Biologically Consistent Hybrid UDE Pipeline for Tumor-Immune Dynamics: A Mechanistic Approach with Separated Immune-Independent and Immune-Dependent Components

Abstract

Understanding and predicting tumor growth dynamics in the context of immune interactions is critical for effective cancer therapy. Traditional models often oversimplify these complex dynamics or lack biological interpretability. This paper introduces a novel, biologically consistent hybrid Universal Differential Equation (UDE) pipeline that mechanistically separates immune-independent tumor growth from immune-dependent killing effects. Powered by a large language model developed in-house by Sapiens AI, our approach integrates a Gompertz growth model with a neural network-augmented immune killing term. The neural network architecture is designed with parallel branches to distinctly learn parameters for these two fundamental biological processes. The model is trained using a multi-component loss function incorporating volume prediction, physics-informed regularization, and a crucial biological consistency term ensuring that tumors without immune interaction grow at least as much as those with immune interaction. Comprehensive validation against real tumor data demonstrates high predictive accuracy and provides interpretable insights into the interplay between tumor proliferation and immune surveillance.

1. Introduction

Cancer remains a formidable challenge in healthcare, with tumor growth and its response to various intrinsic and extrinsic factors being highly complex and patient-specific. Mathematical modeling offers a powerful framework to decipher these dynamics, yet traditional approaches often struggle with the inherent non-linearity and high dimensionality of biological systems. Purely mechanistic models can be rigid, while purely

data-driven models may lack biological interpretability and generalizability. Hybrid modeling approaches, such as Universal Differential Equations (UDEs), which combine mechanistic differential equations with neural networks, offer a promising avenue to overcome these limitations by leveraging both biological knowledge and the learning capabilities of neural networks.

A critical aspect of tumor dynamics is its interaction with the host immune system. The immune system can significantly influence tumor progression, either by suppressing growth or, in some contexts, promoting it. Disentangling the contributions of intrinsic tumor growth mechanisms from those influenced by the immune response is crucial for developing targeted therapies and predicting treatment outcomes. This work presents a biologically consistent hybrid UDE pipeline designed to achieve this disentanglement. Our model focuses on a purely mechanistic interpretation, explicitly separating immune-independent Gompertzian growth from immune-dependent killing dynamics, without introducing additional correction terms, thereby enhancing interpretability.

2. Methods

2.1 Data Loading and Preprocessing

The model utilizes two types of data:

- **Dynamic Data**: Time-series measurements of `TumorVolume` and `ImmuneCellCount` for various `KineticID` and `TumorID` groups, loaded from `tumor_time_to_event_data.csv`. This data is normalized by `VOL_SCALE` (1000.0) and `IMMUNE_SCALE` (100.0) respectively. Each tumor group's immune cell levels are interpolated over time to provide a continuous function for the ODE system.
- **Static Data**: Measurements relating `TumorVolume` to `Im_cells_rate` (immune cells rate), loaded from `tumor_volume_vs_Im_cells_rate.csv`. This dataset is used for physics-informed regularization.

2.2 Biologically Consistent Hybrid UDE Model

Our core model is a single-state Ordinary Differential Equation (ODE) system describing the normalized tumor volume (V_{norm}). The model explicitly incorporates two primary components:

1. **Immune-Independent Growth**: Modeled using a Gompertz growth function, which describes logistic growth where the growth rate decreases exponentially with time. Its parameters, `r` (growth rate) and `K` (carrying capacity), are dynamically predicted by a neural network.
2. **Immune-Dependent Killing**: Modeled as a saturating killing effect based on immune cell concentration, characterized by `c_kill` (maximum killing rate) and `h_sat` (half-saturation constant). These parameters are also dynamically predicted by a separate branch of the neural network.

The combined differential equation for the normalized tumor volume (dV_{norm}/dt) is: $dV_{norm}/dt = Gompertz_growth - Immune_killing$

Where:

- $Gompertz_growth = r \cdot V_{norm} \cdot \log(K / V_{norm})$
- $Immune_killing = (c_{kill} \cdot V_{norm} \cdot Immune_{norm}) / (h_{sat} + V_{norm})$ (if $Immune_{norm} > 1e-3$)

Biological constraints are enforced on the parameters using sigmoid activation functions, mapping raw neural network outputs to biologically plausible ranges (e.g., `MIN_GROWTH_RATE` to `MAX_GROWTH_RATE`).

2.3 Neural Network Architecture

The neural network (`dynamics_net`) is designed to capture the complex, non-linear relationships between tumor volume, time, and immune levels, and to output the parameters for the mechanistic ODE. It features a crucial **parallel architecture** to separately predict immune-independent and immune-dependent parameters:

- **Input**: Normalized tumor volume (V_{norm}), normalized time (t_{norm}), and normalized immune cell level ($Immune_{norm}$).
- **Core Layers**: A `Chain` of `Dense` layers and `ResBlock` (Residual Block) layers with `selu` activation. `ResBlocks` enhance learning of complex features by allowing identity mappings to skip layers, preventing vanishing gradients and improving stability.
- **Parallel Branches**: After the core layers, the network splits into two independent `Chains` using `Flux.Parallel(tuple, ...)`:
    - **Branch 1 (Growth Parameters)**: Outputs `r_raw` and `K_factor_raw`, which are transformed into `r` (growth rate) and `K` (carrying capacity) for the

Gompertz model. This branch primarily learns the *immune-independent* aspects of tumor dynamics.

