

## Article

# A New Fuzzy Reinforcement Learning Method for Effective Chemotherapy

Fawaz E. Alsaadi <sup>1</sup>, Amirreza Yasami <sup>2</sup>, Christos Volos <sup>3,\*</sup>, Stelios Bekiros <sup>4,5,6</sup> and Hadi Jahanshahi <sup>7</sup><sup>1</sup> Communication Systems and Networks Research Group, Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia<sup>2</sup> Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada<sup>3</sup> Nonlinear Laboratory of Nonlinear Systems, Circuits Complexity (LaNSCom), Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece<sup>4</sup> FEMA, University of Malta, MSD 2080 Msida, Malta<sup>5</sup> LSE Health, Department of Health Policy, London School of Economics and Political Science, London WC2A 2AE, UK<sup>6</sup> IPAG Business School (IPAG), 184, bd Saint-Germain, 75006 Paris, France<sup>7</sup> Department of Mechanical Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada

\* Correspondence: volos@physics.auth.gr

**Abstract:** A key challenge for drug dosing schedules is the ability to learn an optimal control policy even when there is a paucity of accurate information about the systems. Artificial intelligence has great potential for shaping a smart control policy for the dosage of drugs for any treatment. Motivated by this issue, in the present research paper a Caputo–Fabrizio fractional-order model of cancer chemotherapy treatment was elaborated and analyzed. A fix-point theorem and an iterative method were implemented to prove the existence and uniqueness of the solutions of the proposed model. Afterward, in order to control cancer through chemotherapy treatment, a fuzzy-reinforcement learning-based control method that uses the State-Action-Reward-State-Action (SARSA) algorithm was proposed. Finally, so as to assess the performance of the proposed control method, the simulations were conducted for young and elderly patients and for ten simulated patients with different parameters. Then, the results of the proposed control method were compared with Watkins’s Q-learning control method for cancer chemotherapy drug dosing. The results of the simulations demonstrate the superiority of the proposed control method in terms of mean squared error, mean variance of the error, and the mean squared of the control action—in other words, in terms of the eradication of tumor cells, keeping normal cells, and the amount of usage of the drug during chemotherapy treatment.

**Keywords:** Caputo-Fabrizio derivative; cancer chemotherapy drug dosing; fuzzy-reinforcement learning; optimal control; SARSA algorithm; artificial intelligence

**MSC:** 93C40; 93C42; 68T05



**Citation:** Alsaadi, F.E.; Yasami, A.; Volos, C.; Bekiros, S.; Jahanshahi, H. A New Fuzzy Reinforcement Learning Method for Effective Chemotherapy. *Mathematics* **2023**, *11*, 477. <https://doi.org/10.3390/math11020477>

Academic Editor: Jakub Nalepa

Received: 17 December 2022

Revised: 9 January 2023

Accepted: 12 January 2023

Published: 16 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cancer is one of the most hazardous and fatal diseases throughout the world [1]. This disease is caused by the abnormal division and spreading of cells that destroy the patient’s body and may lead to death. Though many research efforts are devoted to precisely understanding the interaction between the immune system and tumor cells, the treatment of cancer is still one of the most challenging issues in modern medicine [2,3].

Based on the patient’s conditions and type of cancer, there are several treatments, including radiotherapy, surgery, chemotherapy, immunotherapy, and so forth, to tackle cancer [4]. Cancer treatment can be challenging, as it can have a range of side effects, including fatigue, nausea, and hair loss. Managing these side effects is an important part of cancer care among various treatments. Chemotherapy is one of the most effective treatments

for annihilating cancerous cells. For this reason, in this research the chemotherapy method was chosen to treat cancer. Nonetheless, in all chemotherapy treatments, not only do the applied drugs destroy the cancerous cells, but they also affect the healthy cells and kill some of those cells. Hence, it is crucial to make sure that the patient can tolerate the side effects of the treatment [5]. These sorts of side effects lead to a limitation in the dosage of the drugs [6]. Hence, for prescribing drugs, one aim is to decrease the number of cancerous cells as much as possible and reach the minimum side effects [7].

There are many factors which define the treatment schedule and drug dosing. Some of these factors are the weight of the patient, level of white blood cells, the patient's age, the stage of the tumor, etc.; for this reason, certain established standards are followed by clinicians to define the therapy type and drug dosage for each patient. However, this approach has some limitations which have been approved by scientific and clinical communities [8,9]. Moreover, evaluating the effectiveness of the chemotherapy plan and its feasibility is of significant importance [9]. So as to evaluate the effectiveness of the chemotherapy plans, the clinical trials are a reliable choice, but they have some limitations such as high costs, long trial times, and they are difficult to be conducted. Keeping the above-mentioned limitations in mind, contriving an efficient chemotherapy treatment plan would be desired.

Mathematical models have been used as valuable tools in understanding the transmission dynamics of cancers [10]. Actually, mathematical modeling could play a pivotal role in the understanding of the disease's dynamics [11–15]. One common approach to modeling cancer is to use ordinary differential equations (ODEs), which are used to describe the time evolution of a system. These models can represent the growth and spread of cancer cells and the response of the immune system to the cancer. Researchers can then use these models to study the long-term and short-term dynamics of cancer and to identify potential therapeutic strategies. Finding an accurate model of cancer dynamics could pave the way for the long and short-term prediction of the disease as well as the designing of therapies [16]. This field of study not only aims to anticipate the spread of cancerous cells but also to control the disease as effectively as possible [17,18]. A lot of research has been done using mathematical modeling to understand cancer, and various approaches have been used in this field [19–21].

Fractional calculus is an excellent tool for the description of hereditary properties, and memory of systems has been widely utilized in various fields of study including economics, biology, ecology, and engineering [22–28]. In this regard, the fractional-order model of cancers has started to attract some researchers' attention [29,30]. Although early research studies on fractional-order calculus were based on the Caputo or Riemann–Liouville fractional-order derivative, it has been recently shown that these approaches possess some disadvantages, such as the singularity at the endpoint of an interval of definition [31–33]. In tackling this issue, recently other definitions of fractional derivatives have been introduced [34–39]. Caputo and Fabrizio offered a new derivative with nonsingular kernel [32], namely the Caputo–Fabrizio derivative. The Caputo and Caputo–Fabrizio derivatives are both types of fractional derivatives, which are a generalization of the classical derivative to functions with non-integer order. The Caputo derivative is based on a power law, which means that it is defined in terms of a power of the function's argument. The Caputo–Fabrizio derivative, on the other hand, is based on an exponential decay law, which means that it is defined in terms of the exponential decay of the function over time. One key difference between the two derivatives is that the Caputo derivative is defined in terms of a power of the function's argument, while the Caputo–Fabrizio derivative is defined in terms of an exponential decay law. This means that the two derivatives may behave differently when applied to different types of functions, and they may be used to model different types of physical phenomena.

Caputo and Fabrizio offered a new derivative with nonsingular kernel [32], namely the Caputo–Fabrizio derivative. The principal difference between the Caputo and Caputo–Fabrizio fractional derivative is that the Caputo derivative is based on a power law, while

the Caputo–Fabrizio derivative is achieved with an exponential decay law [32,40]. There are several research studies in the literature that demonstrate the applications of the Caputo–Fabrizio derivative to different systems such as biological processes [41–43]. In Ref. [44], the underlying physical meaning of the nonsingular kernel as well as the applications of fractional differentiation operators with the non-singular kernel were presented.

During for the past several years, many control strategies have been applied to cancer chemotherapy drug dosing.

In [45], targeted chemotherapy was offered for tumor-immune interaction. As it has been studied in [46,47], in addition to the mathematical interpretation of tumor growth, scheduling appropriate treatment strategies has been an important issue, and control-designing strategies can be utilized to optimize drug dosing mathematically. To this end, up to now, many research studies have worked on these subjects. De Pillis and Radunskaya proposed a dynamical model for immunotherapy and chemotherapy schedules in 2001 [48]. In 2003, to control tumor growth, bang-bang type control was employed with chemotherapy [47]. In 2007, De Pillis et al. used linear and quadratic control to a cure tumor with the use of chemotherapy [49]. Also, through Pontryagin’s maximum principle, Ghaffari and Naserifar have designed optimal therapeutic protocols to schedule immunotherapy [50]. On the other hand, the cost-effectiveness of the treatment strategies has been investigated in [51]. Drug resistance, which is a vital phenomenon in cancerous tumors, has been examined in [52]. In addition, the conditions for an siRNA treatment were investigated to eradicate tumor burden in [53], while a combination of chemotherapy and anti-angiogenic agents was utilized to cure the disease in [54]. In [55], robust adaptive control was presented to adjust the drug dosage with an extended Kalman filter observer. An optimal control strategy based on a linear time-varying approximation technique was proposed in [56]. A model-free method for chemotherapy based on reinforcement learning was proposed in [57] using the closed-loop control. To be precise, they develop an optimal controller using the Q-learning algorithm for cancer chemotherapy treatment. In [47], the authors model the chemotherapy treatment based on optimal control theory, and the objective was to minimize the tumor cells while keeping the healthy cells above a fixed level.

The use of state-of-the-art automatic control methods can help to improve the effectiveness and efficiency of various processes and systems, leading to better outcomes and more sustainable solutions [58–67]. Among the stated control strategies, intelligent controllers have significant advantages [68–73]. Artificial intelligence does combine a wide variety of new technologies to give systems an ability to make decisions in new and unfamiliar conditions [74,75]. In some applications, due to the high value of their tasks and their remarkable risks, implementing a reliable controller is the main concern. Where the degree of uncertainty is high, classical methods of control may fail [76,77]. Hence, artificial intelligence-based controllers are rational choices for such systems. Moreover, optimal controls which could be provided based on artificial intelligence are helpful for drug dosage because of their ability to consider various aspects of the biological systems in the optimization function. For example, through intelligent approaches, we are able to design treatments which can take to account the number of healthy and infected cells as well as the cost and side effects of the drugs without having any mathematical model. Reinforcement learning is one of the most popular learning techniques, which explores the system’s response in possible actions; this way, it learns the optimal action by calculating how the last action pushes the system towards desired situations [78–80].

As mentioned, fractional calculus is very advantageous to the modeling of real-world processes [81,82]. Motivated by this background, in the present study, we investigated a Caputo–Fabrizio model of cancer, which is a new model; studies are quite rare in these systems. Also, few studies are dedicated to the evaluation of reinforcement learning methods in chemotherapy drug dosing; to the best of our knowledge, no studies have been done on cancer chemotherapy drug dosing using fuzzy-reinforcement learning-based optimal control. These issues motivated the current study. In this paper, a Caputo–Fabrizio fractional-order model of cancer chemotherapy treatment was analyzed. The existence and uniqueness of

the solutions of the presented model were proven. Then, a fuzzy-reinforcement learning-based control method was proposed in order to control the cancer chemotherapy treatment, and the effective performance of the proposed method was illustrated.

The main contributions of this research paper are as follows: First, a Caputo–Fabrizio fractional-order model of cancer chemotherapy treatment was elaborated and analyzed, which is a novel model for this system. Afterwards, a fix-point theorem and an iterative method were implemented to prove the existence and uniqueness of the solutions of the proposed model. Finally, as the last contribution, a fuzzy-reinforcement learning-based control method that uses the SARSA algorithm was proposed to control cancer through chemotherapy treatment.

The rest of the paper has been organized as follows: In Section 2 the preliminaries for this work are given. Later, in Section 3, the Caputo–Fabrizio fractional model of cancer chemotherapy is elaborated, and then sensitivity analysis is done for the parameters of the system. The fuzzy-reinforcement-learning based controller is described in Section 4. Section 5 is devoted to the numerical simulations; finally, in Section 6, conclusions are made and discussed.

## 2. Preliminaries

As has been shown, the kernels of the Caputo fractional derivative are singular at the endpoint of the interval of integration [83]; therefore, the fractional derivative is not an appropriate kernel to describe the memory effect accurately in real systems. A new fractional derivative has been proposed that has no singularity in its kernels [32]. This section is a summary of the definitions and properties of the Caputo–Fabrizio fractional that have been used in this paper.

Consider  $H^1(a, b) = \{x | x \in L^2(a, b)\}$  and  $x' \in L^2(a, b)$ , where  $L^2(a, b)$  is the space of square integrable functions on interval of  $(a, b)$ .

**Definition 1.** If function  $x$  is  $H^1(t_0, t)$  and derivative order  $\alpha \in (0, 1)$ , then the Caputo–Fabrizio derivative is defined as [32]:

$${}^{CF}D_t^\alpha x(t) = \frac{A(\alpha)}{1-\alpha} \int_{t_0}^t x'(z) e^{-\alpha \frac{t-z}{1-\alpha}} dz, \quad (1)$$

where  $A(\alpha)$  is a normalization function that satisfies  $A(0) = A(1) = 1$ . Furthermore, if  $x \notin H^1(a, b)$ , then the Caputo–Fabrizio derivative is defined as:

$${}^{CF}D_t^\alpha x(t) = \frac{\alpha A(\alpha)}{1-\alpha} \int_{t_0}^t (x(t) - x(z)) e^{-\alpha \frac{t-z}{1-\alpha}} dz, \quad (2)$$

**Remark 1.** [32]. By considering  $\beta = \alpha / (1 - \alpha)$  and  $\beta \in (0, \infty)$ , we have  $\alpha = 1 / (1 + \beta)$  which  $\alpha \in (0, 1)$ . Now, Equation (2) can be written in the following form:

$${}^{CF}D_t^\alpha x(t) = \frac{B(\beta)}{\beta} \int_{t_0}^t x'(z) e^{-\frac{t-z}{\beta}} dz, \quad (3)$$

where  $B(\beta)$  is a normalization function corresponding to  $A(\alpha)$  which  $B(0) = B(\infty) = 1$ .

