Appl. Comput. Math., Vol.25, No.2, Special Issue, 2026, pp.316-331
DOI: 10.30546/1683-6154.25.2.2026.316

DEEP LEARNING ENABLED DIGITAL TWIN IN BREAST CANCER TREATMENT WITH SMART CYBER-PHYSICAL SYSTEM FOR ADAPTIVE THERAPY OPTIMIZATION AND TUMOR EVOLUTION FORECASTING

Y. QIAO[1], X. LI[1], L. ZHAO[1], C. LIU[2], S.B. SUN[1,*]

Abstract. This paper incorporates the Digital Twin (DT) technology, Deep Learning (DL), and Cyber-Physical Systems (CPS) to facilitate change in therapeutic approaches. To predict the evolution of the tumor and response to treatment based on continuously evolving clinical data, a hybrid mechanistic deep learning model is implemented to create a patient-specific DT. According to these predictions, a Deep Reinforcement Learning (DRL) agent is employed to determine optimal, individualized dosing regimens that will maximize tumor ablation and reduce systemic toxicity. The CPS applies such decisions in a closed-loop sensethink act methodology, which allows real-time adaptive control. In silico results demonstrate prolonged time-to-progression and reduced cumulative toxicity compared to conventional treatment strategies, highlighting the potential of the Digital Twin–Cyber-Physical System (DTCPS) systems in advancing personalized oncology.

Keywords: Digital Twin, Deep Learning, Breast Cancer, Cyber-Physical Systems, Adaptive Therapy, Personalized Medicine, Tumor Evolution Forecasting.

AMS Subject Classification: 92C50, 68T07, 93E20, 92C42.

1. Introduction

Breast cancer remains a leading cause of cancer-related mortality among women all over the world mostly because of its extreme heterogeneity. This complexity occurs on various levels, including genetic mutations and molecular subtypes, and differences between the tumor microenvironment of patients (inter-tumor) and within the tumor of individual patients over time (intra-tumor). There are standard therapeutic approaches which commonly use maximum tolerated dose (MTD) chemotherapy administered on fixed schedules derived from population-level data and are literally unprepared towards the dynamic and adaptive nature of individual disease progression. Such an inherent discord between a fixed treatment program and a living biological system often causes the evolution and growth of drug-resistant cell populations, leading to a treatment failure and disease relapse.

The need to defeat these obstacles has initiated a paradigm shift towards precision and personalized medicine, a strategy that aims at tailoring therapeutic interventions to individual patient characteristics. This vision, though, does not only entail a thorough knowledge of the biology of a person, but also the technology needed to track, anticipate, and adjust to the changes in it in real time. The integration of three strong technologies Digital Twins (DTs), DL, and CPS provides a way to achieve this dream.

[1]General Surgery, Hongqi Hospital Affiliated to Mudanjiang Medical University, Mudanjiang 157000, China
e-mail: 19545303079@163.com, starinnight@126.com 342579584@163.com
[2]Department of Pharmacy, The Third Hospital of Heilongjiang, Heihe 164000, China
e-mail: lc8676263@163.com
*Corresponding author e-mail: sun.shaobin@163.com
Manuscript received 14 November 2025.

A Digital Twin in healthcare refers to an active, digitalized model of a patient, a living computational model that is constantly updated with real-world data (e.g., medical imaging, genomic sequencing, clinical biomarkers, etc.) A DT, unlike a static predictive model, offers a simulated, longitudinal, predictive environment of an individual so that a clinician can run in-silico experiments, test therapeutic scenarios of any kind, and predict future health states without putting the real-life patient at risk. DL is the intellectual processing that drives the DT. DL algorithms, especially Convolutional and Recurrent Neural Network architectures, are capable of analyzing of large, multi-modal data unprecedented and the ability to identify complex and non-linear patterns that are not captured by traditional methods of analysis. Combining imaging, genomic, and clinical data, the DL models are capable of learning patient-specific tumor behavior signatures, predict treatment response, and predict disease progression trajectories with higher accuracy.

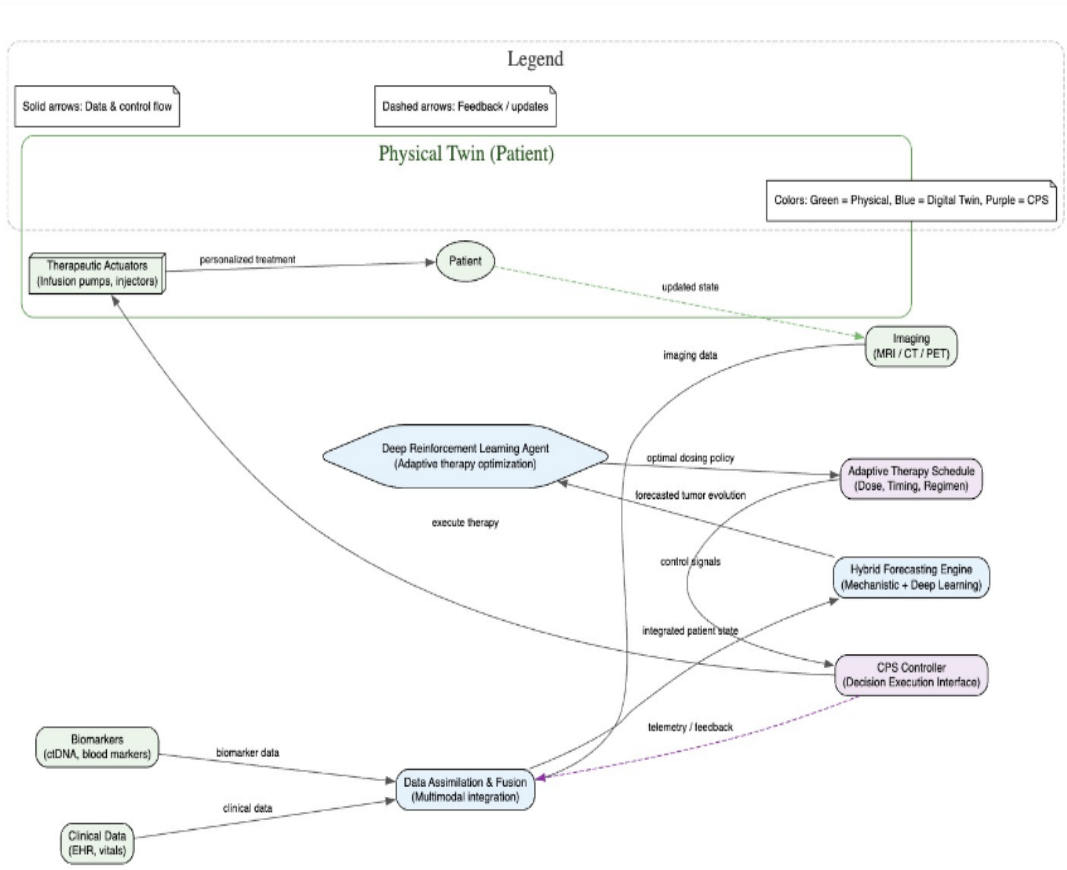


FIGURE 1. DL enabled digital twin in breast cancer treatment with smart CPS for adaptive therapy optimization and tumor evolution forecasting.