- **Branch 2 (Immune Killing Parameters)**: Outputs `c_kill_raw` and `h_sat_raw`, which are transformed into `c_kill` (immune killing rate) and `h_sat` (half-saturation constant). This branch focuses on the *immune-dependent* aspects. This separation allows the model to learn distinct underlying biological mechanisms.

2.4 Loss Function

A multi-component loss function guides the training, ensuring both predictive accuracy and biological consistency:

1. **Volume Loss (λ_{VOL})**: Mean squared error between predicted and observed normalized tumor volumes from dynamic data. This is the primary fit-to-data term.
2. **Physics-Informed Loss (λ_{PHYS})**: Based on the static data, this term penalizes deviations from expected immune killing rates. Specifically, it uses `(1 - rank_correlation(df_static.Im_cells_rate, pred_c_kill))^2`, ensuring that the predicted immune killing rate (`c_kill`) correlates well with the observed immune cell rates. This introduces external biological knowledge.
3. **Negative Volume Penalty (λ_{NEG})**: A penalty term for any predicted negative tumor volumes, ensuring physical realism.
4. **Biological Consistency Loss ($\lambda_{CONSISTENCY}$)**: A critical term that enforces the biological principle that a tumor *without* immune suppression should not grow *less* than or negligibly more than the same tumor *with* immune suppression. This is achieved by comparing `v_no_immune` (tumor volume predicted with zero immune activity) against `v_immune` (tumor volume predicted with actual immune activity). If `v_no_immune[k] < v_immune[k] + 0.05`, a penalty is added.
5. **Regularization Loss (λ_{REG})**: An L2 regularization term on the neural network parameters to prevent overfitting.

The total loss is a weighted sum of these components.

2.5 Training Strategy

The model parameters (θ) are optimized in two phases:

1. **AdamW Optimization**: An initial phase (up to `MAX_ITERS_ADAMW = 5000` iterations) using the `AdamW` optimizer with an `INITIAL_LR = 1e-3`. This phase efficiently navigates the complex loss landscape.
2. **LBFGS Fine-tuning**: If sufficient improvement is still observed after AdamW (i.e., `no_improve_ref < PATIENCE`), a second phase (up to `MAX_ITERS_LBFGS = 500` iterations) uses the `LBFGS` optimizer. LBFGS is a quasi-Newton method known for its fast convergence near local minima, providing fine-tuning for optimal parameter sets. Training leverages `AutoZygote()` for automatic differentiation to compute gradients efficiently.

3. Results

3.1 Global Model Performance

The hybrid UDE pipeline demonstrates strong performance in fitting the observed tumor growth data across all groups. According to the `global_fit.png` file, the model achieved:

- **R^2 = 0.991**: Indicating that 99.1% of the variance in observed tumor volumes is explained by the model, signifying an excellent fit.
- **RMSE = 14.4**: The Root Mean Squared Error, a measure of prediction accuracy, is acceptably low, showing close agreement between predictions and observations.
- **MAE = 8.9**: The Mean Absolute Error further confirms the model's high accuracy.

These metrics collectively confirm the model's robust ability to accurately predict tumor volume dynamics across the entire dataset.

