International Journal of Industrial Electronics, Control and Optimization © 2020 IECO
Vol. 3, No. 3, pp. 353-364, July (2020)

Fault Tolerant Control of Blood Glucose Concentration Using Reinforcement Learning

Amin Noori [1, †], Mohammad Ali Sadrnia [2], and Mohammad Bagher Naghibi-Sistani [3]

[1, 2] Faculty of Electrical and Robotic Engineering, Shahrood University of Technology, Shahrood, Iran
[3] Electrical department, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

ABSTRACT

In this paper, the main focus is on blood glucose level control and the possible sensor and actuator faults which can be observed in a given system. To this aim, the eligibility traces algorithm (a Reinforcement Learning method) and its combination with sliding mode controllers is used to determine the injection dosage. Through this method, the optimal dosage will be determined to be injected to the patient in order to decrease the side effects of the drug. To detect the fault in the system, residual calculation techniques are utilized. To calculate the residual, it is required to predict states of the normal system at each time step, for which, the Radial Basis Function neural network is used. The proposed method is compared with another reinforcement learning method (Actor-Critic method) with its combination with the sliding mode controller. Finally, both RL-based methods are compared with a combinatory method, neural network and sliding mode control. Simulation results have revealed that the eligibility traces algorithm and actor-critic method can control the blood glucose concentration and the desired value can be reached, in the presence of the fault. However, in addition to the reduced injected dosage, the eligibility traces algorithm can provide lower variations about the desired value. The reduced injected dosage will result in the mitigated side effects, which will have considerable advantages for diabetic patients.

Article Info

Keywords:
Fault Tolerant Control, Reinforcement Learning,
Eligibility Traces, Actor Critic, Neural Network, Diabetic Model

Article History:
Received 2019-10-13
Accepted 2020-01-15

I. INTRODUCTION

Currently, diabetes, as a silent epidemic disease, is spreading throughout the world. Main causes of diabetes are namely lack of physical activities and the growing obesity. Recent reports from the World Health Organization (WHO) and diabetes International Federation have stated that diabetes mellitus is an endocrine disease related to the level of glucose and insulin in body. This can be determined by hyperglycemia due to shortage of insulin or insulin operation, or both. Two types of most common diabetes are due to insulin generation reduction (Diabetes type 1) and reduction in body's response to the insulin (Diabetes type 2). Both types of diabetes mellitus will result in hyperglycemia and finally, symptoms of diabetes will appeared. These symptoms include frequent peeing out, become thirstier and increase in the fluid intake, blurred vision, unplanned weight loss and fatigue. Human body needs to preserve the glucose level within 70 and 110 (mg/dL) range. If the glucose level is significantly out of this range, the individual is said to have a plasma glucose problem. Hence, monitoring the blood glucose level for diabetic and non-diabetic individuals' health is vital [1].

It should be noted that the level of blood glucose is the lowest in the morning and before breakfast. Two or three hours after eating a meal, depending on the type of food a person has eaten, the level of blood glucose will increase. For a normal person, the blood glucose level will reach 175 (mg/dL) after each meal. Over time, this value will approach the normal value, which normally happens faster in non-diabetic people comparing to a diabetic patient [2].

Employing medical methods to control the level of blood glucose in diabetic patients is time-consuming. In some cases, this may disturb the level of blood glucose and puts the patient in danger. Due to this reason, employing intelligent methods to control it in diabetic patients have become a necessity. Such methods will accelerate the process of controlling the level of blood glucose. Moreover, this procedure will be performed with minimum dosage injection, mitigating the side effects as

[†]Corresponding Author: Amin_noori@ieee.org; Tel: +98-9155162698
[†]Faculty of Electrical and Robotic Engineering, Shahrood University of Technology, Shahrood, Iran

much as possible. In the following, a case is discussed in which researches performed blood glucose level control employing RL methods.

In [3], Q-learning algorithm -in the RL problem- is used to regulate the blood glucose level in type-1 diabetic patients. The learning agent explores in the environment and learns to select the best action, which is achieving the optimal insulin dosage. Blood hemoglobin changes are considered as states of the environment and classified into 6 states. Using the Q-learning algorithm, the agent receives immediate rewards and over time, it will learn to select the optimal insulin dosage for injection so as to preserve the blood glucose level in diabetic patients within the desired range. In [4], the RL approach is used for regulating the blood glucose level in diabetic type-2 patients. As reported, physical exercise is useful for diabetic patients. In this investigation, diabetic patients are encouraged to do physical exercises, e.g. walking. Using the RL approach, every week a message was sent to 27 patients' cell phones reminding them about their diet and required physical exercises. Results implied that patients treated by this algorithm had more reduction in their blood glucose level. In [5] the glucose level was controlled using Temporal Difference (TD) methods. Indeed, in this paper, the Palumbo model was used in which two dimensional states, including glucose level and insulin, were considered. The drug dosage was considered as the action in RL problem. In [6], the RL method is employed to control the Glycaemia concentration in septic patients. Results have revealed that the proposed method can potentially provide physicians a private glycemic control strategy, based on the optimal policy it has learned. Besides, this method is able to decrease the mortality rate, from 31% to 24.7%. In [7], a hybrid method, involving the RL and feed-forward neural networks, is used to control the blood glucose level and regulate the insulin dosage of injection in patients with diabetes type-1. The Kalman Filter is used to estimate the unmeasurable states of patients. Simulation results have revealed that the proposed controller offered a better performance than the Proportional Integral Derivative (PID) controller in terms of regulating variations in the blood glucose level. Furthermore, the proposed controller can prevent an increase in the level of Hypoglycemia.

