



# Application of reinforcement learning for effective vaccination strategies of coronavirus disease 2019 (COVID-19)

Alireza Beigi<sup>1,a</sup>, Amin Yousefpour<sup>1,b</sup>, Amirreza Yasami<sup>1,c</sup>, J. F. Gómez-Aguilar<sup>2,d</sup> , Stelios Bekiros<sup>3,4,e</sup>, Hadi Jahanshahi<sup>5,f</sup>

<sup>1</sup> School of Mechanical Engineering, College of Engineering, University of Tehran, 14399–57131 Tehran, Iran

<sup>2</sup> CONACyT-Tecnológico Nacional de México/CENIDET, Interior Internado Palmira S/N, Col. Palmira, C.P. 62490 Cuernavaca, Morelos, Mexico

<sup>3</sup> Department of Banking and Finance, FEMA, , University of Malta, Msida MSD 2080, Malta

<sup>4</sup> Department of Economics, European University Institute, Via delle Fontanelle, 18, 50014 Florence, Italy

<sup>5</sup> Department of Mechanical Engineering, University of Manitoba, Winnipeg R3T 5V6, Canada

Received: 5 April 2021 / Accepted: 26 May 2021

© The Author(s), under exclusive licence to Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2021

**Abstract** Since December 2019, the new coronavirus has raged in China and subsequently all over the world. From the first days, researchers have tried to discover vaccines to combat the epidemic. Several vaccines are now available as a result of the contributions of those researchers. As a matter of fact, the available vaccines should be used in effective and efficient manners to put the pandemic to an end. Hence, a major problem now is how to efficiently distribute these available vaccines among various components of the population. Using mathematical modeling and reinforcement learning control approaches, the present article aims to address this issue. To this end, a deterministic Susceptible-Exposed-Infectious-Recovered-type model with additional vaccine components is proposed. The proposed mathematical model can be used to simulate the consequences of vaccination policies. Then, the suppression of the outbreak is taken to account. The main objective is to reduce the effects of Covid-19 and its domino effects which stem from its spreading and progression. Therefore, to reach optimal policies, reinforcement learning optimal control is implemented, and four different optimal strategies are extracted. Demonstrating the efficacy of the proposed methods, finally, numerical simulations are presented.

<sup>a</sup> e-mail: [alireza.beigi@ut.ac.ir](mailto:alireza.beigi@ut.ac.ir)

<sup>b</sup> e-mail: [amin.yousefpour@ut.ac.ir](mailto:amin.yousefpour@ut.ac.ir)

<sup>c</sup> e-mail: [yasamiamirreza@ut.ac.ir](mailto:yasamiamirreza@ut.ac.ir)

<sup>d</sup> e-mail: [jose.ga@cenidet.tecnm.mx](mailto:jose.ga@cenidet.tecnm.mx) (corresponding author)

<sup>e</sup> e-mail: [stelios.bekiros@um.edu.mt](mailto:stelios.bekiros@um.edu.mt)

<sup>f</sup> e-mail: [jahanshahi.hadi90@gmail.com](mailto:jahanshahi.hadi90@gmail.com)

## 1 Introduction

Since the first reported case of coronavirus disease 2019 (COVID-19) in early December 2019 in China, it has resulted in an ongoing crisis that unprecedentedly spreads all around the world [1–4]. Acute respiratory syndrome can occur in patients with serious illness, leading to multiple organ failures and death in some cases. [5, 6]. It has been established that the present pandemic's spread rate is much higher than similar previously reported epidemics in 2003 and 2012, namely SARS coronavirus (SARS-CoV) and MERS coronavirus (MERS-CoV). Until now, the epidemic crisis has resulted in a growing number of deaths all over the globe [7, 8].

Mathematical simulations have long been used to obtain insight into the mechanisms of disease transmission [9–22]. The essence of modeling lies in defining a set of equations that mimic the system's spread or dynamic in reality [23, 24]. From the beginning of the current epidemic, the mathematical models which show its spread have been at the forefront for prediction and control of the novel coronavirus outbreak [25–29]. Through the available data on the reported number of infections and information that we already know about the virus spread, as well as the confirmed number of deaths and hospitalizations, we can get an accurate insight for the future of the virus spread [30, 31].

Up to now, to effectively mitigate the spread of COVID-19, decision makers in all countries have applied various control policies such as mandatory lockdowns, quarantining and isolating infected people, maintaining a minimum social distancing, imposing strict and encouraging and strictly enforcing, avoiding crowded events, and forcing people to use face masks while in public [32–35]. Recently, several effective vaccines have been introduced for battling the pandemic. Some of them have passed all criteria, and now countries are using them. However, now with the advent of confirmed vaccines, governments and decision makers face new challenges. Now, to apply vaccines in effective ways, several questions have to be answered quickly and accurately. Which policies should be taken for vaccination? How can decision making choose different components of people? How can the vaccine be distributed throughout the time? How will the vaccine be able to decrease the risk of being infected? Since the disease's dynamic is complicated, and its spread is affected by several factors, answering these equations requires to be considered as optimization problems, which motivated the current study. The present study aims to solve these questions by proposing reinforcement learning-based optimal policies.

## 2 COVID-19 model with controls

In this study, an extended version of “Susceptible-Exposed-Infectious-Recovered” (SEIR) compartmental model is introduced. In this model, the spread of COVID-19 has been investigated. Using the Markov Chain Monte Carlo (MCMC) method and fitting the proposed model to the real data, the dynamic system's coefficients have been derived.

As mentioned in [36] and [27], the total population is considered as  $N$  which can be classified into eight different epidemiological subclasses: the humans who are not infected but susceptible  $S$ , exposed  $E$ , asymptomatic infected having no clinical symptoms but can infect healthy people  $A$ , infected people showing clinical symptoms  $I$ , the quarantined humans who are not infected but susceptible  $S_q$ , the quarantined humans who are exposed to the infection  $E_q$ , the hospitalized individuals  $H$  and the recovered individuals  $R$ . Under these assumptions, the model is given by defining  $q$  as a rate of quarantined,  $\beta$  as the probability of transmission per contact,  $\rho$  as the likelihood of having symptoms among infected people,

$\sigma$  as the proportion of individuals who move to the infected class,  $\lambda$  as the released rate of the quarantined uninfected contacts and  $c$  as the person-to-person contact rate. The disease-induced death rate of people is  $\alpha$ . In this work,  $\delta_I$  and  $\delta_q$  stand for the transition of infected people and exposed people to the quarantined infected class, respectively. The recovery rate of asymptotically infective patients is  $\gamma_A$  and the  $\gamma_H$  is the rate at which infected individuals get recovery, while  $\gamma_H$  is the rate at which hospitalized individuals get recovery. Based on these coefficients, the epidemic model that proposes the transmission dynamics is given by

$$\begin{aligned}
 \frac{d}{dt} S &= -(\beta c + cq(1 - \beta))S(I + \theta A) + \lambda S_q \\
 \frac{d}{dt} E &= -\beta c(1 - q)S(I + \theta A) - \sigma E \\
 \frac{d}{dt} I &= \sigma \rho E - (\delta_I + \alpha + \gamma_I)I \\
 \frac{d}{dt} A &= \sigma(1 - \rho)E - \gamma_A A \\
 \frac{d}{dt} S_q &= (1 - \beta)cqS(I + \theta A) - \lambda S_q \\
 \frac{d}{dt} E_q &= \beta cqS(I + \theta A) - \delta_q E_q \\
 \frac{d}{dt} H &= \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\
 \frac{d}{dt} R &= \gamma_I I + \gamma_A A + \gamma_H H
 \end{aligned} \tag{1}$$

where

$$c = (c_0 - c_b)e^{-r_1 t} + c_b \tag{2}$$

$$\frac{1}{\delta_I(t)} = \left( \frac{1}{\delta_{I0}} - \frac{1}{\delta_{If}} \right) e^{-r_2 t} + \frac{1}{\delta_{If}} \tag{3}$$

represents the person-to-person contact rate ( $c$ ) and detection rate  $\delta_I$ , respectively. Equation (2) and (3) include six parameters defined as follows:

- $c_0$ : initial contact rate
- $c_b$ : final contact rate that is larger than  $c_0$
- $r_1$ : exponentially decreasing rate of contact rate
- $\delta_{I0}$ : initial diagnosis rate
- $\delta_{If}$ : fastest diagnosis rate
- $r_2$ : exponentially increasing rate of diagnosis rate

It is assumed that the contact rate exponentially decreases over time and the diagnose rate exponentially increases with respect to time. Furthermore, we rewrite system (1) as follows:

$$\dot{x} = f(x)$$

$$f(x(t)) = \begin{bmatrix} -(\beta c + cq(1 - \beta))S(I + \theta A) + \lambda S_q \\ -\beta c(1 - q)S(I + \theta A) - \sigma E \\ \sigma \varrho E - (\delta_I + \alpha + \gamma_I)I \\ \sigma(1 - \varrho)E - \gamma_A A \\ (1 - \beta)cqS(I + \theta A) - \lambda S_q \\ \beta cqS(I + \theta A) - \delta_q E_q \\ \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\ \gamma_I I + \gamma_A A + \gamma_H H \end{bmatrix} \tag{4}$$

where  $x(t) = [S(t), E(t), I(t), A(t), S_q(t), E_q(t), H(t), R(t)]^T \in \mathbb{R}_{0+}^3$  are considered as the state vector. In this paper, this model has been selected because this model can describe the ongoing situation better. Firstly, this model has a higher reproduction rate over the other models [36, 37] that make this compartment model reasonable and superior. To be more specific, in this case, the reproduction rate was found too unstable [37], and some new variants of the novel coronavirus have a higher reproduction rate. Consequently, when we consider a model with a high reproduction rate and impose vaccination as a controlling variable, the optimal controller can be adopted for the worst-case scenario [38, 39]. Moreover, this model could estimate the confirmed case very well from 23 to 29 January 2020 because it considered different parameter variations, and its data collection was performed during intensive social events [39]. Therefore, this model can reflect the real situation better than others. In Sect. 4, we consider vaccination as a control input and discuss the system’s input signal and how to impose the vaccination to the nonlinear system.