Though a DL-powered DT may deliver actionable predictive and prescriptive insights, a mechanism is necessary to translate these insights into timely clinical actions. This mechanism is offered in a smart CPS. A CPS is a designed system that establishes a tight feedback loop between physical processes and computational algorithms. In medical terminology, the physical component is the physiology of the patient and the instruments and equipment used to deliver therapeutic agents to the patient (e.g. infusion pumps), whereas the cyber (computational) center is the DT. This combination creates a closed-loop system capable of autonomous, real-time control that can monitor the condition of the patient, use the DT to determine an optimal course of action and then execute that action upon the patient. The intersection of these technologies makes a basic shift in clinical decision-making, where it becomes not a descriptive paradigm

(what is the tumor doing?), a predictive paradigm (what will the tumor do?), but a prescriptive and ultimately autonomous paradigm (what is the best thing to do, and how can it be done automatically?).

The suggested work proposes a new framework where integrates three key technological components to develop a closed-loop system of adaptive breast cancer therapy. The system uses hybrid mechanistic-DL Digital Twin to predict tumor response to neoadjuvant chemotherapy. This DT is incorporated into a smart CPS, in which a deep reinforcement learning agent has the ability to continuously optimize the therapeutic regimen towards the preclusion of resistance and the reduction of toxicity. The conceptual architecture of the proposed system will be indicated in the graphical abstract presented abowe, Fig.1.

Contribution in Context. Recent DRL-based adaptive therapy studies have demonstrated the potential of reinforcement learning for dose modulation using mechanistic or simulated tumor environments. The present work extends this line of research by embedding the DRL controller within a hybrid Digital Twin framework that continuously personalizes tumor dynamics through deep learning and situates decision-making within a CPS-oriented closed-loop architecture. This integration distinguishes the proposed approach from prior work that primarily focuses on standalone optimization without explicit digital twin or CPS coupling.

Key Contributions. The key contributions of the proposed work are summarized as follows:

- **Integrated DT-CPS Architecture:** The creation of an innovative, fully integrated DT-CPS architecture for closed-loop, adaptive cancer treatment.
- **Hybrid Forecasting Model:** A hybrid forecasting approach that combines the interpretability of mechanistic ordinary differential equations (ODEs) with the adaptive, data-driven capabilities of deep learning to generate personalized tumor evolution predictions.
- **Deep Reinforcement Learning Engine:** The development of a DRL engine for real-time optimization of chemotherapy dosing regimens.
- **Extensive Simulated Validation:** An extensive simulated evaluation of the proposed framework demonstrating improved therapeutic outcomes compared to a simulated standard-of-care treatment.

The subsequent sections of this paper elaborate on the proposed framework in detail. Section 2 assesses the related literature and theoretical background. Section 3 describes the framework's architecture, predictive modeling, and adaptive control design. Section 4 provides in-silico results, compares adaptive and fixed-dose regimens, and discusses the clinical implications. Section 5 concludes with the key findings and directions for future study.

2. Related works

The detailed consideration of the proposed structure requires a review of foundational and recent research in the areas that the structure includes: Digital twins in cancer, medical data analysis with deep learning, Internet of Things in the healthcare sphere, and the concepts of adaptive therapy.

2.1. The Evolution of Digital Twins in Oncology.

The idea of a digital twin, which can be traced back to aerospace and manufacturing and is used to monitor and simulate complex physical assets, has recently become very popular in healthcare [1]. The models are commonly known in oncology as cancer avatars or virtual patients. Initial applications were directed at the development of virtual models based on the baseline patient data including tumor imaging and genomic profiles to recreate the initial reaction to treatment [2, 3, 12]. These model types have been used in different malignancies such as gliomas, lung cancer, and breast cancer, mostly in treatment planning and prognostic measurement.

One of the most important developments in the direction is the differentiation between dynamic and static digital twins. Static digital twins are constructed using one data snapshot at pre-treatment [4]. Although they are useful in early planning, they lack long-term predictive capability because they cannot predict the adaptation of the tumor to therapy or the changes in microenvironment. Conversely, the dynamic digital twins support adaptive decision-making are expected to constantly incorporate longitudinal patient data, including serial imaging scans, circulating tumor DNA (ctDNA) kinetics, and biomarker trends, to re-establish the internal models. This never-ending learning will enable the dynamic twin to reflect the changing condition of the patient so that it can more effectively predict long-term and act out of adaptability [5, 7, 14]. Recent reviews point to the increasing publications of this field, which has become clear evidence of a rapid maturation of the field and its potential to transform personalized medicine by providing the possibility of in-silico clinical trials and real-time decision support.

2.2. **DL in Medical Data Analysis.** Deep learning has emerged as an essential tool to glean useful insights out of the high-dimensional, heterogeneous data of modern medicine. Its uses in oncology are widespread and transformative [6]. Convolutional Neural Networks (CNNs) have become the new standard in medical image analysis, which has the potential to automatically segment tumors from MRI or CT scans and generate quantitative "radiomic" features that are associated with underlying tumor biology and prognosis. Such networks are able to detect faint patterns in imaging data which are generally imperceptible to the human eye, which can give a more detailed characterization of the tumor phenotype [8, 9, 16].

Recurrent Neural Networks (RNNs) and their more advanced implementation, Long Short-Term Memory (LSTM) networks are especially appropriate in modeling dynamic processes, including tumor growth or response to therapy over time. These architectures are created to deal with sequential data, which enables them to learn temporal dependence of longitudinal clinical measurements. One of a major frontier in medical AI is the creation of multi-modal fusion models [10]. These models recognize that no single source of data will be able to capture the full complexity of cancer, as such, combining information in disparate streams, including imaging, genomics, pathology reports and electronic health records, are used to construct a more predictive and robust model [11]. By capturing complex interactions between these various data types, multi-modal architectures have been shown to perform better in tasks such as cancer prognosis and prediction of response to treatment than models that use one source of data.

2.3. **Cyber-Physical Systems in Life-Critical Applications.** CPS represent the tight integration of computation, networking and physical processes. In contrast to the traditional embedded systems, in which computation is a separate aspect, CPS implies the interaction of feedback loops, namely, computational decisions impact physical actions, whereas physical processes impact future computations [17]. A CPS core architecture is composed of sensors that observe the physical world, actuators that alter the physical world, a network of communication and a computational core that implements control logic.