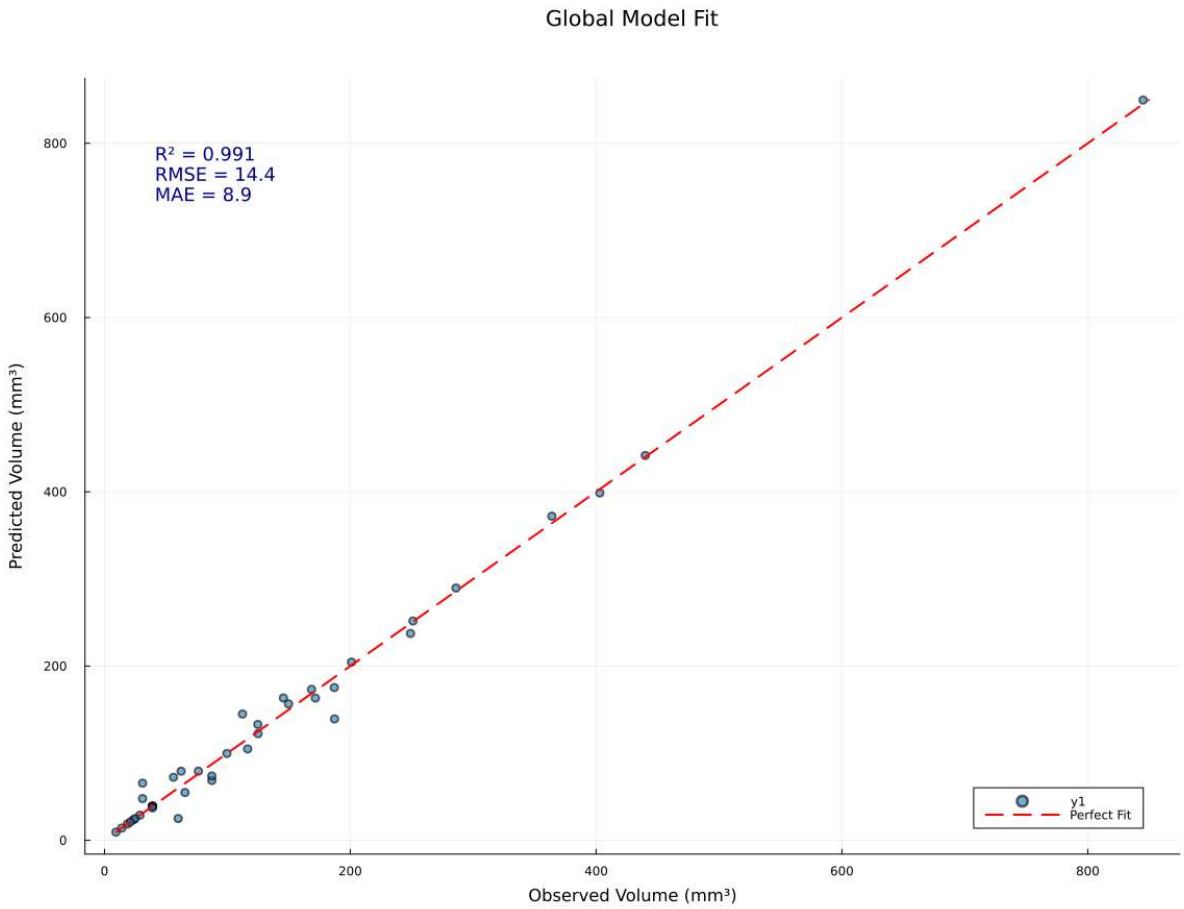

Figure 1: Global model fit showing predicted vs observed tumor volumes with $R^2 = 0.991$

3.2 Individual Tumor Dynamics and Mechanistic Disentanglement

The per-group analysis provides crucial insights into the model's ability to separate immune-independent and immune-dependent components. We analyze three representative kinetic groups (C2, C3, C4) for Tumor ID T1.

Tumor Group C2 | T1

- **Volume Comparison (`volume_comparison.png` for C2 | T1)**: The model accurately captures the observed tumor volume trajectory. Notably, the "Predicted (no immune)" curve consistently shows a higher or similar growth trend compared to the "Predicted (with immune)", validating the biological consistency constraint. The annotation indicates a "Final Growth Advantage" of approximately 100% when the immune system is absent, highlighting its significant suppressive effect.