### 3 Optimal control problem

Consider the system dynamics described by

$$\dot{x} = f(x) + g(x)u \tag{5}$$

with  $x \in \mathbb{R}^n$  denoting the state,  $f(x) \in \mathbb{R}^n$ ,  $g(x) \in \mathbb{R}^{n \times m}$  and the input  $u \in U \subset \mathbb{R}^m$ . Consider  $U$  as a set that is defined for the control input saturation.

**Assumption 1**  $f(\cdot)$  and  $g(\cdot)$  are differentiable in their argument with  $f(0) = 0$  and  $g(0) = 0$ , and they are Lipschitz continuous on their set, so  $f(x) + g(x)u$  is Lipschitz continuous on a set  $\Omega \subseteq \mathbb{R}^n$  containing the origin, so there exists a continuous control function  $u$  such that the dynamics (5) is asymptotically stable on  $\Omega$  and controllable.

**Assumption 2** The control matrix  $g(x)$  and  $f(x)$  are bounded over the compact set;  $\|g(x)\| \leq \Xi_g, \|g(x)\| \leq \Xi_f$

**Definition 1** In this paper, we define infinite horizon integral cost as follows:

$$V(x(t), u(t)) = \int_t^\infty r(x(\tau), u(\tau))d\tau \tag{6}$$

where  $r(x(\tau), u(\tau)) = Q(x(\tau)) + u(x(\tau))^T R u(x(\tau))$  and  $Q(x(\tau))$  is a positive definite monotonically increasing function.  $R$  is a symmetric positive definite matrix and  $Q(0) = 0$ .

**Definition 2 (Admissible Control Policy)** [40, 41]  $u$  is the control policy that can be said to be admissible with respect to the cost function (6) on  $\Omega$ , written as  $u$ , if  $u$  is continuous on a compact set  $\Omega \subset \mathbb{R}^n$  and differentiable on  $\Omega$ ,  $u(0) = 0$ ,  $u$  stabilizes (5) and for every  $x_0 \in \Omega$ , the  $V(x(0), u(0))$  is finite.

According to the differentiability and continuity of cost function, the infinitesimal version of (7) is the nonlinear Lyapunov equation

$$0 = Q(x(\tau)) + u(x(\tau))^T R u(x(\tau)) + (\nabla V)^T (f(x) + g(x)u) \tag{7}$$

with  $(0) = 0$ . In Eq. (6), the notation  $\nabla_x(\cdot)$  means the gradient operator with respect to  $x$  and is equivalent to  $\frac{\partial(\cdot)}{\partial x}$ . Consider the Hamiltonian of (5)

$$H(x, u, \nabla V) = Q(x(\tau)) + u(x(\tau))^T R u(x(\tau)) + (\nabla V)^T (f(x) + g(x)u) \tag{8}$$

The optimal performance index function of (5) can be formulated as

$$V^*(x(t), u(t)) = \int_t^\infty r(x(\tau), u^*(\tau)) d\tau = \min_{u \in U} \int_t^\infty r(x(\tau), u^*(\tau)) d\tau \tag{9}$$

According to the Bellman optimal control theory, the optimal value function  $V^*(x(t), u(t))$  can be obtained by solving the Hamilton–Jacobi–Bellman (HJB) equation:

$$0 = \min_{u \in U} H(x, u, \nabla V^*) \tag{10}$$

Assume that the minimum value on the right-hand side of Eq. (10) exists and is unique. By differentiating the HJB, the optimal control for the given problem can be expressed as

$$u^*(x) = \operatorname{argmin}_{u \in B} H(x, u, \nabla_x V^*(x)) = -\frac{1}{2} R^{-1} g^T(x) \nabla V^*(x) \tag{11}$$

where  $V^*(x(t), u(t))$  is formulated in the following HJB equation

$$\begin{aligned} V^*(x(t), u(t)) &= Q(x(\tau)) + (\nabla V^*)^T f(x) + \frac{1}{4} (\nabla V^*)^T g(x) R^{-1} g^T(x) (\nabla V^*) \\ V^*(0) &= 0 \end{aligned} \tag{12}$$

This nonlinear partial differential HJB equation is extremely difficult to solve and, in general, maybe it is impossible to be computed in some cases. Moreover, complete knowledge of the system’s dynamics is required. According to [42], the IRL algorithm is presented to estimate value function iteratively in the following section.

**Definition 3 (UUB Stability [43, 44])** For nonlinear system (5), with the equilibrium point  $x_e$  its solution is said to be UUB if there exists a compact set  $\Omega \subset \mathbb{R}^n$ , so that for every  $x_0 \in \Omega$ , there exists a positive bound  $\sqrt{\cdot}$  and a time  $T\left(\sqrt{\cdot}, x_0\right) > 0$ , independent of  $t_0$ , such that  $\|x(t) - \sqrt{\cdot}\| \leq x_0$  for  $\forall t \geq t_0 + T$ .

In this article, partially model-free integral reinforcement learning (IRL) has been introduced to obtain the optimal value function approximation  $V^*(x)$  and a continuous optimal control policy  $u^*(x)$ .

### 3.1 Value function approximation using Critic network

The critic control design with neural networks generally is acceptable to determine the optimal approximation for control problems [45, 46]. With the higher-order Weierstrass approximation theorem [47], a single-layer neural network can be utilized to reconstruct the cost function  $V^*(x)$ .

$$V_c^*(x) = w_c^{*T} \phi_c(x) + \varepsilon_c(x) \tag{13}$$

where  $w_c^* \in \mathbb{R}^l$  is suitable coefficients with  $l$  neurons,  $\phi_c(x) \in \mathbb{R}^l$  provides the activation function of the neural network(NN), and  $\varepsilon_c(x) \in \mathbb{R}$  is the reconstruction error. Assuming  $w_c^*$  and  $\phi_c(x)$  are bounded.  $\|w_c^*\| \leq \Xi_{w_c^*}$ ,  $\|\phi_c\| \leq \Xi_{\phi_c}$ . Since  $V_c^*(x)$  is differentiable, its gradient can be approximated as

$$\frac{\partial V_c^*(x)}{\partial x} = \left( \frac{\partial \phi_c(x)}{\partial x} \right)^T w_c^* + \frac{\partial \varepsilon_c}{\partial x} = \nabla \phi_c(x) w_c^* + \nabla \varepsilon_c(x) \tag{14}$$

According to the [48], for  $x \in \Omega$  one can infer that  $\varepsilon_c(x)$  and its gradient  $\nabla \varepsilon_c(x)$  are bounded  $\|\nabla^T \varepsilon_c(x)\| \leq \Xi_{\varepsilon_1}$ . According to the fact that  $\phi_c \leq \Xi_{\phi_c}$ , one can infer  $\nabla \phi_c \leq \Xi_{\nabla \phi_c}$ . While generally, the optimal coefficient  $w_c^*$  is unknown, the estimated value function is given by

$$\widehat{V}_c(x) = \widehat{w}_c^T \phi_c(x) \tag{15}$$

where  $\widehat{w}_c$  denotes estimated weights of these basis functions that are updated through the learning process. The updating rule will be formulated in the following section.

### 3.2 Policy approximation using Actor-network

Zhu et al. [49], have determined policy estimation by considering the fact that if the initial admissible policy is given, the policy function can be expressed by NN. NN approximation is a well-known method for policy estimation in optimal control [50–52]. Therefore, similar to value function, in order to Weirstrass high-order approximation theorem, the smooth policy can be uniformly approximated over a compact set as

$$u_c^*(x) = \Gamma\left(w_a^{*T} \phi_a(x)\right) + \varepsilon_a(x) \tag{16}$$

where  $w_a^* \in \mathbb{R}^{l' \times m}$  is optimal coefficients with  $l'$  neurons,  $\phi_c(x) \in \mathbb{R}^{l'}$  provides the activation function of the neural network, and  $\varepsilon_a(x) \in \mathbb{R}$  is the approximation error that is bounded  $\|\varepsilon_a(x)\| \leq b_a$ .  $\Gamma(\cdot)$  is a continuous activation function.

Assumption 3  $\Gamma : \mathbb{R}^m \cup \{\pm\infty\} \rightarrow U$  is a function that is continuous monotonic bijective. The first derivative of this function is bounded  $\Gamma'(\cdot) = \frac{d\Gamma(\cdot)}{d(\cdot)}$  and  $\Gamma(0) = 0$ .

**Remark 1** *tanH, SQNL* [53], and *softsign* [54] activation functions satisfy Assumption 3. In this case, because the input should be bounded by a constant  $\|u\| \leq \Xi_u$ , *softsign* is employed. Then, the estimated policy function is given by.

$$\widehat{u}_c(x) = \Gamma\left(\widehat{w}_a^T \phi_a(x)\right) \tag{17}$$

where  $\widehat{w}_a$  denotes estimated weights to learn  $w_a^*$ .

### 3.3 Learning rules for actor and critic networks

*Updating Rule for the Critic Network:* By substituting Eqs. (14) in (8), we have

$$H(x, u, w_c^*) = Q(x(\tau)) + u(x(\tau))^T R u(x(\tau)) + (w_c^*)^T \nabla \phi_c(x)(f(x) + g(x)u) = \varepsilon_H \tag{18}$$

, while based on Eq. (7),  $\varepsilon_H$  can be given by

$$\varepsilon_H = -\nabla^T \varepsilon_c(x)(f(x) + g(x)u) \tag{19}$$

**Assumption 4** Under the Lemma 1 that is mentioned in [55] and by using Assumption 7 in [56], the least-squares solution to (17) exists and is unique for any admissible control policy and the number of hidden layer neurons  $N \rightarrow \infty$ ,  $\phi_c(x)$  gives the complete independent basis for  $V_c^*(x)$ .