Medical systems are highly sensitive and their accurate performance is of great importance. Faulty performances in such systems will cause intensive physical harms. Therefore, quick, accurate and on time fault detection and isolation (FDI) and fault tolerant control (FTC) in medical systems have attracted a significant attention in recent years. One of the most widely used methods in fault detection and fault tolerance control is methods using artificial intelligence, which is one of the latest methods is RL, some examples of which are elaborated on the following.

In [8], a fault detection procedure is put into practice in a system with uncertainty using RL. To this end, states of the main system are estimated, the related error is obtained, and then fault detection and identification are implemented. In effect, the eligibility traces algorithm is used for the considered purpose. In [9], the impact of different faults on the system is investigated and the Q-learning algorithm is used for FDI purpose. Results showed that the Q-learning algorithm is more accurate in detecting and identifying system faults compared with the case that no intelligent learning agent is utilized. In [10], a nonlinear Temporal Difference (TD) learning is used for FTC. Since there are noise and disturbance in the considered system, Extended Kalman Filter (EKF) is used to design the observer and estimate states of the system. In [11], FTC is investigated in nonlinear systems. In this regard, actor-critic method of RL and ANNs are utilized. It is shown that the actor-critic method is faster than ANNs in terms of detecting faults and controlling the system in the presence of faults. In [12], for tracking and controlling the linear system against faults more efficiently, residual calculation methods is used in combination with H_∞. In the proposed method, the Q-learning is used for an optimal tracking performance. The main objective of the paper is data-based space identification. For the fault detection process, an adaptive threshold is determined through which false detection of noise is prevented. In [13], assuming there is no information about the applied fault to the system, RL is used for the fault detection and control purposes. For stability analysis of the system, Lyapunov theory is used. Likewise, the actor-critic method, one of the efficient class of methods in RL, is employed. In [14], an advanced method, Auto-Step algorithm, contributed to RL. In this sense, the proposed method is compared with the Recursive Least Squared (RLS) method which is used to estimate states of the system. It is revealed that a combinatory use of RL and Auto-Step algorithm has detected and minimized system faults with a higher accuracy and convergence speed.

A number of studies have focused on fault diagnosis and fault tolerant control of the processes. Some works have investigated the fault diagnosis and fault tolerant control of diabetes in patients such as In [15], where a metabolic model is used for patients with Diabetes type-1. Also, the fault detection procedures were implemented based on sensor faults including the disconnection fault in blood glucose level detection sensors as well as disconnection or leakage faults in insulin injection devices. Here, fault detection is indeed applied within 20 minutes. In [16], the unscented Kalman filter is employed to detect and compensate sensor faults and detect unannounced meals related to the blood glucose tracking in diabetic patients. To this aim, a simulated model of the diabetic patient is used and it is assumed that the system is affected by two sensor faults, drift and Pressure Induced Sensor Attenuation (PISA). Simulation results revealed that the proposed method is able to control the blood glucose level continuously in the presence of both sensor faults. In [17], detecting different faults in the artificial pancreas system in patients with diabetes type-1 is investigated. The main reason for this research is that this can help to determine the insulin dosage. The authors had used remodulated dynamic time warping to synchronize the signal trajectories and have employed the Svitzky-Golay filter for real-time calculations of numerical derivatives in the multiway principal component analysis. To prepare the required data, 4 patients had been tested for 60 hours accompanied by changes in their meals and physical exercises. Results revealed the efficient performance of their real-time method to detect different sensor faults and accurate labeling. The main advantage of this work was a considerable contribution to preventing hypoglycemia. In [18], an overview of different fault detection methods for the blood glucose level control in patients with diabetes is represented. In [19], the importance of the accurate performance of Continuous Glucose Monitoring (CGM) sensors is

emphasized. It is stated that the accurate operation of such sensors affected by the fault is quite vital for artificial pancreas systems in patients suffering from diabetes type-1. In this sense, the sparse recursive kernel filtering algorithm is used for the purpose of fault detection and improvement of the measurement accuracy purposes. The proposed algorithm is designed in a way that the noisy system can detect the fault and keep its operation. In [20], fault detection methods in diabetes management system are studied. Some of prominent methods include ANNs, Deep Learning, Decision Tree and Fuzzy Logic Control.