**Remark 2.** [32]. The following equation is true:

$$\lim_{\beta \rightarrow 0} \frac{1}{\beta} e^{-\frac{t-z}{\beta}} = \delta(z - t), \quad (4)$$

where  $\delta(z - t)$  is the Dirac delta function.

The modified definition of Caputo-Fabrizio has been proposed by Nieto and Losada [31], and it has been defined as

$${}_{t_0}^{\text{CF}} D_t^\alpha x(t) = \frac{(2-\alpha)A(\alpha)}{2(1-\alpha)} \int_{t_0}^t x'(z) e^{-\alpha \frac{t-z}{1-\alpha}} dz, \quad (5)$$

**Definition 2.** The Caputo–Fabrizio fractional integral of order  $\alpha$  of function  $x(t)$  is defined as:

$${}_0^{\text{CF}} I_t^\alpha x(t) = \frac{2(1-\alpha)}{(2-\alpha)A(\alpha)} x(t) + \frac{2}{(2-\alpha)A(\alpha)} \int_0^t x(z) dz, \quad t \geq 0 \quad (6)$$

**Definition 3.** [31]. The fractional Caputo–Fabrizio derivative of order  $\alpha$  of function  $x(t)$  is defined as

$${}_{t_0}^{\text{CF}} D_t^\alpha x(t) = \frac{1}{1-\alpha} \int_{t_0}^t x'(z) e^{-\alpha \frac{t-z}{1-\alpha}} dz, \quad t \geq 0. \quad (7)$$

Moreover, the Caputo–Fabrizio fractional integral of order  $\alpha$  is given as

$${}_0^{\text{CF}} I_t^\alpha x(t) = (1-\alpha)x(t) + \alpha \int_0^t x(z) dz, \quad t \geq 0. \quad (8)$$

In this definition, the normalization function has been considered as  $A(\alpha) = 2/(2-\alpha)$ .

### 3. Caputo–Fabrizio Fractional Model of Cancer Chemotherapy

In this section, in order to derive the Caputo–Fabrizio fractional model of cancer chemotherapy, the nonlinear four-state model given in [47,84–86] has been considered. In this model, the first state is  $N(t)$ , which indicates the number of normal cells; the second variable in this model is  $T(t)$ , and it represents the number of tumor cells. The third state is  $I(t)$ , which represents the number of immune cells. Finally, the last state is the drug concentration  $C(t)$ . The original integer-order model [47,87] is given in Equation (9):

$$\begin{cases} \dot{N}(t) = r_2 N(t)[1 - b_2 N(t)] - c_4 N(t)T(t) - a_3 N(t)C(t), \\ \dot{T}(t) = r_1 T(t)[1 - b_1 T(t)] - c_2 T(t)I(t) - c_3 N(t)T(t) - a_2 T(t)C(t), \\ \dot{I}(t) = s + \left( \frac{\lambda T(t)I(t)}{\gamma + T(t)} \right) - c_1 T(t)I(t) - d_1 I(t) - a_1 I(t)C(t), \\ \dot{C}(t) = -d_2 C(t) + u(t). \end{cases} \quad (9)$$

The initial conditions for this model are assumed to be as follows:

$$N(0) = N_0, T(0) = T_0, I(0) = I_0, C(0) = C_0. \quad (10)$$

in which  $r_1$  and  $r_2$  are the growth rate of the tumor cells and normal cells, respectively;  $b_1$  and  $b_2$  denote the reciprocal carrying capacities of both cells;  $d_1$  is cell death rate,  $d_2$  stands for the decay rate of the injected drug; and the fractional cell kill rate of the immune cells, tumor cells, and normal cells are represented by  $a_1$ ,  $a_2$  and  $a_3$  respectively; also in this model,  $\lambda$  and  $\gamma$  are positive constants which stand for the immune response rate and the immune threshold rate, respectively. The influx rate of the immune cells is represented by  $s$ , and  $u(t)$  is the drug infusion rate.

Replacing the first-order derivatives on the left-hand side of Equation (9) with the Caputo–Fabrizio fractional derivative that has been defined in Equation (7) leads to our

new Caputo–Fabrizio fractional model of cancer chemotherapy. The new model is written in the following equations:

$$\begin{cases} {}_0^{\text{CF}}D_{t_f}^{\alpha_1}N(t) = r_2N(t)[1 - b_2N(t)] - c_4N(t)T(t) - a_3N(t)C(t), \\ {}_0^{\text{CF}}D_{t_f}^{\alpha_2}T(t) = r_1T(t)[1 - b_1T(t)] - c_2T(t)I(t) - c_3N(t)T(t) - a_2T(t)C(t), \\ {}_0^{\text{CF}}D_{t_f}^{\alpha_3}I(t) = s + \frac{\lambda T(t)I(t)}{\gamma + T(t)} - c_1T(t)I(t) - d_1I(t) - a_1I(t)C(t), \\ {}_0^{\text{CF}}D_{t_f}^{\alpha_4}C(t) = -d_2C(t) + u(t). \end{cases} \quad (11)$$

Moreover, the initial condition for this system is:

$$N(0) = N_0, T(0) = T_0, I(0) = I_0, C(0) = C_0 \quad (12)$$

In our new model of cancer chemotherapy, it has been assumed that the fraction-order of each of the state variables is theoretically different, and  $0 < \alpha_i < 1, i = 1, 2, \dots, 4$ .

### 3.1. Existence and Uniqueness of Solutions of the Cancer Chemotherapy Model

This section is devoted to the investigation of the uniqueness and existence of the solutions of our Caputo–Fabrizio fraction model of cancer chemotherapy in Equation (11) with the initial conditions mentioned in Equation (12). To seek this goal, the fix-point theory has been used [88,89].

In the light of the Caputo–Fabrizio fractional order integral operator, defined in Equation (8) and taking the initial conditions in Equation (12) into consideration, in using Equation (12) the following equation will be obtained:

$$\begin{cases} N(t) - N(0) = {}_0^{\text{CF}}I_{t_f}^{\alpha_1}[r_2N(t)[1 - b_2N(t)] - c_4N(t)T(t) - a_3N(t)C(t)], \\ T(t) - T(0) = {}_0^{\text{CF}}I_{t_f}^{\alpha_2}[r_1T(t)[1 - b_1T(t)] - c_2T(t)I(t) - c_3N(t)T(t) - a_2T(t)C(t)], \\ I(t) - I(0) = {}_0^{\text{CF}}I_{t_f}^{\alpha_3}\left[s + \frac{\lambda T(t)I(t)}{\gamma + T(t)} - c_1T(t)I(t) - d_1I(t) - a_1I(t)C(t)\right], \\ C(t) - C(0) = {}_0^{\text{CF}}I_{t_f}^{\alpha_4}[-d_2C(t) + u(t)]. \end{cases} \quad (13)$$

Then, the following kernels are defined:

$$\begin{cases} K_1(t, N(t)) = r_2N(t)[1 - b_2N(t)] - c_4N(t)T(t) - a_3N(t)C(t) \\ K_2(t, T(t)) = r_1T(t)[1 - b_1T(t)] - c_2T(t)I(t) - c_3N(t)T(t) - a_2T(t)C(t), \\ K_3(t, I(t)) = s + \frac{\lambda T(t)I(t)}{\gamma + T(t)} - c_1T(t)I(t) - d_1I(t) - a_1I(t)C(t), \\ K_4(t, C(t)) = -d_2C(t) + u(t). \end{cases} \quad (14)$$

By taking Equation (14) into consideration and then calculating the right-hand side of Equation (13) using the definition of the Caputo–Fabrizio fractional-order integration in Equation (6), the following equation is obtained:

$$\begin{cases} N(t) - N(0) = \Sigma(\alpha_1)K_1(t, N(t)) + \sigma(\alpha_1)\int_0^t K_1(v, N(v))dv, \\ T(t) - T(0) = \Sigma(\alpha_2)K_2(t, T(t)) + \sigma(\alpha_2)\int_0^t K_2(v, T(v))dv, \\ I(t) - I(0) = \Sigma(\alpha_3)K_3(t, I(t)) + \sigma(\alpha_3)\int_0^t K_3(v, I(v))dv, \\ C(t) - C(0) = \Sigma(\alpha_4)K_4(t, C(t)) + \sigma(\alpha_4)\int_0^t K_4(v, I(v))dv \end{cases} \quad (15)$$

where  $\Sigma(\alpha)$  and  $\sigma(\alpha)$  are defined as follows:

$$\Sigma(\alpha) = \frac{2(1 - \alpha)}{(2 - \alpha)A(\alpha)} \quad \text{and} \quad \sigma(\alpha) = \frac{2\alpha}{(2 - \alpha)A(\alpha)}. \quad (16)$$

**Remark 3.** The above-mentioned kernels  $K_1, K_2, \dots, K_4$  satisfy the Lipschitz conditions and are contraction mapping if the following inequality satisfies

$$0 \leq L = \max\{\delta_1, \delta_2, \delta_3, \delta_4\} < 1 \quad (17)$$

In the proof of Remark 3 the following assumption has been made:

$$\|N(t)\| \leq \theta_1, \|T(t)\| \leq \theta_2, \|I(t)\| \leq \theta_3 \text{ and } \|C(t)\| \leq \theta_4. \quad (18)$$

**Proof.** Let  $N$  and  $N_1$  be two different functions; then, by taking the kernel  $K_1$  into consideration, we have

$$\begin{aligned} \|K_1(t, N) - K_1(t, N_1)\| &= \|r_2(N - N_1)[1 - b_2(N - N_1)] - c_4(N - N_1)T(t) \\ &\quad - a_3(N - N_1)C(t)\| \\ &\leq w_1\|(N - N_1)\| - w_2\|(N - N_1)\| - w_3\|(N - N_1)\| \\ &\leq (w_1 - w_2 - w_3)\|(N - N_1)\| \leq \delta_1\|(N - N_1)\| \end{aligned} \quad (19)$$

For the second kernel we have:

$$\begin{aligned} \|K_2(t, T) - K_2(t, T_1)\| &= \|r_1(T - T_1)[1 - b_1(T - T_1)] - c_2(T - T_1)I(t) \\ &\quad - c_3(T - T_1)N(t)\| \\ &\leq a_1\|(T - T_1)\| - a_2\|(T - T_1)\| - a_3\|(T - T_1)\| \\ &\leq (a_1 - a_2 - a_3)\|(T - T_1)\| \leq \delta_2\|(T - T_1)\| \end{aligned} \quad (20)$$

Additionally, for the third and fourth kernels,  $K_3$  and  $K_4$ , the following inequalities can be obtained:

$$\begin{aligned} \|K_3(t, I) - K_3(t, I_1)\| &= \left\| \frac{\lambda T(I - I_1)}{\gamma + T(t)} - c_1T(t)(I - I_1) - d_1(I - I_1) - a_1(I - I_1)C(t) \right\| \\ &\leq \frac{\lambda\theta_2}{\gamma + \theta_2}\|(I - I_1)\| - c_1\theta_2\|(I - I_1)\| - d_1\|(I - I_1)\| \\ &\quad - a_1\theta_4\|(I - I_1)\| \leq (m_1 - m_2 - m_3 - m_4)\|(I - I_1)\| \\ &\leq \delta_3\|(T - T_1)\| \end{aligned} \quad (21)$$

and

$$\|K_4(t, C) - K_4(t, C_1)\| = \|-d_2(C - C_1)\| \leq d_2\|(C - C_1)\| \quad (22)$$

Consequently, the Lipschitz conditions are satisfied for all of the kernels defined in Equation (14). Furthermore, because of  $0 \leq L = \max\{\delta_1, \delta_2, \delta_3, \delta_4\} < 1$ , the kernels are contractions.