Hence,  $\frac{\partial V_c^*(x)}{\partial x}$ ,  $V_c^*(x)$  can be estimated by NNs in view of the following assumption and the Weierstrass higher-order approximation theorem, so as  $N \rightarrow \infty$ ,  $\frac{\partial \varepsilon_c(x)}{\partial x}$ ,  $\varepsilon_c(x)$  approach zero [40]. Motivated by the research in [57], so as to find the updating laws for the critic weights, we define the error function for critic network as

$$e_c = \sigma^T \widehat{w}_c + Q(x) + \widehat{u}_c^T R \widehat{u}_c \tag{20}$$

where  $\sigma = \nabla \phi_c(x)(f(x) + g(x)\widehat{u}_c)$ . To train the critic networks, the squared residual error regarding the critic network training should be minimized.

$$E_c = \frac{1}{2} e_c^2 \tag{21}$$

The weight of the critic network  $\widehat{w}_c$  is updated in a gradient descent algorithm to minimize  $E_c$  with

$$\dot{\widehat{w}}_c = -\alpha_c \frac{\sigma}{(\sigma^T \sigma + 1)^2} \left[ \sigma^T \widehat{w}_c + Q(x) + \widehat{u}_c^T R \widehat{u}_c \right] \tag{22}$$

where  $\alpha_c > 0$  is the critic learning rate. Hence, in the light of (22) and (11) and inspired by [42], the nearly optimal policy can be obtained by

$$u_c(x) = -\frac{1}{2} R^{-1} g^T(x) \nabla \widehat{V}_c(x) = -\frac{1}{2} R^{-1} g^T(x) \widehat{w}_c^T \nabla \phi_c(x) \tag{23}$$

where we used  $\widehat{V}_c(x)$  instead of the optimal value function  $V_c^*(x)$ . This nearly optimal policy is used in the following section to update Actor-Network.

*Updating Rule for the Actor-Network:* Similar to the training of critic network, one may define

$$E_a = \frac{1}{2} e_a^2 \tag{24}$$

where  $e_a$  is the output error between the policy function  $\widehat{u}_c(x)$  and the nearly targeted control policy  $u_c(x)$

$$e_a = \widehat{u}_c(x) - u_c(x) = \Gamma \left( \widehat{w}_a^T \phi_a(x) \right) - u_c(x) \tag{25}$$

To minimize the square actor error  $E_a$ , the weights of the actor-network are tuned by gradient descent rule as follows:

$$\dot{\widehat{w}}_a = -\alpha_a \frac{\partial E_a}{\partial \widehat{w}_a} = -\alpha_a \phi_a(x) \Gamma' \left( \widehat{w}_a^T \phi_a(x) \right) e_a \tag{26}$$

**Theorem 1** Consider the system given by (5) and the updating actor law and critic law given by (22) and (26), respectively. Assume  $\widehat{w}_a = w_a^* - \widetilde{w}_a$ ,  $\widehat{w}_c = w_c^* - \widetilde{w}_c$  and let  $\frac{\sigma}{(\sigma^T \sigma + 1)}$  be persistently exciting (PE) [44]. If Assumptions 1–4 hold, the actor-critic weight estimation errors  $\widetilde{w}_a$  and  $\widetilde{w}_c$  are uniformly ultimately bounded and  $\widehat{w}_a$  and  $\widehat{w}_c$  converge to a residual set in the neighbor of  $w_a^*$  and  $w_c^*$ , respectively.

**Proof** For convenience, we define

$$u_a(x) = \Gamma(w_a^{*T} \phi_a(x)) \tag{27}$$

$$\widehat{D} = \widehat{w}_a^T \phi_a(x) \tag{28}$$

$$D^* = w_a^{*T} \phi_a(x) \tag{29}$$

$$\widetilde{D} = D^* - \widehat{D} \tag{30}$$

The convergence of the actor-critic network during learning is based on Lyapunov analysis. We consider the following Lyapunov candidate

$$L = L_c + L_V + L_a = \frac{1}{2} \widetilde{w}_c^T \alpha_c^{-1} \widetilde{w}_c + V^*(x) + \frac{1}{2} \widetilde{w}_a^T \alpha_a^{-1} \widetilde{w}_a \tag{31}$$

First, let us consider  $L_c = \frac{1}{2} \widetilde{w}_c^T \alpha_c^{-1} \widetilde{w}_c$ . Then,  $\dot{L}_c = -\widetilde{w}_c^T \alpha_c^{-1} \dot{\widehat{w}}_c$  where  $\dot{\widehat{w}}_c$  is given in (22).

$$\dot{L}_c = -\widetilde{w}_c^T \alpha_c^{-1} \dot{\widehat{w}}_c = \widetilde{w}_c^T \frac{\sigma}{(\sigma^T \sigma + 1)^2} \left[ \sigma^T (w_c^* - \widetilde{w}_c) + Q(x) + \widehat{u}_c^T R \widehat{u}_c \right] \tag{32}$$

or equivalently

$$\dot{L}_c = -\frac{\sigma \sigma^T}{(\sigma^T \sigma + 1)^2} \widetilde{w}_c^T \widetilde{w}_c + \frac{\sigma \sigma^T}{(\sigma^T \sigma + 1)^2} \left[ \widetilde{w}_c^T \left( Q(x) + \widehat{u}_c^T R \widehat{u}_c \right) + \widetilde{w}_c^T \sigma^T (w_c^*) \right] \tag{33}$$

The solution of the HJB by considering (12) can be rewritten as

$$\varepsilon_{\text{HJB}} = Q(x(t)) + w_c^{*T} \nabla \phi_c(x) f(x) + \frac{1}{4} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla^T \phi_c(x) w_c^* \tag{34}$$

where  $\varepsilon_{\text{HJB}}$  is the residual error due to the function approximation error, which is

$$\begin{aligned} \varepsilon_{\text{HJB}} = & -\nabla^T \varepsilon_c(x) f(x) + \frac{1}{2} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla \varepsilon_c(x) \\ & + \frac{1}{4} \nabla^T \varepsilon_c(x) g(x) R^{-1} g^T(x) \nabla \varepsilon_c(x) \end{aligned} \tag{35}$$

It is now desired to show that this error converges uniformly to zero as the number of hidden layer units  $N$  increases [40]. Hence, it can be demonstrated that  $\varepsilon_{\text{HJB}}$  is bounded.  $\|\varepsilon_{\text{HJB}}\| \leq \Xi_1$

By considering (34), one can derive

$$\begin{aligned} \dot{L}_c = & -\frac{\sigma \sigma^T}{(\sigma^T \sigma + 1)^2} \widetilde{w}_c^T \widetilde{w}_c \\ & + \frac{\sigma \sigma^T}{(\sigma^T \sigma + 1)^2} \left[ \widetilde{w}_c^T \left( \varepsilon_{\text{HJB}} - w_c^{*T} \nabla \phi_c(x) f(x) \right. \right. \\ & \left. \left. - \frac{1}{4} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla^T \phi_c(x) w_c^* + \widehat{u}_c^T R \widehat{u}_c \right) + \widetilde{w}_c^T \sigma^T (w_c^*) \right] \end{aligned} \tag{36}$$



Based on (25),  $\sigma = \nabla\phi_c(x)(f(x) + g(x)\hat{u}_c)$  and making use of  $\bar{\sigma} = \frac{\sigma\sigma^T}{(\sigma^T\sigma+1)^2}$

$$\begin{aligned} \dot{L}_c &= -\bar{\sigma} \tilde{w}_c^T \tilde{w}_c \\ &+ \bar{\sigma} \left[ \tilde{w}_c^T \left( \varepsilon_{HJB} - w_c^{*T} \nabla\phi_c(x) f(x) - \frac{1}{4} w_c^{*T} \nabla\phi_c(x) g(x) R^{-1} g^T(x) \nabla^T\phi_c(x) w_c^* \right. \right. \\ &\quad \left. \left. + \hat{u}_c^T R \hat{u}_c \right) + \tilde{w}_c^T w_c^{*T} \nabla\phi_c(x) (f(x) + g(x)\hat{u}_c) \right] \end{aligned} \tag{37}$$

where  $\bar{\sigma}$  is bounded.  $\|\bar{\sigma}\| \leq \Xi_{\bar{\sigma}}$  Then,

$$\begin{aligned} \dot{L}_c &= -\bar{\sigma} \tilde{w}_c^T \tilde{w}_c + \bar{\sigma} \left[ \tilde{w}_c^T \left( \varepsilon_{HJB} - \frac{1}{4} w_c^{*T} \nabla\phi_c(x) g(x) R^{-1} g^T(x) \nabla^T\phi_c(x) w_c^* + \hat{u}_c^T R \hat{u}_c \right) \right. \\ &\quad \left. + \tilde{w}_c^T w_c^{*T} \nabla\phi_c(x) (g(x)\hat{u}_c) \right] \end{aligned} \tag{38}$$

The second term is,

$$\begin{aligned} \dot{L}_V &= \left( w_c^{*T} \nabla\phi_c(x) + \nabla^T \varepsilon_c(x) \right) (f(x) + g(x)\hat{u}_c) \\ &= \left( w_c^{*T} \nabla\phi_c(x) + \nabla^T \varepsilon_c(x) \right) \left( f(x) + g(x)\Gamma(\hat{w}_a^T \phi_a(x)) \right) \end{aligned} \tag{39}$$

Combining with (34),

$$\begin{aligned} \dot{L}_V &= \nabla^T \varepsilon_c(x) \left( f(x) + g(x)\Gamma(\hat{w}_a^T \phi_a(x)) \right) \\ &+ w_c^{*T} \nabla\phi_c(x) f(x) + w_c^{*T} \nabla\phi_c(x) g(x)\Gamma(\hat{w}_a^T \phi_a(x)) \\ &= \nabla^T \varepsilon_c(x) \left( f(x) + g(x)\Gamma(\hat{w}_a^T \phi_a(x)) \right) \\ &+ \varepsilon_{HJB} - Q(x(t)) - \frac{1}{4} w_c^{*T} \nabla\phi_c(x) g(x) R^{-1} g^T(x) \nabla^T\phi_c(x) w_c^* \end{aligned} \tag{40}$$

For the second term of (31), one can write

$$\begin{aligned} \dot{L}_a &= \tilde{w}_a^T \alpha_a^{-1} \dot{\tilde{w}}_a = -\tilde{w}_a^T \alpha_a^{-1} \dot{\tilde{w}}_a = \tilde{w}_a^T \left( \phi_a(x)\Gamma'(\hat{w}_a^T \phi_a(x)) e_a \right) \\ &= \tilde{w}_a^T \left( \phi_a(x)\Gamma'(\hat{w}_a^T \phi_a(x)) \left( \Gamma(\hat{w}_a^T \phi_a(x)) + \frac{1}{2} R^{-1} g^T(x) \hat{w}_c^T \nabla\phi_c(x) \right) \right) \end{aligned} \tag{41}$$