In the medical field, such systems are referred to as Medical Cyber-Physical Systems (MCPS) and they can be life-critical. Representative applications include are artificial closed-loop pancreas to manage diabetes, which continuously monitors blood glucose levels, and automatically delivers insulin, and modern anesthesia systems, which analyze physiological parameters in real-time and automatically adjust the amount of drug delivered [13]. CPS use in the medical field is likely to improve patient safety, automatize routine operations, and allow more specific control of the therapeutic process. An MCPS development is however quite a challenge. These entail maintaining high assurance software reliability, ensuring interoperability between devices of various manufacturers, high-level security against cyber threats, which may cause physical damage, and coming up with verifiable safety measures of autonomous decision-making. The proposed framework is based on this MCPS paradigm by modeling the interaction between the patient and the therapy as a closed-loop control problem. In this study, the CPS framework is presented as a

conceptual and architectural abstraction rather than a fully deployed clinical system. Practical real-world integration would require validated sensing modalities (e.g., imaging, ctDNA, wearable toxicity monitoring), certified actuation mechanisms (e.g., programmable infusion pumps), secure communication protocols, and clinician-in-the-loop supervision layers. These components form a roadmap for future translational implementation of the proposed DTCPS paradigm.

2.4. **Adaptive Therapy and Tumor Evolution Modeling.** Adaptive therapy (AT) is grounded in evolutionary game theory. The high and sustained selective pressure of standard MTD chemotherapy is very effective in killing cancer cells sensitive to drugs but unintentionally causes the growth of pre-existing or newly acquired resistant cells [15]. Instead, adaptive therapy tries to control and not eliminate the tumor. It uses intermittent or modulated dose to preserve a population of drug sensitive cells at an appreciable level. These sensitive cells are usually more fit without the drug, and can competitively inhibit the growth of the less fit resistant population. The purpose of this strategy is to extend the therapeutic control by averting or postponing the occurrence of extensive resistance.

Mathematical models that would be able to describe the dynamics of tumor evolution are essential in the design of effective AT strategies. Many models of cell interactions between various cell populations (e.g., sensitive, resistant), between the immune system and therapeutic agents are often modeled in a mechanistic form: as a system of ordinary or partial differential equations (ODEs/PDEs), equations. These models give a cause-effect interpretable design of simulating tumor growth and response. More recently, machine learning methods, especially DRL, have been investigated in order to learn the best AT policies directly out of data or simulations. By exploring a very large library of possible dosing schedules, DRL agents can discover strategies which are complex and non-intuitive (not founded on intuitive rules), thus outperforming simple, rule-based adaptive strategies on a long-term clinical goal, such as time to progression [18]. The proposed methodology is a combination of these methods; whereby, a mechanistic model is used as the base and a DRL agent is employed as the optimization engine.

TABLE 1. Synthesis of enabling technologies in precision oncology.

Technology/Method	Core Features	Key Limitations
Convolutional Neural Networks (CNNs)	Spatial feature extraction from grid-like data (e.g., images) [9]. High performance in segmentation, classification, and radiomics [1,13].	Requires large, labeled datasets for training. Can be computationally expensive. Often considered a "black box" with limited interpretability.
Recurrent Neural Networks (LSTMs)	Models temporal dependencies in sequential data (e.g., time-series biomarkers, clinical text) [8]. Captures long-range patterns.	Can be difficult to train (vanishing/exploding gradients). Performance depends on the quality and frequency of time-series data.
Mechanistic Models (ODEs/PDEs)	Provides a causal, interpretable representation of biological processes, such as evolutionary game theory [5,18]. Parameters have direct physical meaning. Allows for *in-silico* hypothesis testing.	May oversimplify complex biological reality. Parameterization can be challenging and patient-specific.
Deep Reinforcement Learning (DRL)	Learns optimal decision-making policies (e.g., for adaptive therapy) through trial-and-error interaction with an environment [15]. Capable of discovering complex, long-term strategies.	Requires a well-defined reward function and an accurate simulation environment. Training can be computationally intensive and unstable.
Cyber-Physical Systems (CPS)	Enables real-time, closed-loop interaction between computational intelligence and the physical world [10]. Facilitates automation and autonomous control.	High development complexity. Requires stringent safety, security, and reliability verification. Interoperability standards are still emerging.

For this hybrid model to be operationalized in a clinically relevant and adaptive manner, an integrated combination of these models within a DT paradigm provides a promising approach. A DT represents a continuously updated computational representation of a patients tumor, based on multi-modal clinical, genomic, and imaging data. As part of a Medical Cyber-Physical System (MCPS) architecture, the DT can function in a closed-loop, real-time manner with the patient, using patient-derived data predicting the tumor progression with the mechanistic-deep learning hybrid model, and utilizing the DRL-based control engine to offer recommendations for the optimal drug dosing for the patient. This real-time interactivity between the virtual representation of the tumor and the patient creates a level of personalization and adaptiveness that will not be accomplished through static treatment paradigms, working towards evidence-based intelligent cancer treatment.

Table 1 summarizes the key technologies supporting the proposed DT-CPS frameworkfrom deep learning (CNNs, LSTMs) to mechanistic modeling, reinforcement learning, and CPSdetailing their core functions, benefits, and limitations. Each technology contributes a distinct capability, such as the predictive intelligence of deep learning, the interpretability of mechanistic models, or the closed-loop adaptability of CPS. Together, they enable the overall goal of real-time, adaptive, and personalized treatment optimization.

3. Problem statement and research objectives

Problem Context. Common, off-the-shelf therapies for breast cancer, typically based on fixed-dose and predetermined schedules, are ill-adapted to address the dynamic and adaptive nature of cancer evolution. This mismatch often leads to the development of therapeutic resistance, unnecessary toxicity, and suboptimal patient outcomes. Addressing this gap is of critical importance due to the absence of a therapeutic paradigm capable of perceiving, predicting, and responding to patient-specific tumor dynamics in real-time.

3.1. **Research Objectives.**

To address this challenge, the following research objectives are proposed:

(1) **Predictive Modeling:** To create and implement a hybrid mathematicaldeep learning model capable of predicting the evolution of breast cancer tumors in individual patients undergoing neoadjuvant chemotherapy.
(2) **Adaptive Therapy Optimization:** To develop an adaptive therapy optimization engine, based on DRL, that dynamically identifies optimal drug dosing regimens with the dual objectives of maximizing tumor control and minimizing toxicity.
(3) **Digital Twin Integration:** To integrate the prediction and control components into a unified DT framework embedded within a smart CPS architecture for closed-loop therapeutic control.
(4) **In-Silico Validation:** To evaluate the effectiveness of the proposed DTCPS framework through in-silico simulations, comparing its efficacy against a simulated standard-of-care MTD regimen.