By considering Equation (16), the following recursive formula can be obtained:

$$\left. \begin{aligned} N_n(t) &= \Sigma(\alpha_1)K_1(t, N_{n-1}(t)) + \sigma(\alpha_1)\int_0^t K_1(v, N_{n-1}(v))dv, \\ T_n(t) &= \Sigma(\alpha_2)K_2(t, T_{n-1}(t)) + \sigma(\alpha_2)\int_0^t K_2(v, T_{n-1}(v))dv, \\ I_n(t) &= \Sigma(\alpha_3)K_3(t, I_{n-1}(t)) + \sigma(\alpha_3)\int_0^t K_3(v, I_{n-1}(v))dv, \\ C_n(t) &= \Sigma(\alpha_4)K_4(t, C_{n-1}(t)) + \sigma(\alpha_4)\int_0^t K_4(v, I_{n-1}(v))dv. \end{aligned} \right\}. \quad (23)$$

Moreover, the initial components for the above-mentioned recursive formal are considered to be as the following equation:

$$N_0(t) = N_0, T_0(t) = T_0, I_0(t) = I_0, C_0(t) = C_0 \quad (24)$$



Equation (23) can be written in the following formation:

$$N_n(t) = \sum_{i=1}^n \mu_i(t), T_n(t) = \sum_{i=1}^n \xi_i(t), I_n(t) = \sum_{i=1}^n \omega_i(t), C_n(t) = \sum_{i=1}^n \kappa_i(t). \quad (25)$$

where  $\mu_i(t)$ ,  $\xi_i(t)$ ,  $\omega_i(t)$  and  $\kappa_i(t)$  are defined as follows:

$$\left. \begin{aligned} \mu_i(t) &= N_i(t) - N_{i-1}(t), \xi_i(t) = T_i(t) - T_{i-1}(t), \\ \omega_i(t) &= I_i(t) - I_{i-1}(t), \kappa_i(t) = C_i(t) - C_{i-1}(t). \end{aligned} \right\} \quad (26)$$

Now, the following inequalities can be derived for the functions defined in Equation (26):

$$\begin{aligned} \|\mu_n(t)\| &= \|N_i(t) - N_{i-1}(t)\| \\ &\leq \Sigma(\alpha_1) \| (K_1(t, N_{n-1}(t)) - K_1(t, N_{n-2}(t))) \| \\ &\quad + \sigma(\alpha_1) \int_0^t \| (K_1(v, N_{n-1}(v)) - K_1(v, N_{n-2}(v))) \| dv \end{aligned} \quad (27)$$

Equation (19) showed that the kernel  $K_1$  satisfies the Lipschitz condition. Therefore, for Equation (27) we have:

$$\begin{aligned} \|\mu_n(t)\| &\leq \Sigma(\alpha_1) \delta_1 \|N_{n-1}(t) - N_{n-2}(t)\| + \sigma(\alpha_1) \delta_1 \int_0^t \|N_{n-1}(v) - N_{n-2}(v)\| dv \\ &\leq \Sigma(\alpha_1) \delta_1 \|\mu_{n-1}(t)\| + \sigma(\alpha_1) \delta_1 \int_0^t \|\mu_{n-1}(v)\| dv \end{aligned} \quad (28)$$

The same is true for the other defined functions in Equation (26); for this reason, we can obtain the following inequalities readily:

$$\left. \begin{aligned} \|\xi_n(t)\| &\leq \Sigma(\alpha_2) \delta_2 \|\xi_{n-1}(t)\| + \sigma(\alpha_2) \delta_2 \int_0^t \|\xi_{n-1}(v)\| dv, \\ \|\omega_n(t)\| &\leq \Sigma(\alpha_3) \delta_3 \|\omega_{n-1}(t)\| + \sigma(\alpha_3) \delta_3 \int_0^t \|\omega_{n-1}(v)\| dv, \\ \|\kappa_n(t)\| &\leq \Sigma(\alpha_4) \delta_4 \|\kappa_{n-1}(t)\| + \sigma(\alpha_4) \delta_4 \int_0^t \|\kappa_{n-1}(v)\| dv. \end{aligned} \right\} \quad (29)$$

□

**Remark 4.** The Caputo–Fabrizio fractional model of cancer chemotherapy (Equation (11)) has a system of solutions if following inequalities hold at a time such that  $t_1 > 0$ :

$$\Sigma(\alpha_i) \delta_i + \sigma(\alpha_i) \delta_i t_1 < 1 \quad \text{for } i = 1, 2, 3, 4 \quad (30)$$

**Proof.** In this part, we have proven the existence and smoothness of the defined functions in Equation (25), the existence of a system of solutions for the model in Equation (11), and Equation (12) has been illustrated.

It has been assumed that  $N(t)$ ,  $T(t)$ ,  $I(t)$  and  $C(t)$  are bounded  $\|N(t)\| \leq \theta_1$ ,  $\|T(t)\| \leq \theta_2$ ,  $\|I(t)\| \leq \theta_3$  and  $\|C(t)\| \leq \theta_4$ . Moreover, we have proven that each of the defined kernels satisfies the Lipschitz condition. Consequently, we can obtain the following results using Equations (28) and (29):

$$\left. \begin{aligned} \|\mu_n(t)\| &\leq \|N(0)\| [\Sigma(\alpha_1) \delta_1 + \sigma(\alpha_1) \delta_1 t]^n, \\ \|\xi_n(t)\| &\leq \|T(0)\| [\Sigma(\alpha_2) \delta_2 + \sigma(\alpha_2) \delta_2 t]^n, \\ \|\omega_n(t)\| &\leq \|I(0)\| [\Sigma(\alpha_3) \delta_3 + \sigma(\alpha_3) \delta_3 t]^n, \\ \|\kappa_n(t)\| &\leq \|C(0)\| [\Sigma(\alpha_4) \delta_4 + \sigma(\alpha_4) \delta_4 t]^n. \end{aligned} \right\} \quad (31)$$



Now, it has been demonstrated that the functions  $N_n(t), T_n(t), I_n(t)$  and  $C_n(t)$  defined in Equation (25) converge to solutions of the model (Equation (11)) with the initial condition (Equation (12)), by defining remainder terms after  $n$  iteration as follows:

$$\left. \begin{aligned} N(t) - N(0) &= N_n(t) - X_n(t), \\ T(t) - T(0) &= T_n(t) - Y_n(t), \\ I(t) - I(0) &= I_n(t) - Z_n(t), \\ C(t) - C(0) &= C_n(t) - V_n(t). \end{aligned} \right\}. \quad (32)$$

To prove that the functions  $N_n(t), T_n(t), I_n(t)$  and  $C_n(t)$  converge to the solutions of the model (Equation (11)), we must show that as  $n \rightarrow \infty$ , then the reminder term converges to zero. Using the Lipschitz condition for the kernels defined in Equation (14) and triangle inequality, the following results will be obtained:

$$\begin{aligned} \|X_n(t)\| &\leq \Sigma(\alpha_1) \|K_1(t, N(t)) - K_1(t, N_{n-1}(t))\| \\ &\quad + \sigma(\alpha_1) \int_0^t \|K_1(v, N(v)) - K_1(v, N_{n-1}(v))\| dv \\ &\leq \Sigma(\alpha_1) \delta_1 \|N_{n-1}(t) - N_{n-2}(t)\| + \sigma(\alpha_1) \delta_1 \|N_{n-1}(t) - N_{n-2}(t)\| t. \end{aligned} \quad (33)$$

Continuing the above process will lead us to the following inequality:

$$\|X_n(t)\| \leq [\Sigma(\alpha_1) \delta_1 + \sigma(\alpha_1) \delta_1 t]^{n+1} \theta_1 \quad (34)$$

Now, by taking Equation (30) into account and when  $n \rightarrow \infty$  at time  $t_1$ , by taking a limit on both sides of Equation (34) the following result will be acquired:

$$\lim_{n \rightarrow \infty} \|X_n(t)\| \leq \lim_{n \rightarrow \infty} [\Sigma(\alpha_1) \delta_1 + \sigma(\alpha_1) \delta_1 t]^{n+1} \theta_1. \quad (35)$$

It can be seen that the right-hand side of Equation (35) converges to zero; for this reason, it can be concluded that  $\|X_n(t)\| \rightarrow 0$  when  $n \rightarrow \infty$ . Using the same manner for other remainder terms defined in Equation (32), the following results will be obtained:

$$\lim_{n \rightarrow \infty} \|Y_n(t)\| = 0, \quad \lim_{n \rightarrow \infty} \|Z_n(t)\| = 0, \quad \lim_{n \rightarrow \infty} \|V_n(t)\| = 0 \quad (36)$$

Equations (35) and (36) demonstrate the existence of a system of solutions for the model in Equations (11) and (12).  $\square$

**Remark 5.** The system of solutions of the model in Equations (11) and (12) will be unique if the following inequality is satisfied:

$$\Sigma(\alpha_i) \delta_i + \sigma(\alpha_i) \delta_i t < 1 \quad \text{for } i = 1, 2, 3, 4. \quad (37)$$

**Proof.** To prove that the model described in Equation (11) with initial conditions in Equation (12) has a unique system of solutions, we have assumed that the first system of solutions of the model is  $\{N(t), T(t), I(t), C(t)\}$ . Furthermore, we have considered another system of solutions for the model, which is denoted by  $\{N_1(t), T_1(t), I_1(t), C_1(t)\}$ . Then, using Equation (16), we have

$$\begin{aligned} N(t) - N_1(t) &= \Sigma(\alpha_1) (K_1(t, N(t)) - K_1(t, N_1(t))) \\ &\quad + \sigma(\alpha_1) \int_0^t (K_1(v, N(v)) - K_1(v, N_1(v))) dv \end{aligned} \quad (38)$$

Then, we have

$$\begin{aligned} \|N(t) - N_1(t)\| &= \\ &\leq \Sigma(\alpha_1) \|K_1(t, N(t)) - K_1(t, N_1(t))\| \\ &\quad + \sigma(\alpha_1) \left\| \int_0^t (K_1(v, N(v)) - K_1(v, N_1(v))) dv \right\|. \end{aligned} \quad (39)$$

As we proved, the kernels satisfy the Lipschitz conditions; as a result,

$$\|N(t) - N_1(t)\| \leq \Sigma(\alpha_1)\delta_1\|N(t) - N_1(t)\| + \sigma(\alpha_1)\delta_1\|N(t) - N_1(t)\|t \quad (40)$$

By subtracting the right-hand side of Equation (40) from both side of this inequality, the following inequality is given:

$$(1 - \Sigma(\alpha_1)\delta_1 - \sigma(\alpha_1)\delta_1 t)\|N(t) - N_1(t)\| \leq 0 \quad (41)$$

Considering Equation (37) and the fact that the output of the norm function is non-negative, the following result will be obtained:

$$\|N(t) - N_1(t)\| = 0 \quad (42)$$

In the same way for  $T(t), I(t), C(t)$ , the following equations have been obtained:

$$\left. \begin{aligned} \|T(t) - T_1(t)\| &= 0, \\ \|I(t) - I_1(t)\| &= 0, \\ \|C(t) - C_1(t)\| &= 0. \end{aligned} \right\} \quad (43)$$

Equations (42) and (43) imply

$$N(t) = N_1(t), \quad T(t) = T_1(t), \quad I(t) = I_1(t), \quad C(t) = C_1(t) \quad (44)$$

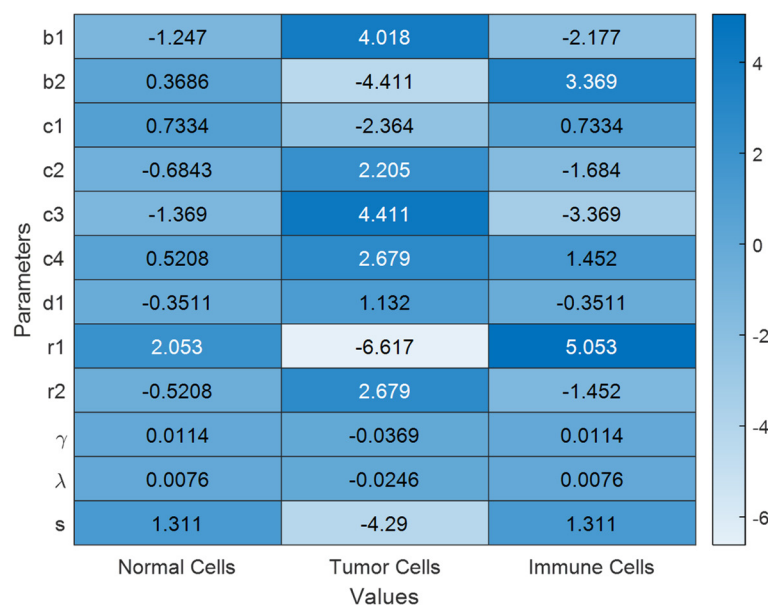
Equation (44) shows that the system of solution for the model in Equation (11) with the initial conditions in Equation (12) is unique, and this is the end of the proof.  $\square$

### 3.2. Sensitivity Analysis

In this section, a sensitivity analysis for the chemotherapy drug-using system given in Equation (11) is conducted. The parameters of the system which have been used for analysis are given in Table 1. The results of the sensitivity analysis are given in Figure 1.