Consequently,

$$\begin{aligned} \dot{L}_a &= \tilde{w}_a^T \phi_a(x)\Gamma'(\hat{w}_a^T \phi_a(x))\Gamma(\hat{w}_a^T \phi_a(x)) \\ &+ \tilde{w}_a^T \phi_a(x)\Gamma'(\hat{w}_a^T \phi_a(x)) \frac{1}{2} R^{-1} g^T(x) \hat{w}_c^T \nabla\phi_c(x) \end{aligned} \tag{42}$$

Based on (28)–(30), (41) becomes

$$\begin{aligned} \dot{L}_a &= \tilde{D}^T \Gamma'(\hat{D})\Gamma(\hat{D}) + \frac{1}{2} \tilde{D}^T \Gamma'(\hat{D}) R^{-1} g^T(x) \hat{w}_c^T \nabla\phi_c(x) \\ &= \tilde{D}^T \Gamma'(\hat{D})\Gamma(\hat{D}) - \tilde{D}^T \Gamma'(\hat{D}) u_c \\ &= \tilde{D}^T \Gamma'(\hat{D})\Gamma(\hat{D}) - \tilde{D}^T \Gamma'(\hat{D}) (\Gamma(\hat{D}) - e_a) = \tilde{D}^T \Gamma'(\hat{D}) e_a \end{aligned} \tag{43}$$

Then,  $u_c^*(x)$  be can be written

$$u_c^*(x) = \Gamma(D^*) + \varepsilon_a(x) = -\frac{1}{2}R^{-1}g^T(x)w_c^{*T}\nabla\phi_c(x) \tag{44}$$

Based on (25),  $e_a$  become

$$\begin{aligned} e_a &= \hat{u}_c(x) - u_c(x) \\ &= \Gamma(\hat{D}) + \frac{1}{2}R^{-1}g^T(x)(w_c^{*T} - \tilde{w}_c^T)\nabla\phi_c(x) \\ &= \Gamma(\hat{D}) - \Gamma(D^*) - \varepsilon_a(x) - \frac{1}{2}R^{-1}g^T(x)\tilde{w}_c^T\nabla\phi_c(x) \\ &= -\Gamma(\tilde{D}) - \varepsilon_a(x) - \frac{1}{2}R^{-1}g^T(x)\tilde{w}_c^T\nabla\phi_c(x) \end{aligned} \tag{45}$$

Then, using (45) in (42)

$$\begin{aligned} \dot{L}_a &= -\tilde{D}^T\Gamma'(\hat{D})\left(\Gamma(\tilde{D}) + \varepsilon_a(x) + \frac{1}{2}R^{-1}g^T(x)\tilde{w}_c^T\nabla\phi_c(x)\right) \\ &= -\tilde{D}^T\Gamma'(\hat{D})\varepsilon_a(x) - \tilde{D}^T\Gamma'(\hat{D})\left(\Gamma(\tilde{D})\right) - \tilde{D}^T\Gamma'(\hat{D})\frac{1}{2}R^{-1}g^T(x)\tilde{w}_c^T\nabla\phi_c(x) \end{aligned} \tag{46}$$

By applying the Young inequality to the second term, we have

$$\begin{aligned} &-\tilde{D}^T\Gamma'(\hat{D})\frac{1}{2}R^{-1}g^T(x)\nabla\phi_c(x)\tilde{w}_c \\ &\leq \frac{1}{4}\tilde{w}_c\nabla^T\phi_c(x)g(x)R^{-1}g^T(x)\nabla\phi_c(x)\tilde{w}_c \\ &\quad + \tilde{D}^T\Gamma'(\hat{D})\Gamma'(\hat{D})^T\tilde{D} \end{aligned} \tag{47}$$

Then, we rewrite (46)

$$\begin{aligned} \dot{L}_a &\leq -\tilde{D}^T\Gamma'(\hat{D})\left(-\Gamma'(\hat{D})^T\tilde{D} + \varepsilon_a(x) + \left(\Gamma(\tilde{D})\right)\right) \\ &\quad + \frac{1}{4}\tilde{w}_c\nabla^T\phi_c(x)g(x)R^{-1}g^T(x)\nabla\phi_c(x)\tilde{w}_c \end{aligned} \tag{48}$$

By considering softsign as the activation function,  $\Gamma(y) = \frac{y}{(1+|y|)}$  therefore  $\Gamma'(y) = \frac{1}{(1+|y|)^2}$ . With these functions, one can then show that the first term of (48) can be given

$$\begin{aligned} &-\tilde{D}^T\Gamma'(\hat{D})\left(-\Gamma'(\hat{D})^T\tilde{D} + \varepsilon_a(x) + \left(\Gamma(\tilde{D})\right)\right) \\ &= -\tilde{D}^T\Gamma'(\hat{D})\left(\frac{-\tilde{D}}{(1+\|\hat{D}\|)^2} + \varepsilon_a(x) - \frac{\hat{D}}{(1+\|\hat{D}\|)} + \frac{D^*}{(1+\|D^*\|)}\right) \\ &= -\tilde{D}^T\Gamma'(\hat{D})\left(\frac{-(D^* - \hat{D}) - \hat{D}(1+\|\hat{D}\|) + D^*(1+\|\hat{D}\|)}{(1+\|\hat{D}\|)^2}\right) \\ &= -\tilde{D}^T\Gamma'(\hat{D})\frac{\|\hat{D}\|\tilde{D}}{(1+\|\hat{D}\|)^2} \end{aligned} \tag{49}$$

Substituting (49) in (48)

$$\dot{L}_a \leq -\tilde{D}^T \Gamma'(\tilde{D}) \frac{\|\tilde{D}\| \tilde{D}}{(1 + \|\tilde{D}\|)^2} + \frac{1}{4} \tilde{w}_c \nabla^T \phi_c(x) g(x) R^{-1} g^T(x) \nabla \phi_c(x) \tilde{w}_c \tag{50}$$

Next, using (38), (40), and (50) to rewrite (30), we get

$$\begin{aligned} \dot{L} = & -Q(x(t)) - \bar{\sigma} \tilde{w}_c^T \tilde{w}_c - \tilde{D}^T \Gamma'(\tilde{D}) \frac{\|\tilde{D}\| \tilde{D}}{(1 + \|\tilde{D}\|)^2} \\ & - \frac{1}{4} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla^T \phi_c(x) w_c^* - \tilde{D}^T \Gamma'(\tilde{D}) \left( \Gamma(\tilde{D}) \right) \\ & + \nabla^T \varepsilon_c(x) \left( f(x) + g(x) \Gamma(\hat{w}_a^T \phi_a(x)) \right) \\ & + \varepsilon_{HJB} + \bar{\sigma} \left[ \tilde{w}_c^T \left( \varepsilon_{HJB} - \frac{1}{4} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla^T \phi_c(x) w_c^* + \hat{u}_c^T R \hat{u}_c \right) \right. \\ & \left. + \tilde{w}_c^T w_c^{*T} \nabla \phi_c(x) (g(x) \hat{u}_c) \right] + \frac{1}{4} \varepsilon_K \tilde{w}_c \nabla^T \phi_c(x) g(x) R^{-1} g^T(x) \nabla \phi_c(x) \tilde{w}_c \tag{51} \end{aligned}$$

Let  $\Upsilon_0 = \frac{\Gamma'(\tilde{D}) \|\tilde{D}\|}{(1 + \|\tilde{D}\|)^2}$ ,  $\Upsilon_1 = \frac{1}{4} \varepsilon_K \nabla^T \phi_c(x) g(x) R^{-1} g^T(x) \nabla \phi_c(x)$  and  $\Upsilon_2 = \frac{\bar{\sigma}}{4} w_c^{*T} \nabla \phi_c(x) g(x) R^{-1} g^T(x) \nabla^T \phi_c(x) w_c^*$  positive definites. Based on definitions of  $\hat{u}_c$ ,  $\varepsilon_{HJB}$ ,  $\nabla^T \varepsilon_c$ ,  $\Gamma(\hat{w}_a^T \phi_a(x))$ ,  $\bar{\sigma}$ ,  $w_c^*$ ,  $\nabla \phi_c(x)$  and Assumption 2 and 3, one can conclude

$$\begin{aligned} \|\varepsilon_{HJB}\| & \leq \Xi_1 \\ \|\nabla^T \varepsilon_c(x) \left( f(x) + g(x) \Gamma(\hat{w}_a^T \phi_a(x)) \right)\| & \leq \Xi_{\varepsilon_1} (\Xi_f + \Xi_g \Xi_u) = \Xi_2 \\ \|\bar{\sigma} \hat{u}_c^T R \hat{u}_c\| & \leq \Xi_{\bar{\sigma}} \Xi_u^2 R = \Xi_3 \\ \|\bar{\sigma} w_c^{*T} \nabla \phi_c(x) (g(x) \hat{u}_c)\| & \leq \Xi_{\bar{\sigma}} \Xi_{w_c^*} \Xi_{\nabla \phi_c} \Xi_g \Xi_u = \Xi_4 \tag{52} \end{aligned}$$

where  $\Xi_u$ ,  $\Xi_1$ ,  $\Xi_{\varepsilon_1}$ ,  $\Xi_{\bar{\sigma}}$ ,  $\Xi_{w_c^*}$ ,  $\Xi_{\nabla \phi_c}$ ,  $\Xi_f$  and  $\Xi_g$  are the upper bound of  $\hat{u}_c$ ,  $\varepsilon_{HJB}$ ,  $\nabla^T \varepsilon_c$ ,  $\Gamma(\hat{w}_a^T \phi_a(x))$ ,  $\bar{\sigma}$ ,  $w_c^*$ ,  $\nabla \phi_c(x)$ ,  $f(x)$  and  $g(x)$ , respectively. Consequently, we can obtain

$$\dot{L} \leq -\lambda_Q x^2 - \tilde{w}_c^T (\bar{\sigma} - \Upsilon_1) \tilde{w}_c - \tilde{D}^T \Upsilon_0 \tilde{D} - \tilde{w}_c^T (\Upsilon_2 - \Xi_3 - \Xi_4) + \Xi_1 + \Xi_2 \tag{53}$$

where  $\lambda_Q$  is a positive constant such that  $Q(x(t)) > x^T \lambda_Q x$  for every  $x \in \Omega$ . If we choose  $\varepsilon_K$  and  $R$  such that  $(\bar{\sigma} - \varepsilon_1)$  bigger than zero, then  $\dot{L}$  yields

$$\dot{L} \leq -\|\bar{Z}\| \left\| \lambda_{\min}(\mathcal{G}) + \|q\| \|\bar{Z}\| \right\| + \Sigma \tag{54}$$

where  $\bar{Z} = \begin{bmatrix} x \\ \tilde{D} \\ \tilde{w}_c \end{bmatrix}$  and

$$\begin{aligned} \mathcal{G} & = \begin{bmatrix} Q & 0 & 0 \\ 0 & \Upsilon_0 & 0 \\ 0 & 0 & (\bar{\sigma} - \Upsilon_1) \end{bmatrix} \\ \mathcal{G} & = \begin{bmatrix} 0 \\ 0 \\ -(\Upsilon_2 - \Xi_3 - \Xi_4) \end{bmatrix} \\ \Sigma & = \Xi_1 + \Xi_2 \end{aligned}$$