In the following, a review of papers on fault detection and isolation in different systems is represented.

In [21], the performance of wind turbines based on Doubly-Fed Induction Generators (DFIG) in the presence of the fault is investigated. Here, an adaptive programmable controller is employed, which is able to control the turbines against the occurring faults. In [22], the fault detection and control is investigated in photovoltaic systems using Sliding mode control method. In [23], the Local Model Network (LMN) method is utilized to model the system and identify the occurring faults. The main reason for adopting this method is high dimensionality of inputs and number of parameters. The adopted method can considerably reduce and optimize the exploration space of the considered problem. In [24], the fault detection procedure is practiced in discrete-time nonlinear systems in time-delayed networks (Patri nets).

A significant issue that should be noted is the fact that after the learning phase, the policies learned by the agent can be applied to real diabetic patients. However, in this case, the RL agent will have a reduced and negligible trial and error on the real diabetic patient. In this regard, the agent will adopt the optimal policy based on the intrinsic characteristics of the patient. This is known as drug personalization in the new literature and considered as a brand-new concept for biomedical engineering researchers and scholars.

If the optimal control method is used for controlling the level of blood glucose, the system must be linearized. When linearizing the system, an estimation of the main system will be obtained, which will not show a precise behavior of the main system. Hence, classic methods cannot reveal the specifications of a suitable performance when facing a real patient. Moreover, this method highly depends on the mathematical model of the system. Consequently, in this paper, the model-free approach of Eligibility Traces Algorithm is employed. Opposite to the optimal control approach, the eligibility traces algorithm is model-free and can be implemented and adapted to all nonlinear systems. Due to the lack of access to the real patients, a time-delay mathematical model of a diabetic patient is used.

To describe the operation of this method, first, several trials and errors are taken by the agent on the mathematical model. Hence, the agent learns to determine different dosages (dosage to be injected) for different states of the patient. Using the eligibility traces algorithm to determine the optimal injected dosage for each individual patient- drug personalization- is considered as one of the prominent important aspects of this algorithm. Drug personalization is one of the most recent subjects, which is quite vital and emphasized by different researchers. For realizing the drug personalization using eligibility traces algorithm, the obtained Q-table is used and learning is performed for each real patient, separately. In this case, less trails and errors are taken on the real patients to determine the optimal injected dosage, which has its own prominent advantages.

The remainder of this paper is organized as follows. In section 2, the mathematical model used in this study is represented and described. In section 3, RL, the eligibility traces algorithm and the actor-critic method are explained briefly. In section 4, the proposed method for blood glucose level regulation and control is given. Finally, the proposed method is simulated in MATLAB Software and simulation results are represented, discussed and concluded.

II. MATHEMATICAL MODEL

Using appropriate models to express the biological behavior of Glucose-insulin diabetic patients is an important issue. Different mathematical models have been proposed to describe diabetes [2, 25-33]. The model of Diabetes we use in this paper was introduced by Palumbo et. al. [34], which has some advantages over other mathematical models. The most important advantages are considering delaying time (τ_g) and Plasma insulin concentrations decline index (K_{xi}) . Considering the delay in nonlinear model reduces the systems degree, but the analysis would be more complex. This model takes the form of

$$\frac{dG(t)}{dt} = -K_{xgi}I(t)G(t) + \frac{T_{gh}}{Vg} \quad (1)$$

$$\frac{dI(t)}{dt} = -K_{xi}I(t) + \frac{T_{igmax}}{v_i} f\left(G(t-\tau_g)\right) + u(t) \quad (2)$$

where G(t) [mM], and I(t) [pM] are plasma glycemia and insulinemia, respectively. The variable *u* represents the control unit. In this model, parameters are:

- $K_{xgi}[min^{-1}PM^{-1}]$: Saved rate of glucose-insulin dependent manufacturing.
- $T_{gh}\left[min^{-1}\left(\frac{mmol}{KgBW}\right)\right]$: index of hepatic glucose and glucose intake.
- $V_g\left[\frac{L}{KgBW}\right]$: Distribution rate of glucose.
- $K_{xi}[min^{-1}]$: Plasma insulin concentrations decline index.
- $T_{igmax}\left[min^{-1}\left(\frac{pmol}{KgBW}\right)\right]$: The maximum rate of insulin secretion in the second phase.
- $V_i\left[\frac{L}{KgBW}\right]$: Plasma insulin distribution rate.

Nonlinear function *f(G(t))* which describes the rate of insulin delivery is:

$$f\left(G(t-\tau_g)\right) = \frac{\left(\frac{G(t-\tau_g)}{G^*(t-\tau_g)}\right)^{\delta}}{1+\left(\frac{G(t-\tau_g)}{G^*(t-\tau_g)}\right)^{\delta}} \quad (3)$$

where

- δ : Positive constant parameter, which describes the ability of the pancreas to the cycle of glucose in plasma. If δ = 0, the pancreas would not reply to glucose circulation at all; if δ = 1, the pancreas would reply according to a Michaelis-Menten dynamics [34].