**Table 1.** The simulated patient's parameters [47].

Parameter	Description (Unit)	Value
$a_1$	Fractional immune cell kill rate ( $\text{mg}^{-1} \text{day}^{-1}$ )	0.2
$a_2$	Fractional tumor cell kill rate ( $\text{mg}^{-1} \text{day}^{-1}$ )	0.3
$a_3$	Fractional normal cell kill rate ( $\text{mg}^{-1} \text{day}^{-1}$ )	0.1
$b_1$	Reciprocal carrying capacity of tumor cells ( $\text{cell}^{-1}$ )	1
$b_2$	Reciprocal carrying capacity of normal cells ( $\text{cell}^{-1}$ )	1
$c_1$	Immune cell competition term (competition between immune and tumor cells) ( $\text{cell}^{-1} \text{day}^{-1}$ )	1
$c_2$	Tumor cell competition term (competition between immune and tumor cells) ( $\text{cell}^{-1} \text{day}^{-1}$ )	0.5
$c_3$	Tumor cell competition term (competition between normal and tumor cells) ( $\text{cell}^{-1} \text{day}^{-1}$ )	1
$c_4$	Normal cell competition term (competition between normal and tumor cells) ( $\text{cell}^{-1} \text{day}^{-1}$ )	1
$d_1$	Immune cell death rate ( $\text{day}^{-1}$ )	0.2
$d_2$	Decay rate of injected drug ( $\text{day}^{-1}$ )	1
$r_1$	Per unit growth rate of tumor cells ( $\text{day}^{-1}$ )	1.5
$r_2$	Per unit growth rate of normal cells ( $\text{day}^{-1}$ )	1
$s$	Immune cell influx rate ( $\text{cell day}^{-1}$ )	0.33
$\gamma$	Immune threshold rate (cell)	0.3
$\lambda$	Immune response rate ( $\text{day}^{-1}$ )	0.01



**Figure 1.** Sensitivity analysis.

The relationship between the patient's parameter variations is elucidated in Figure 1. Conspicuously, per-unit growth rate of tumor cells ( $r_1$ ), immune cell influx rate ( $s$ ), and tumor cell competition term (competition between normal and tumor cells) ( $c_3$ ) are three parameters whose variations affect the number of normal cells the most. There are four parameters whose variations have the most impact on the number of tumor cells, of which three of them are the same as the ones for normal cells, and the other parameter is the reciprocal carrying capacity of normal cells ( $b_2$ ). Moreover, the number of immune cells is highly affected by the variation of the per-unit growth rate of tumor cells ( $r_1$ ), the reciprocal carrying capacity of normal cells ( $b_2$ ), and the tumor cell competition term (competition between normal and tumor cells) ( $c_3$ ).

#### 4. Methodology

In this section, we have described the proposed fuzzy-reinforcement-learning based controller (FRLC) whose aim is to control the number of tumor cells; in other words, the aim of the proposed controller is to reach the desired number of tumor cells  $T(t) = T_{desire}$  from a non-zero initial number of tumor cells  $T(0) > 0$ .

Expected-SARSA is a variation of SARSA wherein the variance in its update rule is decreased in comparison with that of the SARSA algorithm, and it has better performance than SARSA for online applications that can be considered as an on-policy version of Q-Learning. On-policy methods have some advantages over off-policy learning algorithms, such as stronger convergence guarantees in cases in which it is combined with function approximation, while the off-policy methods can diverge in those cases [90–92]. They have outstanding performance in online applications because the policy that is estimated in on-policy ways will be improved iteratively, and the agent will behave based on this policy.

The advantage of data-driven controllers is that an accurate model of the system is not required, while selecting the best action for each rule in the fuzzy interface system demands accurate knowledge about the system. However, the proposed control method learns the best action for each rule; therefore, it has the advantages of data-driven controllers and fuzzy controllers simultaneously.

##### *Fuzzy Controller Based on Expected SARSA Learning (FESL)*

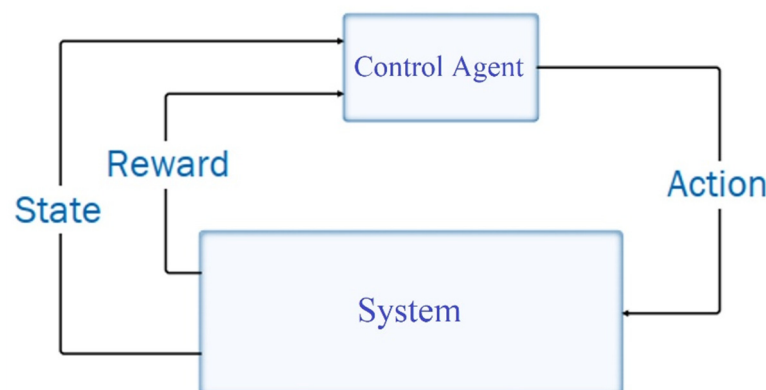
In this paper, in order to control the number of tumor cells, a fuzzy controller is proposed of which its learning process is based on the expected SARSA algorithm. The fuzzy controller uses Expected-SARSA algorithms in order to find the best map between

state and actions. In other words, using the Expected-SARSA algorithm, the best action for each of the fuzzy system inference's rules will be obtained. Consider the rules in a fuzzy inference system be as follows:

$R_i$ : if  $x_1$  is  $L_{1i}$  and ... and  $x_n$  is  $L_{ni}$  then  $(a_i \text{ is } a_{i1})$  or  $(a_i \text{ is } a_{i2})$  or ... or  $(a_i \text{ is } a_{ik})$ ,

in which  $R_i$  is the  $i$ -th rule of the fuzzy inference system, and  $s$  is the vector of the  $n$ -dimensional input state and is defined as  $s = x_1 \times x_2 \times \dots \times x_n$ ;  $L_i$  is the  $n$ -dimensional strictly convex and normal fuzzy set of the  $i$ -th rule with a unique center and is defined as  $L_i = L_{i1} \times L_{i2} \times \dots \times L_{in}$ ;  $a_i$  is the consequent action for the  $i$ -th rule;  $a_{i1}$  is the first candidate action for the  $i$ -th rule;  $a_{i2}$  is the second candidate action; and finally,  $a_{ik}$  is the  $k$ th candidate action for the  $i$ -th rule. The aim of the learning is to find the optimal action for each of the rules of the fuzzy inference system. So as to seek this aim, a value action matrix has been defined and is denoted by  $Q$ , of which its elements are the value action of each candidate action for each of the rules. The optimal action for each of the rules is the candidate action for that rule, of which its value action is the most among all the candidate actions of that rule.

As aforementioned, the best consequent action for each of the rules will be obtained using reinforcement learning methods. Figure 2 depicts the structure of reinforcement learning. In this paper, the reinforcement learning algorithm that has been implemented to update value actions is the Expected-SARSA method.



**Figure 2.** Structure of reinforcement learning.

The algorithm used in this paper has been delineated in Algorithm 1. The function that is used for calculating the reward of the agent for a transition from state  $s_k$  to state  $s_{k+1}$  is as follows:

$$r_{k+1} = \begin{cases} \frac{e(kT) - e((k+1)T)}{e(kT)}, & e((k+1)T) < e(kT) \\ 0 & e((k+1)T) \geq e(kT) \end{cases} \quad (45)$$

where  $e(t), t \geq 0$  is defined as:

$$e(t) = \beta(T(t) - T_d(t)) + (1 - \beta)(N(t) - N_d(t)). \quad (46)$$

The goal of the control agent in reinforcement learning (RL) is to find an optimal policy  $\pi^*$ ; using this policy, the expected discount reward would be maximum. The discount reward is defined as follows:

$$R_t = r_{t+1} + \delta r_{t+2} + \delta^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \delta^k r_{t+k+1} \quad (47)$$

where  $\delta$  is discount factor and is a non-negative constant, which satisfies  $\delta \leq 1$ .

The output action of the fuzzy inference system is calculated as

$$a = \frac{\sum_{i=1}^n \eta_i a_{ip}}{\sum_{i=1}^n \eta_i} \quad (48)$$

where  $n$  is the number of fuzzy rules in the fuzzy inference,  $\eta_i$  is the firing strength of the  $i$ -th rule, and  $a_{ip}$  is the selected action using the  $\varepsilon$ -greedy method for the  $i$ -th rule among all possible actions. Updating the rule for FESL is considered to be as

$$Q_t[i, p] \leftarrow Q_{t-1}[i, p] + \alpha \frac{\eta_i}{\sum_{i=1}^n \eta_i} \gamma_t \quad (49)$$

where  $\alpha$  denotes the learning rate of the algorithm. In Equation (49),  $\gamma_t$  is defined as:

$$\gamma_t = r_{t+1} + \delta \left[ \sum_{k=1}^n \frac{\eta_k}{\sum_{t=1}^n \eta_t} \sum_{j=1}^m \pi_{s'}[k, j] Q_t[k, j] \right] - Q_{t-1}[i, p]. \quad (50)$$

in which  $\delta$  is a discount factor,  $\pi_{s'}[k, j]$  is the probability of selecting the  $j$ -th action for the  $k$ -th rule that has been obtained using  $\varepsilon$ -greedy method, and  $Q_t[k, j]$  is an approximation of the value of the  $j$ -th action for the  $k$ -th rule.

---

**Algorithm 1** Fuzzy Expected SARSA

---

```

1: Initialize Q-table
2: Loop {for all of episodes}
3:   Initialize state  $s$ 
4:   repeat {for each step in episode}
5:     calculate firing strength of each rule ( $\alpha_i$ ) at state  $s$ 
6:     choose action ( $\alpha_i$ ) for each of the rules at state  $s$  using policy  $\pi$ 
7:     calculate action  $a$  at state  $s$ 
8:     take action  $a$ , observe reward  $r$  and next state  $s'$ 
9:     calculate state value of state  $s'$ 
10:    update Q-table
11:  until  $s$  is terminal state
12: end loop

```

---

One of the key features of FRL-based controllers is their ability to handle uncertainty and imprecision in the system. Fuzzy logic allows the controller to make decisions based on approximate rather than precise values, which can be helpful in situations where the system is not well understood or where there is a high degree of variability. In addition to its ability to handle uncertainty, an FRL-based controller can also learn from its past experiences and adapt its behavior to better achieve the desired outcomes. This learning process is known as reinforcement learning, and it allows the controller to optimize its performance over time by adjusting its control inputs based on the consequences of those inputs. Overall, the combination of fuzzy logic and reinforcement learning in an FRL-based controller can make it well-suited for handling unexpected situations and adapting to a range of conditions.

## 5. Numerical Simulations

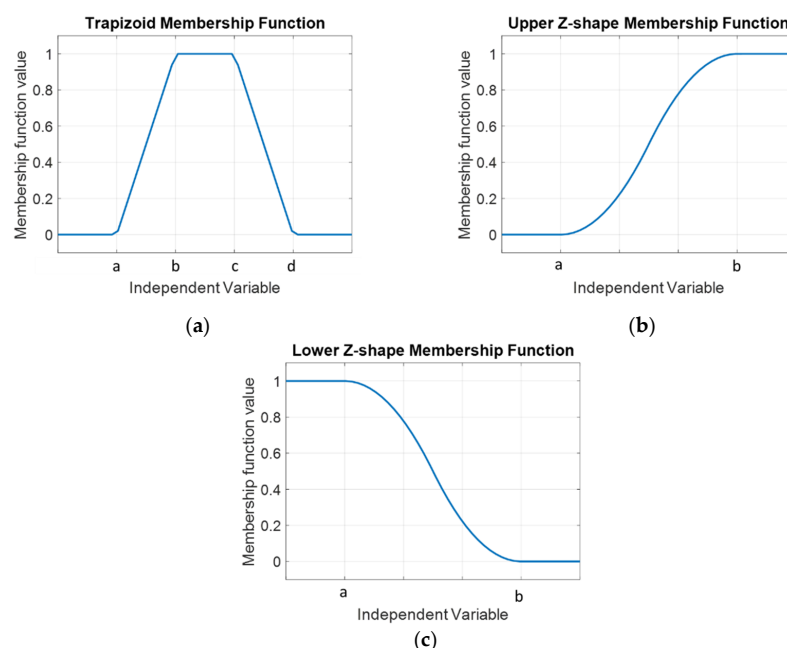
In this section, the performance of the FESL method for controlling the closed-loop system of cancer chemotherapy drug dosing, of which its input variable is considered to be a chemotherapy drug such as carboplatin, is assessed through simulations that are done in MATLAB. Oncologists consider different factors when they determine the drug dosage for a cancer patient. Age and gender are two examples of those factors. As has been shown in [87], the growth rate of normal cells and immune cells depends on age, and this rate will be larger for younger patients rather than elderly patients. Therefore, for young patients, oncologists prefer to eradicate cancerous cells as soon as possible without

taking the damage to normal cells into consideration, which avoids cancer metastasis. In the current study, we assumed the value of fractional derivative as is mentioned in Table 1. However, estimating the fractional derivative parameter in the same way as the other parameters helps to ensure that the model is as accurate and reliable as possible. This can be done using a variety of techniques such as fitting the model to experimental data or using statistical methods to optimize the model parameters. By taking the time to accurately estimate all of the model parameters, researchers can be more confident in the results of their study and the conclusions they draw from the data.

For the cases in which the patient is elderly, the patient suffers from other diseases, or the cancer is in a vital organ such as the brain, the degeneration of the normal cells is not desirable, and the normal cells should be kept undamaged. The above-mentioned conditions have been taken into consideration by the proposed control method through the defining error in Equation (46).

For this reason, two scenarios have been considered to evaluate the performance of the proposed control method in the abovementioned cases. In scenario “A”, it has been considered that the patient is young. On the other hand, the simulation of scenario “B” has been conducted for an elderly patient. Finally, the results of the simulations for the proposed control method have been compared with those of Watkin’s Q-learning method for both scenarios.