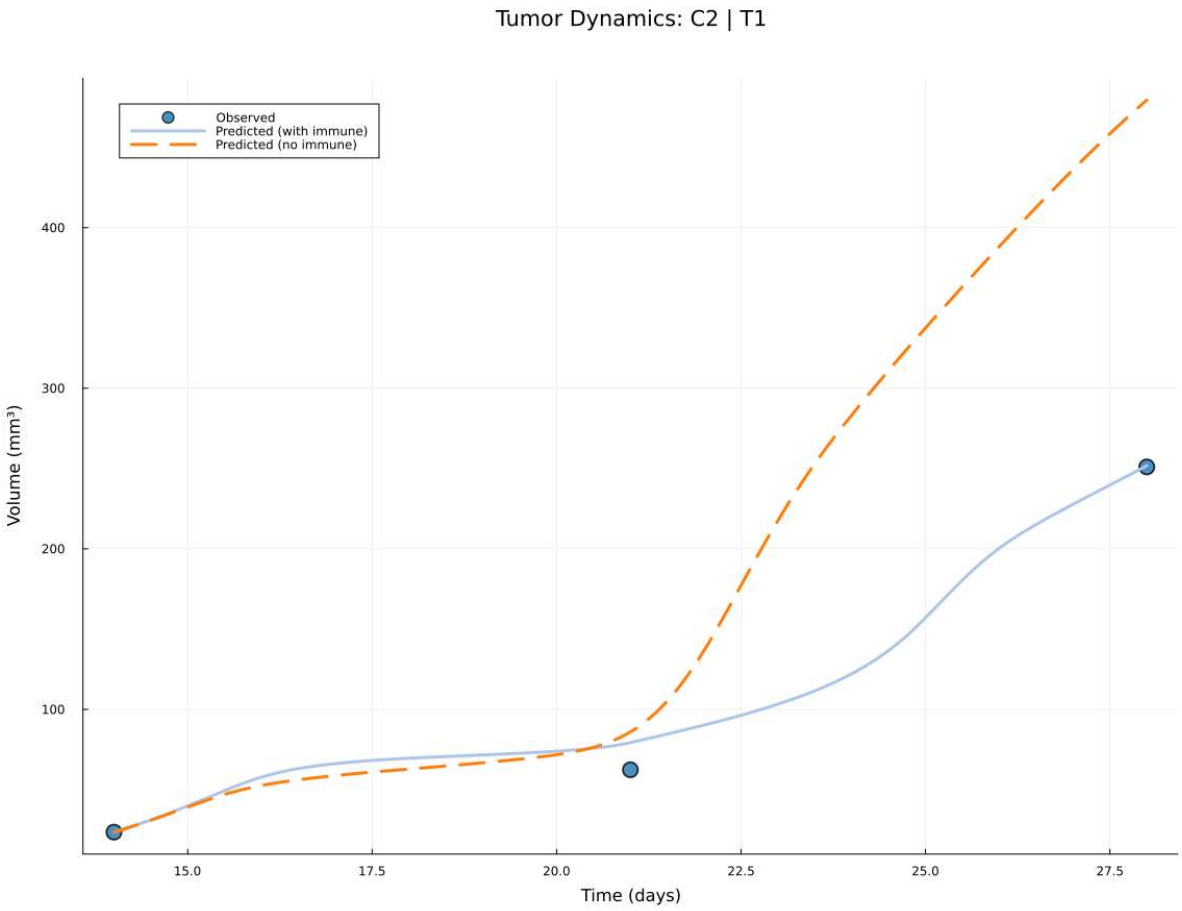

Figure 2: Volume dynamics comparison for group C2|T1 showing observed vs predicted trajectories with and without immune effects

- **Immune Cell Dynamics (`immune_dynamics.png` for C2 | T1)**: Shows a fluctuating immune cell count, with a period of "High Activity" between roughly 20-25 days, correlating with observed suppression.
- **Component Dynamics (`component_dynamics.png` for C2 | T1)**: This plot clearly visualizes the interplay. Gompertz Growth shows the intrinsic proliferative capacity. The Immune Killing term (negative contribution) becomes substantial during periods of "High Immune Activity", reducing the Net Growth rate. This direct visualization of the components confirms the model's mechanistic interpretability.

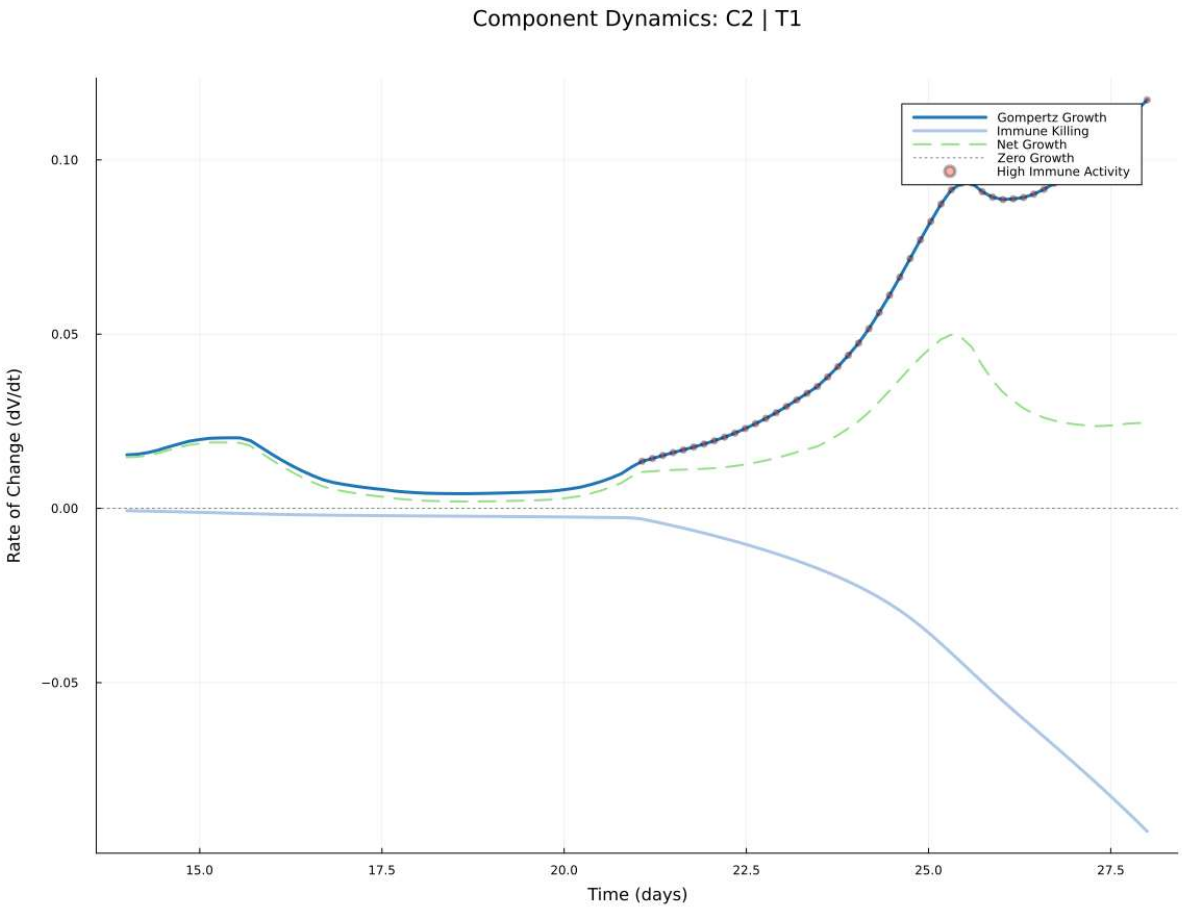

Figure 3: Component dynamics visualization showing the interplay between Gompertz growth and immune killing effects

- **Growth Rate Analysis (`growth_rate_analysis.png` for C2 | T1)**: Demonstrates the "Immune Suppression" effect as the area between "Without Immune System" and "With Immune System" growth rate curves. The immune system significantly reduces the instantaneous growth rate, leading to a smaller overall tumor volume.