4. Methodology

The proposed approach is designed as a multi-layer architecture based on the sensethinkact concept of a CPS. The architecture consists of three main components:

(1) **The Digital Twin Core** a hybrid, personalized forecasting model of the patients tumor dynamics.
(2) **The Adaptive Therapy Optimization Engine** a DRL module that computes optimal treatment actions.

(3) **The Smart CPS Loop** orchestrates the flow of information and control between the cyber and physical domains.

4.1. **The Digital Twin Core: A Hybrid Forecasting Model.** At the heart of the digital twin is a hybrid model that leverages both the interpretability of mechanistic modeling and the adaptive capacity of deep learning - providing a causal biologic basis, while allowing real time personalization using incoming data streams.

The mechanistic backbone is a system of coupled Ordinary Differential Equations (ODEs), which models the population dynamics of sensitive and resistant tumor cells, as well as drug pharmacokinetics and a simplified immune response:

$$\frac{dS}{dt} = \rho_S S\left(1 - \frac{S+R}{K}\right) - d_S C(t) S - k_{IS} I S, \tag{1}$$

$$\frac{dR}{dt} = \rho_R R\left(1 - \frac{S+R}{K}\right) - d_R C(t) R - k_{IR} I R, \tag{2}$$

$$\frac{dC}{dt} = U(t) - \delta_C C, \tag{3}$$

$$\frac{dI}{dt} = \sigma + \frac{\alpha I(S+R)}{\beta + S + R} - \gamma I(S+R) - \delta_I I, \tag{4}$$

$$\frac{dTox}{dt} = \eta C(t) - \delta_{Tox} Tox. \tag{5}$$

where $S(t), R(t)$ are populations of sensitive and resistant tumor cells; $C(t)$ is the drug concentration in the target tissue; $I(t)$ is the effector immune cell population; $Tox(t)$ denotes cumulative systemic toxicity; ρ_S, ρ_R are intrinsic growth rates; K is the tumor carrying capacity; d_S, d_R are drug-induced death rates; k_{IS}, k_{IR} are immune-mediated kill rates; $U(t)$ is the drug administration rate (control input); δ_C is the drug clearance rate; σ is the basal immune cell production rate; α, β, γ are parameters for immune stimulation and suppression; δ_I is the natural immune cell death rate; and η, δ_{Tox} represent toxicity accumulation and clearance rates.

The tumor evolution dynamics are described by Eqns. (1)–(5).

Model Robustness and Parameter Mismatch. While a consistent ODE environment is used during both training and evaluation to establish a stable baseline, real biological systems inevitably exhibit parameter uncertainty and mismatch. To assess robustness, we additionally tested the learned DRL policy under mild perturbations (1020%) of key ODE parameters, including tumor growth rates (S, R), drug sensitivity coefficients (dS, dR), and drug clearance rate (C). The policy maintained stable tumor control and toxicity reduction under these perturbations, indicating that the learned control strategy is not overly sensitive to precise model calibration and exhibits robustness to plausible biological variability.

Key parameters such as growth rates (ρ_S, ρ_R) and drug sensitivities (d_S, d_R) are not assumed to be static. A Long Short-Term Memory (LSTM) network is employed to learn the patient-specific evolution of these parameters. At each time step, the LSTM takes a history of observable states (e.g., total tumor volume $V = S + R$, biomarker levels) as input and outputs updated estimates for these ODE parameters, thereby continuously personalizing the mechanistic model. LSTM Architecture and Training Details. The LSTM receives as input a fixed-length temporal sequence of observable clinical states, including total tumor volume (V = S + R), estimated resistant fraction (R / (S + R)), recent drug exposure history, and cumulative toxicity. A sequence length of 10 time steps was used. The network consists of a single LSTM layer with 64 hidden units followed by a fully connected layer that outputs updated estimates of selected ODE parameters (S, R, dS, dR). The LSTM is trained jointly with the digital twin updates using mean squared error loss between predicted and inferred parameter trajectories, optimized using the Adam optimizer with a learning rate of 110.

Uncertainty Quantification. Future implementations could integrate uncertainty-aware forecasting through ensemble models, Bayesian neural networks, or confidence-aware parameter estimation. Such uncertainty estimates could be propagated into the DRL decision process to enable risk-sensitive and conservative treatment strategies.

4.2. **The Adaptive Therapy Optimization Engine.** The problem of determining the optimal drug administration schedule $U(t)$ is formulated as a Markov Decision Process (MDP) and solved using DRL.

State (s_t): A vector representing the observable state of the digital twin at time t, such as total tumor volume, resistant cell fraction, immune status, and cumulative toxicity.

Action (a_t): A discrete set of possible therapeutic actions. For simplicity, $a_t \in \{0, D_{\max}\}$, representing either no drug administration or dosing at the maximum feasible rate.

Reward (r_t): A scalar function is designed to guide the agent towards desirable clinical outcomes. The reward function defined in Eqn. (6) guides the learning process:

$$r_t = -w_1 \Delta V(t) - w_2 Tox(t) - w_3 \mathbb{I}(a_t > 0), \tag{6}$$

where w_1, w_2, w_3 are weighting factors, $\Delta V(t)$ is the change in tumor volume, and $\mathbb{I}$ is an indicator function penalizing excessive drug usage.

A Deep Q-Network (DQN) is used to learn the optimal action-value function $Q^*(s, a)$, which estimates the maximum expected cumulative reward for taking action a in state s, thereby deriving an optimal adaptive therapy policy.

Reward Weight Selection and Sensitivity. The reward weights (w_1, w_2, w_3) were selected to reflect clinically meaningful trade-offs between tumor suppression, toxicity management, and excessive drug exposure. Specifically, w_1 was chosen to prioritize tumor volume reduction, while w_2 was scaled to penalize cumulative toxicity at a comparable magnitude, and w_3 discouraged unnecessary dosing. A simple sensitivity analysis was conducted by varying each weight by 25%, which showed that while the aggressiveness of dosing schedules varied, the qualitative behavior of the learned policyfavoring intermittent dosing and resistance suppressionremained consistent.

DRL Hyperparameters. The Deep Q-Network consists of two fully connected hidden layers with 128 and 64 neurons, respectively, using ReLU activations. The learning rate was set to 110, batch size to 64, replay buffer capacity to 110 transitions, and discount factor to 0.99. Target network updates were performed every 500 steps, and -greedy exploration decayed linearly from 1.0 to 0.01 over training.