The simulated patient’s parameters are considered to be the same as the ones given in Table 1 [87]; in [93], it was shown how the parameters that are given in Table 1 can be obtained. The number of episodes that have been considered for simulations is considered to be 500, of which each of the episodes is defined as a set of transitions from the initial state to the terminal state. For the simulations, 20 fuzzy-sets have been considered for the fuzzy system, and the membership functions used for the fuzzy sets are considered to be trapezoidal-shape and z-shape membership functions, which have been shown in Figure 3. Table 2 implicates the type of membership function used for each fuzzy-set and their features. Furthermore, in the simulations, we have set  $\delta = 0.9$ , and the learning rate is considered to be  $\alpha = 0.2$  and should decrease as the number of iterations and episodes increase so as to guarantee the convergence of the learning algorithm.



**Figure 3.** Membership functions used for fuzzy-sets. (a) Trapezoidal-shape membership function. (b) Z-shape membership function for upper bound. (c) Z-shape membership function for lower bound.

**Table 2.** Fuzzy-sets and their membership functions.

#	Type	[a, b, c, d]	#	Type	[a, b, c, d]
1	zmfl	[0.0055, 0.0068, ~, ~]	11	trapmf	[0.3495, 0.3505, 0.3995, 0.4005]
2	trapmf	[0.0058, 0.0068, 0.0120, 0.0130]	12	trapmf	[0.3995, 0.4005, 0.4495, 0.4505]
3	trapmf	[0.0120, 0.0130, 0.0245, 0.0255]	13	trapmf	[0.4495, 0.4505, 0.4995, 0.5005]
4	trapmf	[0.0245, 0.0255, 0.0395, 0.0405]	14	trapmf	[0.4995, 0.5005, 0.5495, 0.5505]
5	trapmf	[0.0395, 0.0405, 0.0495, 0.0505]	15	trapmf	[0.5495, 0.5505, 0.5995, 0.6005]
6	trapmf	[0.0495, 0.0505, 0.0995, 0.1005]	16	trapmf	[0.5995, 0.6005, 0.6495, 0.6505]
7	trapmf	[0.0995, 0.1005, 0.1995, 0.2005]	17	trapmf	[0.6495, 0.6505, 0.6995, 0.7005]
8	trapmf	[0.1995, 0.2005, 0.2495, 0.2505]	18	trapmf	[0.6995, 0.7005, 0.7995, 0.8005]
9	trapmf	[0.2495, 0.2505, 0.2995, 0.3005]	19	trapmf	[0.7995, 0.8005, 0.8995, 0.9005]
10	trapmf	[0.2995, 0.3005, 0.3495, 0.3505]	20	zmfh	[0.8995, 0.9005, ~, ~]

In Table 2, zmfl denotes the z-shape membership function for the lower bound; trapmf and zmfh stand for trapezoidal-shape and z-shape membership function for the upper bound respectively. Variables “a”, “b”, “c”, and “d” are illustrated in Figure 3.

During simulations, learning algorithms should explore the best actions in order to find the best consequent action for each rule. Keeping this in mind, the  $\varepsilon$  value has been considered to be a small number; then, as the number of step-times and the number of episodes increases, the  $\varepsilon$  value in the  $\varepsilon$ -greedy method increases in order to reach a greedy policy. The  $\varepsilon$  has been calculated using the following formula:

$$\varepsilon = 0.3 \left( 0.5 + \frac{I}{MNI} \right) + 0.25 \left( 1 + 3 \frac{E}{MNE} \right) \quad (51)$$

where  $I$  is the iteration number or the time-step number,  $MNI$  is the maximum number of iterations,  $E$  is the episode number, and  $MNE$  is the maximum number of episodes; in the simulations, the maximum number of iterations is chosen as 500. The candidate actions for each of the rules are considered to be  $a_i \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, 10\}$ . Moreover, the initial value of the state variables of the system model in Equation (11) are considered to be

$$N(0) = 0.6, T(0) = 0.8, I(0) = 0.3, C(0) = 0.2. \quad (52)$$

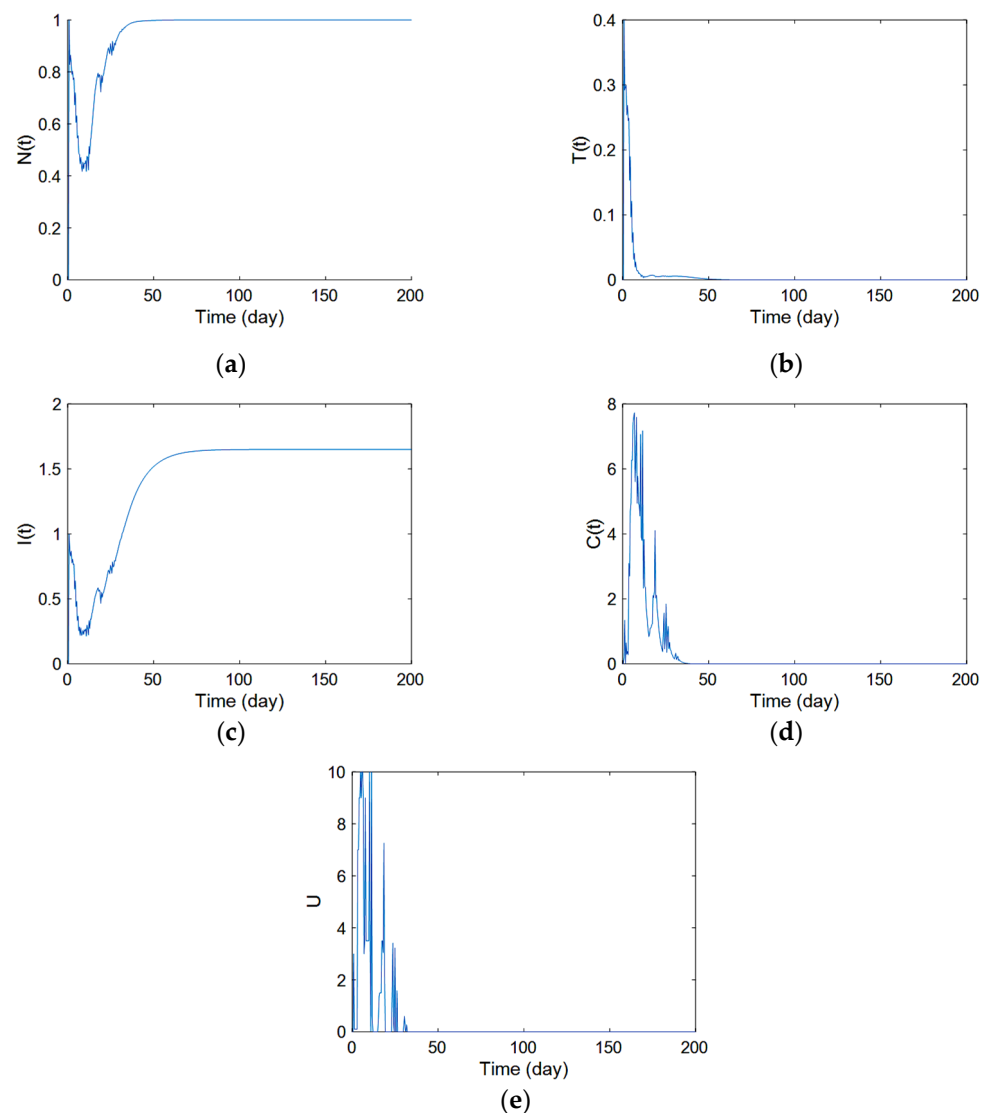
### 5.1. Scenario A

In this scenario, first, the simulations have been conducted for the patient system, which has no uncertainty. Furthermore, the aim is to reach  $T = 0$ ; to put it differently, the controller must take the system from the initial state to the first fuzzy set. Therefore, for the simulation, we have set  $\beta = 1$ , which means in this scenario the error will be  $(t) = (T(t) - T_d(t))$ .

The result of the simulation is given in Figure 4 and is comprised of the number of normal cells in (a), the number of tumor cells in (b), the number of immune cells in (c), the concentration of the chemotherapeutic drug in blood (d), and the amount of chemotherapeutic drug executed for scenario “A” (d). As is evident, the number of tumor cells decreased consistently until it reached zero, which means that the proposed controller was able to obliterate all of the tumor cells. However, due to injecting a chemotherapeutic drug into the body of the patient, the number of normal cells and immune cells decreased



at the beginning of the chemotherapy. Nonetheless, later, the numbers of normal cells and immune cells increased.



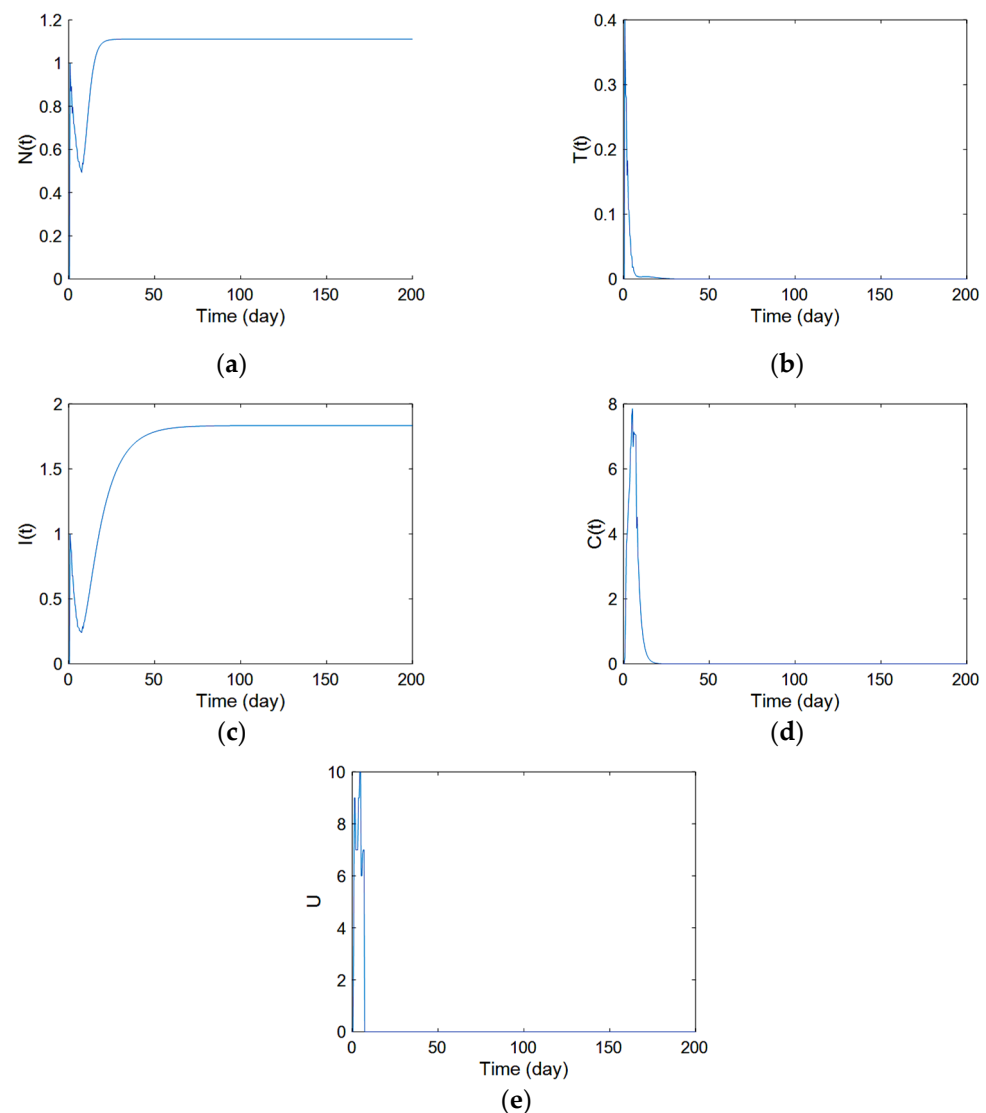
**Figure 4.** The result of simulation using the proposed control method for young patient without uncertainty. (a) Normal-cells (b) Tumor-cells (c) Immune-cells (d) Concentration of cells (e) Control action.

The simulation confirms that the chemotherapy treatment was successful in eliminating all of the tumor cells, but it caused a decrease in the number of normal cells and immune cells at the beginning of treatment. This is a common occurrence with chemotherapy, as the drugs used to kill cancer cells can also damage healthy cells. However, it is also possible for the body to recover and for the number of normal cells and immune cells to increase again after treatment.

It is important to carefully consider the potential benefits and risks of chemotherapy treatment, as it can be an effective way to treat cancer; however, it can also cause significant side effects. It is generally recommended to undergo chemotherapy under the supervision of a medical professional who can help to monitor the patient's condition and adjust the treatment plan as needed.

So as to prove the robustness of the proposed control method, after training the control agent using the reinforcement learning algorithm, the parameters of the patients were varied during simulation. The parameters of the modified patient system were considered

to have a 10% variation from their original values, which are given in Table 1. The result of the simulation is depicted in Figure 5.



**Figure 5.** The result of simulation using the proposed control method for young patient with uncertainty. (a) Normal-cells (b) Tumor-cells (c) Immune-cells (d) Concentration of cells (e) Control action.