If the parameters are selected such that  $\Pi$  and  $\mathcal{G}$  are positive, then the Lyapunov derivative is negative if

$$\left\| \frac{\dot{Z}}{Z} \right\| \frac{\|q\| + \sqrt{\|q\|^2 + 4\lambda_{\min}(\mathcal{G})\Sigma}}{2\lambda_{\min}(\mathcal{G})} \quad (55)$$

By mathematical induction, it is now desired to show if we can bound the Lyapunov function, then for sufficiently large  $\left\| \frac{\dot{Z}}{Z} \right\|$   $\dot{L}$  is negative. Therefore, by using the standard Lyapunov extension theorem [58], it follows that the system state and the weights error are UUB, which completes the proof.

**Remark 2** If Assumptions 1–4 hold, then the assumption of the nonlinear Lyapunov equation solution of (12) can be relaxed without loss of the stability performance, and the equilibrium point of (5) remains UUB.

Remark 3: By utilizing gradient descent rules (22), (26), and the backpropagation rule, the HJB error  $E_c = \frac{1}{2}e_c^2$  and the point-wise control error  $E_a = \frac{1}{2}e_a^2$  update the critic and actor networks. The error  $E_c$  is convex with respect to  $\hat{w}_c$ . Hence, the critic network's weight converges to its global optimal point by applying the updating rule (22). The error  $E_a$  is non-convex with respect to  $\hat{w}_a$  so the weights of the actor-network converge to a locally optimal point by applying the updating rule (26).

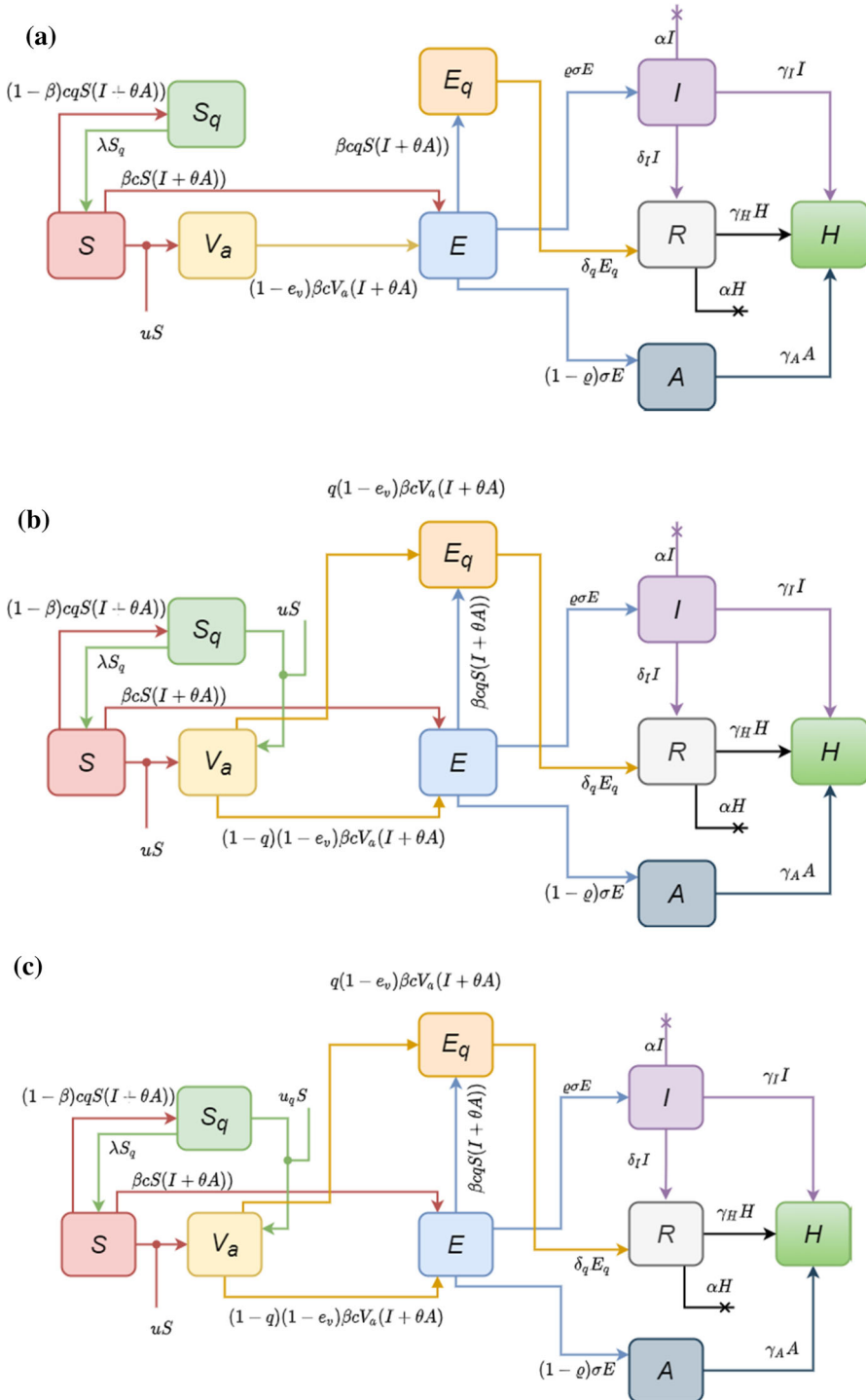
## 4 Optimal vaccination strategies

In this study, several optimal vaccination strategies are proposed. In each strategy, we consider four cost functions to formulate the main concerns and characterizations that should be considered and noted. Then, we define the appropriate model for each of them. It should be noted that the main objective is to reduce the effects of Covid-19 and its domino effects which stem from its spreading and progression. In this section, the optimal control principle provides an optimized approach that can hinder the impacts of Covid-19 outbreak. Therefore, to minimize defined objective functions, we implement reinforcement learning optimal control to reach the optimal policy. Here, by using the HJB equation which is described in each subsection, the necessary condition for optimality is satisfied.

### 4.1 Strategy 1

In view of the above discussion, we suppose the optimal control problem with vaccination as input control. Motivated from proposed model in [59] and [60], we introduce a mathematical model with end-point state constraint, control input inspired of [37], and considering vaccine efficiency. This model is shown in Fig. 1 and can be driven as follows:

$$\dot{x} = f(x) + g(x)u$$



**Fig. 1** Transmission diagram of dynamics of COVID-19 spread, by the implementation of **a** vaccination by strategy 1 or 2, **b** strategy 3, **c** strategy 4

$$\begin{aligned}
 f(x(t)) &= \begin{bmatrix} -(\beta c + cq(1 - \beta))S(I + \theta A) + \lambda S_q \\ \beta c(1 - q)S(I + \theta A) - \sigma E + (1 - e_v)\beta c V_a(I + \theta A) \\ \sigma E - (\delta_I + \alpha + \gamma_I)I \\ \sigma(1 - \gamma)E - \gamma_A A \\ (1 - \beta)cqS(I + \theta A) - \lambda S_q \\ \beta cqS(I + \theta A) - \delta_q E_q \\ \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\ \gamma_I I + \gamma_A A + \gamma_H H \\ -(1 - e_v)\beta c V_a(I + \theta A) \end{bmatrix} \\
 g(x) &= [-S, 0, 0, 0, 0, 0, 0, 0, 0, S]^T \\
 x(t) &= [S(t), E(t), I(t), A(t), S_q(t), E_q(t), H(t), R(t), V_a(t)]^T \tag{56}
 \end{aligned}$$

where  $u$  denotes the ratio of susceptible individuals vaccinated per day ( $0 \leq u \leq 1$ ) and  $e_v$  is the vaccine efficiency. This parameter demonstrates the effectiveness of vaccine which means that if this factor  $e_v = 1$ , the vaccine is fully effective. Note that the term  $(1 - e_v)\beta c V_a(I + \theta A)$  represents the vaccinated people who can be infected due to vaccine incompleteness and expresses the fact that no vaccination is 100% effective. Moreover, in the introduced model,  $V_a$  indicates the vaccinated sub-population. In this strategy, we wish to reduce the objective function which considers the infected individuals and the ratio of vaccinated people

$$V^* = \min \left( \int_0^{t_f} \Gamma_{I,s_1} I^2(t) + \Gamma_{u,s_1} u^2 dt \right) \tag{57}$$

where  $\Gamma_{I,s_1}$  and  $\Gamma_{u,s_1}$  are the relative weight factors selected to balance the objective function over innervation time  $t_f$ .

### 4.2 Strategy 2

The objective function (55) can be enhanced by assuming exposed individuals as a population that should be considered to be minimized. In fact, by employing a strategy whose aim is to reduce exposed people, we can find a better solution to minimize the number of infected individuals and the cost of vaccination. More precisely, we seek that the optimal control consists of minimizing the objective functional

$$V^* = \min \left( \int_0^{t_f} \Gamma_{I,s_2} I^2(t) + \Gamma_{E,s_2} E^2(t) + \Gamma_{u,s_2} u^2 dt \right) \tag{58}$$

subject to the mathematical model proposed by (56). Similar to (57),  $\Gamma_{I,s_2}$ ,  $\Gamma_{E,s_2}$  and  $\Gamma_{u,s_2}$  are positive weights for balancing cost function.