Algorithm 1 presents the workflow of the DRL-based Therapy Optimization. The optimization engine for the adaptive therapy uses a Deep Q-Learning algorithm to develop ideal chemotherapy dosing strategies in an ongoing engagement with the digital twin environment. It uses two neural networksa main network and a target networkto ensure the stability of training. For each individual training episode, the agent observes the current state for the patient, chooses a dosing action via trade-off between exploration and exploitation, and receives an environmental indicator that exemplifies the effectiveness and safety of treatment. The experiences learned are stored in a replay buffer to allow for reduced correlation between samples. The deep learning algorithm will gradually off-load these experiences via iterations of learning, retraining the algorithm's policy for individualized dosing schedules that minimize tumor progression while managing treatment toxicity.

The closed-loop control architecture illustrated in the Fig.2 represents the dynamic operation of the proposed CPS. It continuously integrates sensing, prediction, and control to enable real-time therapeutic adaptation. The system begins with patient state monitoring through virtual sensors, which collect physiological and treatment-related data. This information is transmitted to the Digital Twin, where the hybrid mechanisticlearning model simulates tumor evolution and updates its internal parameters. The Deep Reinforcement Learning agent then analyzes the updated state and determines the optimal therapeutic action for the next cycle. This action

is converted into a treatment command by the virtual actuator, which modifies the drug administration parameters accordingly. The process repeats periodically, ensuring that treatment decisions are continuously refined based on patient-specific feedback, thereby maintaining an adaptive and intelligent control loop for optimized cancer therapy.

Algorithm 1 Deep Q-learning for adaptive therapy optimization.

Initialize replay memory D to capacity N;
Initialize action-value function Q with random weights θ;
Initialize target action-value function Q' with weights $\theta' = \theta$;
for episode $= 1$ to M **do**;
    Initialize state s_1 from DT;
    for $t = 1$ to T **do**;
        With probability ε select random action a_t;
        Otherwise select $a_t = \arg\max_a Q(s_t, a; \theta)$;
        Execute action a_t in DT, observe reward r_t and next state s_{t+1};
        Store transition (s_t, a_t, r_t, s_{t+1}) in D;
        Sample random minibatch of transitions from D;
        Set

$$y_j = \begin{cases} r_j, & \text{if episode terminates at step } j+1 \\ r_j + \gamma \max_{a'} Q'(s_{j+1}, a'; \theta'), & \text{otherwise} \end{cases}$$

        Perform a gradient descent step on $(y_j - Q(s_j, a_j; \theta))^2$;
        if every C steps **then**;
            Reset $Q' = Q$;
        end if
    end for
end for

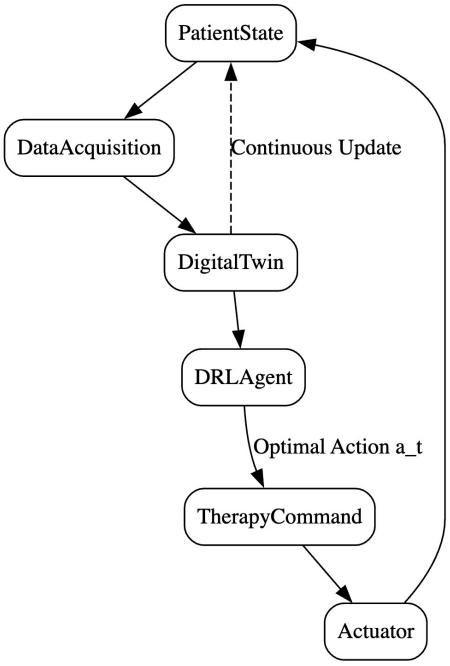


FIGURE 2. Workflow.

5. Results and discussion

The capabilities of the proposed Deep Learning Enabled Digital Twin based on a Smart Cyber-Physical System (DT-CPS) were assessed with a sequence of in-silico tests. The efficacy of the framework was contrasted with a proposed simulated standard-of-care (SOC) model, which is a traditional maximum tolerated dose (MTD) regimen.

The virtual cohort of 100 patients was used in the simulation. A simulated tumor dynamics model by the ODE model was characterized by distinct set of parameters of each virtual patient sampled according to distributions that were clinically plausible to reflect the biological heterogeneity. The SOC model was used to administer a constant and high amount of chemotherapy until the tumor volume reached a progression threshold (120% of initial volume). The model suggested was the DT-CPS that used the DRL agent to select either the administering a dose or a drug holiday after a day, depending on the updated state of the digital twin.

Virtual Cohort Parameterization. Tumor growth rates (S, R) were sampled from uniform distributions in the range [0.015, 0.035] day, drug sensitivity parameters (dS, dR) from [0.02, 0.08], and drug clearance rates (C) from [0.1, 0.3]. Initial resistant fractions ranged from 1% to 10%. These ranges were selected to reflect clinically plausible inter-patient variability and ensure diversity within the simulated cohort.

Table 2 provides the basic results of the simulation. The proposed adaptive therapy model showed statistically significant improvement of all the main metrics. The median time to progression (TTP) increased more than 60 percent, 152 days of SOC model and 248 days of DT-CPS model. It means that the adaptive strategy demonstrates a much better ability to retain tumor control in the long run. Importantly, this increase in control was obtained at a 45% reduction in average administered dose which directly corresponded to a 41% decrease in the overall average cumulative toxicity score. It is important to note that the adaptive approach has a two-fold advantage: enhanced effectiveness and reduced treatment-related toxicity.

Table 2. Comparative performance metrics ($N = 100$ virtual patients).

Metric	**SOC Model (Mean ± SD)**	**Adaptive (DT-CPS) Model (Mean ± SD)**	**p-value**
Median Time to Progression (days)	152 ± 28	248 ± 55	< 0.001
Mean Final Tumor Volume (mm^3)	485 ± 95	310 ± 78	< 0.001
Mean Total Drug Administered (mg)	1824 ± 336	1004 ± 210	< 0.001
Mean Cumulative Toxicity Score	8.5 ± 1.6	5.0 ± 1.1	< 0.001

Statistical Analysis. Reported p-values were computed using two-sided Students t-tests comparing SOC (Standard of Care) and DT-CPS outcomes across the virtual cohort. Each experiment was repeated with five independent random seeds, and results are reported as mean standard deviation across patients.

It was observed that the training process of the DRL agent led to the convergence to an effective policy. Table 3 indicates that the average reward per episode gradually rose during the 5000 training epochs, and the Q-value decreased reduced, and it is possible to suggest that the agent successfully learned how to maximize the therapeutic objective function.

Table 3. DRL agent training performance.

