processes. Capable of embedding priors (i.e., mechanistic models) in its structure, our system can speed up its computation, through a data-driven model refinement similar in nature with the unsupervised learning process. Basically, in order to combine the learning process with a mechanistic model, the only update will be done in the factorization of the weight update in Eq. 6.

A third aspect relevant to the analysis is the stability and robustness of the obtained representation. The representation of the hidden relation (1) can be encoded in a weight matrix Cook et al. (2010) and Weber and Wermter (2007) such that, after learning, given new input, the representation is continuously refined to accommodate new inputs; (2) can be fixed in vector directions of random variables requiring a new iterative algorithm run from initial conditions to accommodate new input (Mandal and Cichocki, 2013); or (3) can be obtained as an optimization process given the new available input signals (Champion et al., 2019). Given initial conditions, prior knowledge and an optimization criteria (Mandal and Cichocki, 2013) or a recurrent relaxation process toward a point attractor (Cook et al., 2010; Weber and Wermter, 2007; Champion et al., 2019) are required to reach a desired tolerance. Our system exploits the temporal regularities among tumor growth data covariates, to learn the governing relations using a robust distributed representation of each data quantity. The choice of a distributed representation to encode and process the input data gives out the system an advantage in terms of explainability for clinical adoption. As shown in Figure 3, each scalar quantity can be projected in a high dimension where the shape of the distribution can be inferred. Such insights can support the decisions of the system by explaining its predictions.

The capability to handle noisy data is an important aspect concerning the applicability in real-world scenarios. Using either computational mechanisms for denoising (Cook et al., 2010; Weber and Wermter, 2007), iterative updates to minimize a distance metric Mandal and Cichocki (2013), or optimization Champion et al. (2019), each method is capable to cope with moderate amounts of noise. Despite this, some methods have intrinsic methods to cope with noisy data intrinsically, through their dynamics, by recurrently

propagating correct estimates and balancing new samples (Cook et al., 2010). The distributed representation used in our system ensures that the system is robust to noise, and the local learning rules ensure fast convergence on real-world data—as our experiments demonstrated.

Another relevant feature is the capability to infer (i.e., predict/anticipate) missing quantities once the mathematical relation is learned. The capability to use the learned relations to determine missing quantities is not available in all presented systems, such as the system of Mandal and Cichocki (2013). This is due to the fact that the divergence and correlation coefficient expressions might be noninvertible functions that support a simple pass-through of available values to extract missing ones. On the other side, using either the learned co-activation weight matrix (Cook et al., 2010; Weber and Wermter, 2007) or the known standard deviations of the canonical variants (Champion et al., 2019), some systems are able to predict missing quantities. Our system stores learned mathematical relations in the Hebbian matrix, which can be used bidirectionally to recover missing quantities on one side of the input given the other available quantity. This feature is crucial for the predictive aspects of our system. Basically, in its typical operation, the system learns from sets of observations the underlying relations among quantities describing the tumor's state (e.g., growth curve, phenotypic stage, extracellular drug concentration). For prediction purposes, the system is fed with only one quantity (e.g., time index) and, given the learned relation, will recover the most plausible value for the correlated quantity that was trained with (e.g., growth curve) for the next step.

Finally, because of the fact that all methods reencode the real-world values in new representation, it is important to study the capability to decode the learned representation and subsequently measure the precision of the learned representation. Although not explicitly treated in the presented systems, decoding the extracted representations is not trivial. Using a tiled mapping of the input values along the neural network representations, the system of Cook et al. (2010) decoded the encoded value in activity patterns by simply computing the distribution of the input space over the neural population units, whereas Weber and Wermter (2007) used a simple WTA readout, given that the representation was constrained to have a uniquely defined mapping. Given that the model learns the relations in data space through optimization processes, as in the system of Champion et al. (2019), one can use learned curves to simply project available sensory values through the learned function to get the second value, as the scale is preserved. Albeit its capability to precisely extract nonlinear relations from high-dimensional random datasets, the system of Mandal and Cichocki (2013) cannot provide any readout mechanisms to support a proper decoded representation of the extracted relations. This is due to the fact that the method cannot recover the sign and scale of the relations. The human-understandable relation learned by our system is efficiently decoded from the Hebbian matrix back to real-world values. As our experiments demonstrate, the approach introduced through our system

excels in capturing the peculiarities that clinical data carry. Contributing to the explainability features of our system, the read-out mechanism is able to turn the human-understandable visual representation of the learned relation (**Figure 4**) into a function providing the most plausible values of the queried quantities.