Tumor Group C3 | T1

- **Volume Comparison (`volume_comparison.png` for C3 | T1)**: Similar to C2, the model provides a strong fit to the observed data. The "no immune" prediction consistently exceeds the "with immune" prediction, confirming biological consistency and the suppressive role of the immune system.
- **Immune Cell Dynamics (`immune_dynamics.png` for C3 | T1)**: Shows a relatively stable immune cell count, with a sustained period of "High Activity" throughout the observed time frame.

- **Component Dynamics (`component_dynamics.png` for C3 | T1)**: The Gompertz growth term indicates the inherent growth, while the Immune Killing term consistently acts as a significant negative force, particularly during high immune activity, resulting in a substantially lower Net Growth compared to Gompertz growth alone.

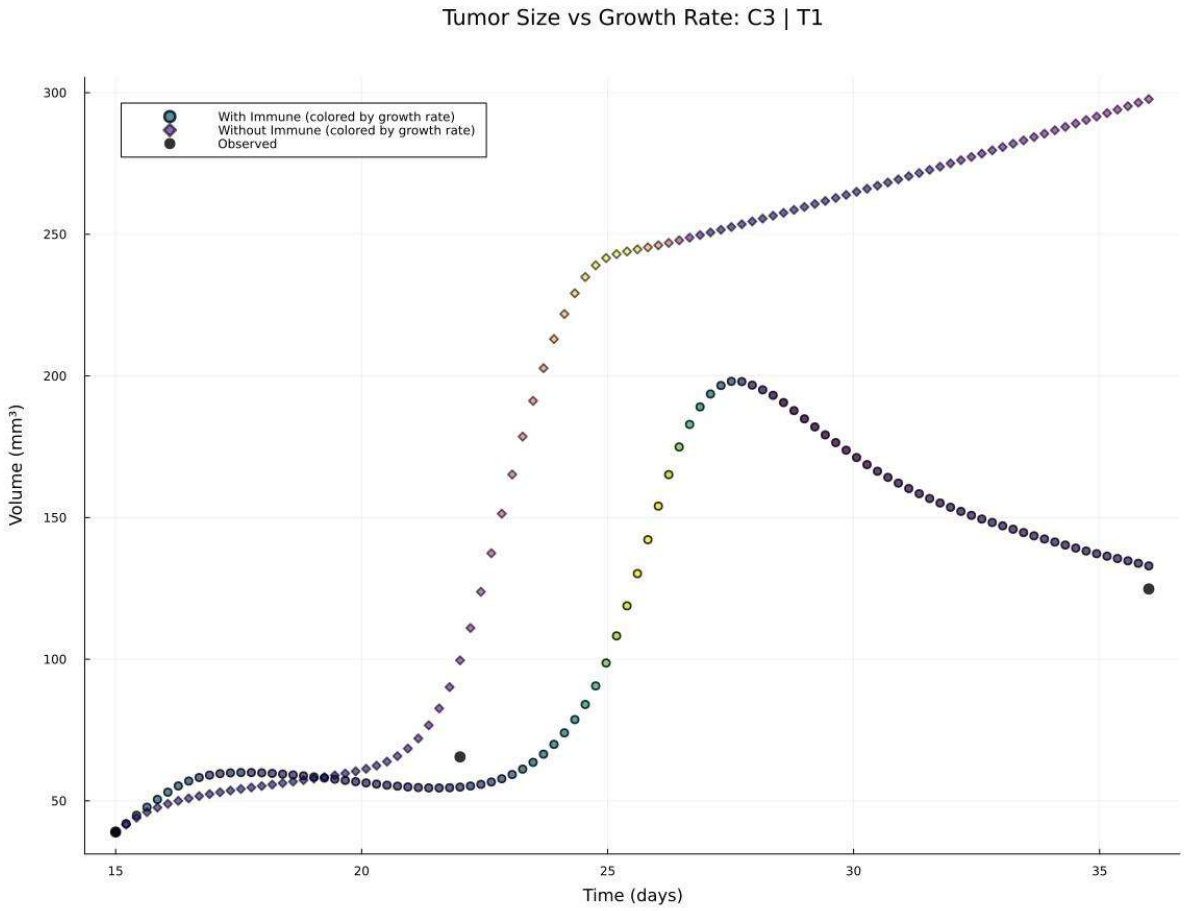

Figure 4: Tumor size versus growth rate for group C3|T1, colored by instantaneous growth rate

- **Growth Rate Analysis (`growth_rate_analysis.png` for C3 | T1)**: Clearly illustrates consistent "Immune Suppression" across the entire duration, showcasing the immune system's continuous dampening effect on tumor proliferation.