- $G^*[mM]$: is the glycemia level, at which the insulin release is the half of its maximum rate; at a glycemia equal to G∗ corresponds an insulin secretion equal to$\frac{T_{igmax}}{2}$

The structure of the equations guarantees non-negative solutions for state variables. System constants are shown in Table I.

TABLE I
SYSTEM CONSTANTS [34]

Value	Parameter	Value	Parameter
0.187	V_g	6.14	G_b
1.211×10^{-2}	K_{xi}	93.669	I_b
0.003	T_{gh}	1.573	T_{igmax}
0.25	V_i	3.205	δ

III. REINFORCEMENT LEARNING

In RL, two components play key roles, agent and the environment. At each moment, the agent acquires new information from the environment, through the trial and error and updates its performance. In fact, RL is neither a supervised nor an unsupervised learning. The best action is not determined for the learning agent. After a sufficient number of iterations, through trial and error in the environment and receiving rewards or penalties, the agent will find out the best action. Conceptually, the best action will lead to the optimal policy. In solving an optimization problem in the RL framework, three important components have also key roles, namely states, actions and rewards. At each moment, the learning agent observes the current state of the environment (s_t), takes the action (a_t), and transits to the next state (s_{t+1}) receiving the immediate reward (r_{t+1}) from the environment. This reward will affect the previous states or states-actions pairs [35]. Fig.1 illustrates the state-action chain in the RL framework.

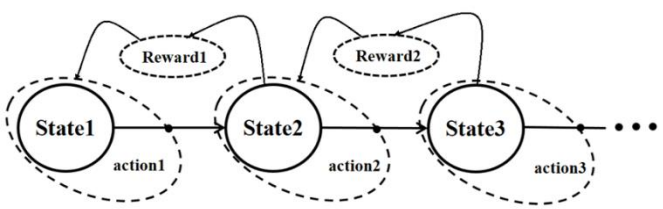

Fig. 1. State-action chain in RL

Equ. (4) represents the state value estimation under the policy π denoted by V^π.

$$V^\pi(s) = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\} \quad (4)$$

where γ is the discount rate, within [0,1] and k denote the time step. The value of each state under the policy π equals the expected value of sum of the discount rate multiplied by rewards received from the environment, from the current moment (s_t) to the end of the path. At each moment, multiplying $\gamma^k$by the value of the current state (s_t), illustrates the effect of received rewards on the value of the current state. Equ. (5) gives the same concept for a pair of state-action at each time step.

$$Q^\pi(s,a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\} \quad (5)$$

Here, the value of each state-action pair under a specific policy π is represented. As represented, this can be obtained by sum of the discount rate multiplication by the value of received rewards, from the current moment to the end of the episode.

A. Eligibility Traces Algorithm

The eligibility traces algorithm can be considered as an interaction between two solving methods of RL, Monte-Carlo (MC) and Temporal Difference (TD). In this method, the value of each state-action pair is updated at every step. Beside the forward view in each state (s_t), there exists a backward view, as well. So, there is a bilateral view; that is the value of state-action pair at the next time step as well as the total value of previous state-action pairs affect updating the value of the current state-action pair. This will accelerate the convergence and in our specific application, this will accelerate the process of controlling the reduction in the level of blood glucose, through determining the optimal dosage. Moreover, the side effects will be significantly decreased [35]. Fig.2 illustrates the backward view in the eligibility traces algorithm.

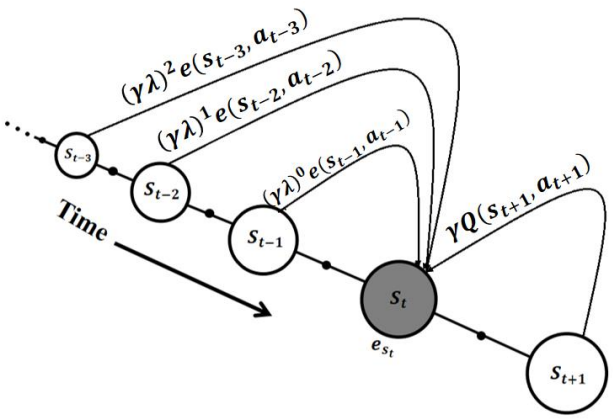

Fig. 2. Backward view in the eligibility traces algorithm

The eligibility of a state-action pair is obtained using Equ. (6).

$$e(s_t, a_t) = e(s_t, a_t) + 1 \quad (6)$$

For updating the value of the state-action pair Equ. (7) is proposed, which enables faster convergence than previous methods

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \sum_a \pi(s_t, a) Q(s_{t+1}, a) - Q(s_t, a_t)]\, e(s_t, a_t) \quad (7)$$

where α and γ are learning rate and forgetting factor respectively. r_{t+1} denotes the received reward and $Q(s_{t+1}, a_{t+1})$ represents the value of the state-action pair at the next time step (t+1).