### 4.3 Strategy 3

In this strategy, we consider quarantined vaccination as a method that can effectively be taken into account to reduce the cost of vaccination and the number of infected and exposed individuals. Hence, the vaccination control variable is imposed on quarantined individuals who are susceptible to the virus. Then, the epidemic model with two control input imposed is given by

$$\dot{x} = f(x) + g(x)u$$

$$\begin{aligned}
 f(x(t)) &= \begin{bmatrix}
 -(\beta c + c q(1 - \beta))S(I + \theta A) + \lambda S_q \\
 \beta c(1 - q)S(I + \theta A) - \sigma E + (1 - q)(1 - e_v)\beta c V_a(I + \theta A) \\
 \sigma E - (\delta_I + \alpha + \gamma_I)I \\
 \sigma(1 - \cdot)E - \gamma_A A \\
 (1 - \beta)c q S(I + \theta A) - \lambda S_q \\
 \beta c q S(I + \theta A) - \delta_q E_q + q(1 - e_v)\beta c V_a(I + \theta A) \\
 \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\
 \gamma_I I + \gamma_A A + \gamma_H H \\
 -(1 - e_v)\beta c V_a(I + \theta A)
 \end{bmatrix} \\
 g(x) &= [-S, 0, 0, 0, -S_q, 0, 0, 0, (S + S_q)]^T \\
 x(t) &= [S(t), E(t), I(t), A(t), S_q(t), E_q(t), H(t), R(t), V_a(t)]^T \tag{59}
 \end{aligned}$$

Indeed, in this model, the vaccination will be distributed equally between the quarantined and non-quarantined individuals. By considering this fact that quarantined individuals are considered as one group which are vaccinated, the quadratic objective functional (56) is defined as

$$V^* = \min \left( \int_0^{t_f} \Gamma_{I,s_2} I^2(t) + \Gamma_{E,s_2} E^2(t) + \Gamma_{E_q,s_2} E_q^2(t) + \Gamma_{u,s_2} u^2 dt \right) \tag{60}$$

where  $\Gamma_{E_q,s_2} E_q^2(t)$  is the quadratic terms of quarantined exposed individuals representing the population that we wish to minimize besides infected and non-quarantined exposed individuals.

#### 4.4 Strategy 4

Here, instead of the uniform allocation of vaccine proposed in Strategy 3, we use two independent control variables for the propagation control of the coronavirus. The resulting control model, after incorporating the aforementioned control variables, is formulated via the following system:

$$\begin{aligned}
 f(x(t)) &= \begin{bmatrix}
 -(\beta c + c q(1 - \beta))S(I + \theta A) + \lambda S_q \\
 \beta c(1 - q)S(I + \theta A) - \sigma E + (1 - q)(1 - e_v)\beta c V_a(I + \theta A) \\
 \sigma E - (\delta_I + \alpha + \gamma_I)I \\
 \sigma(1 - \cdot)E - \gamma_A A \\
 (1 - \beta)c q S(I + \theta A) - \lambda S_q \\
 \beta c q S(I + \theta A) - \delta_q E_q + q(1 - e_v)\beta c V_a(I + \theta A) \\
 \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\
 \gamma_I I + \gamma_A A + \gamma_H H \\
 -(1 - e_v)\beta c V_a(I + \theta A)
 \end{bmatrix} \\
 g(x) &= \begin{bmatrix}
 -S, 0, 0, 0, 0, 0, 0, 0, (S) \\
 0, 0, 0, 0, -S_q, 0, 0, 0, (S_q)
 \end{bmatrix}^T \\
 x(t) &= [S(t), E(t), I(t), A(t), S_q(t), E_q(t), H(t), R(t), V_a(t)]^T \\
 x(t) &= [S(t), E(t), I(t), A(t), S_q(t), E_q(t), H(t), R(t), V_a(t)]^T \tag{61}
 \end{aligned}$$

$$\mathcal{V} = \begin{bmatrix} u \\ u_q \end{bmatrix}$$

In this strategy,  $u_q$  is the input variable which represents the fraction of quarantined susceptible individuals who vaccinated. By modifying *Strategy 3* in the control variables assumption, the updated objective functional (60) is given as

$$V^* = \min \left( \int_0^{t_f} \Gamma_{I,s_2} I^2(t) + \Gamma_{E,s_2} E^2(t) + \Gamma_{E_q,s_2} E_q^2(t) + \Gamma_{u_q,s_2} u_q^2 + \Gamma_{u,s_2} u^2 dt \right) \quad (62)$$

where the  $\Gamma_{u_q,s_2} u_q^2$  stands for minimization of quarantined vaccination. This objective illustrates the importance of vaccination optimization. Moreover, the constant  $\Gamma_{u_q,s_2}$  similar to the proposed strategies is the balancing factor, which measures the relative cost of quarantined vaccination. Figure 1 shows these strategies. In the next section, we will present the result of each strategy and compare the numerical results of their optimal solution.

These strategies are designed based on this assumption that susceptible people have been determined. Moreover, in this article, the susceptible people are considered as the only group of people who should be prioritized for getting the vaccine because they are more likely to be infected by the infection, and their infection will be more severe than the other people. After identification of the susceptible individuals, they should be classified. In this case, susceptibility can be discerned through potential risk factors such as age or pregnancy. Consequently, it will be necessary to monitor the susceptible individuals and prioritize them regarding their conditions.

## 5 Numerical results

In this section, we simulate the epidemiological model with vaccination based on the obtained data from the laboratory-confirmed case of 2019-nCoV that occurred in mainland China which is proposed by [36]. It should be noted that their research was based on a collected dataset and surveys until January 22, 2020. They employed the Markov Chain Monte Carlo to estimate the model parameters and their baselines. Based on these parameters, we implement these four strategies in python. In each strategy, the balancing factors are considered to countervail the imbalances between the magnitude of objectives' value. In this simulation, the embedded Runge–Kutta (RK5(4)) [61] has been used to model the dynamics of the epidemiological system. According to the research in [62], we assume vaccine efficacy  $e_v = 0.90$ , and the parameters in the optimal control framework are taken as

$$\phi_c(x) = [S^2, E^2, I^2, A^2, S_q^2, E_q^2, H^2, R^2, S.E, S.I, S.A, S.S_q, S.E_q, S.H, S.R],$$

$$\phi_a(x) = [S, E, I, A, S_q, E_q, H, R]$$

The initial values of weights are as follows:

$$\widehat{w}_c = [10^{-5}, \dots, 10^{-5}]$$

$$\widehat{w}_a = [10^{-4}, \dots, 10^{-4}]$$

Based on [36], the model's baselines and initial values are given in Tables 1 and 2. Here, we use them as baselines of the model and the initial values. Next, according to the defined cost functions (57), (58), (60), and (62), reinforcement learning optimal control has been applied as a feedback controller. Now, the time evolution of respective subpopulations and vaccination efforts are shown in Figs. 2, 3 and 4. In Fig. 2, the outcome of different optimal control strategies on the population of "stratified groups of people" is shown. First, in Fig. 2, the time evolution of the subpopulations illustrates that by using the vaccination strategies, susceptible,



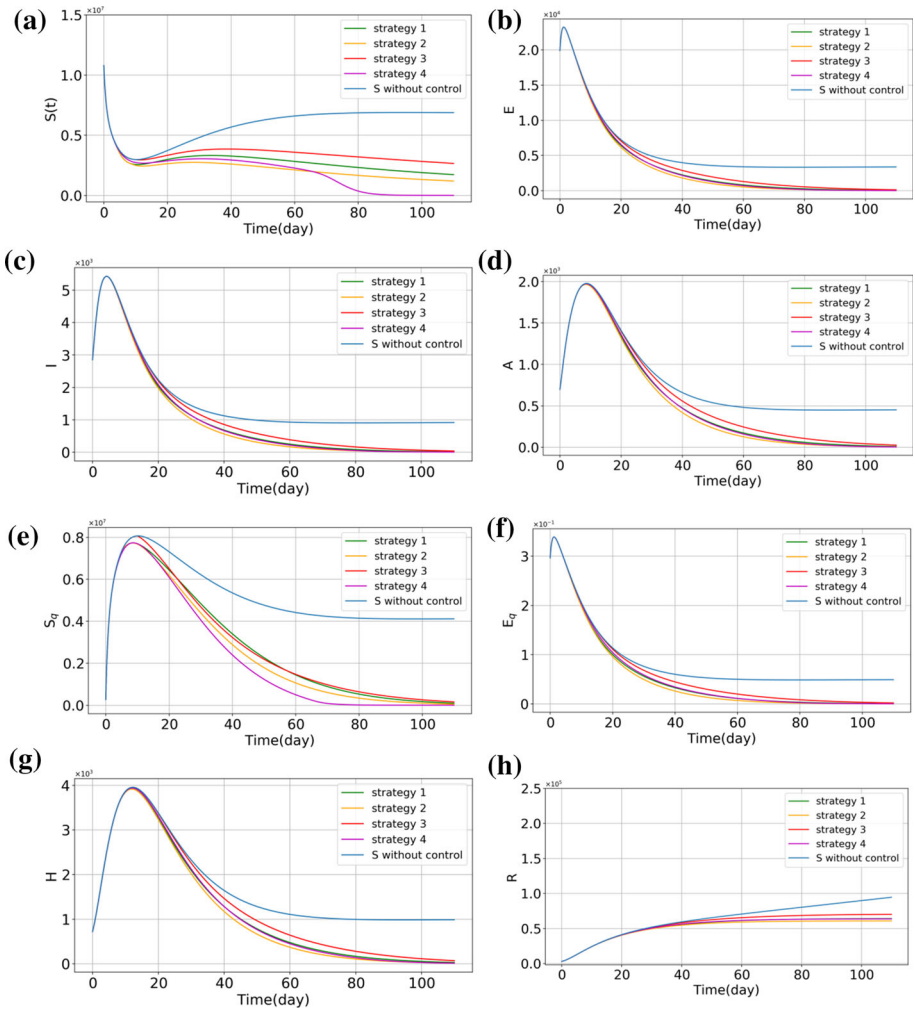
**Table 1** Parameter estimates for COVID-19 in Wuhan, China [36]