5 CONCLUSION

Data-driven approaches to improve decision-making in clinical oncology are now going beyond diagnosis. From early detection of infiltrating tumors to unperturbed tumor growth phenotypic staging, and from pharmacokinetics-dictated therapy planning to treatment outcome, data-driven tools capable of learning hidden correlations in the data are now taking the foreground in mathematical and computational oncology. Our study introduces a novel framework and versatile system capable of learning physical and mathematical relations in heterogeneous oncology data. Together with a lightweight and transparent computational substrate, our system provides human-understandable solutions. This is achieved by capturing the distribution of the data in order to achieve superior fit and prediction capabilities between and within cancer types. Supported by an exhaustive evaluation on *in vitro* and *in vivo* data, against state-of-the-art machine learning and DL systems, the proposed system stands out as a promising candidate for clinical adoption. Mathematical and computational oncology is an emerging field where efficient, transparent, and understandable data-driven systems hold the promise of paving the way to individualized therapy. But this can only be achieved by capturing the peculiarities of a patient's tumor across scales and data types.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://gitlab.com/akii-microlab/chimera/-/tree/master/datasets>.

AUTHOR CONTRIBUTIONS

DK designed the research, collected clinical datasets, and performed data analysis of clinical studies used in the experiments. CS developed the source code for the experiments and the analysis. CA designed the research and developed the source code for the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2021.713690/full#supplementary-material>

