Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 22(9): 2486–2505.
DOI: 10.3934/mbe.2025091
Received: 08 May 2025
Revised: 02 July 2025
Accepted: 08 July 2025
Published: 28 July 2025

Research article

Investigating glucose-lactate metabolism in glioblastoma multiforme via universal physics-informed neural networks

Shadi Vandvajdi[1], Yuannong Mao[2], Mahla Poudineh[1] and Mohammad Kohandel[2,*]

[1] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

[2] Department of Applied Mathematics, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

* **Correspondence:** Email: kohandel@uwaterloo.ca.

Abstract: Understanding the metabolic adaptations of cancer cells is crucial for uncovering potential therapeutic targets and improving treatment strategies. In this study, we present a hybrid modeling framework that combines Physics-Informed Neural Networks (PINNs) and Universal PINNs (UPINNs) to investigate glucose-lactate metabolism in glioblastoma cell lines. We first employed PINNs to infer critical model parameters governing glucose uptake and phenotypic switching in tumor cells, demonstrating high accuracy using synthetic data. We then extended this framework using UPINNs to uncover hidden metabolic dynamics that could not be explicitly modeled, introducing a latent variable W to represent unknown functional behavior in glycolytic processes. Our approach was validated for both synthetic and experimental datasets for two glioblastoma cell lines (LN18 and LN229) with distinct metabolic phenotypes. The UPINN framework not only captured cell-type-specific behaviors but also remained robust in the presence of moderate experimental noise. Furthermore, we explored the sensitivity of the model to the trade-off between data fidelity and mechanistic constraints, showing that the choice of loss term weighting significantly impacts predictive performance. While our application centered on cancer metabolism, the proposed method was general and applicable to a wide range of systems described by differential equations, including problems in biology, engineering, and physical sciences. This work demonstrates the potential of UPINNs as a powerful and interpretable tool for data-driven discovery in partially observed dynamical systems.

Keywords: universal physics-informed neural networks; deep learning; glioblastoma multiforme; glucose-lactate metabolism

1. Introduction

Cancer is a multifaceted disease marked by unchecked cell growth and the ability to spread, primarily resulting from genetic alterations that impair normal regulatory mechanisms [1, 2]. As these malignant cells multiply, they can form tumors that infiltrate surrounding tissues or spread throughout the body [3, 4]. Despite advances in research and treatment, cancer remains a major public health challenge worldwide [5]. Among the many defining features of cancer, one key biological hallmark is metabolic reprogramming [2,6]. Cancer cells frequently reprogram their metabolism by favoring aerobic glycolysis over oxidative phosphorylation, a process widely referred to as the Warburg effect, enabling them to sustain growth even under oxygen-limited conditions [7, 8]. This metabolic adaptation involves elevated glucose uptake and significant lactate production [9].

Lactate plays a central role in shaping the tumor microenvironment. It acidifies the extracellular space, suppresses immune responses, and modulates surrounding normal cells [10]. Additionally, lactate can be used as a metabolic fuel via the lactate shuttle, where glycolytic and oxidative tumor cells exchange lactate through MCT transporters to optimize energy usage and maintain pH balance. This metabolic symbiosis enhances tumor robustness and survival in hypoxic niches.

Glioblastoma multiforme (GBM) represents the most malignant type of primary brain cancer in adults [11]. GBM displays rapid proliferation, infiltrative behavior, and resistance to therapy, with patients typically surviving only 12–15 months after diagnosis. Understanding its metabolic adaptations is crucial to identifying vulnerabilities and designing targeted treatments.

Mathematical modeling offers a powerful tool for quantifying the metabolic dynamics in cancer. Models based on systems of differential equations have helped simulate shifts in energy production, including glucose and oxygen consumption, lactate generation, and ATP production [12]. These frameworks enable researchers to probe the influence of microenvironmental factors and test hypothetical interventions computationally [13]. However, detailed mechanistic models face limitations due to incomplete knowledge of system parameters and dynamics. To address this, researchers have integrated machine learning (ML) approaches with mechanistic models [14]. ML is particularly effective in analyzing complex omics data and uncovering hidden patterns. Comparative studies show that hybrid modeling, combining ML's predictive power with mechanistic interpretability, can enhance insight into cancer metabolism [15].

Physics-Informed Neural Networks (PINNs) integrate differential equations directly into neural network training [16–18], enabling reliable learning from limited or noisy data. Universal Physics-Informed Neural Networks (UPINNs) extend this approach by also learning unknown functional components, making them well-suited for partially understood biological systems such as cancer metabolism [19].

In this study, we apply PINNs to infer unknown parameters in a metabolic model of glucose and lactate dynamics based on the work of Mendoza-Juez et al. [20]. We validate the approach on synthetic data before applying it to experimental measurements of glucose, lactate, and cancer cell populations. To further enhance the model's flexibility and biological insight, we employ UPINNs to learn unknown regulatory terms and investigate hypotheses related to metabolic switching, such as transitions from glycolysis to lactate-supported respiration [18, 21].

2. Related work

A growing body of research shows that researchers employ mathematical and multi-scale computational models to explore tumor metabolism, including the Warburg effect, lactate shuttling, and glutamine addiction. Shan et al. [22] developed a multi-scale agent-based model to examine how tumor growth is influenced by metabolic strategies, showing that both Warburg and reverse Warburg effects can provide selective advantages under spatial constraints [7, 8, 23]. Almeida et al. [24] used integro-differential models to study MCT1 regulation via lactate signaling under glucose deprivation.

Roy and Finley [25] proposed a multiscale model integrating intracellular metabolite dynamics with tissue-scale growth in tumor spheroids. Jayathilake et al. [26] and McGillen et al. [27] developed agent-based and spatial models simulating glucose-lactate symbiosis between tumor cell populations. Both studies highlighted how metabolic cooperation reduces glucose depletion and supports tumor viability.

Yang et al. [28] developed and validated a mechanistic model to predict tumor cell proliferation as a function of glucose availability. Building on Mendoza-Juez et al. [20], they simplified the model to a single tumor cell population and used time-resolved microscopy data from breast cancer cell lines to calibrate a system of ODEs capturing glucose consumption, proliferation, and death. Their model achieved high predictive accuracy and identified subtype-specific metabolic parameters. Notably, the researchers addressed limitations of earlier models by explicitly calibrating to experimental data and analyzing dependence of model parameters on initial glucose and confluence conditions.

Experimental studies by Sonveaux et al. [29] provided the biological foundation for lactate-fueled respiration in oxygenated tumor cells, which has inspired several modeling frameworks [30, 31]. These include game-theoretical and agent-based models that explore evolutionary dynamics and spatial heterogeneity in tumor metabolism.

We build on this literature using UPINNs to flexibly incorporate known physical principles while learning unknown dynamics directly from data. This positions our framework as a hybrid modeling approach that balances predictive accuracy, mechanistic interpretability, and adaptability to complex biological systems.

3. Methods

3.1. Glucose-lactate metabolism model

A mathematical model for tumor cell nutrient uptake driven by glucose and lactate [20], containing the total tumor cell population $P(t)$, with fractions undergoing oxidative phosphorylation $P_o(t)$ and glycolysis $P_g(t)$, alongside the nutrient concentrations of glucose $G(t)$ and lactate $L(t)$, all evolve based on nutrient availability and acidosis. Its non-dimensionalized equation system is expressed as

$$\frac{dp_o}{dt} = \frac{1}{\tau_o}(1 - p_o - p_g)p_o + \frac{1}{\tau_{go}}\chi_L(L)p_g - \frac{1}{\tau_{og}}\chi_L^*(L)\chi_G(G)p_o \tag{3.1}$$

$$\frac{dp_g}{dt} = \frac{1}{\tau_g}(1 - p_o - p_g)p_g - \frac{1}{\tau_{go}}\chi_L(L)p_g + \frac{1}{\tau_{og}}\chi_L^*(L)\chi_G(G)p_o \tag{3.2}$$

$$\frac{dG}{dt} = -\kappa_o\frac{G}{G + \lambda L + n_*}p_o - \kappa_G\frac{G}{G + G_*}p_g \tag{3.3}$$

$$\frac{dL}{dt} = -\kappa_L \frac{L}{G/\lambda + L + m_*} p_o + 2\kappa_G \frac{G}{G + G_*} p_g \tag{3.4}$$

where the switch function $\chi_L(L)$ is the form:

$$\chi_L(L) = \frac{1}{2}\left[1 + \tanh\left(\gamma(L - L_*)\right)\right] \tag{3.5}$$

models the transition from glycolytic to oxidative metabolism based on lactate concentration. It remains near zero for low lactate levels and approaches one for high ones. Parameter γ controls the sharpness of this transition, reflecting cell sensitivity to acidity changes.