Figure 5 has the same sub-figures as Figure 4. As can be seen, the system has shown the same behavior as the system without uncertainty. Obviously, the controller has shown its ability to eradicate the tumor cells in the presence of the uncertainty in the parameters.

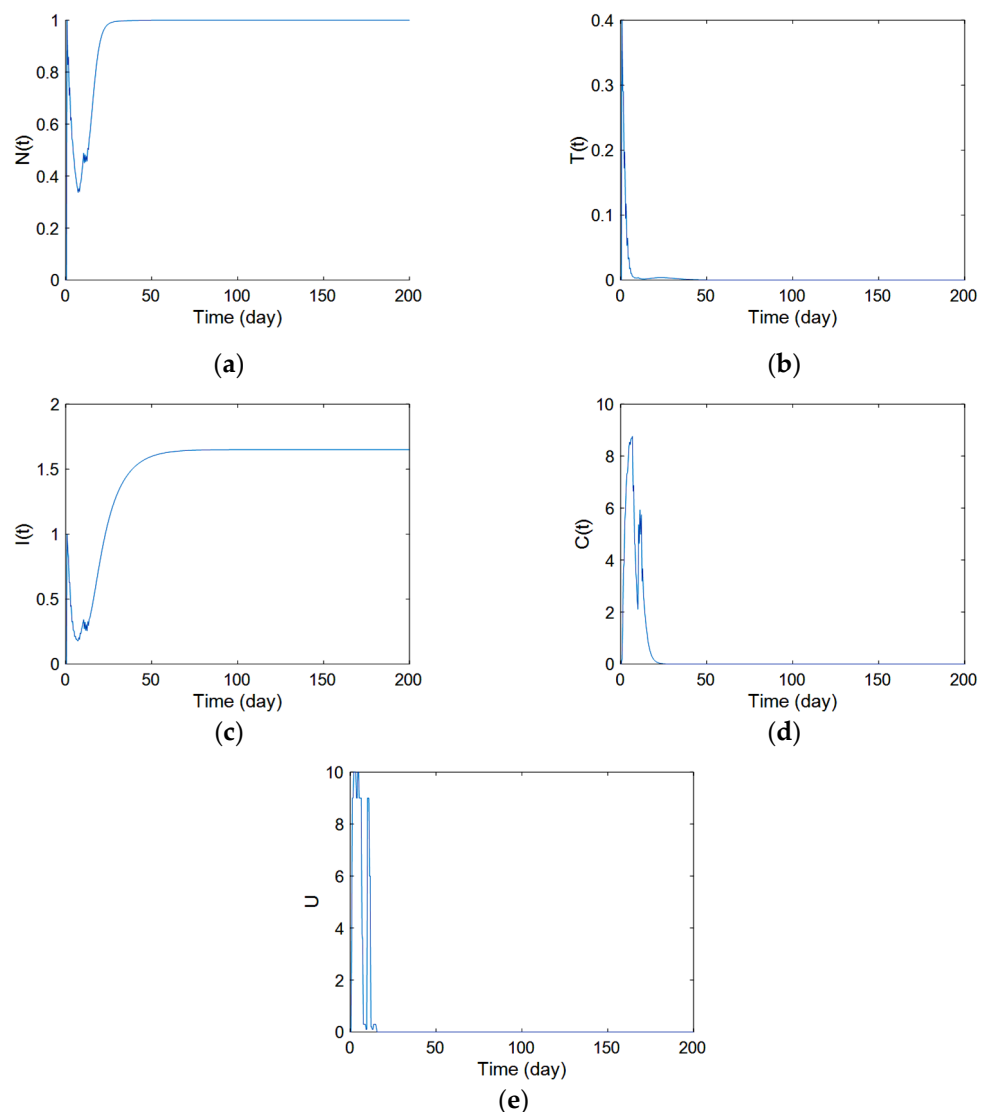
The obtained results are remarkable because the following main achievements:

- The effectiveness of the proposed controller in eliminating the tumor cells: The results show that the controller was able to successfully reduce the number of tumor cells to zero, which indicates that the treatment was effective in destroying the cancer cells.
- The temporary decrease in normal cells and immune cells: The chemotherapy used in the treatment caused a temporary decrease in the number of normal cells and immune cells in the body. This is a common side effect of chemotherapy, as the drugs used can also harm healthy cells.
- The recovery of normal cells and immune cells: Despite the initial decrease in normal cells and immune cells, the numbers of these cells eventually increased over time. This suggests that the body was able to recover and rebuild healthy cells after the chemotherapy treatment.

- The importance of monitoring the effects of chemotherapy: The results of the simulation highlight the importance of carefully monitoring the effects of chemotherapy to ensure that it is being administered effectively and safely. This includes monitoring the number of cancer cells, normal cells, and immune cells in the body, as well as the concentration of chemotherapy drugs in the blood.

### 5.2. Scenario B

This scenario has been considered to assess the performance of the proposed control method in control of the tumor-cell state in old patients. In this scenario, the aim is to control both tumor-cells and normal-cells. Consequently, in this scenario, it has been considered that  $\beta = 0.95$ , which means the error in this scenario will be  $e(t) = 0.95(T(t) - T_d(t)) + 0.05(N(t) - N_d(t))$ ; as can be seen, the number of normal cells has been taken into consideration in this scenario. The result of the simulations for this scenario is illustrated in Figure 6.



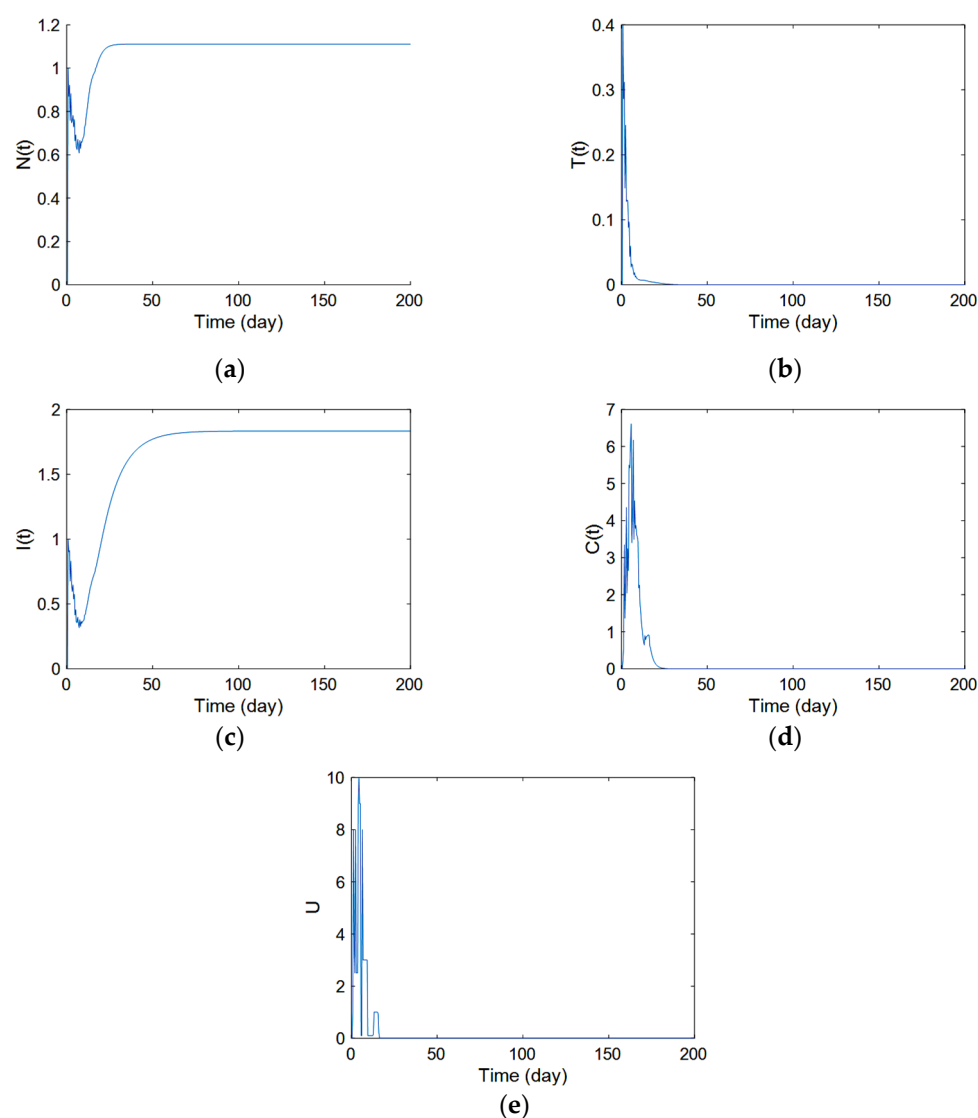
**Figure 6.** The result of simulation using the proposed control method for elderly patient without uncertainty. (a) Normal-cells (b) Tumor-cells (c) Immune-cells (d) Concentration of cells (e) Control action.

By comparing Figures 4 and 6, it can be confirmed that in the case in which the patient is elderly, the number of normal cells has reached its maximum faster. Moreover, the

2-norm of the amount of drug executed to the patient's body in the case that the patient is elderly decreased by 5.5 percent.

It is remarkable that the proposed control method was able to reach the maximum number of normal cells and use a smaller amount of chemotherapy drugs more quickly in the case of an elderly patient compared to a non-elderly patient. This suggests that the control method was able to minimize the negative effects of the chemotherapy on healthy cells while still effectively treating the cancer. This is important because chemotherapy can have significant side effects on the body, and minimizing these effects is especially important for elderly patients who may be more sensitive to these effects.

In addition, in this scenario, in order to confirm the robustness of the proposed control method for elderly patients, the simulation was performed for the patient system with 10% uncertainty for each of the system parameters, and the results are given in Figure 7.



**Figure 7.** The result of simulation using the proposed control method for elderly patient with uncertainty. (a) Normal-cells (b) Tumor-cells (c) Immune-cells (d) Concentration of cells (e) Control action.

It is notable that the control method was able to maintain its effectiveness even when the system parameters were varied by 10%. This indicates that the control method is robust and can be applied effectively to elderly patients with a range of different characteristics. This is important because elderly patients may have a wide range of health conditions and characteristics that can affect their response to chemotherapy.

Figure 7 demonstrates the robustness of the proposed control method against the system's parameter variation for an elderly patient. The results of the simulations for the elderly patient in both cases show the effectiveness of the proposed control method. Overall, these findings suggest that the proposed control method could be a valuable tool for improving the treatment of elderly patients with chemotherapy. They could help to minimize the negative effects of chemotherapy on healthy cells and provide a more effective treatment for cancer.

### 5.3. Scenario C

In this scenario, the proposed control method and Watkin's Q-learning method were exerted to 10 different patients. So as to evaluate the performance of the proposed control method, the simulation results for the proposed control method and the Watkin's Q-learning method are elucidated in Table 3. The patient's parameters are considered to be as follows:  $a_i \in (0.1, 0.5]$  for  $i = 1, 2, 3$ ,  $a_i \in [0.3, 1]$ ,  $i = 1, 2, 3, 4$ ,  $d_1 \in [0.15, 0.3]$ ,  $r_1 \in [1.2, 1.6]$ ,  $r_1 \in [0.3, 0.5]$ ,  $\gamma \in [0.3, 0.5]$  and  $\lambda \in [0.01, 0.05]$ ; other parameters are the same as ones given in Table 1. Furthermore, the parameters must hold the following conditions [47]:  $a_3 \leq a_1 \leq a_2$  and  $b_1 \geq b_2 = 1$ .

**Table 3.** Result of simulations.

	Young Patients			Old Patients		
	$M(\ E\ _2)$	$M(\text{var}(E))$	$M(\ U\ _2)$	$M(\ E\ _2)$	$M(\text{var}(E))$	$M(\ U\ _2)$
FESL	1.9633	0.0011	130.1827	3.0863	0.0011	123.0201
Q-Learning	3.0206	0.0079	131.7332	4.1083	0.0066	137.0524

where  $M(\cdot)$ ,  $\text{var}(\cdot)$  and  $\|\cdot\|_2$  denote the mean function, variance function, and the norm function, respectively.

Conspicuously, as can be seen in Table 3, the 2-norm of the input signal for the proposed control method is less than that of the Watkin's Q-learning method; that means that the proposed control method used less drug than the conventional controller. In any case, the number of episodes for the proposed control method is less than 500 episodes, and it is significantly less than the required number of episodes for the Watkins Q-learning algorithm's convergence, which is 50,000 [57]; therefore, the proposed control method is 100 times faster than the Watkins Q-learning algorithm in term of convergence rate.

Table 3 implies that the 2-norm of the error in the case in which the proposed control method was used decreased by 35 percent for scenario "A" and 24 percent for scenario "B", in comparison with the case in which the Watkin's Q-learning algorithm was used. In any case, the 2-norm of the variance of the error and amount of drug usage decreased by 86 percent and 1 percent, respectively, for scenario "A" and 83 percent and 10 percent for scenario "B".