Parameter	Estimated mean value	Parameter	Estimated Mean value
$\beta$	$2.1011 \times 10^{-8}$	$\delta_q$	0.1259
$q$	$1.8887 \times 10^{-7}$	$\gamma_I$	0.33029
$\sigma$	1/7	$\gamma_A$	0.13978
$\lambda$	1/14	$\gamma_H$	0.11624
$\varrho$	0.86834	$\alpha$	$1.7826 \times 10^{-5}$

**Table 2** Initial values estimation for COVID-19 in Wuhan, China [37]

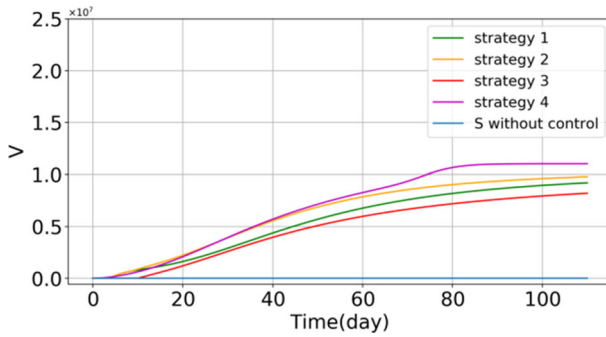
Initial values	Value	Initial values	Value
$S(0)$	10,893,000	$S_q(0)$	167,000
$E(0)$	16,000	$E_q(0)$	0
$I(0)$	2000	$H(0)$	1000
$A(0)$	1000	$R(0)$	2000

exposed, infected individuals (with or without a sign of disease), hospitalized, and recovered population fall. Moreover, the number of hospitalized individuals is reduced compared with the no-control strategy. It can be considered as a secondary effect of vaccination. Furthermore, at the beginning of public vaccination, it can successfully reduce the number of infected people, and as a consequence, people are less likely to be exposed to the infected people who can spread the disease. Therefore, the need for hospitalization will decrease in the long term. From Fig. 2a, it is precisely shown that the population of susceptible individuals declines more in Strategy 4, in which vaccine has been considered for both quarantined and non-quarantined susceptible individuals. This can explain that quarantined vaccination is one of the best options for eradicating the disease in the long run. In Fig. 3, the vaccinated population is shown in each control policy. As shown in this figure, the total population of vaccinated individuals in Strategy 3 is lower than the other optimal control strategies; however, Fig. 2c illustrates that the time evolution of infected individuals in all optimal strategies are close to each other. Thus, one can see after 110 days, the population of infected individuals in each strategy is similar to the others. Therefore, Strategy 3 will be suggested to be taken into account if restriction exists in vaccine supplements. In view of Fig. 2a, one can obtain that the tenth day can be supposed as the perfect trigger time for vaccination. On this day, the population of susceptible exceeds the minimum itself, and after that, the population of this compartment will rise gradually. In this context, if the cost of vaccination is important to governments, they can follow Strategy 3, which is the best option to bring down the cost of vaccination and reduce the number of infected people simultaneously. Based on this strategy, it would be better for governments and authorities to begin the public vaccination when the population of susceptible people reaches its minimum. Figure 4 shows the number of vaccinated population in each strategy per day presented to compare the time evolution of the vaccinated population in four different strategies. From Fig. 4a, we compare the control profiles of each strategy. One can see that in Strategy 4, the number of susceptible people declines more than the others; this strategy requires more vaccination effort. In this case, this fact is shown that from the aspect of vaccination cost, Strategy 3 can be more satisfying than the other strategies. Note that in Strategy 4, the vaccination distributes among quarantined and non-quarantined

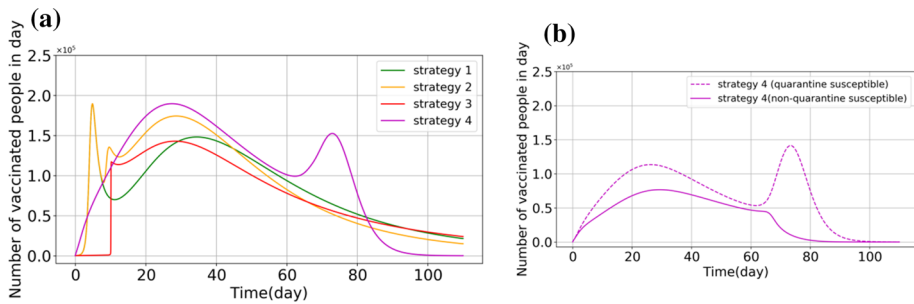


**Fig. 2** Number of individuals with different values of derivative order, **a** susceptible people, **b** exposed people, **c** symptomatic infected people, **d** asymptomatic infected people, **e** quarantined susceptible people, **f** quarantined exposed people, **g** quarantined infected people, **h** recovered people

susceptible individuals. The allocation of the vaccine in this strategy is shown in Fig. 4b. One can infer from this figure that in the primary phase, the authorities should give top priority to the quarantined susceptible individuals, although the non-quarantined susceptible individuals also should be considered for vaccination. As mentioned in the previous section, these strategies are formulated to be performed just for susceptible individuals. As a result, it should be noted that before performing, the susceptible people should be identified, stratified, and prioritized. This stratification can be performed based on their risk factor and their vulnerability. Moreover, ring vaccination is another strategy to control the outbreak [63, 64]. To be more specific, utilizing smart surveillance monitoring can provide the authorities and governors with a great tool to identify the susceptible people. This method can be taken into



**Fig. 3** Number of vaccinated people



**Fig. 4** Compared solutions vaccination for Covid-19 by different strategies

action to reduce transmission earlier by vaccination and immunization of the susceptible ring. Therefore, these proposed strategies can provide effective protection.

In this sense, in Fig. 4a, *Strategy 2* shows that its performance is better than the performance of *Strategy 1* because this strategy reduces the number of susceptible people more than other strategies. This figure also demonstrates that if the exposed people are considered in the objective function, the optimal controller can perform better regarding the reduction in both susceptible and exposed populations. It should be noted that the more exposed the population decrease, the less susceptible individuals can be infected. Due to this fact, one can infer that both susceptible and exposed populations should be considered in objective functions.

Also, from a practical viewpoint, let us denote that the reinforcement learning optimal control can introduce a better policy for vaccination distribution regarding Pontryagin’s minimum principle. For example, in [59, 65, 66], the proposed optimal controls suggest the time evolution vaccination whose initial proportion is high and significant. This high initial value of vaccine usage makes Pontryagin’s minimum principle approach impractical and too harsh in the real world, but as presented in this article, reinforcement learning optimal control can propose a policy with smooth starting that provides functionality and practicality for public vaccination.

Graphical results depict the importance of vaccine allocation. In these graphical interpretation shows that if the vaccination is taken into account, the severity of infection can be reduced gradually. In the presented model, vaccination plays a vital role in the reduction in susceptible individuals. Consequently, one can see that when susceptible individuals who can transmit the virus and be infected start to fall, the number of infected people can decline in number. Decreasing the number of symptomatically infected people will reduce the expo-

sure of uninfected people to infected people. Therefore, it can reduce the probability of being infected through the disease transmission too. As a result, the number of infected people is decreased significantly, which can end up with the elimination of the disease in society. It should be noted that due to the slow dynamic behavior of the epidemic model, it seems that the vaccine does not affect the population of infected people, but over time, the significance of vaccination effectiveness can be observable. Hence, this simulation highly suggests that governments and authorities should not be obsessed with the number of infected people during the early stage of vaccination because vaccines take time to induce immunity.

## 6 Conclusion

In this research, the significant challenge regarding vaccination strategies for COVID-19 has been investigated. Based on data from confirmed cases of 2019-nCoV in mainland China, a new deterministic SEIR model with additional vaccination components was developed. Following that, based on the reinforcement learning method, an optimal control was developed to discover the best policies. By implementing the dynamic model of the epidemiological system, numerical results for four different control strategies obtained by the proposed technique were demonstrated. The feasibility of the recommended method for designing optimal vaccine plans was clearly shown by these findings. As a future study, it would be useful to look at any of the behavioral or emotional side effects of quarantine, such as depression, which may impact depression or even suicide rate in society. Such investigations lead us to find an optimal trade-off for quarantine decisions.

## References

1. K. Yuki, M. Fujiogi, S. Koutsogiannaki, COVID-19 pathophysiology: a review. *Clin. Immunol.* **2020**, 108427 (2020)
2. P. Xu, Q. Zhou, J. Xu, Mechanism of thrombocytopenia in COVID-19 patients. *Ann. Hematol.* **99**, 1205–1208 (2020)
3. M. Ciotti, S. Angeletti, M. Minieri, M. Giovannetti, D. Benvenuto, S. Pascarella et al., COVID-19 outbreak: an overview. *Chemotherapy* **64**, 215–223 (2019)
4. T.P. Velavan, C.G. Meyer, The COVID-19 epidemic. *Trop. Med. Int. Health* **25**, 278 (2020)
5. L. Gattinoni, S. Coppola, M. Cressoni, M. Busana, S. Rossi, D. Chiumello, COVID-19 does not lead to a “typical” acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **201**, 1299–1300 (2020)
6. Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **8**, 420–422 (2020)
7. J.A. Al-Tawfiq, Asymptomatic coronavirus infection: MERS-CoV and SARS-CoV-2 (COVID-19). *Travel Med. Infect. Dis.* **35**, 101608 (2020)
8. D. Giannis, I.A. Ziogas, P. Gianni, Coagulation disorders in coronavirus infected patients: COVID-19, SARS-CoV-1, MERS-CoV and lessons from the past. *J. Clin. Virol.* **127**, 104362 (2020)
9. N. Gul, R. Bilal, E.A. Algehyne, M.G. Alshehri, M.A. Khan, Y.-M. Chu et al., The dynamics of fractional order Hepatitis B virus model with asymptomatic carriers. *Alex. Eng. J.* **60**, 3945–3955 (2021)
10. A. Ali, F.S. Alshammari, S. Islam, M.A. Khan, S. Ullah, Modeling and analysis of the dynamics of novel coronavirus (COVID-19) with Caputo fractional derivative. *Results Phys.* **20**, 103669 (2021)
11. H. Alrabaiah, M.A. Safi, M.H. DarAssi, B. Al-Hdaibat, S. Ullah, M.A. Khan et al., Optimal control analysis of hepatitis B virus with treatment and vaccination. *Results Phys.* **19**, 103599 (2020)
12. Y.-M. Chu, A. Ali, M.A. Khan, S. Islam, S. Ullah, Dynamics of fractional order COVID-19 model with a case study of Saudi Arabia. *Results Phys.* **21**, 103787 (2021)
13. M.A. Khan, A. Atangana, Modeling the dynamics of novel coronavirus (2019-nCov) with fractional derivative. *Alex. Eng. J.* **59**, 2379–2389 (2020)
14. E.O. Alzahrani, W. Ahmad, M.A. Khan, S.J. Malebary, Optimal control strategies of Zika virus model with mutant. *Commun. Nonlinear Sci. Numer. Simul.* **93**, 105532 (2021)