REFERENCES

- Abler, D., Büchler, P., and Rockne, R. C. (2019). "Towards Model-Based Characterization of Biomechanical Tumor Growth Phenotypes," in *Mathematical and Computational Oncology*. Editors G. Bebis, T. Benos, K. Chen, K. Jahn, and E. Lima (Cham: Springer International Publishing), 75–86. doi:10.1007/978-3-030-35210-3_6
- Antonelli, A., Noort, W. A., Jaques, J., de Boer, B., de Jong-Korlaar, R., Brouwers-Vos, A. Z., et al. (2016). Establishing Human Leukemia Xenograft Mouse Models by Implanting Human Bone Marrow-like Scaffold-Based Niches. *Blood J. Am. Soc. Hematol.* 128, 2949–2959. doi:10.1182/blood-2016-05-719021
- Axenie, C., and Kurz, D. (2020a). "Chimera: Combining Mechanistic Models and Machine Learning for Personalized Chemotherapy and Surgery Sequencing in Breast Cancer," in *International Symposium on Mathematical and Computational Oncology* (Springer), 13–24. doi:10.1007/978-3-030-64511-3_2
- Axenie, C., and Kurz, D. (2021). "Glueck: Growth Pattern Learning for Unsupervised Extraction of Cancer Kinetics," in *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V* (Springer International Publishing), 171–186. doi:10.1007/978-3-030-67670-4_11
- Axenie, C., and Kurz, D. (2020b). "Princess: Prediction of Individual Breast Cancer Evolution to Surgical Size," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) (IEEE)*, 457–462. doi:10.1109/cbms49503.2020.00093
- Axenie, C., and Kurz, D. (2020c). "Tumor Characterization Using Unsupervised Learning of Mathematical Relations within Breast Cancer Data," in *International Conference on Artificial Neural Networks* (Springer), 838–849. doi:10.1007/978-3-030-61616-8_67
- Ben-Shmuel, A., Biber, G., and Barda-Saad, M. (2020). Unleashing Natural Killer Cells in the Tumor Microenvironment-The Next Generation of Immunotherapy?. *Front. Immunol.* 11, 275. doi:10.3389/fimmu.2020.00275
- Benzekry, S., Lamont, C., Weremowicz, J., Beheshti, A., Hlatky, L., and Hahnfeldt, P. (2019). *Tumor Growth Kinetics of Subcutaneously Implanted Lewis Lung Carcinoma Cells*. Zenodo, 3572401. doi:10.5281/zenodo.3572401
- Benzekry, S. (2020). Artificial Intelligence and Mechanistic Modeling for Clinical Decision Making in Oncology. *Clin. Pharmacol. Ther.* 108, 471–486. doi:10.1002/cpt.1951
- Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J. M. L., Hlatky, L., et al. (2014). Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth. *Plos Comput. Biol.* 10 (8), e1003800. doi:10.1371/journal.pcbi.1003800
- Berg, J., and Nyström, K. (2019). Data-driven Discovery of Pdes in Complex Datasets. *J. Comput. Phys.* 384, 239–252. doi:10.1016/j.jcp.2019.01.036
- Brent, R. P. (2013). *Algorithms for Minimization without Derivatives*. Chelmsford, United States: Courier Corporation.
- Champion, K., Lusch, B., Kutz, J. N., and Brunton, S. L. (2019). Data-driven Discovery of Coordinates and Governing Equations. *Proc. Natl. Acad. Sci. USA* 116, 22445–22451. doi:10.1073/pnas.1906995116
- Chamseddine, I. M., and Rejniak, K. A. (2020). Hybrid Modeling Frameworks of Tumor Development and Treatment. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 12, e1461. doi:10.1002/wsbm.1461
- Chen, N., Li, Y., Ye, Y., Palmisano, M., Chopra, R., and Zhou, S. (2014). Pharmacokinetics and Pharmacodynamics of Nab -paclitaxel in Patients with Solid Tumors: Disposition Kinetics and Pharmacology Distinct from Solvent-based Paclitaxel. *J. Clin. Pharmacol.* 54, 1097–1107. doi:10.1002/jcph.304
- Chen, Z., Haykin, S., Eggermont, J. J., and Becker, S. (2008). *Correlative Learning: A Basis for Brain and Adaptive Systems*. Vol. 49. John Wiley & Sons.
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv preprint arXiv:1412.3555.
- Comen, E., Gilewski, T. A., and Norton, L. (2016). *Tumor Growth Kinetics*. Hoboken, United States: Holland-Frei Cancer Medicine, 1–11.
- Cook, M., Jug, F., Krautz, C., and Steger, A. (2010). "Unsupervised Learning of Relations," in *International Conference on Artificial Neural Networks* (Springer), 164–173. doi:10.1007/978-3-642-15819-3_21
- Cornish, A. J., and Markowetz, F. (2014). Santa: Quantifying the Functional Content of Molecular Networks. *Plos Comput. Biol.* 10, e1003808. doi:10.1371/journal.pcbi.1003808