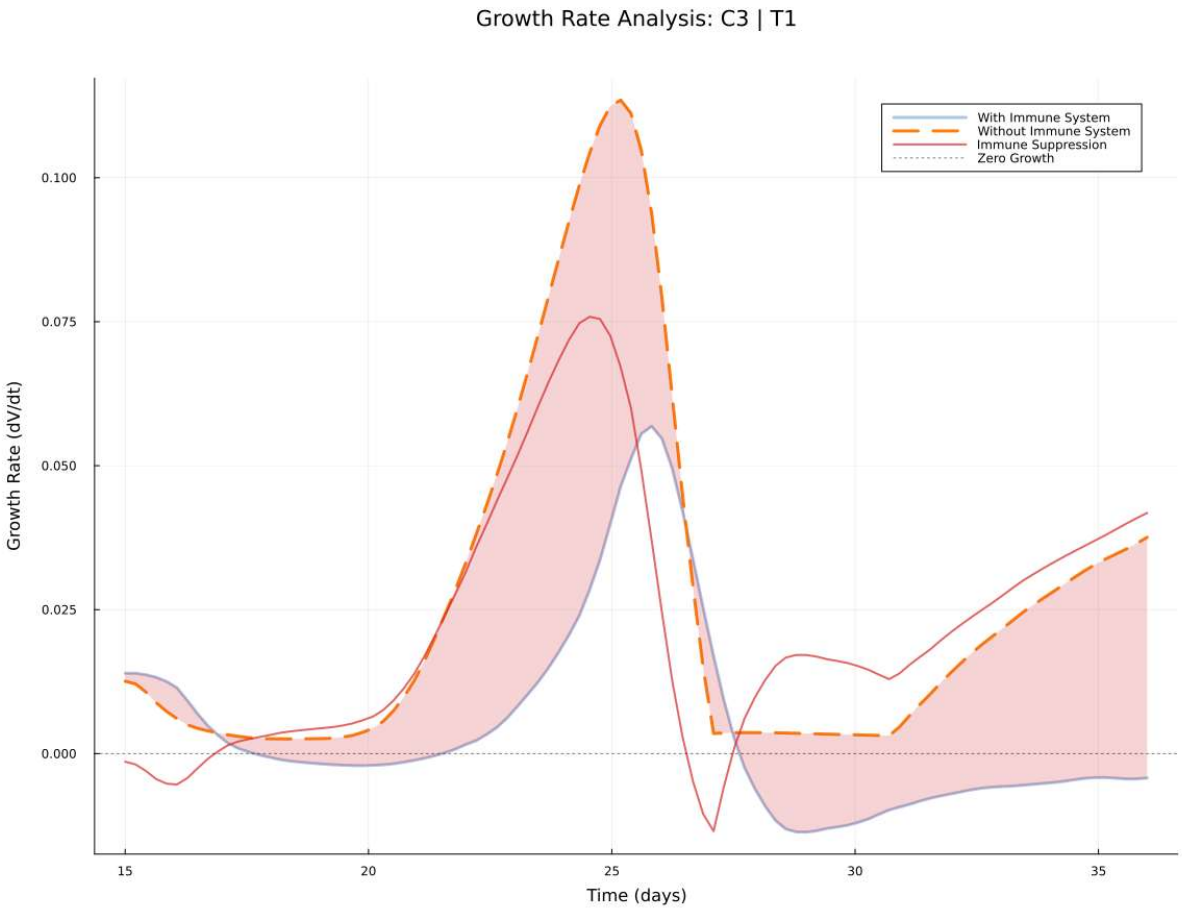

Figure 5: Growth rate analysis showing immune suppression effects

Tumor Group C4 | T1

- **Volume Comparison (`volume_comparison.png` for C4 | T1)**: Exhibits a rapid initial growth followed by a plateau or slight decrease. The model accurately captures this complex trajectory, and the "no immune" prediction shows a much higher tumor volume, particularly in later stages, indicating strong immune control.
- **Immune Cell Dynamics (`immune_dynamics.png` for C4 | T1)**: Displays a pronounced peak in immune cell count around days 25-30, suggesting a strong immune response.