The inverse transition from oxidative to glycolytic metabolism is governed by $\chi_L^*(L) = 1 - \chi_L(L)$, defined as a decreasing function of lactate concentration, This function ensures that cells maintain oxidative metabolism at low lactate levels and switch back to glycolysis when lactate is sufficiently reduced. Additionally, the function $\chi_G(G)$ is given by:

$$\chi_G(G) = \begin{cases} 0, & \text{if } G \le G_{\min}, \\ 1, & \text{if } G > G_{\min}. \end{cases} \tag{3.6}$$

which accounts for glucose availability, preventing cells from switching to glycolysis without glucose. It is modeled as a step function, where glycolytic metabolism is activated only when glucose exceeds a minimum threshold $G_{\min}$.

To capture the hidden dynamics governing nutrient exchange within tumor cells, the model incorporates specific terms representing glucose uptake and lactate production, reflecting the intricate balance of metabolic processes observed in cancer. In particular, the term $-\kappa_G \frac{G}{G+G_*} p_g$ quantifies the consumption of glucose by the glycolytic cell population p_gg, where the saturation function $\frac{G}{G+G_*}$ embodies the diminishing incremental effect of glucose availability as concentrations exceed the characteristic threshold G_*. This formulation, with κ_G serving as the consumption rate constant, emphasizes that higher populations of glycolytic cells will proportionally increase the rate of glucose uptake, although uptake is limited by saturation kinetics. Complementarily, lactate production is modeled by the term $+2\kappa_G \frac{G}{G+G_*} p_g$, where the multiplicative factor of 2 results from the stoichiometric conversion of one molecule of glucose to two molecules of lactate during glycolysis. The identical saturation term indicates that lactate production is inextricably linked to the available glucose concentration, reinforcing the biological phenomenon known as the Warburg effect, whereby tumor cells preferentially convert glucose to lactate even under aerobic conditions. Combining empirical insights with core biological principles enables the two-term model to capture nutrient-metabolism interactions in tumors, providing a suitable structure for UPINNs to infer hidden dynamic components in a transparent and data-efficient manner.

3.2. Physics-informed neural networks

Let $\vec{u}(t)$ denote an unknown real-valued function governed by a known differential operator $\mathcal{N}$:

$$\frac{d}{dt}\vec{u}(t) = \mathcal{N}\vec{u}(t), \quad t \in [0, T]. \tag{3.7}$$

Given a set of observational data, the goal of a PINN is to approximate $\vec{u}(t)$ using a neural network that both fits the available data and satisfies the governing physical equations. This is achieved by

minimizing a composite loss function composed of two key components. The first is the data loss term, $\mathcal{L}_{\text{data}}$, which quantifies the discrepancy between the predicted and observed data. The second is the physical loss term, $\mathcal{L}_{\text{physics}}$, which enforces consistency with the underlying differential equation using automatic differentiation. The total loss is given by:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{physics}}. \tag{3.8}$$

In addition to solving forward problems, PINNs are also highly effective for inverse problems, such as inferring unknown parameters in dynamical models by fitting to experimental measurements. This is particularly valuable in biomedical systems, where key biological parameters are often inaccessible through direct experimentation. The physics-based regularization in PINNs helps manage noise and data sparsity, which is critical for real-world biomedical applications with imperfect observations.

3.3. Universal physics-informed neural networks

Universal Physics-Informed Neural Networks (UPINNs) extend the standard PINN framework by not only incorporating known physical laws but also learning latent biological processes that cannot be directly encoded within existing mechanistic models. This capability is particularly valuable in biomedical systems, where critical regulatory mechanisms are often only partially understood or entirely unknown. In our framework, UPINNs enable the discovery of such hidden dynamics through a neural representation of an unknown functional term, enabling the model to capture complex, nonlinear metabolic behaviors, such as dose-dependent glucose uptake or emergent glycolytic responses, without requiring these relationships to be explicitly defined a priori. By combining mechanistic constraints with the expressive power of neural networks, UPINNs provide a data-driven approach to uncovering biologically relevant patterns embedded within sparse and noisy experimental data.

Let $\vec{u}(t) \in \mathbb{R}^m$ represent the state vector of the system, and assume the true dynamics are governed by a differential operator $\mathcal{N}$ that can be decomposed into a known component $\mathcal{N}_K$ and an unknown functional component $\mathcal{F}$. The governing equation is then written as:

$$\frac{d}{dt}\vec{u}(t) = \mathcal{N}_K[\vec{u}](t) + \mathcal{F}[\vec{u}](t), \quad t \in [0, T], \tag{3.9}$$

subject to the initial condition:

$$\vec{u}(0) = \vec{u}_0.$$

The UPINN method trains two neural networks: One, $U(t)$, to approximate the solution $\vec{u}(t)$, and another, $\mathcal{F}_\theta[\vec{u}]$, to represent the hidden dynamics $\mathcal{F}$. The training involves minimizing a composite loss function:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{physics}}, \tag{3.10}$$

where the data loss $\mathcal{L}_{\text{MSE}}$ ensures fidelity to observed data $\{\vec{u}_i\}_{i=1}^n$:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n} \left\| U(t_i) - \vec{u}_i \right\|^2, \tag{3.11}$$

and the physical loss $\mathcal{L}_{\text{physics}}$ ensures that the learned solution satisfies the modified dynamics:

$$\mathcal{L}_{\text{physics}} = \frac{1}{k}\sum_{i=1}^{k} \left\| \frac{d}{dt}U(t_i) - (\mathcal{N}_K[U](t_i) + \mathcal{F}_\theta[U](t_i)) \right\|^2. \tag{3.12}$$

This formulation enables the model not only to fit sparse and noisy data, but also to infer symbolic representations of unknown dynamics. Notably, symbolic regression tools such as AI Feynman can be applied post hoc to extract interpretable equations from $\mathcal{F}_\theta$ [19].

We employ this framework to infer unknown metabolic interactions in cancer models, where full mechanistic knowledge is either unavailable or uncertain. This includes identifying non-linear or dose-dependent effects in tumor growth and treatment response dynamics [19]. UPINNs thus provide a powerful hybrid approach; anchored in physical laws yet flexible enough to learn new biology from data.

4. Results

4.1. Parameter estimation based on physics-informed neural networks

Using the PINN framework, we estimate critical model parameters including κ_o, κ_G, τ_{og}, and τ_{go}, which are typically challenging to measure directly through experimental methods. The parameter estimates and their corresponding Mean Squared Error (MSE) values are summarized in Table 1.

Table 1. Estimated parameters and their MSE values using PINN. The true values listed here are those reported in Mendoza-Juez et al. [20], selected to ensure consistency with experimental observations.

Cell line	Parameter	Estimated value	True value	MSE
LN18	κ_o	1.067	1.5	1.87×10^{-1}
	κ_G	7.473	10.0	6.384
	τ_{og}	2.2×10^{-2}	$1/24$	4.0×10^{-4}
	τ_{go}	8.88×10^{-1}	1.0	1.2×10^{-2}
LN229	κ_o	1.000	1.5	2.5×10^{-1}
	κ_G	2.101	2.5	1.59×10^{-1}
	τ_{og}	2.0×10^{-2}	$1/24$	5.0×10^{-4}
	τ_{go}	1.014	1.0	2.0×10^{-4}

The values estimated using PINN are in close agreement with the true reference values, demonstrating the method's effectiveness in recovering biologically meaningful parameters. For the LN18 cell line, κ_o is estimated as 1.067 versus the true value of 1.5, with an MSE of 0.187. Similarly, τ_{og} and τ_{go} are estimated as 0.022 and 0.888, respectively, closely matching their true values of $1/24$ and 1.0, and yielding low MSEs. These results suggest that the network effectively captures the dynamics associated with oxidative metabolism and phenotypic switching.