- Cristini, V., Koay, E., and Wang, Z. (2017). *An Introduction to Physical Oncology: How Mechanistic Mathematical Modeling Can Improve Cancer Therapy Outcomes*. Boca Raton: CRC Press.
- de Silva, B. M., Higdon, D. M., Brunton, S. L., and Kutz, J. N. (2020). Discovery of Physics from Data: Universal Laws and Discrepancies. *Front. Artif. Intell.* 3, 25. doi:10.3389/frai.2020.00025
- Edgerton, M. E., Chuang, Y.-L., Macklin, P., Yang, W., Bearer, E. L., and Cristini, V. (2011). A Novel, Patient-specific Mathematical Pathology Approach for Assessment of Surgical Volume: Application to Ductal Carcinoma In Situ of the Breast. *Anal. Cell. Pathol.* 34, 247–263. doi:10.1155/2011/803816
- Gaddy, T. D., Wu, Q., Arnheim, A. D., and Finley, S. D. (2017). Mechanistic Modeling Quantifies the Influence of Tumor Growth Kinetics on the Response to Anti-angiogenic Treatment. *Plos Comput. Biol.* 13, e1005874. doi:10.1371/journal.pcbi.1005874
- Gerlee, P. (2013). The Model Muddle: in Search of Tumor Growth Laws. *Cancer Res.* 73, 2407–2411. doi:10.1158/0008-5472.can-12-4355
- Griffon-Etienne, G., Boucher, Y., Brekken, C., Suit, H. D., and Jain, R. K. (1999). Taxane-induced Apoptosis Decompresses Blood Vessels and Lowers Interstitial Fluid Pressure in Solid Tumors: Clinical Implications. *Cancer Res.* 59, 3776–3782. doi:10.1158/0008-5472
- Jansen, T., Geleijnse, G., Van Maaren, M., Hendriks, M. P., Ten Teije, A., and Moncada-Torres, A. (2020). "Machine Learning Explainability in Breast Cancer Survival," in *30th Medical Informatics Europe Conference, MIE 2020 (Amsterdam, Netherlands: IOS Press)*, 307–311.
- Kohonen, T. (1982). Self-organized Formation of Topologically Correct Feature Maps. *Biol. Cybern.* 43, 59–69. doi:10.1007/bf00337288
- Kondylakis, H., Axenie, C., Bastola, D., Katehakis, D. G., Kouroubali, A., Kurz, D., et al. (2020). Status and Recommendations of Technological and Data-Driven Innovations in Cancer Care: Focus Group Study. *J. Med. Internet Res.* 22, e22034. doi:10.2196/22034
- Kuh, H. J., Jang, S. H., Wientjes, M. G., and Au, J. L. (2000). Computational Model of Intracellular Pharmacokinetics of Paclitaxel. *J. Pharmacol. Exp. Ther.* 293, 761–770.
- Kurz, D., and Axenie, C. (2020). "Perfecto: Prediction of Extended Response and Growth Functions for Estimating Chemotherapy Outcomes in Breast Cancer," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE)*, 609–614. doi:10.1109/bibm49941.2020.9313551
- Lamy, J.-B., Sekar, B., Guezennec, G., Bouaud, J., and Séroussi, B. (2019). Explainable Artificial Intelligence for Breast Cancer: A Visual Case-Based Reasoning Approach. *Artif. intelligence Med.* 94, 42–53. doi:10.1016/j.artmed.2019.01.001
- Long, Z., Lu, Y., Ma, X., and Dong, B. (2018). "Pde-net: Learning Pdes from Data," in *International Conference on Machine Learning (Brookline: Microtome Publishing)*, 3208–3216.
- Macklin, P., Edgerton, M. E., Thompson, A. M., and Cristini, V. (2012). Patient-calibrated Agent-Based Modelling of Ductal Carcinoma In Situ (Dcis): from Microscopic Measurements to Macroscopic Predictions of Clinical Progression. *J. Theor. Biol.* 301, 122–140. doi:10.1016/j.jtbi.2012.02.002
- Mandal, A., and Cichocki, A. (2013). Non-linear Canonical Correlation Analysis Using Alpha-Beta Divergence. *Entropy* 15, 2788–2804. doi:10.3390/e15072788
- Markowetz, F., and Troyanskaya, O. G. (2007). Computational Identification of Cellular Networks and Pathways. *Mol. Biosyst.* 3, 478–482. doi:10.1039/b617014p
- Mastri, M., Tracz, A., and Ebos, J. M. (2019). Population Modeling of Tumor Growth Curves and the Reduced Gompertz Model Improve Prediction of the Age of Experimental Tumors. *PLoS Comput. Biol.* 16, e1007178. doi:10.1371/journal.pcbi.1007178
- Nathanson, S. D., and Nelson, L. (1994). Interstitial Fluid Pressure in Breast Cancer, Benign Breast Conditions, and Breast Parenchyma. *Ann. Surg. Oncol.* 1, 333–338. doi:10.1007/bf03187139
- Nia, H. T., Liu, H., Seano, G., Datta, M., Jones, D., Rahbari, N., et al. (2016). Solid Stress and Elastic Energy as Measures of Tumour Mechanopathology. *Nat. Biomed. Eng.* 1, 1–11. doi:10.1038/s41551-016-0004
- Nia, H. T., Munn, L. L., and Jain, R. K. (2020). Physical Traits of Cancer. *Science* 370, eaaz0868. doi:10.1126/science.aaz0868
- Raissi, M. (2018). Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations. *J. Machine Learn. Res.* 19, 932–955.