15. H. Jahanshahi, Smooth control of HIV/AIDS infection using a robust adaptive scheme with decoupled sliding mode supervision. *Eur. Phys. J. Spec. Top.* **227**, 707–718 (2018)
16. H. Jahanshahi, K. Shanazari, M. Mesrizadeh, S. Soradi-Zeid, J.F. Gómez-Aguilar, Numerical analysis of Galerkin meshless method for parabolic equations of tumor angiogenesis problem. *Eur. Phys. J. Plus* **135**, 1–23 (2020)
17. T.H. Zhao, O. Castillo, H. Jahanshahi, A fuzzy-based strategy to suppress the novel coronavirus (2019-nCoV) massive outbreak. *Appl Comput Math.* **20**, 160–176 (2021)
18. H. Jahanshahi, J.M. Munoz-Pacheco, S. Bekiros, N.D. Alotaibi, A fractional-order SIRD model with time-dependent memory indexes for encompassing the multi-fractional characteristics of the COVID-19. *Chaos Solitons Fract.* **143**, 110632 (2021)
19. H. Wang, H. Jahanshahi, M.-K. Wang, S. Bekiros, J. Liu, A.A. Aly, A caputo-fabrizio fractional-order model of HIV/AIDS with a treatment compartment: sensitivity analysis and optimal control strategies. *Entropy* **23**, 610 (2021)
20. P. Pandey, Y.-M. Chu, J.F. Gómez-Aguilar, H. Jahanshahi, A.A. Aly, A novel fractional mathematical model of COVID-19 epidemic considering quarantine and latent time. *Results Phys.* **221**, 104286 (2021)
21. N.C. Grassly, C. Fraser, Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487 (2008)
22. S.-B. Chen, F. Rajaei, A. Yousefpour, R. Alcaraz, Y.-M. Chu, J.F. Gómez-Aguilar et al., Antiretroviral therapy of HIV infection using a novel optimal type-2 fuzzy control strategy. *Alex. Eng. J.* **60**, 1545–1555 (2021)
23. M.A. Khan, C. Alfiniyah, E. Alzahrani, Analysis of dengue model with fractal-fractional Caputo-Fabrizio operator. *Adv. Differ. Equ.* **2020**, 1–23 (2020)
24. S. Eubank, H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczka et al., Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004)
25. S. Ullah, M.A. Khan, Modeling the impact of non-pharmaceutical interventions on the dynamics of novel coronavirus with optimal control analysis with a case study. *Chaos Solitons Fract.* **139**, 110075 (2020)
26. M. Awais, F.S. Alshammari, S. Ullah, M.A. Khan, S. Islam, Modeling and simulation of the novel coronavirus in Caputo derivative. *Results Phys.* **19**, 103588 (2020)
27. A. Yousefpour, H. Jahanshahi, S. Bekiros, Optimal policies for control of the novel coronavirus disease (COVID-19) outbreak. *Chaos Solitons Fract.* **136**, 109883 (2020)
28. F. Ndaïrou, I. Area, J.J. Nieto, D.F.M. Torres, Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fract.* **135**, 109846 (2020)
29. N.H. Tuan, H. Mohammadi Rezapouridi, A mathematical model for COVID-19 transmission by using the Caputo fractional derivative. *Chaos Solitons Fract.* **140**, 110107 (2020)
30. M.S. Alqarni, M. Alghamdi, T. Muhammad, A.S. Alshomrani, M.A. Khan, *Mathematical Modeling for Novel Coronavirus (COVID-19) and Control* (Methods Partial Differ. Equ, Numer, 2020). <https://doi.org/10.1002/num.22695>
31. J. Panovska-Griffiths, *Can Mathematical Modelling Solve the Current Covid-19 Crisis?* (Springer, Berlin, 2020).
32. M.A. Khan, A. Atangana, E. Alzahrani, The dynamics of COVID-19 with quarantined and isolation. *Adv. Differ. Equ.* **2020**, 1–22 (2020)
33. M.A.A. Oud, A. Ali, H. Alrabaiah, S. Ullah, M.A. Khan, S. Islam, A fractional order mathematical model for COVID-19 dynamics with quarantine, isolation, and environmental viral load. *Adv. Differ. Equ.* **2021**, 1–19 (2021)
34. M.A. Khan, S. Ullah, S. Kumar, A robust study on 2019-nCoV outbreaks through non-singular derivative. *Eur. Phys. J. Plus* **136**, 1–20 (2021)
35. D. Dunford, B. Dale, N. Stylianou, E. Lowther, M. Ahmed, A.I. De la Torre, Coronavirus: the world in lockdown in maps and charts. *BBC News* **9**, 462 (2020)
36. B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao et al., Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.* **9**, 462 (2020)
37. B. Tang, N.L. Bragazzi, Q. Li, S. Tang, Y. Xiao, J. Wu, An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). *Infect. Dis. Model.* **5**, 248–255 (2020)
38. E. Volz, S. Mishra, M. Chand, J.C. Barrett, R. Johnson, L. Geidelberg et al., Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England. *Nature* **593**, 266–269 (2021)
39. N.G. Davies, S. Abbott, R.C. Barnard, C.I. Jarvis, A.J. Kucharski, J.D. Munday et al., Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science* **372**, eabg3055 (2021)
40. M. Abu-Khalaf, F.L. Lewis, Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica* **41**, 779–791 (2005)
41. R. Beard, Improving the closed-loop performance of nonlinear systems, Ph.D. dissertation, Rensselaer Polytech. Inst., Troy, NY (1995)

42. F.A. Yaghmaie, D.J. Braun, Reinforcement learning for a class of continuous-time input constrained optimal control problems. *Automatica* **99**, 221–227 (2019)
43. P.G. Drazin, P.D. Drazin, *Nonlinear Systems* (Cambridge University Press, Cambridge, 1992).
44. P.A. Ioannou, J. Sun, *Robust Adaptive Control* (Courier Corporation, London, 2012).
45. X. Zhong, H. He, An event-triggered ADP control approach for continuous-time system with unknown internal states. *IEEE Trans. Cybern.* **47**, 683–694 (2016)
46. Wei Q, Zhang H, *A New Approach to Solve a Class of Continuous-Time Nonlinear Quadratic Zero-Sum Game Using ADP*. *IEEE*. pp. 507–512.
47. V.R. Konda, J.N. Tsitsiklis, *Actor-Critic Algorithms*, Citeseer. pp. 1008–10014.
48. K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* **3**, 551–560 (1990)
49. Y. Zhu, D. Zhao, X. Li, Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics. *IET Control Theory Appl.* **10**, 1339–1347 (2016)
50. G. Xiao, H. Zhang, Y. Luo, Q. Qu, General value iteration based reinforcement learning for solving optimal tracking control problem of continuous-time affine nonlinear systems. *Neurocomputing* **245**, 114–123 (2017)
51. B. Kiumarsi, F.L. Lewis, Actor-critic-based optimal tracking for partially unknown nonlinear discrete-time systems. *IEEE Trans. Neural Netw. Learn. Syst.* **26**, 140–151 (2014)
52. X. Yang, D. Liu, B. Luo, C. Li, Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning. *Inf. Sci.* **369**, 731–747 (2016)
53. A. Wuraola, N. Patel, SQNL: A new computationally efficient activation function. *IEEE*, pp. 1–7.
54. D.L. Elliott, A better activation function for artificial neural networks. *ISR technical report TR 93-8*, University of Maryland (1993)
55. K.G. Vamvoudakis, F.L. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica* **46**, 878–888 (2010)
56. S. Bhasin, R. Kamalapurkar, M. Johnson, K.G. Vamvoudakis, F.L. Lewis, W.E. Dixon, A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica* **49**, 82–92 (2013)
57. J. Shi, D. Yue, X. Xie, Adaptive optimal tracking control for nonlinear continuous-time systems with time delay using value iteration algorithm. *Neurocomputing* **396**, 172–178 (2020)
58. F.W. Lewis, S. Jagannathan, A. Yesildirak, *Neural Network Control of Robot Manipulators and Non-Linear Systems* (CRC Press, London, 2020).
59. Y. Yang, S. Tang, X. Ren, H. Zhao, C. Guo, Global stability and optimal control for a tuberculosis model with vaccination and treatment. *Discrete Contin. Dyn. Syst. B* **21**, 1009 (2016)
60. D. Rostamy, E. Mottaghi, Stability analysis of a fractional-order epidemics model with multiple equilibriums. *Adv. Differ. Equ.* **2016**, 1–11 (2016)
61. J.R. Dormand, P.J. Prince, A reconsideration of some embedded Runge–Kutta formulae. *J. Comput. Appl. Math.* **15**, 203–211 (1986)
62. F.P. Polack, S.J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart et al., Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020)
63. W. Xu, S. Su, S. Jiang, Ring vaccination of COVID-19 vaccines in medium-and high-risk areas of countries with low incidence of SARS-CoV-2 infection. *Clin. Transl. Med.* **11**, 2 (2021)
64. A.J. Kucharski, R.M. Eggo, C.H. Watson, A. Camacho, S. Funk, W.J. Edmunds, Effectiveness of ring vaccination as control strategy for Ebola virus disease. *Emerg. Infect. Dis.* **22**, 105 (2016)
65. R.A. Sari, U. Habibah, A. Widodo, Optimal control on model of SARS disease spread with vaccination and treatment. *J. Exp. Life Sci.* **7**, 61–68 (2017)
66. M.D. Ahmad, M. Usman, A. Khan, M. Imran, Optimal control analysis of Ebola disease with control strategies of quarantine and vaccination. *Infect. Dis. Poverty* **5**, 1–12 (2016)