However, κ_G exhibited a larger deviation. It was estimated at 7.473 compared to the true value of 10.0, resulting in a relatively higher MSE of 6.384. This discrepancy may stem from the increased sensitivity of glycolytic dynamics to data noise, combined with fewer observations capturing glucose-driven transitions. These findings indicate that more informative data or additional regularization might be needed to constrain this parameter more accurately.

For the LN229 cell line, the estimates similarly demonstrate strong agreement with the true values. κ_o and κ_G were estimated at 1.000 and 2.101, respectively, while the true values are 1.5 and 2.5. Although slightly underestimated, these values remain within an acceptable range. The estimated τ_{og} and τ_{go}, 0.020 and 1.014, respectively, are nearly identical to the ground truth, supported by very low MSEs of

0.0005 and 0.0002, respectively.

These comparisons highlight the utility of PINNs for parameter estimation in complex biological models using sparse and potentially noisy data. The low MSEs for most parameters underscore the model's reliability. The variation observed in κ_G identifies a direction for further refinement; either through enhanced data collection strategies or more expressive neural architectures. Overall, the results support the application of PINNs as a robust alternative to direct experimental measurement in parameter estimation tasks.

4.2. Identifying cancer cell metabolic dynamics using synthetic data

We apply the UPINN framework to the glucose–lactate metabolism model (Eqs 3.1–3.4) to identify the regulatory mechanisms underlying glucose uptake and lactate production, and to investigate how these metabolic processes differ between LN18 and LN229 glioblastoma cell lines. As discussed, the terms $-\kappa_G \frac{G}{G+G_*} p_g$ and $+2\kappa_G \frac{G}{G+G_*} p_g$ reflect the stoichiometry of glycolysis, where one molecule of glucose is converted into two molecules of lactate.

To simplify the modeling of unknown dynamics while preserving biological meaning, we introduce a latent function W, which enables a simplified yet biologically meaningful representation of glycolytic activity, consolidating glucose uptake and lactate production into a single term that can be learned from data, and define:

$$F = W = \kappa_G \frac{G}{G + G_*} p_g, \tag{4.1}$$

so that the expressions $-\kappa_G \frac{G}{G+G_*} p_g$ and $+2\kappa_G \frac{G}{G+G_*} p_g$ can be rewritten as $-W$ and $2W$, respectively. In this setting, κ_G remains an unknown parameter to be inferred, while all other terms are assumed known and included in the operator $\mathcal{N}_K[U]$.

We adopt all the parameter values listed in Table 2 from Mendoza-Juez et al. [20]. The differences between LN18 and LN229 capture variations in metabolic phenotype, such as glycolytic rate and lactate threshold. Initial conditions are chosen to reflect the Warburg effect, with approximately 85% of cells in a glycolytic state and 15% in an oxidative state. Accordingly, the initial conditions for LN18 and LN229 are defined as

$$y_0^{\text{LN18}} = y_0^{\text{LN229}} = [0.1, 0.0176, 9.444, 1.124],$$

based on [20], and the time interval considered is $t \in [0, 4]$. Synthetic datasets $\{t_i, p_{o,i}, p_{g,i}, G_i, L_i\}$ are generated by numerically integrating the ODE system using a classic fourth-order Runge–Kutta method.

In experimental settings [20], only three quantities are observable: Total cell accumulation ($P_o + P_g$), glucose concentration, and lactate concentration. Since the MSE loss in the UPINN framework must correspond to these observable quantities, we redefine the data loss as:

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \Big(\big| P_o(t_i) + P_g(t_i) - P_i \big|^2 + |G(t_i) - G_i|^2 + |L(t_i) - L_i|^2 \Big), \tag{4.2}$$

where P_i, G_i, and L_i are the synthetic ground truth values. This ensures a direct mapping between model predictions and measurable quantities. The residual ODE loss, $\mathcal{L}_{ODE}$, remains unchanged, continuing to enforce the consistency of the model with the governing dynamics through automatic differentiation.

Table 2. Model parameters used for LN18 and LN229 glioblastoma cell lines.

Notation	Parameter (units)	LN18	LN229
τ_o	Oxidative cell proliferation (days)	$1/\log 2$	$1/\log 2$
τ_g	Glycolytic cell proliferation (days)	$1/\log 2$	$1/\log 2$
P_*	Confluence ($\times 10^3$ cells)	850	750
τ_{og}	Oxidative $\rightarrow$ Glycolytic switch time (days)	1/24	1/24
τ_{go}	Glycolytic $\rightarrow$ Oxidative switch time (days)	1	1
γ	Phenotype switching sensitivity (1/mM)	100	100
L_*	Lactate threshold (mM)	10	4
κ_o	Glucose consumption by oxidative cells (mM/day)	1.5	1.5
κ_L	Lactate consumption by oxidative cells (mM/day)	5	5
κ_G	Glucose consumption by glycolytic cells (mM/day)	10	2.5
λ	Preference for lactate	1	1
n_*	Glucose oxidation threshold (mM)	1	1
m_*	Lactate oxidation threshold (mM)	1	1
G_*	Michaelis-Menten constant (mM)	0.5	0.5

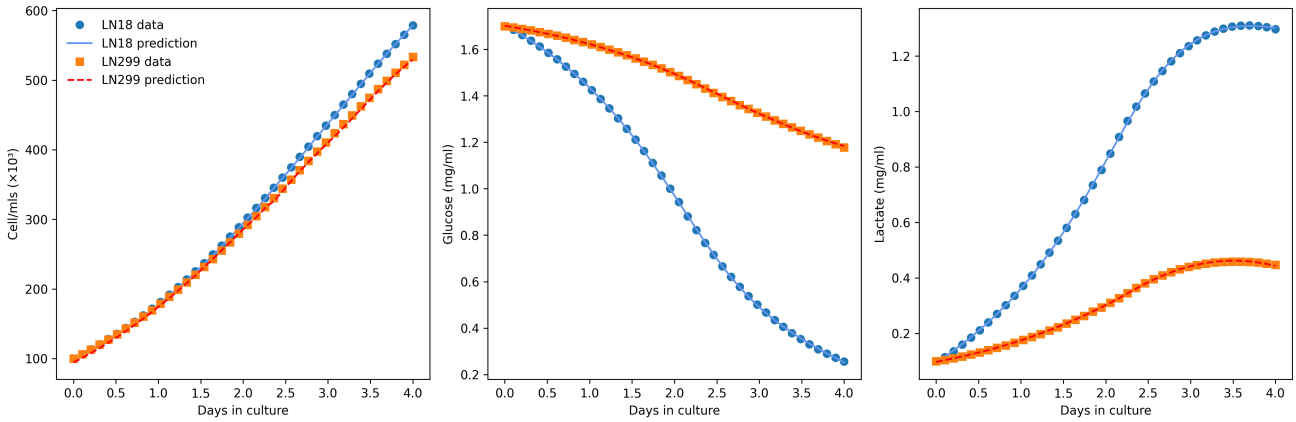
4.3. UPINN implementation and performance on synthetic data

The UPINN framework employs a deep neural network (DNN) as a surrogate model to approximate the solution $U(t)$, in conjunction with auxiliary networks to represent the unknown dynamics encapsulated by the latent function W. The surrogate network is a fully connected DNN comprising eight hidden layers, each with 128 neurons, utilizing hyperbolic tangent (Tanh) activation functions and input normalization to improve training stability. The auxiliary networks, which share the same architecture, receive the relevant state variables as inputs and output predictions for the hidden dynamics.