It is noteworthy to mention that the proposed control method can be exerted by clinicians to find the optimal amount of drug based on the patient's state to eradicate tumor cells. To this aim, clinicians must first obtain the patients' parameters by measuring the patient's state and fit the logged data to the mathematical model of the patient given in Equation (11). Then, using the mathematical model, the control agent should be pre-trained. Afterwards, the control agent would be able to offer the optimal dosage of the drug for each day. Keeping that in mind, the training process of the control agent goes on during the chemotherapy process. For this reason, the control framework performance would not be affected by varying the patient's parameters during chemotherapy.

## 6. Conclusions

This study was aimed at the investigation of a Caputo–Fabrizio fractional-order model of cancer chemotherapy treatment. At first, the existence and uniqueness of the solutions of the proposed model were proven through a fix-point theorem and an iterative method. After that, since applying optimal policies for drug dosage is crucial, we proposed to for-

mulate the control problem of the chemotherapy treatment as an optimization problem and find optimal actions using a fuzzy reinforcement learning algorithm. The significant features of the designed fuzzy reinforcement learning-based method are its model-free approach as well as its optimal performance. Finally, three scenarios were considered to evaluate the performance of the proposed control technique. The results for the proposed control method demonstrate the effectiveness of the controller in the annihilation of tumor cells for young patients and elderly patients. The control method has the ability to bring the level of normal cells back to its normal range. In addition, the results implicate the robustness of the proposed control method against the uncertainties of the patients' parameters. Moreover, because of being real-time, the performance of the proposed control method would not be affected even if the patients' parameters had been changed during chemotherapy treatment. Also, the presented comparison with the Watkin's Q-learning method conspicuously demonstrated the superiority of the proposed method in terms of the annihilation of the tumor cells in simulated patients and the drug usage during chemotherapy treatment. While in the current study it was shown that the fuzzy reinforcement learning algorithm is also able to achieve a good level of optimization in this task, deep reinforcement learning has the potential to achieve even better results by taking into account a greater amount of data and making more sophisticated decisions. However, it is important to note that this is an area of active research, and it is not yet clear how well deep reinforcement learning will perform in this specific application. Further research and development will be needed to determine whether this approach is practical and effective for scheduling cancer chemotherapy drug dosing. Furthermore, it would be helpful to run a statistical test using variance to determine whether there are statistically significant differences between two or more categorical groups.

**Author Contributions:** Methodology, F.E.A., A.Y., C.V., S.B., and H.J.; Validation, F.E.A., A.Y., C.V., S.B., and H.J.; Investigation, F.E.A., A.Y., C.V., S.B., and H.J.; Writing—original draft, F.E.A., A.Y., C.V., S.B., and H.J.; Supervision, F.E.A., A.Y., C.V., S.B., and H.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 148-611-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 148-611-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

FESL	fuzzy controller based on Expected-SARSA learning
FRLC	Fuzzy reinforcement learning based controller
RL	reinforcement learning
trapmf	trapezoidal-shape membership function for upper bound
SARSA	State-Action-Reward-State-Action
zmfl	z-shape membership function for lower bound
zmfh	z-shape membership function for upper bound

## References

1. Miller, K.D.; Nogueira, L.; Mariotto, A.B.; Rowland, J.H.; Yabroff, K.R.; Alfano, C.M.; Jemal, A.; Kramer, J.L.; Siegel, R.L. Cancer treatment and survivorship statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 363–385. [[CrossRef](#)] [[PubMed](#)]
2. Dos Santos, A.I.F.; De Almeida, D.R.Q.; Terra, L.F.; Baptista, M.c.S.; Labriola, L. Photodynamic therapy in cancer treatment—an update review. *J. Cancer Metastasis Treat* **2019**, *5*, 25. [[CrossRef](#)]

3. Thallinger, C.; Füreder, T.; Preusser, M.; Heller, G.; Müllauer, L.; Höller, C.; Prosch, H.; Frank, N.; Swierzewski, R.; Berger, W. Review of cancer treatment with immune checkpoint inhibitors. *Wien. Klin. Wochenschr.* **2018**, *130*, 85–91. [CrossRef]
4. Yagawa, Y.; Tanigawa, K.; Kobayashi, Y.; Yamamoto, M. Cancer immunity and therapy using hyperthermia with immunotherapy, radiotherapy, chemotherapy, and surgery. *J. Cancer Metastasis Treat* **2017**, *3*, 218. [CrossRef]
5. Cui, C.; Yang, J.; Li, X.; Liu, D.; Fu, L.; Wang, X. Functions and mechanisms of circular RNAs in cancer radiotherapy and chemotherapy resistance. *Mol. Cancer* **2020**, *19*, 1–16. [CrossRef]
6. Coates, A.; Abraham, S.; Kaye, S.B.; Sowerbutts, T.; Frewin, C.; Fox, R.M.; Tattersall, M.H.N. On the receiving end—Patient perception of the side-effects of cancer chemotherapy. *Eur. J. Cancer Clin. Oncol.* **1983**, *19*, 203–208. [CrossRef]
7. Toker, D.; Sommer, F.T.; D'Esposito, M. A simple method for detecting chaos in nature. *Commun. Biol.* **2020**, *3*, 11. [CrossRef]
8. Chen, T.; Kirkby, N.F.; Jena, R. Optimal dosing of cancer chemotherapy using model predictive control and moving horizon state/parameter estimation. *Comput. Methods Programs Biomed.* **2012**, *108*, 973–983. [CrossRef]
9. Sbeity, H.; Younes, R. Review of optimization methods for cancer chemotherapy treatment planning. *J. Comput. Sci. Syst. Biol.* **2015**, *8*, 74. [CrossRef]
10. Michor, F. Mathematical models of cancer stem cells. *J. Clin. Oncol.* **2008**, *26*, 2854–2861. [CrossRef]
11. Granata, D.; Lorenzi, L. An Evaluation of Propagation of the HIV-Infected Cells via Optimization Problem. *Mathematics* **2022**, *10*, 2021. [CrossRef]
12. Mokhtare, Z.; Vu, M.T.; Mobayen, S.; Rojsiraphisal, T. An adaptive barrier function terminal sliding mode controller for partial seizure disease based on the Pinsky–Rinzel mathematical model. *Mathematics* **2022**, *10*, 2940. [CrossRef]
13. Jahanshahi, H.; Munoz-Pacheco, J.M.; Bekiros, S.; Alotaibi, N.D. A fractional-order SIRD model with time-dependent memory indexes for encompassing the multi-fractional characteristics of the COVID-19. *Chaos Solitons Fractals* **2021**, *143*, 110632. [CrossRef] [PubMed]
14. Jahanshahi, H. Smooth control of HIV/AIDS infection using a robust adaptive scheme with decoupled sliding mode supervision. *Eur. Phys. J. Spec. Top.* **2018**, *227*, 707–718. [CrossRef]
15. Jahanshahi, H.; Shanazari, K.; Mesrizadeh, M.; Soradi-Zeid, S.; Gómez-Aguilar, J.F. Numerical analysis of Galerkin meshless method for parabolic equations of tumor angiogenesis problem. *Eur. Phys. J. Plus* **2020**, *135*, 866. [CrossRef]
16. Bachmann, J.; Raue, A.; Schilling, M.; Becker, V.; Timmer, J.; Klingmüller, U. Predictive mathematical models of cancer signalling pathways. *J. Intern. Med.* **2012**, *271*, 155–165. [CrossRef] [PubMed]
17. Wilkie, K.P. A review of mathematical models of cancer–immune interactions in the context of tumor dormancy. In *Systems Biology of Tumor Dormancy*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 201–234.
18. Brady, R.; Enderling, H. Mathematical models of cancer: When to predict novel therapies, and when not to. *Bull. Math. Biol.* **2019**, *81*, 3722–3731. [CrossRef]
19. Ira, J.I.; Islam, M.S.; Misra, J.C. Mathematical Modelling of the Dynamics of Tumor Growth and its Optimal Control. *Preprints* **2020**, 2020040391. Available online: <https://www.preprints.org/manuscript/202004.0391/v2> (accessed on 24 December 2022).
20. Eisen, M. *Mathematical Models in Cell Biology and Cancer Chemotherapy*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 30.
21. Schättler, H.; Ledzewicz, U. *Optimal Control for Mathematical Models of Cancer Therapies*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 42.
22. Kumar, D.; Singh, J.; Baleanu, D. On the analysis of vibration equation involving a fractional derivative with Mittag-Leffler law. *Math. Methods Appl. Sci.* **2020**, *43*, 443–457. [CrossRef]
23. Soradi-Zeid, S.; Jahanshahi, H.; Yousefpour, A.; Bekiros, S. King algorithm: A novel optimization approach based on variable-order fractional calculus with application in chaotic financial systems. *Chaos Solitons Fractals* **2020**, *132*, 109569. [CrossRef]
24. Kumar, D.; Singh, J.; Al Qurashi, M.; Baleanu, D. A new fractional SIRS-SI malaria disease model with application of vaccines, antimalarial drugs, and spraying. *Adv. Differ. Equ.* **2019**, *2019*, 278. [CrossRef]
25. Srivastava, H.M.; Dubey, V.P.; Kumar, R.; Singh, J.; Kumar, D.; Baleanu, D. An efficient computational approach for a fractional-order biological population model with carrying capacity. *Chaos Solitons Fractals* **2020**, *138*, 109880. [CrossRef]
26. Chen, S.-B.; Jahanshahi, H.; Abba, O.A.; Solís-Pérez, J.E.; Bekiros, S.; Gómez-Aguilar, J.F.; Yousefpour, A.; Chu, Y.-M. The effect of market confidence on a financial system from the perspective of fractional calculus: Numerical investigation and circuit realization. *Chaos Solitons Fractals* **2020**, *140*, 110223. [CrossRef]
27. Chen, S.-B.; Soradi-Zeid, S.; Jahanshahi, H.; Alcaraz, R.; Gómez-Aguilar, J.F.; Bekiros, S.; Chu, Y.-M. Optimal Control of Time-Delay Fractional Equations via a Joint Application of Radial Basis Functions and Collocation Method. *Entropy* **2020**, *22*, 1213. [CrossRef] [PubMed]
28. Singh, J.; Kumar, D.; Baleanu, D. A new analysis of fractional fish farm model associated with Mittag-Leffler-type kernel. *Int. J. Biomath.* **2020**, *13*, 2050010. [CrossRef]
29. Morales-Delgado, V.F.; Gómez-Aguilar, J.F.; Saad, K.; Escobar Jiménez, R.F. Application of the Caputo-Fabrizio and Atangana-Baleanu fractional derivatives to mathematical model of cancer chemotherapy effect. *Math. Methods Appl. Sci.* **2019**, *42*, 1167–1193. [CrossRef]
30. Dokuyucu, M.A.; Celik, E.; Bulut, H.; Baskonus, H.M. Cancer treatment model with the Caputo-Fabrizio fractional derivative. *Eur. Phys. J. Plus* **2018**, *133*, 1–6. [CrossRef]
31. Losada, J.; Nieto, J.J. Properties of a new fractional derivative without singular kernel. *Progr. Fract. Differ. Appl.* **2015**, *1*, 87–92.