The eligibility of the state-action pair at each time step can be obtained using Equ.(8).

$$e_t(s,a) = \gamma\lambda \times e_t(s,a) \quad (8)$$

where λ is a constant value, ranges within [0,1] and weighting rewards from the current step till the end of the episode. Based on this relation, the eligibility of the current state-action pair is obtained by multiplying the eligibility of the state-action pair at previous time step by γλ factor [35]. It should be noted that the action selection is performed by the Softmax method in this paper.

B. Actor-Critic method

In this method a separate memory structure is considered to represent policies which are independent from the value function. The structure of the policy indicates the actor and actions are generated by the actor. The estimated value function indicates the critic, which criticizes the actions made by the actor [35]. The general structure of the actor-critic method is shown in Fig. 3.

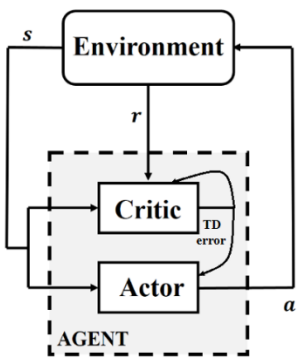

Fig. 3. The actor-critic architecture

Evaluating the action is done by the critic, as Equ. (9)

$$\delta_t = r_{t+1} + \gamma v(s_{t+1}) - v(s_t) \quad (9)$$

where v is the current value function implemented by the critic and r_{t+1} indicates the reward received by the agent. The action-selection probability can be calculated as Equ. (10).

$$\pi_t(s,a) = \Pr\{a_t = a | s_t = s\} = \frac{e^{p(s,a)}}{\sum_b e^{p(s,b)}} \quad (10)$$

where $p(s,a)$ are the values at time step t of the modifiable policy parameters of the actor and can be calculated as Equ. (11):

$$p(s_t,a_t) = p(s_t,a_t) + \beta\delta_t(1-\pi_t(s_t,a_t)) \quad (11)$$

where β is a constant positive value, $\beta \in [0,1]$. If δ_t is positive, then the selected action will be good and suggested for the next steps. In contrast, if δ_t is negative, then the agent will not tend to select this action in the future. In this method, the update of the state-action value at the present time step can be formulated as Equ. (12):

$$Q(s_t,a_t) = Q(s_t,a_t) + \alpha[r_{t+1} + \gamma\sum_a \pi(s_t,a)Q(s_{t+1},a) - Q(s_t,a_t)] \quad (12)$$

as shown in Equ. (12), the value of the state-action pair at each time step depends on the value of the state-action pair at that time step and the sum of the product of the probability of selecting each action in that state and value of the next state-action pair. α indicates the learning rate, which is a constant positive value.

C. Proposed Control Strategy in the Presence of Fault

In this section, for the diabetes system, an RL structure is introduced with the aim of fault tolerant control. At this point, it is assumed that sensor and actuator faults have occurred in the system. In this study, the eligibility traces algorithm is employed to control the blood glucose concentration in diabetic patients. Fig.4 depicts the blood glucose control process using the eligibility traces algorithm. In this method, RBF is used to estimate the states of the system while Sliding Mode Controllers are employed. Estimated values from RBF neural network are compared with real outputs of the diabetic system, and the level of residual is then calculated. Sliding Mode Control (SMC) is a simple method to control nonlinear systems, which is robust against noise and disturbance. To design the controller by SMC, a sliding surface should be defined. System states should converge to this sliding surface [36].

The functionality of the control architecture shown in Fig. 4 is as follows. First, The RBF neural network is used to obtain the model of the normal system. When the fault occurs in the system, the blood glucose level will be controlled following a certain procedure. In the first step, the value of both variables (blood glucose and insulin) are given to the RL-based controller, as inputs. The intelligent agent sends the selected action (u^R) to the output. Based on whether the blood glucose level control is effective or not, the immediate reward is allocated. At each time step, values of mentioned two variables are estimated, using the RBF neural network. At each time step, estimated values are the same as measured values, in the absence of the fault and disturbance in the system. Therefore, the difference between the estimated values and measured ones, known as residuals, will be given as inputs to the sliding mode controller. On the other hand, the difference between estimated values and desired values, known as errors, will be given to the sliding mode controller. These values have effect, when calculating the drug dosage for the blood glucose level control. Corresponding relations are represented in the following.

As shown in Fig.4, the considered system is unknown. However, since there was no access to real patients, the mathematical model of the diabetic patient is employed. It should be noted that the adopted approach is able to be implemented on the real patient.