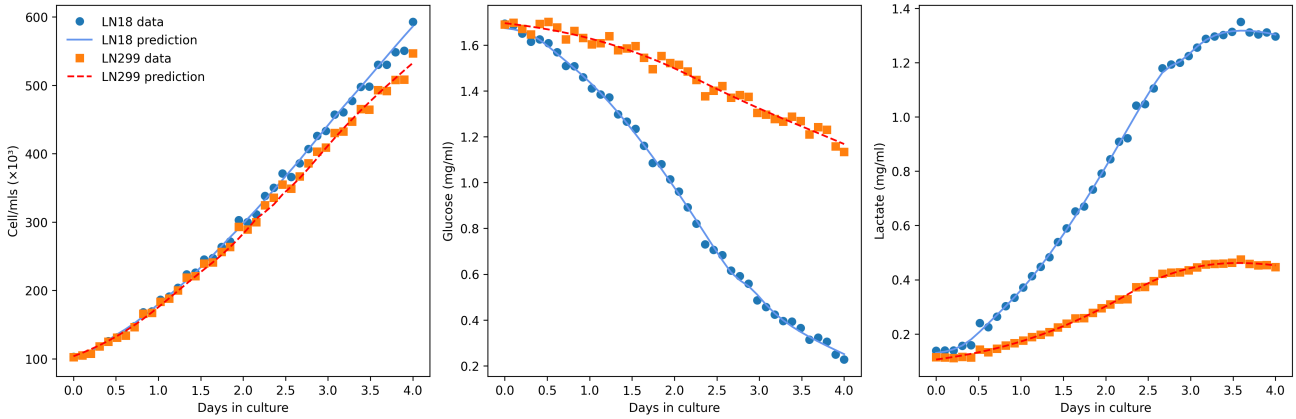
Training is performed using a two-stage optimization process. The model is first trained with the Adam optimizer to provide a coarse solution, followed by fine-tuning using the quasi-Newton L-BFGS optimizer, which enhances convergence and overall accuracy.

We first evaluate UPINN on synthetic noise-free data to assess its ability to reconstruct the hidden dynamics represented by W, which governs glucose consumption and lactate production in glycolytic tumor cells. Figure 1(a) shows the comparison between UPINN predictions and the ground truth solutions for LN18 and LN229. The subplots show, from left to right, total cell accumulation, glucose concentration, and lactate concentration.

Both cell lines exhibit similar proliferation patterns, with initial seeding at 10^5 cells/mL. Consistent with their glycolytic phenotype, both lines demonstrate preferential glucose usage, along with comparable oxidative nutrient uptake. Given the stoichiometry of glycolysis, where one glucose molecule (approximately 180 Da) yields two lactate molecules (each approximately 89 Da), the temporal trends in lactate production closely reflect those of glucose consumption. Notably, lactate accumulation slows as it approaches the tolerance threshold, with this effect more pronounced in LN229, which exhibits a more gradual metabolic adaptation compared to LN18.



(a) Comparison between UPINN predictions and synthetic noise-free solutions. From left to right: Total cell population, glucose concentration, and lactate concentration for LN18 and LN229.



(b) Comparison between UPINN predictions and synthetic solutions with 3% added noise. From left to right: Total cell population, glucose concentration, and lactate concentration for LN18 and LN229.

Figure 1. UPINN prediction performance for synthetic data under noise-free and noisy conditions.

Table 3 summarizes the MSE values between the UPINN predictions and the ground truth for each of the four observable quantities: Total cell accumulation, glucose, lactate, and the latent variable W. In the absence of noise, all MSE values are below 10^{-4}, indicating that the UPINN can precisely capture the underlying metabolic dynamics.

Figure 4 shows the learned W functions under both noise-free and noisy conditions. In the ideal setting (left panels), the network accurately captures the trajectory of W, including the peak amplitude and temporal dynamics. The estimated MSEs for W are 1.117×10^{-4} for LN18 and 3.097×10^{-4} for LN229, confirming the model's precision. While noise degrades prediction accuracy, the core features of the hidden dynamics remain identifiable, indicating the UPINN's robustness in the presence of moderate observational uncertainty.

To further assess the robustness of the model, we introduce Gaussian noise into the synthetic dataset

Table 3. Mean squared error (MSE) for UPINN predictions under noise-free and 3% noisy conditions.

Condition	Cell Line	Cell Accumulation	G	L	W
Noise-free	LN18	2.482×10^{-7}	4.519×10^{-5}	2.003×10^{-4}	1.117×10^{-4}
	LN229	1.078×10^{-5}	2.191×10^{-5}	6.488×10^{-5}	3.097×10^{-4}
Noisy (3%)	LN18	6.829×10^{-5}	9.257×10^{-3}	2.163×10^{-2}	9.707×10^{-2}
	LN229	9.342×10^{-5}	1.737×10^{-2}	4.707×10^{-3}	1.388×10^{-3}

using the following formulation:

$$\mathbf{x}_{\text{noisy}} = \mathbf{x}_{\text{true}} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, 0.03^2\mathbf{I}), \tag{4.3}$$

and re-evaluate the UPINN predictions (Figure 1(b), bottom row). As expected, the addition of noise lead to increased MSE values across all output variables (Table 3). For the LN18 cell line, glucose and lactate concentrations exhibit the most significant increases in prediction error, with the MSE for glucose rising from 4.519×10^{-5} to 9.257×10^{-3}, and for lactate from 2.003×10^{-4} to 2.163×10^{-2}. In contrast, LN229 showed relatively smaller increases, with glucose MSE increasing from 2.191×10^{-5} to 1.737×10^{-2}, and lactate MSE from 6.488×10^{-5} to 4.707×10^{-3}.

Despite the increase in error, Figure 1(b) illustrates that UPINN successfully captures the overall dynamics of cell accumulation, glucose depletion, and lactate production under noisy conditions. This highlights the model's robustness and its capacity to recover key metabolic trends even in the presence of moderate experimental noise.

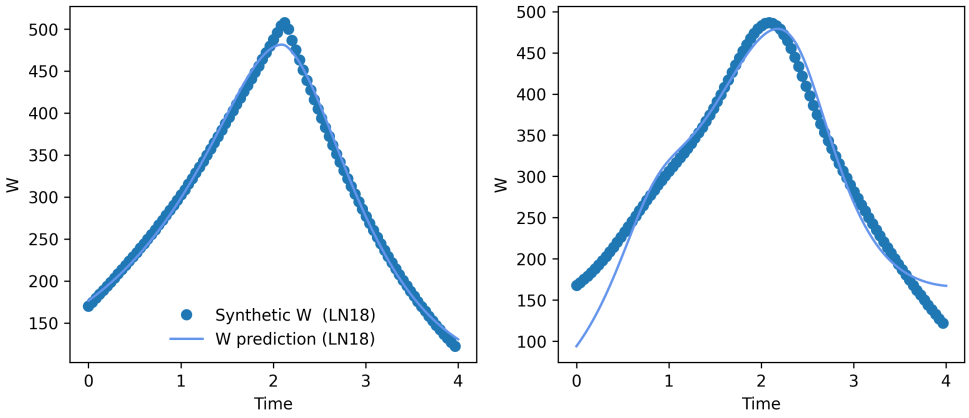
Overall, these results demonstrate that UPINN is a reliable method for identifying metabolic parameters governing glucose uptake and lactate production in glycolytic cells. The observed differences between LN18 and LN229 reflect their distinct metabolic phenotypes: LN18, which exhibits greater reliance on glycolysis, is more sensitive to noise perturbations, while LN229 displays more stable behavior. This further emphasizes the importance of modeling cell-line-specific metabolic traits when interpreting experimental data or designing targeted interventions.