32. Caputo, M.; Fabrizio, M. A new definition of fractional derivative without singular kernel. *Progr. Fract. Differ. Appl.* **2015**, *1*, 73–85.
33. Alsaedi, A.; Nieto, J.J.; Venkatesh, V. Fractional electrical circuits. *Adv. Mech. Eng.* **2015**, *7*, 1687814015618127. [\[CrossRef\]](#)
34. Singh, J.; Kumar, D.; Baleanu, D. New aspects of fractional Biswas–Milovic model with Mittag-Leffler law. *Math. Model. Nat. Phenom.* **2019**, *14*, 303. [\[CrossRef\]](#)
35. Kumar, D.; Tchier, F.; Singh, J.; Baleanu, D. An efficient computational technique for fractal vehicular traffic flow. *Entropy* **2018**, *20*, 259. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Owolabi, K.M.; Atangana, A. Analysis and application of new fractional Adams–Bashforth scheme with Caputo–Fabrizio derivative. *Chaos Solitons Fractals* **2017**, *105*, 111–119. [\[CrossRef\]](#)
37. Kumar, D.; Singh, J.; Al Qurashi, M.; Baleanu, D. Analysis of logistic equation pertaining to a new fractional derivative with non-singular kernel. *Adv. Mech. Eng.* **2017**, *9*, 1687814017690069. [\[CrossRef\]](#)
38. Tateishi, A.A.; Ribeiro, H.V.; Lenzi, E.K. The role of fractional time-derivative operators on anomalous diffusion. *Front. Phys.* **2017**, *5*, 52. [\[CrossRef\]](#)
39. Atangana, A.; Alkahtani, B. Analysis of the Keller–Segel model with a fractional derivative without singular kernel. *Entropy* **2015**, *17*, 4439–4453. [\[CrossRef\]](#)
40. Diethelm, K. *The Analysis of Fractional Differential Equations: An Application-Oriented Exposition Using Differential Operators of Caputo Type*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
41. Baleanu, D.; Jajarmi, A.; Mohammadi, H.; Rezapour, S. A new study on the mathematical modelling of human liver with Caputo–Fabrizio fractional derivative. *Chaos Solitons Fractals* **2020**, *134*, 109705. [\[CrossRef\]](#)
42. Bushnaq, S.; Khan, S.A.; Shah, K.; Zaman, G. Mathematical analysis of HIV/AIDS infection model with Caputo–Fabrizio fractional derivative. *Cogent Math. Stat.* **2018**, *5*, 1432521. [\[CrossRef\]](#)
43. Ullah, S.; Khan, M.A.; Farooq, M.; Hammouch, Z.; Baleanu, D. A fractional model for the dynamics of tuberculosis infection using Caputo–Fabrizio derivative. *Discret. Contin. Dyn. Syst.-S* **2020**, *13*, 975–993. [\[CrossRef\]](#)
44. Hristov, J. Derivatives with non-singular kernels from the Caputo–Fabrizio definition and beyond: Appraising analysis with emphasis on diffusion models. *Front. Fract. Calc.* **2017**, *1*, 270–342.
45. Liu, P.; Liu, X. Dynamics of a tumor-immune model considering targeted chemotherapy. *Chaos Solitons Fractals* **2017**, *98*, 7–13. [\[CrossRef\]](#)
46. Moradi, H.; Sharifi, M.; Vossoughi, G. Adaptive robust control of cancer chemotherapy in the presence of parametric uncertainties: A comparison between three hypotheses. *Comput. Biol. Med.* **2015**, *56*, 145–157. [\[CrossRef\]](#)
47. De Pillis, L.G.; Radunskaya, A. The dynamics of an optimally controlled tumor model: A case study. *Math. Comput. Model.* **2003**, *37*, 1221–1244. [\[CrossRef\]](#)
48. De Pillis, L.G.; Radunskaya, A. A mathematical tumor model with immune resistance and drug therapy: An optimal control approach. *Comput. Math. Methods Med.* **2001**, *3*, 79–100. [\[CrossRef\]](#)
49. de Pillis, L.G.; Gu, W.; Fister, K.R.; Head, T.A.; Maples, K.; Murugan, A.; Neal, T.; Yoshida, K. Chemotherapy for tumors: An analysis of the dynamics and a study of quadratic and linear optimal controls. *Math. Biosci.* **2007**, *209*, 292–315. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Ghaffari, A.; Naserifar, N. Optimal therapeutic protocols in cancer immunotherapy. *Comput. Biol. Med.* **2010**, *40*, 261–270. [\[CrossRef\]](#)
51. Pang, L.; Zhao, Z.; Song, X. Cost-effectiveness analysis of optimal strategy for tumor treatment. *Chaos Solitons Fractals* **2016**, *87*, 293–301. [\[CrossRef\]](#)
52. Ledzewicz, U.; Schättler, H. Drug resistance in cancer chemotherapy as an optimal control problem. *Discret. Contin. Dyn. Syst. -B* **2006**, *6*, 129. [\[CrossRef\]](#)
53. Arciero, J.C.; Jackson, T.L.; Kirschner, D.E. A mathematical model of tumor-immune evasion and siRNA treatment. *Discret. Contin. Dyn. Syst.-B* **2004**, *4*, 39.
54. Letellier, C.; Sasmal, S.K.; Draghi, C.; Denis, F.; Ghosh, D. A chemotherapy combined with an anti-angiogenic drug applied to a cancer model including angiogenesis. *Chaos Solitons Fractals* **2017**, *99*, 297–311. [\[CrossRef\]](#)
55. Rokhforoz, P.; Jamshidi, A.A.; Sarvestani, N.N. Adaptive robust control of cancer chemotherapy with extended Kalman filter observer. *Inform. Med. Unlocked* **2017**, *8*, 1–7. [\[CrossRef\]](#)
56. Itik, M.; Salamci, M.U.; Banks, S.P. Optimal control of drug therapy in cancer treatment. *Nonlinear Anal. Theory Methods Appl.* **2009**, *71*, e1473–e1486. [\[CrossRef\]](#)
57. Padmanabhan, R.; Meskin, N.; Haddad, W.M. Reinforcement learning-based control of drug dosing for cancer chemotherapy treatment. *Math. Biosci.* **2017**, *293*, 11–20. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Yousefpour, A.; Haji Hosseinloo, A.; Reza Hairi Yazdi, M.; Bahrami, A. Disturbance observer-based terminal sliding mode control for effective performance of a nonlinear vibration energy harvester. *J. Intell. Mater. Syst. Struct.* **2020**, *31*, 1495–1510. [\[CrossRef\]](#)
59. Yousefpour, A.; Yasami, A.; Beigi, A.; Liu, J. On the development of an intelligent controller for neural networks: A type 2 fuzzy and chatter-free approach for variable-order fractional cases. *Eur. Phys. J. Spec. Top.* **2022**, *231*, 2045–2057. [\[CrossRef\]](#)
60. Yousefpour, A.; Jahanshahi, H. Fast disturbance-observer-based robust integral terminal sliding mode control of a hyperchaotic memristor oscillator. *Eur. Phys. J. Spec. Top.* **2019**, *228*, 2247–2268. [\[CrossRef\]](#)

61. Yousefpour, A.; Jahanshahi, H.; Bekiros, S.; Muñoz-Pacheco, J.M. Robust adaptive control of fractional-order memristive neural networks. In *Mem-Elements for Neuromorphic Circuits with Artificial Intelligence Applications*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 501–515.
62. Yousefpour, A.; Jahanshahi, H.; Gan, D. Fuzzy integral sliding mode technique for synchronization of memristive neural networks. In *Mem-Elements for Neuromorphic Circuits with Artificial Intelligence Applications*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 485–500.
63. Jahanshahi, H.; Sajjadi, S.S.; Bekiros, S.; Aly, A.A. On the development of variable-order fractional hyperchaotic economic system with a nonlinear model predictive controller. *Chaos Solitons Fractals* **2021**, *144*, 110698. [\[CrossRef\]](#)
64. Jahanshahi, H.; Yousefpour, A.; Wei, Z.; Alcaraz, R.; Bekiros, S. A financial hyperchaotic system with coexisting attractors: Dynamic investigation, entropy analysis, control and synchronization. *Chaos Solitons Fractals* **2019**, *126*, 66–77. [\[CrossRef\]](#)
65. Yao, Q.; Jahanshahi, H.; Bekiros, S.; Mihalache, S.F.; Alotaibi, N.D. Gain-Scheduled Sliding-Mode-Type Iterative Learning Control Design for Mechanical Systems. *Mathematics* **2022**, *10*, 3005. [\[CrossRef\]](#)
66. Yao, Q.; Jahanshahi, H.; Batrancea, L.M.; Alotaibi, N.D.; Rus, M.-I. Fixed-Time Output-Constrained Synchronization of Unknown Chaotic Financial Systems Using Neural Learning. *Mathematics* **2022**, *10*, 3682. [\[CrossRef\]](#)
67. Alsaadi, F.E.; Yasami, A.; Alsubaie, H.; Alotaibi, A.; Jahanshahi, H. Control of a Hydraulic Generator Regulating System Using Chebyshev-Neural-Network-Based Non-Singular Fast Terminal Sliding Mode Method. *Mathematics* **2023**, *11*, 168. [\[CrossRef\]](#)
68. Jahanshahi, H.; Yao, Q.; Khan, M.I.; Moroz, I. Unified neural output-constrained control for space manipulator using tan-type barrier Lyapunov function. *Adv. Space Res.* **2022**; *in press*.
69. Jahanshahi, H.; Zambrano-Serrano, E.; Bekiros, S.; Wei, Z.; Volos, C.; Castillo, O.; Aly, A.A. On the dynamical investigation and synchronization of variable-order fractional neural networks: The Hopfield-like neural network model. *Eur. Phys. J. Spec. Top.* **2022**, *231*, 1757–1769. [\[CrossRef\]](#)
70. Jahanshahi, H.; Yousefpour, A.; Soradi-Zeid, S.; Castillo, O. A review on design and implementation of type-2 fuzzy controllers. *Math. Methods Appl. Sci.* **2022**. [\[CrossRef\]](#)
71. Yao, Q.; Jahanshahi, H.; Bekiros, S.; Mihalache, S.F.; Alotaibi, N.D. Indirect neural-enhanced integral sliding mode control for finite-time fault-tolerant attitude tracking of spacecraft. *Mathematics* **2022**, *10*, 2467. [\[CrossRef\]](#)
72. Alsaade, F.W.; Yao, Q.; Bekiros, S.; Al-zahrani, M.S.; Alzahrani, A.S.; Jahanshahi, H. Chaotic attitude synchronization and anti-synchronization of master-slave satellites using a robust fixed-time adaptive controller. *Chaos Solitons Fractals* **2022**, *165*, 112883. [\[CrossRef\]](#)
73. Yao, Q.; Jahanshahi, H.; Moroz, I.; Bekiros, S.; Alsaafi, M.O. Indirect neural-based finite-time integral sliding mode control for trajectory tracking guidance of Mars entry vehicle. *Adv. Space Res.* **2022**; *in press*.
74. Chen, X. Research on application of artificial intelligence model in automobile machinery control system. *Int. J. Heavy Veh. Syst.* **2020**, *27*, 83–96. [\[CrossRef\]](#)
75. Das, P.; Chanda, S.; De, A. Artificial Intelligence-Based Economic Control of Micro-grids: A Review of Application of IoT. In *Computational Advancement in Communication Circuits and Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 145–155.
76. Hamet, P.; Tremblay, J. Artificial intelligence in medicine. *Metabolism* **2017**, *69*, S36–S40. [\[CrossRef\]](#)
77. Lim, C.W.; Zhang, G.; Reddy, J.N. A higher-order nonlocal elasticity and strain gradient theory and its applications in wave propagation. *J. Mech. Phys. Solids* **2015**, *78*, 298–313. [\[CrossRef\]](#)
78. Mao, Y.; Wang, J.; Jia, P.; Li, S.; Qiu, Z.; Zhang, L.; Han, Z. A reinforcement learning based dynamic walking control. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; pp. 3609–3614.
79. Qiao, J.; Hou, Z.; Ruan, X. Application of reinforcement learning based on neural network to dynamic obstacle avoidance. In Proceedings of the 2008 International Conference on Information and Automation, Changsha, China, 20–23 June 2008; pp. 784–788.
80. Wei, C.; Zhang, Z.; Qiao, W.; Qu, L. Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6360–6370. [\[CrossRef\]](#)
81. Sabatier, J.; Farges, C.; Merveillaut, M.; Feneteau, L. On observability and pseudo state estimation of fractional order systems. *Eur. J. Control* **2012**, *18*, 260–271. [\[CrossRef\]](#)
82. Wang, S.; He, S.; Yousefpour, A.; Jahanshahi, H.; Repnik, R.; Perc, M. Chaos and complexity in a fractional-order financial system with time delays. *Chaos Solitons Fractals* **2020**, *131*, 109521. [\[CrossRef\]](#)
83. Kilbas, A.A.; Srivastava, H.M.; Trujillo, J.J. *Theory and Applications of Fractional Differential Equations*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 204.
84. Moore, B.L.; Pyeatt, L.D.; Kulkarni, V.; Panousis, P.; Padrez, K.; Doufas, A.G. Reinforcement learning for closed-loop propofol anesthesia: A study in human volunteers. *J. Mach. Learn. Res.* **2014**, *15*, 655–696.
85. Martín-Guerrero, J.D.; Gomez, F.; Soria-Olivas, E.; Schmidhuber, J.; Clemente-Martí, M.; Jiménez-Torres, N.V. A reinforcement learning approach for individualizing erythropoietin dosages in hemodialysis patients. *Expert Syst. Appl.* **2009**, *36*, 9737–9742. [\[CrossRef\]](#)
86. Padmanabhan, R.; Meskin, N.; Haddad, W.M. Closed-loop control of anesthesia and mean arterial pressure using reinforcement learning. *Biomed. Signal Process. Control* **2015**, *22*, 54–64. [\[CrossRef\]](#)
87. Batmani, Y.; Khaloozadeh, H. Optimal chemotherapy in cancer treatment: State dependent Riccati equation control and extended Kalman filter. *Optim. Control Appl. Methods* **2013**, *34*, 562–577. [\[CrossRef\]](#)

88. Hunter, J.K.; Nachtergaele, B. *Applied Analysis*; World Scientific Publishing Company, Toh Tuck Link: Singapore, 2001.
89. Kreyszig, E. *Introductory Functional Analysis with Applications*; Wiley: New York, NY, USA, 1978; Volume 1.
90. Baird, L. Residual algorithms: Reinforcement learning with function approximation. In *Machine Learning Proceedings 1995*; Elsevier: Amsterdam, The Netherlands, 1995; pp. 30–37.
91. Gordon, G.J. Stable function approximation in dynamic programming. In *Machine Learning Proceedings 1995*; Elsevier: Amsterdam, The Netherlands, 1995; pp. 261–268.
92. Boyan, J.A.; Moore, A.W. Generalization in reinforcement learning: Safely approximating the value function. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1995; pp. 369–376.
93. Kuznetsov, V.A.; Makalkin, I.A.; Taylor, M.A.; Perelson, A.S. Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. *Bull. Math. Biol.* **1994**, *56*, 295–321. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.