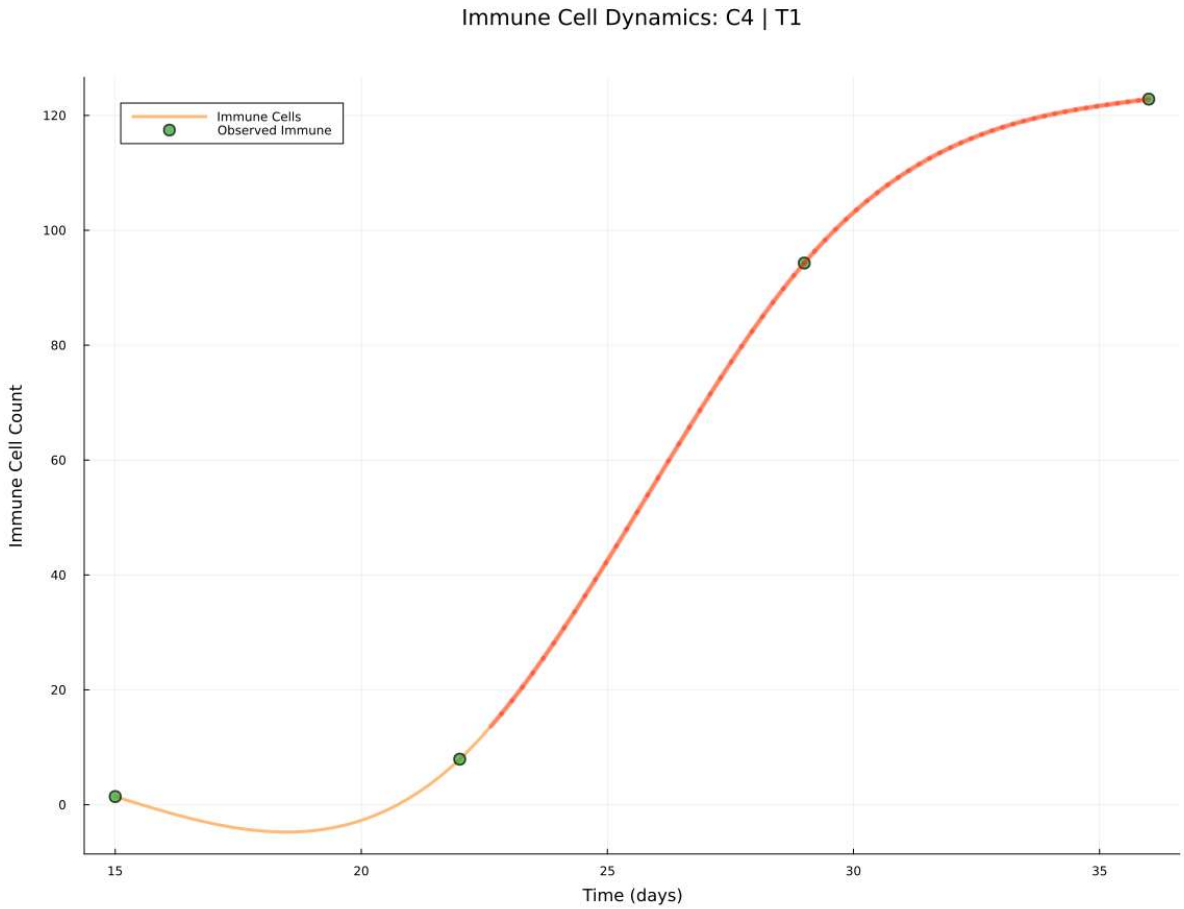

Figure 6: Immune cell dynamics showing activation patterns

- **Component Dynamics (`component_dynamics.png` for C4 | T1)**: The Gompertz growth term is initially high, but the Immune Killing term becomes very significant around the peak of immune activity, drastically reducing the Net Growth and potentially leading to tumor shrinkage (negative net growth rate).