4.4. Identifying cancer cell metabolic dynamics using experimental data

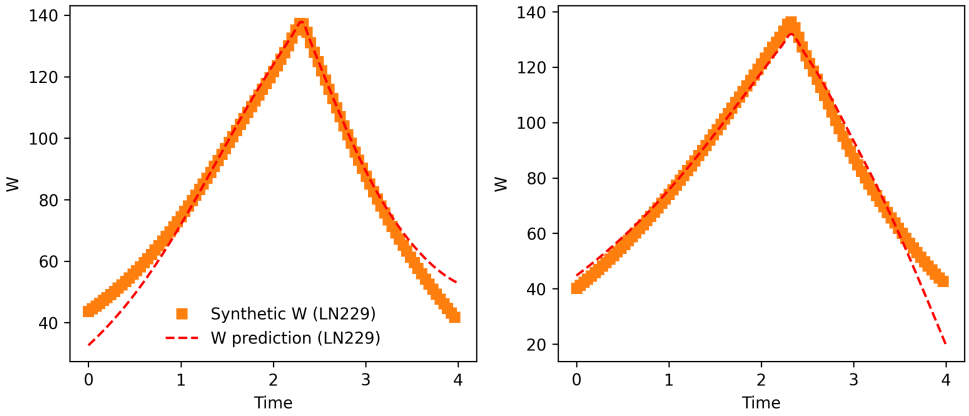
Having demonstrated the effectiveness of UPINN in capturing metabolic dynamics using synthetic data, we next apply the framework to experimental measurements reported in [32], which examine the proliferation and nutrient consumption of LN18 and LN229 glioblastoma cell lines in culture. Unlike synthetic datasets, real-world measurements are subject to biological variability and noise, and often lack explicit ground truth for unobservable variables. Consequently, discrepancies between model predictions and experimental observations must be interpreted with the inherent uncertainty and complexity of in vitro experiments in mind.

To improve predictive accuracy under such conditions, we incorporate measurement uncertainty directly into the UPINN training process. Specifically, we integrate experimental error bars for total cell accumulation ($P_o + P_g$), glucose concentration (G), and lactate concentration (L) into the data loss term. The modified loss function weighs each data point according to its associated variance σ_i^2:

$$\mathcal{L}_{\text{MSE}} = \sum_{i=1}^{N} \frac{(f_\theta(x_i) - y_i)^2}{\sigma_i^2}. \tag{4.4}$$



(a) Comparison of W for LN18 under noise-free (left) and 3% noise (right) conditions.



(b) Comparison of W for LN229 under noise-free (left) and 3% noise (right) conditions.

Figure 2. Reconstruction of the hidden dynamics W using UPINN.

This formulation reduces the influence of noisy or uncertain measurements while giving more weight to reliable data points. As a result, the model is guided toward fitting trends supported by high-confidence observations, thereby improving its ability to identify the underlying metabolic dynamics.

Figure 3 compares UPINN predictions (solid lines) to experimental data (markers with error bars) for LN18 and LN229 over a four-day culture period. The left panel displays total cell accumulation, the middle shows glucose concentration, and the right shows lactate concentration. For LN18, the model closely tracks the steep decline in glucose and the corresponding rise in lactate, consistent with a strongly glycolytic phenotype. In contrast, LN229 exhibits a more gradual decline in glucose and a slower lactate accumulation, reflecting its comparatively lower glycolytic activity.

Table 4. MSE values for LN18 and LN229 using experimental data.

Cell line	Cell accumulation	G	L
LN18	4.09×10^{-4}	5.02×10^{-3}	1.28×10^{-3}
LN229	3.66×10^{-4}	5.15×10^{-3}	1.34×10^{-3}

The MSE values for each variable, summarized in Table 4, confirm that UPINN maintains low

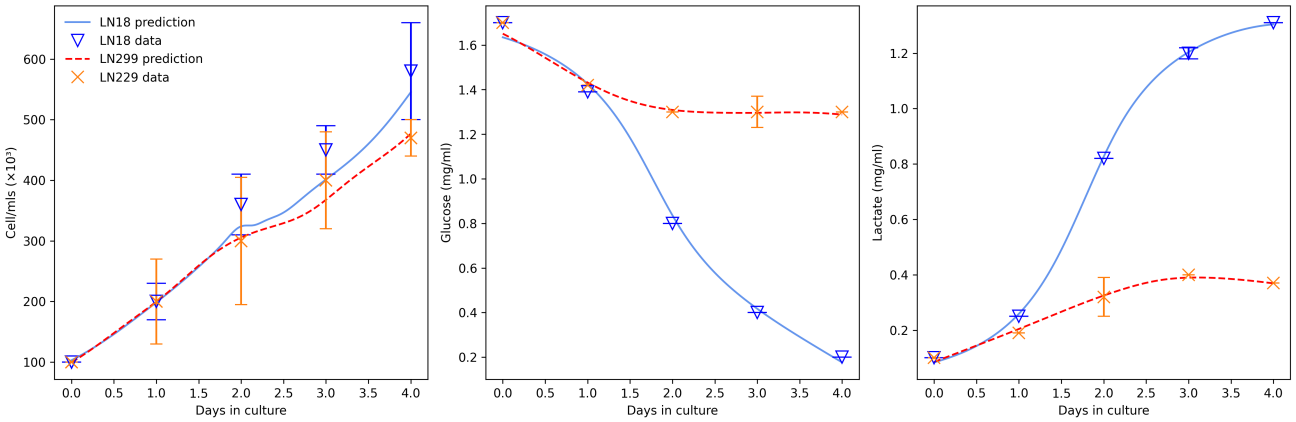


Figure 3. (a) Total cell accumulation ($P_o + P_g$) for LN18 and LN229 cells. (b) Glucose concentration. (c) Lactate concentration. Error bars indicate experimental variability; solid lines show UPINN predictions.

prediction errors across both cell lines. Notably, the model captures the key features of each metabolic profile with consistent precision.

Figure 4 shows the inferred hidden terms W for both LN18 (top) and LN229 (bottom) under experimental conditions. In both cases, the UPINN framework successfully reconstructs the latent metabolic dynamics, revealing differences in glycolytic activity consistent with prior biological characterization of the two cell lines. The resulting W profiles further validate the model's capability to extract meaningful biological insights from noisy, real-world measurements.

Figure 4 illustrates the inferred hidden metabolic variable W for each cell line, capturing unobserved dynamics underlying glucose uptake and lactate production. LN18 exhibits a higher and earlier peak in W, indicating a rapid glycolytic response, consistent with the sharp drop in glucose and rise in lactate observed experimentally (Figure 3). In contrast, LN229 displays a lower, more gradually decaying W profile, reflecting its comparatively slower glucose consumption and lactate accumulation. These differences in W align with the distinct metabolic phenotypes previously reported for the two cell lines.

Quantitatively, Table 4 summarizes the model's performance, with the lowest MSE values observed for total cell accumulation ($\sim 4 \times 10^{-4}$), while glucose and lactate predictions yield slightly higher MSEs on the order of 10^{-3}. These differences likely reflect the higher variability and measurement noise typically associated with metabolic assays, as compared to the more reproducible cell counting.

Interestingly, when experimental uncertainties are explicitly incorporated into the data loss function, UPINN predictions for LN18 reveal a pronounced inflection point at $t = 2$ days, particularly in the glucose and lactate profiles. This behavior contrasts with the smoother trajectories observed in the synthetic datasets. One plausible explanation lies in the discrepancy between idealized synthetic data and real-world experimental observations, which include biological variability, noise, and potential systematic errors. By weighting data points according to their uncertainty, the model is encouraged to balance fidelity to uncertain measurements with adherence to the mechanistic constraints imposed by the governing differential equations. The network may learn to introduce new features (e.g., inflection points) to reduce the total loss.

To investigate how these inflection points arise, we conduct an additional set of experiments exploring the effect of varying the relative weights of the data and physical loss terms. Specifically, we introduce

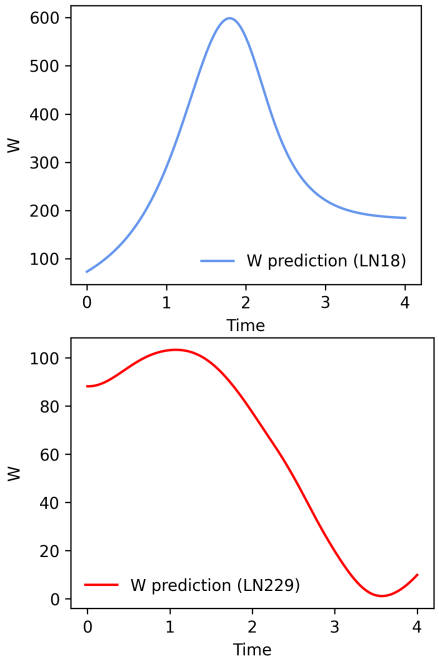


Figure 4. Learned hidden dynamics W for LN18 (top) and LN229 (bottom) using experimental data.

two weighting parameters, λ_a and λ_b, to modulate the influence of the observational and mechanistic components of the total loss:

$$\mathcal{L} = \lambda_a \mathcal{L}_{\text{MSE}} + \lambda_b \mathcal{L}_{\text{physics}}. \tag{4.5}$$

We evaluate a range of λ_a and λ_b combinations and analyze their impact on model predictions, demonstrating that glucose and lactate predictions are highly sensitive to the choice of these weights (see Appendix).