- Rodallec, A., Giacometti, S., Ciccolini, J., and Fanciullino, R. (2019). *Tumor Growth Kinetics of Human MDA-MB-231 Cells Transfected with dTomato Lentivirus*. Springer. doi:10.5281/zenodo.3593919
- Roland, C. L., Dineen, S. P., Lynn, K. D., Sullivan, L. A., Dellinger, M. T., Sadegh, L., et al. (2009). Inhibition of Vascular Endothelial Growth Factor Reduces Angiogenesis and Modulates Immune Cell Infiltration of Orthotopic Breast Cancer Xenografts. *Mol. Cancer Ther.* 8, 1761–1771. doi:10.1158/1535-7163.MCT-09-0280
- Rouvière, O., Melodelima, C., Hoang Dinh, A., Bratan, F., Pagnoux, G., Sanzalone, T., et al. (2017). Stiffness of Benign and Malignant Prostate Tissue Measured by Shear-Wave Elastography: a Preliminary Study. *Eur. Radiol.* 27, 1858–1866. doi:10.1007/s00330-016-4534-9
- Sarapata, E. A., and de Pillis, L. G. (2014). A Comparison and Catalog of Intrinsic Tumor Growth Models. *Bull. Math. Biol.* 76, 2010–2024. doi:10.1007/s11538-014-9986-y
- Schaeffer, H. (2017). Learning Partial Differential Equations via Data Discovery and Sparse Optimization. *Proc. R. Soc. A.* 473, 20160446. doi:10.1098/rspa.2016.0446
- Simpson-Herren, L., and Lloyd, H. H. (1970). Kinetic Parameters and Growth Curves for Experimental Tumor Systems. *Cancer Chemother. Rep.* 54, 143–174.
- Stage, T. B., Bergmann, T. K., and Kroetz, D. L. (2018). Clinical Pharmacokinetics of Paclitaxel Monotherapy: an Updated Literature Review. *Clin. Pharmacokinet.* 57, 7–19. doi:10.1007/s40262-017-0563-z
- Tan, G., Kasuya, H., Sahin, T. T., Yamamura, K., Wu, Z., Koide, Y., et al. (2015). Combination Therapy of Oncolytic Herpes Simplex Virus Hf10 and Bevacizumab against Experimental Model of Human Breast Carcinoma Xenograft. *Int. J. Cancer* 136, 1718–1730. doi:10.1002/ijc.29163
- Uzhachenko, R. V., and Shanker, A. (2019). Cd8+ T Lymphocyte and Nk Cell Network: Circuitry in the Cytotoxic Domain of Immunity. *Front. Immunol.* 10, 1906. doi:10.3389/fimmu.2019.01906
- Van de Wiel, M., Dockx, Y., Van den Wyngaert, T., Stroobants, S., Tjalma, W. A. A., and Huizing, M. T. (2017). Neoadjuvant Systemic Therapy in Breast Cancer: Challenges and Uncertainties. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 210, 144–156. doi:10.1016/j.ejogrb.2016.12.014
- Volk, L. D., Flister, M. J., Chihade, D., Desai, N., Trieu, V., and Ran, S. (2011). Synergy of Nab-Paclitaxel and Bevacizumab in Eradicating Large Orthotopic Breast Tumors and Preexisting Metastases. *Neoplasia* 13, 327–IN14. doi:10.1593/neo.101490
- Weber, C., and Wermter, S. (2007). A Self-Organizing Map of Sigma-Pi Units. *Neurocomputing* 70, 2552–2560. doi:10.1016/j.neucom.2006.05.014
- Werner, B., Scott, J. G., Sottoriva, A., Anderson, A. R. A., Traulsen, A., and Altmann, P. M. (2016). The Cancer Stem Cell Fraction in Hierarchically Organized Tumors Can Be Estimated Using Mathematical Modeling and Patient-specific Treatment Trajectories. *Cancer Res.* 76, 1705–1713. doi:10.1158/0008-5472.can-15-2069
- White, F. M., Gatenby, R. A., and Fischbach, C. (2019). The Physics of Cancer. *Cancer Res.* 79, 2107–2110. doi:10.1158/0008-5472.can-18-3937
- Yee, D., DeMichele, A. M., Yau, C., Isaacs, C., Symmans, W. F., Albain, K. S., et al. (2020). Association of Event-free and Distant Recurrence-free Survival with Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer: Three-Year Follow-Up Analysis for the I-Spy2 Adaptively Randomized Clinical Trial. Chicago: JAMA.
- Zaheed, M., Wilcken, N., Willson, M. L., O'Connell, D. L., and Goodwin, A. (2019). *Sequencing of Anthracyclines and Taxanes in Neoadjuvant and Adjuvant Therapy for Early Breast Cancer*. Hoboken: John Wiley & Sons Ltd.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kurz, Sánchez and Axenie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.