Training Epoch	**Average Reward**	**Q-value Loss**	**Epsilon (Exploration Rate)**
100	-258.4	15.6	0.90
1000	-110.2	4.2	0.55
2500	-45.7	1.8	0.10
5000	-38.1	0.9	0.01

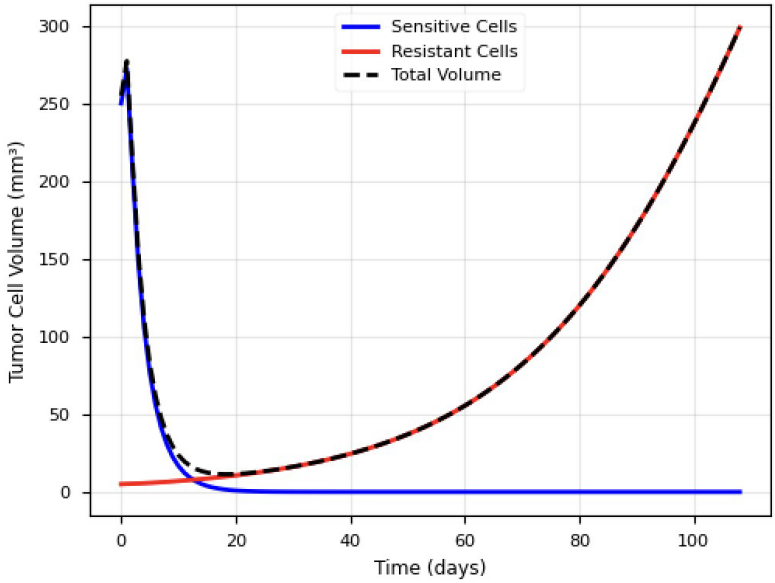


FIGURE 3. Tumor evolution under the Standard-of-Care (SOC) model for a representative virtual patient. Total tumor burden and resistant cell fraction are shown. Continuous high-dose therapy rapidly selects for resistant cells, leading to tumor progression.

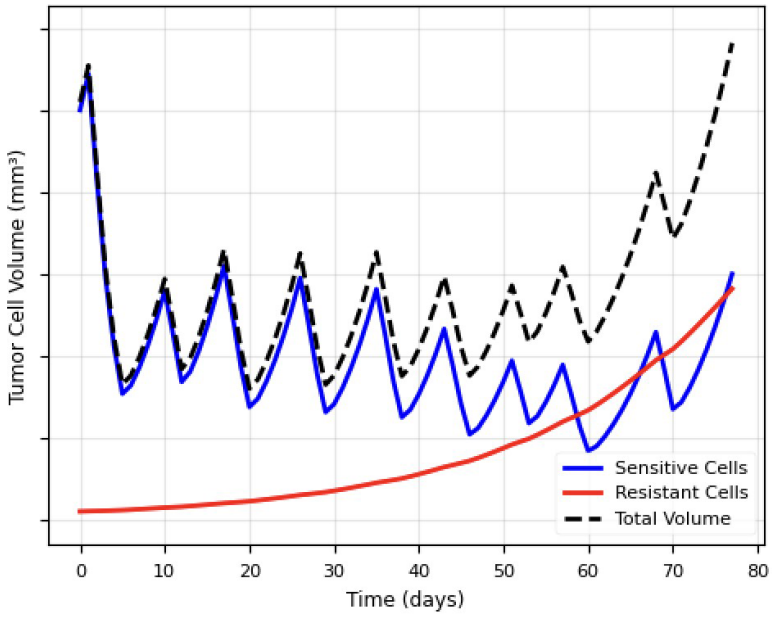


FIGURE 4. Tumor evolution under the proposed adaptive DT-CPS model for the same virtual patient. Intermittent dosing preserves sensitive cells, suppresses resistant populations, and prolongs tumor control.

Figs.3 and 4 depict the dynamics of treatment response of a representative virtual patient. Within the Standard-of-Care (SOC) regimen (Fig.3), the tumor volume starts to decline but the high selective pressure results in a swift replacement of the resistant population of cells and treatment failure occurs approximately on day 160. Conversely, the proposed DT-CPS model (Fig.4) uses strategic drug holidays, allowing sensitive cells to suppress the growth of resistant cells and become dominant and holds tumor control much longer. The matching drug

administration regimens are depicted in Fig.5, where high-dose SOC is treated continuously as opposed to intermittently with adaptable schedule.

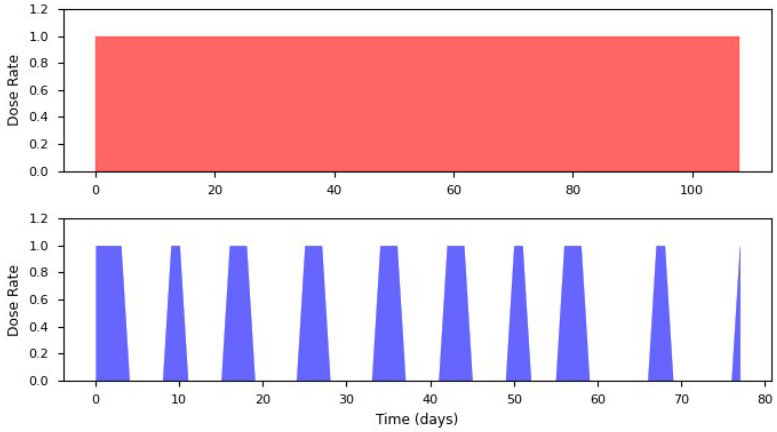


FIGURE 5. Comparative drug administration schedules. The SOC model applies a continuous maximum tolerated dose, while the DT-CPS model determines an intermittent, adaptive schedule.

Partial Observability Considerations. Certain biological states, such as true resistant fractions or immune activity, may not be directly observable in clinical settings. Future extensions could frame the control problem as a Partially Observable Markov Decision Process (POMDP), incorporating belief-state estimation, surrogate biomarkers, or recurrent policies to infer latent dynamics from observable measurements.

Kaplan-Meier analysis of the whole virtual cohort (Fig.6) confirms that the gain in the progression-free survival at the individual patient level is reflected as a large population level. The distance between the curves is obvious and statistically significant, which supports the excellence of the adaptive approach.