In this paper, the sliding surface is defined on the basis of the errors and the residuals. The sliding surface for the normal system can be stated as Equ. (13) [11]:

$$s^N = \binom{1}{0} e_2{}^N + \binom{1}{1} \lambda e_1{}^N \quad (13)$$

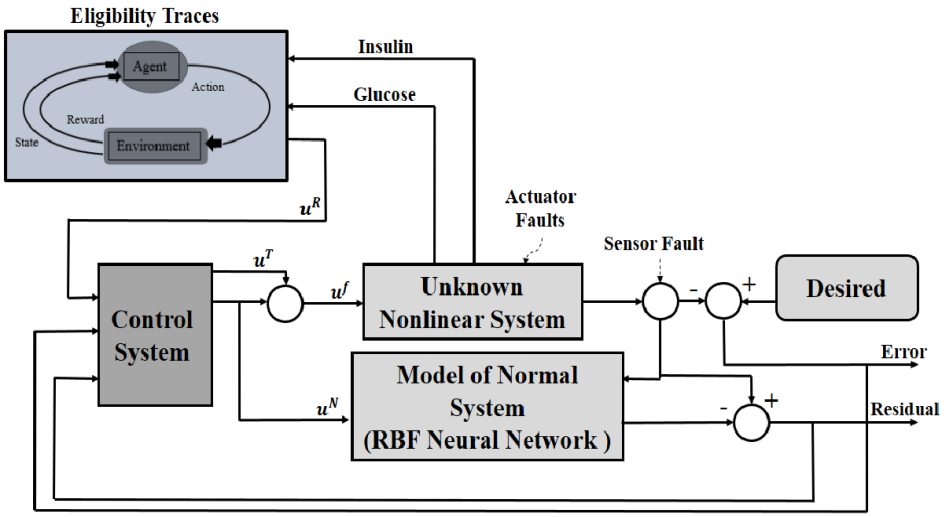

Fig. 4. Blood glucose concentration control using the eligibility traces control, in the presence of faults

where s^N denotes the sliding surface definition on the normal system and s^R is the definition of the sliding surface on the residual value. In Eq. (13), e denotes the difference between the desired value of glucose and insulin and output of the main system, at each time step. In Eq. (14), the error is calculated based on the difference between the main output and the desired value.

$$e = y_{new} - x_d \tag{14}$$

where y_{new} denotes the output value at the current time step and x_d represents the output desired value. Also, λ represents a strictly constant value. The sliding surface is calculated based on the residue, as Eq. (15)

$$s^R = \binom{1}{0} R_2 + \binom{1}{1} \lambda R_1 \tag{15}$$

In Eq. (15), R is the residual and represents the difference between the output of the main system and the estimated system. where the residual can be obtained based on Eq. (16). The residual is the difference between the main output and the estimated value of the main system.

$$R = y_{new} - \hat{y}_n \tag{16}$$

where $\hat{y}_n$ is the estimated value related to the output of the main system.

The control input in the considered system, u^N can be defined as Eq.(17).

$$u^N = \eta_N sgn(s^N) - \binom{1}{0}\left(\hat{f}_N(\hat{X}^N) - {x_d}^{(n)}\right) - \binom{1}{1}\lambda^1 {e_2}^N - \binom{2}{2}\lambda^2 {e_2}^N \tag{17}$$

where η_N is a positive constant value. $\hat{f}_N(\hat{X}^N)$ is the estimation of states in the normal system at each moment, and x_d is the desired output.

For the eligibility traces part, considered states are two dimensional, including the blood glucose level and insulin level. The reward function for the eligibility traces algorithm is determined based on Eq.(18) [11].

$$r(t) = -\lambda_R(s^R \dot{s}^R) \tag{18}$$

where λ_R is a positive constant value. $\dot{s}^R$ is the derivative of the function s^R and can be calculated as Eq.(19).

$$\dot{s}^R = \binom{1}{0}\left(f(x^f) - f(x^N)\right) + \binom{1}{0}\left(F^{(n)} + d + u^f + u^N\right) + \binom{1}{1}\lambda^1 R_2 \tag{19}$$

where $f(x^f)$ and $f(x^N)$ are output of the faulty system and output of the normal system respectively. $F^{(n)}$ Indicates a system fault that is estimated by the neural network. According to Eq. (18), when $s^R \dot{s}^R$ is positive, the energy of the system has increased and the stability has decreased. So, the reward will be negative. Otherwise, the energy of the system has decreased and the stability has increased. Therefore, the reward will be positive [11]. Mathematically proof of this argument is fully stated in [11].

The control input u^T is calculated based on the action that the agent will select. This can be described as Eq. (20).

$$u^T = -u^R - (\eta_R sgn(s^R)) \tag{20}$$

where η_R is a constant, positive value and u^R is the practical input, selected by the agent. Finally, the input of the system is the dosage of the drug injected to the diabetic patient, denoted by u^f and can be calculated as Eq.(21).

$$u^f = u^N + u^T \tag{21}$$

The blood glucose level control using the eligibility traces algorithm is shown in Fig. 5.