For LN229, reducing λ_a from 1.0 to 0.1 while keeping $\lambda_b = 1.0$, results in a doubling of glucose MSE from 0.038 to 0.078, and a moderate increase in lactate MSE from 0.066 to 0.080. In contrast, LN18 is significantly more sensitive: The same reduction in λ_a leads to a dramatic increase in glucose MSE from 0.029 to 1.144 and in lactate MSE from 0.008 to 7.018. Notably, the accuracy of total cell accumulation predictions remain stable, with MSEs consistently between 10^{-4} and 10^{-3} across all tested configurations.

These results highlight the importance of carefully selecting weighting parameters when applying UPINN to experimental data. Overemphasis on data fidelity may lead to artifacts such as spurious inflection points, while overemphasis on mechanistic consistency may compromise the model's ability to accurately fit the data. Optimal performance requires a balanced weighting scheme that accounts for the nature and uncertainty of the measurements and the biological context of the system under investigation.

5. Discussion and conclusions

In this work, we developed a unified framework leveraging PINNs and UPINNs to investigate the metabolic dynamics of glioblastoma cell lines. Our objective was twofold: First, to infer unknown parameters within a mechanistic model using PINNs; and second, to extend this capability using UPINNs to uncover hidden functional dynamics that are not explicitly defined in the model structure.

We began by employing PINNs to estimate key kinetic parameters involved in glucose and lactate metabolism. Using synthetic data generated under controlled conditions, the model achieved high accuracy in recovering parameters such as glucose consumption rates and phenotype switching times, with low MSEs validating the approach. This phase demonstrated the effectiveness of PINNs in solving inverse problems where the governing equations are known, but parameters remain unmeasured or uncertain.

Building on this, we transitioned to UPINNs to discover unknown components of the governing dynamics, specifically, the hidden term W that modulates glucose uptake and lactate production in glycolytic tumor cells. By learning W directly from data, UPINNs enabled a more flexible and expressive representation of the system, capturing metabolic behaviors that could not be fully explained by parameterized models alone. In both noise-free and noisy synthetic settings, UPINNs recovered biologically consistent trends and revealed differences between the LN18 and LN229 cell lines, supporting their known metabolic phenotypes.

The latent function W, learned through the UPINN framework, captures nonlinear metabolic dynamics that are otherwise difficult to model explicitly. While introduced as a flexible neural representation, W retains biological interpretability: Its temporal profile reflects the aggregate rate of glucose uptake and lactate production by glycolytic cells, enabling us to distinguish metabolic phenotypes across glioblastoma cell lines. For instance, a sharper and higher peak in W corresponds to more aggressive glycolytic activity, as observed in the LN18 cell line. These variations are consistent with experimentally reported differences in metabolic behavior and validate W as a proxy for hidden biological processes. Furthermore, since W is learned as a continuous and differentiable function, it can be post-processed using symbolic regression techniques to potentially recover interpretable, mechanistically grounded expressions, offering a path toward deeper insight into tumor metabolism. Importantly, the UPINN framework is not limited to this specific application; it is a general approach that can be applied to infer other unknown components in differential equation models across a wide range of biological, physical, and engineering systems where partial mechanistic knowledge is available.

The application of the framework to experimental data introduced new challenges, including biological variability, measurement noise, and the lack of ground truth for hidden variables. To address this, we incorporated experimental uncertainty into the loss function by weighting data points based on their associated variance. This adjustment improved model robustness and helped mitigate the influence of unreliable measurements. However, we observed the emergence of inflection points in the predicted profiles, especially in glucose and lactate concentrations for LN18, highlighting the sensitivity of the model to the trade-off between data fidelity and physical consistency.

To better understand this behavior, we introduced the weighting parameters λ_a and λ_b to control the relative contribution of the loss of data and the residual loss of ODE. Through a systematic study, we found that improper weighting can introduce artifacts (by overfitting noisy data) or degrade predictive performance (by overly constraining the model with rigid dynamics). These findings emphasize the importance of carefully tuning the model based on the nature of the experimental data and the biological

system under study.

Overall, our results show that the combined use of PINNs and UPINNs provides a powerful and interpretable framework for modeling complex biological systems. PINNs are well-suited for parameter estimation when the model structure is known, while UPINNs offer greater flexibility for uncovering hidden biological mechanisms. Together, they enable data-driven discovery constrained by physical laws, enabling meaningful insights even when direct measurements are incomplete or uncertain.

Beyond our application, the generality of UPINNs has also been demonstrated in other biomedical domains. For instance, Podina et al. [33] applied UPINNs to pharmacokinetic and pharmacodynamic modeling of chemotherapy drug action, successfully identifying unknown dose- and time-dependent effects of doxorubicin from both synthetic and in-vitro data. Their approach showed that UPINNs can learn drug mechanisms directly from experimental observations, even when the model structure is only partially known. This highlights the potential of UPINNs to complement or extend traditional quantitative systems pharmacology (QSP) models and supports their application to a wide range of systems biology problems involving incomplete mechanistic knowledge.

While the chosen UPINN architecture yielded robust results in both synthetic and experimental settings, we acknowledge the absence of a systematic ablation study on the impact of network depth and width. Due to computational constraints and limited biological replicates, we prioritized demonstrating the feasibility and accuracy of the framework under a fixed architecture. In future work, we aim to conduct a more comprehensive ablation analysis to evaluate model sensitivity to architectural choices and to assess overfitting risk. Furthermore, coupling these computational studies with biochemical validation will enable us to quantitatively evaluate the biological plausibility of the learned latent dynamics.

In addition, this study demonstrates the feasibility of UPINNs for modeling hidden metabolic dynamics using synthetic and limited experimental data, but we acknowledge that the interpretation of the latent variable W remains largely phenomenological. To enhance its biological plausibility, future work should incorporate targeted perturbation experiments, such as MCT1 inhibition or glucose withdrawal, and employ biochemical assays to empirically test the inferred dynamics. These approaches will be crucial for validating the model's predictions and ensuring that the learned latent functions reflect true physiological behavior.

Moreover, we focus on modeling glucose-lactate metabolism in two glioblastoma cell lines (LN18 and LN229), UPINN framework is general and can be applied to other biological systems governed by differential equations. As a next step, we plan to extend this approach to additional cancer cell lines such as HCT116 or MCF-7, which have well-characterized metabolic behaviors and publicly available glucose/lactate time-course data. Applying the model to these systems will provide a robust test of its generalizability and further validate its ability to uncover cell-line-specific metabolic dynamics across tumor phenotypes.

Also, future work will entail on modeling glucose-lactate metabolism in two glioblastoma cell lines (LN18 and LN229), UPINN framework is general and can be applied to other biological systems governed by differential equations. As a next step, we plan to extend this approach to additional cancer cell lines such as HCT116 or MCF-7, which have well-characterized metabolic behaviors and publicly available glucose/lactate time-course data. Applying the model to these systems will provide a robust test of its generalizability and further validate its ability to uncover cell-line-specific metabolic dynamics across tumor phenotypes.on extending this framework to spatial models involving partial differential

equations (PDEs), improving interpretability through symbolic regression of learned dynamics, and applying the method to other biological systems beyond cancer metabolism. As datasets become richer and more diverse, these tools will be invaluable for integrating experimental data with mechanistic understanding to build more accurate and predictive models of disease.