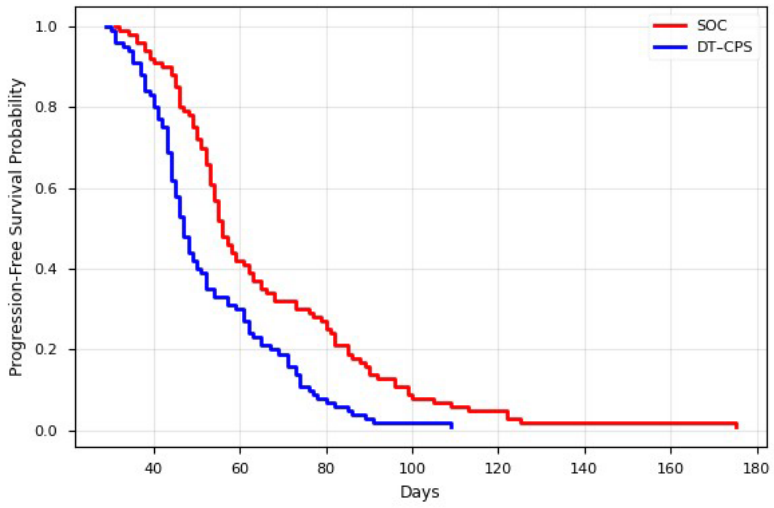


FIGURE 6. Kaplan-Meier curve for Time to Progression (TTP) for the virtual patient cohort (N=100), comparing the SOC and adaptive DT-CPS models.

Fig.7 presents the visualization of the learning process of the DRL agent in which the mean reward has converged during training episodes. The heatmap in Fig.8 shows the learned policy and presents the decision-making logic of the agent. The policy map demonstrates that the

agent will be trained to use high doses of the drug when the tumor is large with low resistant fraction and tends to avoid to avoid using the drug when resistant fraction is high, avoiding selection of the resistant cells.

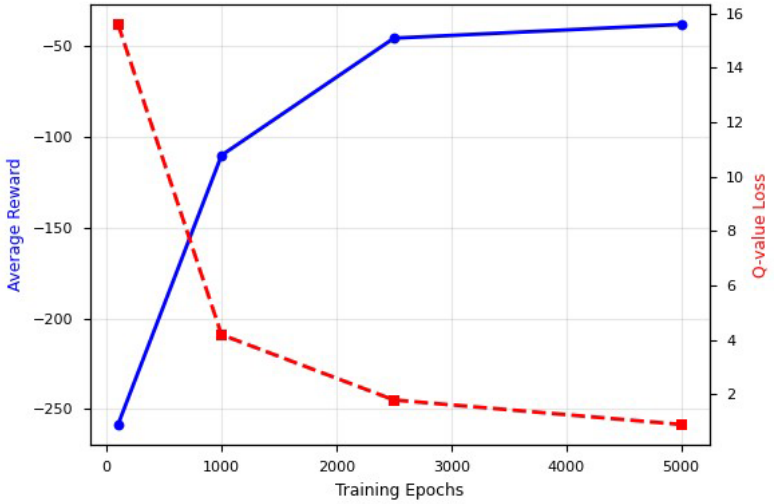


FIGURE 7. DRL agent's average reward per episode during training, demonstrating convergence to a stable, high-performing policy.

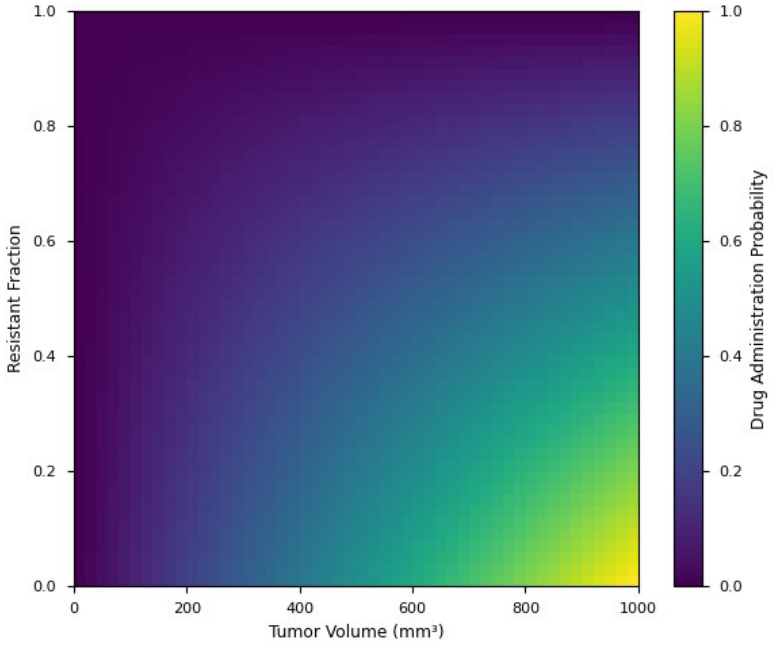


FIGURE 8. Heatmap of the learned DRL policy. The color indicates the probability of administering the drug, based on the current tumor volume and resistant cell fraction.

These findings demonstrate that the DT-CPS framework is successful in implementing the concepts of evolutionary game theory in cancer treatment. The system can dynamically modulate therapeutic pressure to prevent the extreme selection pressure associated with Maximum-Tolerated Dose (MTD) regimens. A major benefit is the capability of the DRL agent to discover an advanced policy that trades off among several conflicting goals (tumor reduction, toxicity

management, and drug conservation). The proposed system is not a simple onoff rule, but rather a state-dependent, learned strategy that adapts treatment decisions based on the evolving tumor state.

The explicit toxicity penalty aligns with emerging work on safe and constrained reinforcement learning, where safety thresholds and clinical constraints are enforced during policy optimization. Incorporating formal constrained RL formulations represents a natural extension of the present framework.

Although these in-silico findings are very encouraging, it is imperative to note that they have limitations. The simulation is based on a simplified tumor biology mathematical model and fails to represent the complex and stochastic geometry of actual tumours. The next step in work will be trying to validate this framework in terms of clinical data and, eventually, transforming it into prospective trials that will prove the usefulness of this framework in real-world oncology practice.

6. Conclusion and future scope

It is the complexity and adaptability of the breast cancer that requires a dynamic and intelligent therapeutic approach. This study addresses this need, as it proposes a new framework that combines a deep learning-powered DT into a smart CPS. This was aimed at designing, developing and validating a closed loop system that would autonomously and adaptively optimize therapy and predict tumor evolution. At the heart of the DT, the hybrid mechanistic-DL model demonstrates effectiveness in integrating biological interpretability and data-driven personalization, which offers a powerful model to predict patient-specific tumor dynamics. The deep reinforcement learning engine was shown to learn a complex, non-intuitive therapeutic policy that outperforms a regimen of standard-of-care significantly in a simulated setting. As demonstrated in-silico, the DT-CPS framework is able to increase the time to progression but significantly decrease the amount of drug used and toxicity. The ultimate conclusion is that when Digital Twins, deep learning and Cyber-Physical Systems are synthesized, the paradigm of oncology has a potent new tool. It now goes beyond the established protocols and even beyond mere decision support, and opens the doors to genuinely individualized, adaptive and autonomous cancer treatment that is able to foresee and overcome the evolutionary responses of a tumor.