Eligibility Trace Algorithm

States $S = \{1, \dots, n_s\}$ *Glucose and Insulin*
Actions $A = \{1, \dots, n_a\}$ (u^R)
Reward function R: $S \times A \to \mathbb{R}$ $r(t) = -\lambda_R(s^R \dot{s}^R)$
Black-box (probabilistic) transition function $T: S \times A \to S$

Learning rate $\alpha \epsilon [0,1]$, typically $\alpha = 0.5$

Discounting factor $\gamma \epsilon [0.1]$, typically $\alpha = 0.9$

Trade-off between TD and MC: $\lambda \epsilon [0,1]$
Procedure QLEARNING (S, A, R, α, γ, λ)
Initialize Q: $S \times A \to \mathbb{R}$ arbitrarily
Initialize e: $S \times A \to \mathbb{R}$ with 0 ⊳ eligibility trace
While Q is not converged **do**
Select $(s, a) \epsilon S \times A$ arbitrarily
While s is not terminal **do**
$r \leftarrow R(s, a)$
$s' \leftarrow T(s, a)$ ⊳ Receive the new state
Calculate π based on Q ($\varepsilon - greedy$)
$a' \leftarrow \pi(s')$
$e(s, a) \leftarrow e(s, a) + 1$
$\delta \leftarrow r + \gamma . Q(s', a') - Q(s, a)$
for $(\tilde{s}, \tilde{a}) \epsilon S \times A$ ***do***
$Q(\tilde{s}, \tilde{a}) \leftarrow Q(\tilde{s}, \tilde{a}) + \alpha . \delta . e(\tilde{s}, \tilde{a})$
$e(\tilde{s}, \tilde{a}) \leftarrow \gamma . \lambda . e(\tilde{s}, \tilde{a})$
$s \leftarrow s'$
$a \leftarrow a'$
return Q

Fig. 5. Blood glucose level control by using the eligibility traces algorithm

IV. SIMULATION

In this paper, the blood glucose level control is investigated in the presence of faults in the system. To this aim, the eligibility traces algorithm is used. All simulations are performed by MATLAB Software. Initial conditions are considered for diabetic patient are as follows: $G_0 = 6.1915\, mM, I_0 = 98.6056 PM$. The optimal value (which finally should be reached) for the blood glucose level is 5.6 mM and for the insulin level is 96 PM. The injected dosage of the drug, denoted by u^f, should be determined such that the proposed method can control the blood glucose level and preserve it at the desired level, even in the case the fault has occurred in the system.

In this paper, the proposed method is used to control the blood glucose concentration in patients with diabetes and make a comparison with the actor-critic (represented in [11] for a class of unknown nonlinear systems). At first, proposed method of [11] is simulated on the diabetes model and compared with our proposed eligibility traces based approach.

In this paper, to demonstrate a successful performance of the eligibility traces algorithm for controlling the blood glucose level, a comparison is made with the ANNs method. All mathematical relations related with the ANNs approach to control the fault are adopted from [11].

Actuator and sensor faults applied to the system are shown in Fig. 6 and 7.

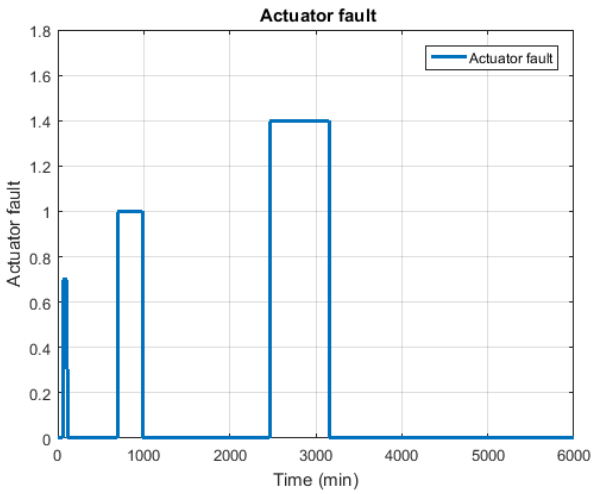

Fig. 6. Actuator fault applied to the system

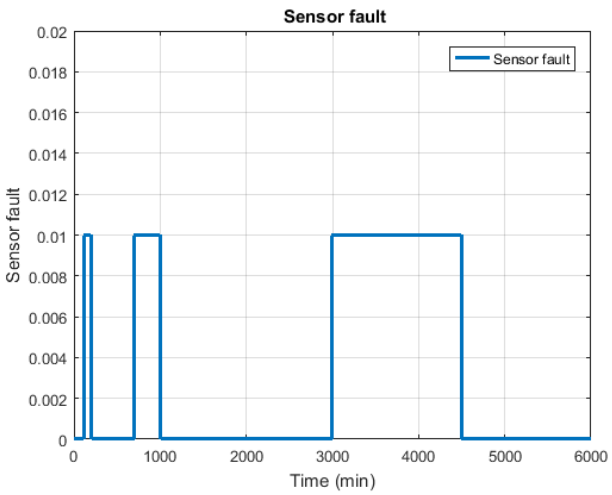

Fig. 7. Sensor fault applied to the system

In this section, the sensor fault and actuator fault are investigated separately. Fig. 8 and Fig. 9 show the blood glucose level and insulin level in diabetic patients, in the presence of the sensor fault and actuator fault, respectively. In both cases, eligibility traces algorithm, actor-critic and neural network methods are compared.