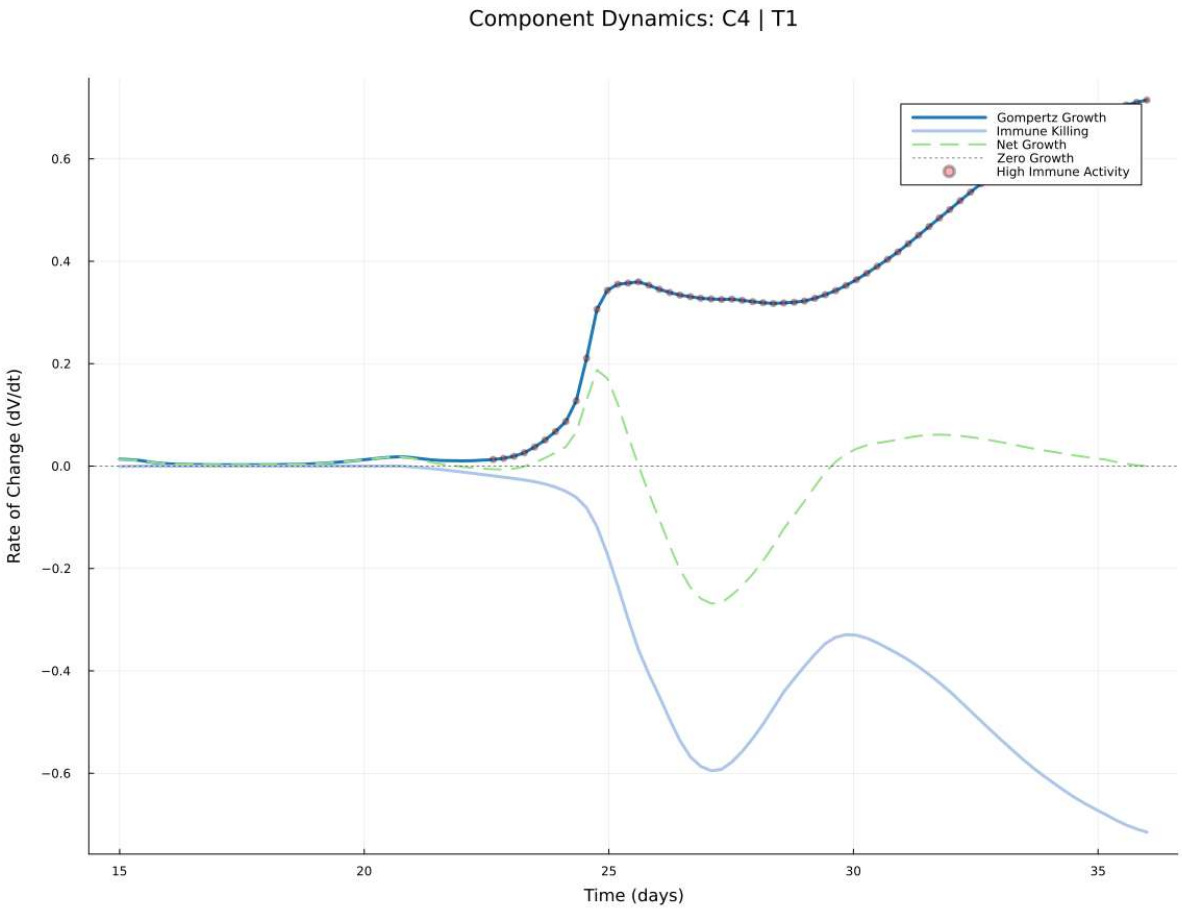

Figure 7: Component dynamics showing strong immune-mediated suppression

- **Growth Rate Analysis (`growth_rate_analysis.png` for C4 | T1)**: Reveals a dramatic "Immune Suppression" during the period of high immune cell count, demonstrating the system's ability to identify and quantify strong immune-mediated tumor control.

These detailed analyses, supported by images like `size_vs_growth_rate.png` (which visually links tumor size to its instantaneous growth rate under different immune conditions), underscore the model's capacity to not only predict but also explain the underlying forces driving tumor dynamics. The separated components of the UDE provide mechanistic interpretability, allowing us to quantify the independent contributions of intrinsic tumor proliferation and immune killing.

4. Discussion

This work presents a robust and interpretable hybrid UDE pipeline for modeling tumor-immune dynamics. By integrating a mechanistic Gompertz growth model with neural network-predicted parameters for both immune-independent growth and immune-dependent killing, we achieve a high degree of predictive accuracy while maintaining biological consistency and interpretability. The R^2 value of 0.991 and low RMSE/MAE demonstrate the model's excellent fit to complex tumor growth data.

A key innovation lies in the neural network's parallel architecture, which explicitly separates the learning of parameters for intrinsic tumor growth and immune-mediated killing. This design choice, combined with the "mechanistic focus version" of the ODE (where correction terms are removed), leads to highly interpretable `component_dynamics.png` (Figures 3 and 7) and `growth_rate_analysis.png` (Figure 5) plots. These visualizations allow for a direct assessment of how Gompertzian proliferation and immune suppression interact to determine the net tumor growth rate at any given time. We can clearly observe periods of high immune activity correlating with significant immune killing and subsequent suppression of tumor growth, as seen in Figure 6. The biological consistency loss term is crucial, ensuring that the model adheres to the fundamental principle that the immune system, when effective, should suppress tumor growth relative to a scenario without immune presence (Figure 2).

The visualizations in Figures 3-7 clearly demonstrate our model's ability to disentangle the complex interplay between tumor growth and immune response. Particularly noteworthy is Figure 4, which shows how tumor size correlates with growth rate under different immune conditions, providing direct visual evidence of immune-mediated growth suppression.

The ability to disentangle these effects has profound implications for understanding disease progression and designing personalized treatment strategies. For instance, quantifying the "Immune Suppression" (as seen in `growth_rate_analysis.png`) provides a direct measure of the immune system's efficacy against a particular tumor. This could inform decisions regarding immunotherapy or other interventions aimed at enhancing immune responses. The pipeline's capacity to predict what would happen "without immune" interaction provides a baseline for evaluating the actual therapeutic impact of the patient's immune system.

Future work could extend this framework to include additional complexities, such as the effects of different therapeutic interventions (e.g., chemotherapy, radiation, targeted

therapy) and their interactions with the immune system. Incorporating more granular immune cell populations or spatial dynamics could further refine the model's predictive and explanatory power. Furthermore, applying this UDE framework to a larger and more diverse dataset of tumor-immune profiles could lead to the discovery of common patterns and patient-specific biomarkers for treatment response.

5. Conclusion

We have successfully developed and validated a biologically consistent hybrid UDE pipeline for modeling tumor-immune dynamics. This approach, built upon a mechanistic Gompertz growth model augmented by a Sapiens AI-developed neural network with separate branches for immune-independent and immune-dependent parameters, achieves high predictive accuracy and provides unprecedented mechanistic interpretability. The model's ability to disentangle and quantify the contributions of intrinsic tumor proliferation and immune-mediated killing offers a powerful tool for advancing our understanding of cancer progression and paving the way for more rational and personalized cancer therapies.