While we focus on demonstrating the utility of UPINNs within a biologically relevant context, modeling glucose-lactate metabolism in glioblastoma cell lines, we acknowledge the importance of benchmarking this framework against alternative modeling approaches. In future work, we plan to systematically evaluate the performance of UPINNs against these approaches, particularly in settings where experimental data are available. The combined use of PINNs and UPINNs offers a flexible framework for integrating sparse or noisy observational data with mechanistic models, making it suitable for problems in biology, physics, engineering, and beyond. For example, this approach can be adapted to study ecological population dynamics, pharmacokinetics, epidemiological models, fluid transport, or any system where governing equations are known or partially known but parameters or dynamics are uncertain. By embedding physical or biological constraints into neural networks, the framework ensures interpretability, robustness, and generalizability, making it a powerful tool for data-driven discovery in diverse scientific domains.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article. AI tools were used solely to assist with minor language editing. All analyses, interpretations, and conclusions were developed independently by the authors.

Acknowledgments

We thank Lena Podina for generously providing the original implementation of the UPINN framework, which served as the foundation for our simulations. We are also grateful to Shelev Manor for insightful discussions that contributed to the development of this work. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant (MK).

Conflict of interest

The authors declare there is no conflict of interest.

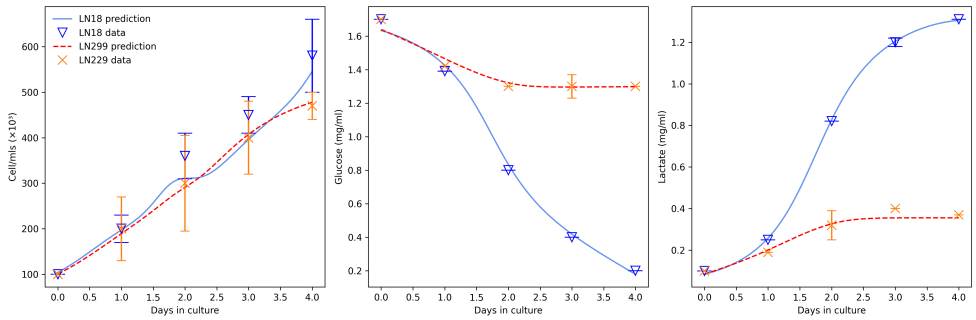
References

1. National Cancer Institute, *What is Cancer?*, *National Institutes of Health*, 2021. Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.
2. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation, *Cell*, **144** (2011), 646–674. https://doi.org/10.1016/j.cell.2011.02.013
3. D. Hanahan, Hallmarks of cancer: new dimensions, *Cancer Discovery*, **12** (2022), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059
4. R. A. Weinberg, *The Biology of Cancer*, WW Norton, 2013

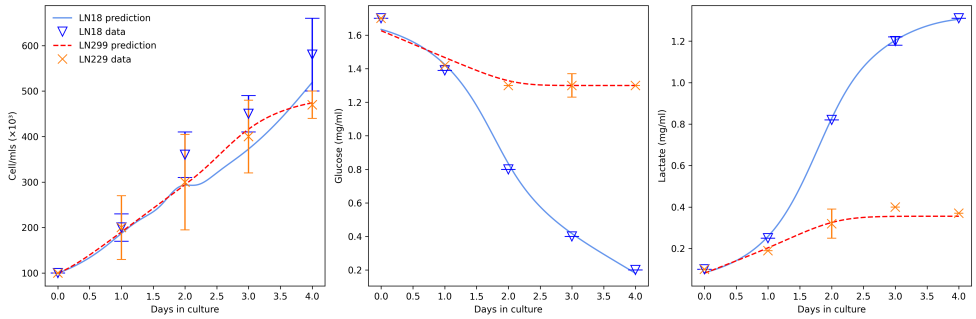
5. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J. Clin.*, **72** (2022), 7–33. https://doi.org/10.3322/caac.21708
6. D. Hanahan, R. A. Weinberg, The hallmarks of cancer, *Cell*, **100** (2000), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9
7. O. Warburg, On the origin of cancer cells, *Science*, **123** (1956), 309–314. https://doi.org/10.1126/science.123.3191.309
8. M. G. Vander Heiden, L. C. Cantley, C. B. Thompson, Understanding the Warburg effect: the metabolic requirements of cell proliferation, *Science*, **324** (2009), 1029–1033. https://doi.org/10.1126/science.1160809
9. N. N. Pavlova, C. B. Thompson, The emerging hallmarks of cancer metabolism, *Cell Metab.*, **23** (2016), 27–47. https://doi.org/10.1016/j.cmet.2015.12.006
10. R. Pérez-Tomás, I. Pérez-Guillén, Lactate in the tumor microenvironment: an essential molecule in cancer progression and treatment, *Cancers*, **12** (2020), 3244. https://doi.org/10.3390/cancers12113244
11. Q. T. Ostrom, G. Cioffi, H. Gittleman, N. Patil, K. Waite, C. Kruchko, et al., CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016, *Neuro-Oncology*, **21** (2019), v1–v100. https://doi.org/10.1093/neuonc/noz150
12. R. A. Gatenby, R. J. Gillies, Why do cancers have high aerobic glycolysis?, *Nat. Rev. Cancer*, **4** (2004), 891–899. https://doi.org/10.1038/nrc1478
13. N. Raghunand, R. A. Gatenby, R. J. Gillies, Microenvironmental and cellular consequences of altered blood flow in tumours, *Br. J. Radiol.*, **76** (2003), S11–S22. https://doi.org/10.1259/bjr/12913493
14. J. Metzcar, C. R. Jutzeler, P. Macklin, A. Köhn-Luque, S. C. Brüningk, A review of mechanistic learning in mathematical oncology, *Front. Immunol.*, **15** (2024), 1363144. https://doi.org/10.3389/fimmu.2024.1363144
15. J. Yang, J. Virostko, J. Liu, A. M. Jarrett, D. A. Hormuth, T. E. Yankeelov, Comparing mechanism-based and machine learning models for predicting the effects of glucose accessibility on tumor cell proliferation, *Sci. Rep.*, **13** (2023), 10387. https://doi.org/10.1038/s41598-023-37493-3
16. M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, **378** (2019), 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
17. A. D. Jagtap, K. Kawaguchi, G. E. Karniadakis, Extended physics-informed neural networks (XPINNs): a generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations, *Commun. Comput. Phys.*, **28** (2020), 2002–2041. https://doi.org/10.4208/cicp.OA-2020-0165
18. J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating physics-based modeling with machine learning: a survey, preprint, arXiv:2003.04919, 2020. https://arxiv.org/abs/2003.04919
19. L. Podina, B. Eastman, M. Kohandel, Universal physics-informed neural networks: symbolic differential operator discovery with sparse data, in *International Conference on Machine Learning*, (2023), 27948–27956.
20. B. Mendoza-Juez, A. Martínez-González, G. F. Calvo, V. M. Pérez-García, A mathematical model for the glucose-lactate metabolism of *in vitro* cancer cells, *Bull. Math. Biol.*, **74** (2012), 1125–1142. https://doi.org/10.1007/s11538-011-9705-1