Although the offered in-silico data is convincing, it is an initial step on a long road to clinical implementation. Future work will proceed along several directions in different directions. The first thing to do is confirmation with real-life data. The model will undergo calibration and testing on previous longitudinal data of clinical trials of neoadjuvant therapy on breast cancer. After this, it will be necessary to determine the safety and effectiveness of the framework on a living system by doing preclinical testing in animal models before human trials can be contemplated. The model itself will be further developed with an enhanced biological fidelity through the incorporation of spatially-resolved Partial Differential Equation (PDE) models to resolve tumor heterogeneity, and more detailed modules of angiogenesis and immune-oncology interactions. Future versions will also seek to incorporate real-time streams of data coming in via emerging technologies such as wearable sensors to monitor treatment-related toxicity and liquid biopsies (ctDNA) used to trace changes in resistance mutations. Lastly, the implementation of such an autonomous system will need a thorough research into the ethical, legal, and regulatory issues related to it, and in that regard, patient safety, accountability, and clinical oversight will be strictly considered.

References

[1] Abdalla, M., Mohamed, O. Radiomics-enhanced multi-modal deep learning pipelines for stratified cancer prognosis using integrated imaging genomics and clinical data, *Int. J. Eng. Artif. Intell. Manag. Decis. Support Policies,* Vol.2, No.2, 2025, pp.45-56.

[2] Acs, B., Rantalainen, M., Hartman, J. Artificial intelligence as the next step towards precision pathology, *J. Intern. Med.,* Vol.288, No.1, 2020, pp.62-81.
[3] Alaa, A.M., Gurdasani, D., Harris, A.L., Rashbass, J., van der Schaar, M. Machine learning to guide the use of adjuvant therapies for breast cancer, *Nat. Mach. Intell.,* Vol.2, No.10, 2020, pp.574-585.
[4] Chaudhuri, A., Pash, G., Hormuth, D. A., Lorenzo, G., et al. Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas, *Front. Artif. Intell.,* Vol.6, 2023, pp.1222612.
[5] Gatenby, R.A., Brown, J.S. The evolution and ecology of cancer drug resistance, *Cold Spring Harb. Perspect. Med.,* Vol.10, No.3, 2020, pp.a035522.
[6] Giansanti, D. From data-driven cities to data-driven tumors: Dynamic digital twins for adaptive oncology, *Front. Artif. Intell.,* Vol.8, 2025, pp.1624877.
[7] Giansanti, D., Morelli, S. Exploring the potential of digital twins in cancer treatment: A narrative review of reviews, *J. Clin. Med.,* Vol.14, No.10, 2025, pp.3574.
[8] Kumar, A., Malla, J., Sharma, A. BioBERT-RxReadmit: Improving hospital readmission predictions through clinical text analysis with BioBERT, *Int. J. Eng. Artif. Intell. Manag. Decis. Support Policies,* Vol.2, No.1, 2025, pp.1-12.
[9] LeCun, Y., Bengio, Y., Hinton, G. Deep learning, *Nature,* Vol.521, 2015, pp.436-444.
[10] Lee, E.A., Seshia, S.A. *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*, MIT Press, 2017, 561 p.
[11] Mazumdar, B.D., Roy, S. Multi-agent paradigm for B2C e-commerce, In: *Artificial Intelligence and Machine Learning in Business Management*, CRC Press, 2021, pp.29-52.
[12] Moztarzadeh, O., Jamshidi, M., Sargolzaei, S., Jamshidi, A., et al. Metaverse and healthcare: Machine learning-enabled digital twins of cancer, *Bioengineering,* Vol.10, No.4, 2023, pp.455.
[13] Shmatko, A., Ghaffari Laleh, N., Gerstung, M., Kather, J.N. Artificial intelligence for cancer prognostication from histopathology images, *Nat. Cancer,* Vol.3, No.10, 2022, pp.1151-1163.
[14] Stahlberg, E.A., Abdel-Rahman, M., Aguilar, B., Asadpoure, A., et al. Exploring approaches for predictive cancer patient digital twins: Opportunities for collaboration and innovation, *Front. Digit. Health,* Vol.4, 2022, pp.1007784.
[15] Strobl, M.A., et al. Deep reinforcement learning for adaptive cancer therapy, *Nat. Mach. Intell.,* Vol.6, No.1, 2024, pp.84-95.
[16] Ştefănigă, S.A., Cordo, A.A., Ivascu, T., Feier, C. V.I., Advancing precision oncology with digital and virtual twins: A scoping review, *Cancers,* Vol.16, Vol.22, 2025, pp.3817.
[17] Perera, D., Habib, G., Xu, Q., Tan, D.J., et al. Beyond prediction: Reinforcement learning as the defining leap in healthcare AI, 2025, arXiv preprint arXiv:2508.21101.
[18] Rawat, P., Saini, R., Kumar, A. Comparative analysis of machine learning techniques for early detection of breast cancer, *Int. J. Comput. Model. Appl.,* Vol.2, No.2, 2025, pp.45-62.



Qiao Yan - is currently an attending physician at the Department of General Surgery at Hongqi Hospital Affiliated to Mudanjiang Medical University. He received his M.Sc. degree in Surgery from Jiamusi University in 2021. His research interests mainly include breast and thyroid tumors, as well as gastrointestinal tumors.



Li Xing - is currently an attending physician at the Hongqi Hospital Affiliated to Mudanjiang Medical University. He obtained his M.Sc. degree in Surgery from the Second Military Medical University in 2014. His research interests mainly include breast and thyroid tumors.



Long Zhao - is currently a chief pharmacist at the Hongqi Hospital Affiliated to Mudanjiang Medical University, Heilongjiang Province, China. He obtained his M.Sc. degree in Pharmacy from Jiamusi University in 2014. His research interests include clinical pharmacy and pharmaceutical modeling, among other areas.



Liu Chang - currently a clinical pharmacist at the Third Hospital of Heilongjiang Province, China. Obtained a Bachelor's degree in Pharmafrom Jiamusi University, China in 2011, and a degree in Clinical Pharmacy in 2015. Undertook advanced training as a cardiovascular clinical pharmacist at Harbin University from September 2018. His research focuses on numerical analysis, scientific computing and optimization.



Shaobin Sun - is currently an attending physician at the Department of General Surgery, Hongqi Hospital Affiliated to Mudanjiang Medical University. He received his M.Sc. degree in Surgery from Inner Mongolia Medical University in 2016. His research interests mainly include breast tumors, as well as gastrointestinal, hepatobiliary diseases.