21. G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.*, **3** (2021), 422–440. https://doi.org/10.1038/s42254-021-00314-5
22. M. Shan, D. Dai, A. Vudem, J. D. Varner, A. D. Stroock, Multi-scale computational study of the Warburg effect, reverse Warburg effect and glutamine addiction in solid tumors, *PLoS Comput. Biol.*, **14** (2018), e1006584. https://doi.org/10.1371/journal.pcbi.1006584
23. S. Pavlides, D. Whitaker-Menezes, R. Castello-Cros, N. Flomenberg, A. K. Witkiewicz, P. G. Frank, et al., The reverse Warburg effect: aerobic glycolysis in cancer associated fibroblasts and the tumor stroma, *Cell Cycle*, **8** (2009), 3984–4001. https://doi.org/10.4161/cc.8.23.10238
24. L. Almeida, J. A. Denis, N. Ferrand, T. Lorenzi, A. Prunet, M. Sabbah, et al., Evolutionary dynamics of glucose-deprived cancer cells: insights from experimentally informed mathematical modelling, *J. R. Soc. Interface*, **21** (2024), 2023.0587. https://doi.org/10.1098/rsif.2023.0587
25. M. Roy, S. D. Finley, Metabolic reprogramming dynamics in tumor spheroids: insights from a multicellular, multiscale model, *PLoS Comput. Biol.*, **15** (2019), e1007053. https://doi.org/10.1371/journal.pcbi.1007053
26. P. G. Jayathilake, P. Victori, C. E. Pavillet, C. H. Lee, D. Voukantsis, A. Miar, et al., Metabolic symbiosis between oxygenated and hypoxic tumour cells: an agent-based modelling study, *PLoS Comput. Biol.*, **20** (2024), e1011944. https://doi.org/10.1371/journal.pcbi.1011944
27. J. B. McGillen, C. J. Kelly, A. Martínez-González, N. K. Martin, E. A. Gaffney, P. K. Maini, et al., Glucose–lactate metabolic cooperation in cancer: insights from a spatial mathematical model and implications for targeted therapy, *J. Theor. Biol.*, **361** (2014), 19–33. https://doi.org/10.1016/j.jtbi.2014.09.018
28. J. Yang, J. Virostko, D. A. Hormuth, J. Liu, A. Brock, J. Kowalski, et al., An experimental-mathematical approach to predict tumor cell growth as a function of glucose availability in breast cancer cell lines, *PLoS One*, **16** (2021), e0240765. https://doi.org/10.1371/journal.pone.0240765
29. P. Sonveaux, F. Végran, T. Schroeder, M. C. Wergin, J. Verrax, Z. N. Rabbani, et al., Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice, *J. Clin. Invest.*, **118** (2008), 3930–3942. https://doi.org/10.1172/JCI36843
30. D. Basanta, H. Hatzikirou, A. Deutsch, Evolutionary game theory elucidates the role of glycolysis in glioma progression and invasion, *Cell Proliferation*, **41** (2008), 980–987. https://doi.org/10.1111/j.1365-2184.2008.00560.x
31. L. You, H. Zhang, X. Zhu, Y. Xing, J. Zhang, An agent-based model of glucose-lactate metabolic cooperation in cancer: insights into the Warburg effect, *PLoS One*, **12** (2017), e0177325. https://doi.org/10.1371/journal.pone.0177325
32. R. L. Elstrom, D. E. Bauer, M. Buzzai, R. Karnauskas, M. H. Harris, D. R. Plas, et al., Akt stimulates aerobic glycolysis in cancer cells, *Cancer Res.*, **64** (2004), 3892–3899. https://doi.org/10.1158/0008-5472.CAN-03-2904
33. L. Podina, A. Ghodsi, M. Kohandel, Learning chemotherapy drug action via universal physics-informed neural networks, *Pharm. Res.*, **42** (2025), 593–612. https://doi.org/10.1007/s11095-025-03521-8

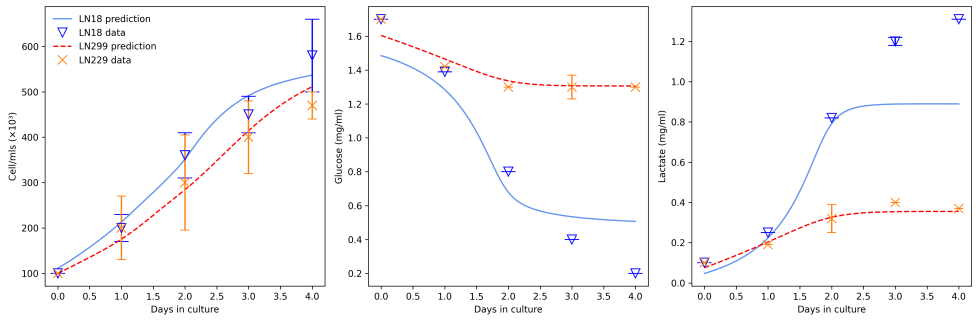
Appendix



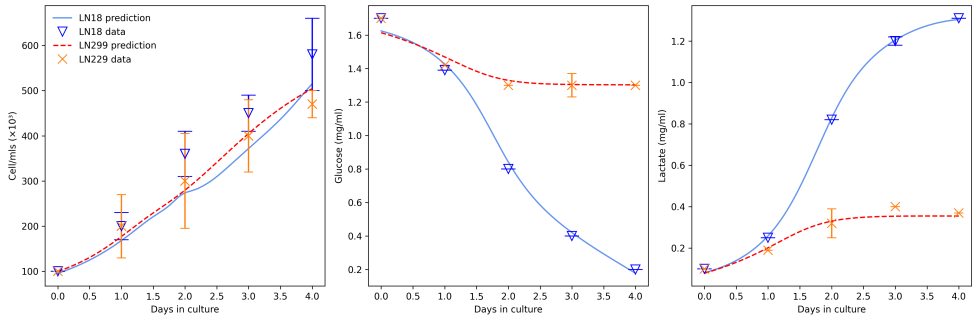
(a) data loss factor = 1.0, pinn loss factor = 1.0



(b) data loss factor = 0.5, pinn loss factor = 1.0



(c) data loss factor = 0.1, pinn loss factor = 1.0



(d) data loss factor = 0.5, pinn loss factor = 2.0

Figure A1. UPINN predictions for total cell accumulation, glucose con- centration, and lactate concentration under a specific weighting configuration for LN229 and LN18 cell lines.

Effect of loss term weighting on UPINN predictions

To further investigate the sensitivity of UPINN predictions to the relative weighting of data fidelity and physical constraints, we conducted a series of experiments by varying the weights associated with the data loss term (λ_a) and the ODE residual loss term (λ_b). The results, summarized in Figure A1 and Table A1, illustrate how different loss factor combinations impact the model's ability to fit experimental data for both LN229 and LN18 cell lines.

Each subfigure in Figure A1 presents UPINN predictions for total cell accumulation, glucose concentration, and lactate concentration under a specific weighting configuration. A progressive decrease in the data loss factor λ_a from 1.0 to 0.1 (with λ_b fixed at 1.0) leads to visibly degraded fits to experimental data, especially in the metabolic variables (glucose and lactate). Conversely, increasing the weight of the ODE constraint term (e.g., $\lambda_b = 2.0$) slightly mitigates the loss of accuracy but cannot fully compensate for low λ_a values.

Table A1 quantitatively confirms these trends. For LN229, reducing λ_a from 1.0 to 0.1 results in nearly doubling the glucose MSE and increasing the lactate MSE, while the MSE for total cell accumulation remains relatively stable. LN18, which is more glycolytically active and thus more sensitive to metabolic modeling, exhibits a sharp increase in error, particularly with $\lambda_a = 0.1$, where the glucose MSE reaches 1.14 and lactate MSE exceeds 7.0. This emphasizes that underweighting the data term can lead to substantial degradation in prediction quality, especially for biologically dynamic variables.

These findings reinforce the need for a well-balanced weighting scheme when applying UPINN to experimental data. The trade-off between data fidelity and mechanistic consistency must be tuned carefully to preserve both the predictive accuracy and the physical plausibility of the model.

Table A1. MSE values obtained using various data and PINN loss factor combinations for LN229 and LN18 cell lines.

Cell line	Loss factors (data/PINN)	MSE (Cell Accumulation)	MSE (G)	MSE (L)
LN229	1.0 / 1.0	1.11×10^{-4}	3.82×10^{-2}	6.60×10^{-2}
	0.5 / 1.0	1.48×10^{-4}	5.19×10^{-2}	7.09×10^{-2}
	0.1 / 1.0	9.84×10^{-4}	7.81×10^{-2}	7.99×10^{-2}
	0.5 / 2.0	7.56×10^{-4}	6.62×10^{-2}	7.63×10^{-2}
LN18	1.0 / 1.0	2.52×10^{-3}	2.93×10^{-2}	7.87×10^{-3}
	0.5 / 1.0	3.93×10^{-3}	4.43×10^{-2}	1.17×10^{-2}
	0.1 / 1.0	1.00×10^{-3}	1.14×10^{0}	7.02×10^{0}
	0.5 / 2.0	5.76×10^{-3}	5.82×10^{-2}	1.44×10^{-2}