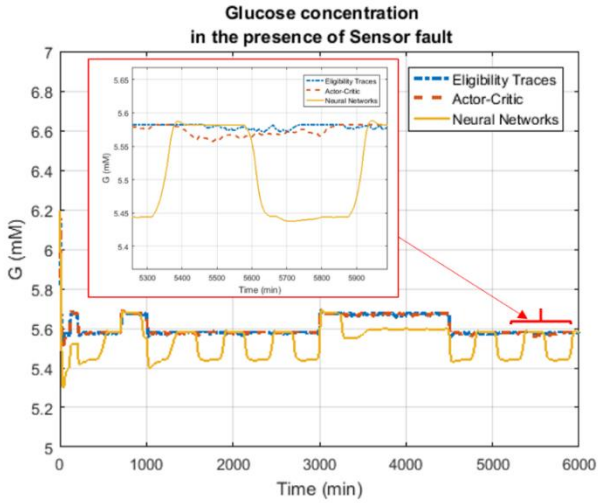

Fig. 8 (a)

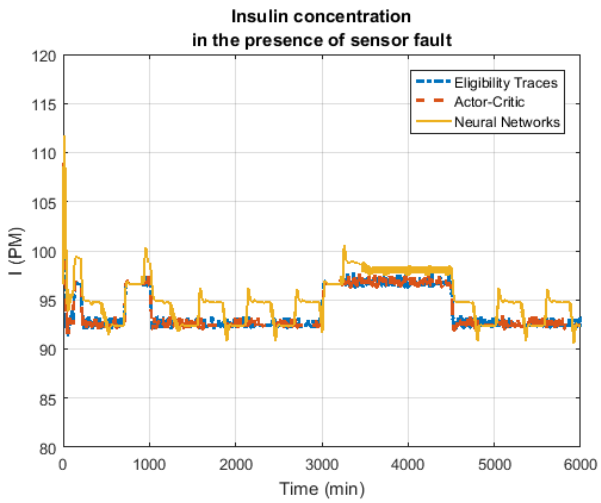

Fig. 8 (b)

Fig. 8. Simulation results in the presence of sensor fault; (a) blood glucose concentration; (b) insulin concentration

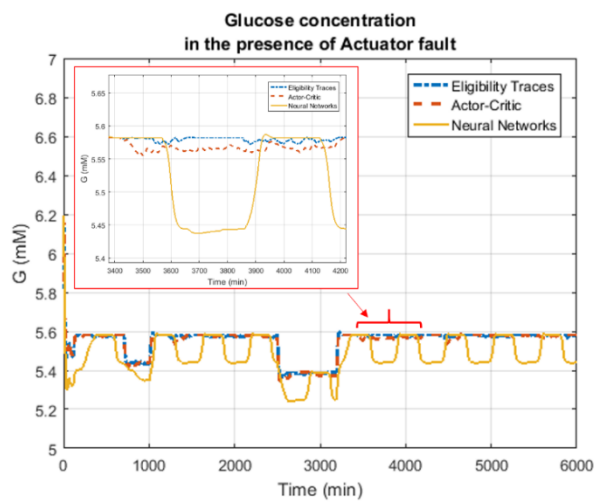

Fig. 9 (a)

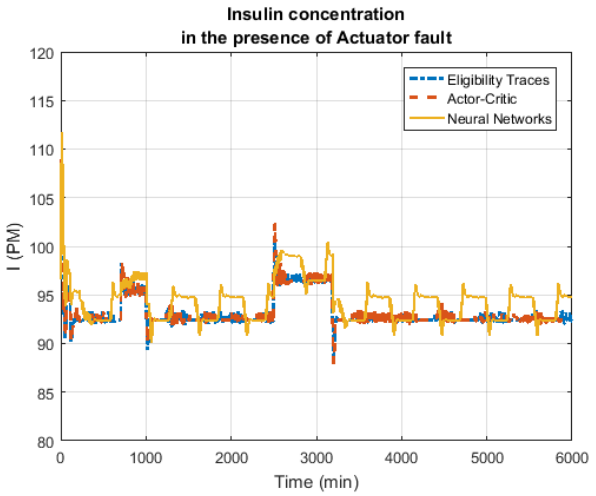

Fig. 9 (b)

Fig. 9. Simulation results in the presence of actuator fault; (a) blood glucose concentration; (b) insulin concentration

As can be observed in Fig.8, the eligibility traces algorithm can control the blood glucose level and bring it to the desired level under the sensor fault, when compared with other two methods. The proposed method uses a lower dosage of the drug and results in reduced side effects. Also, in the case the ANNs method is used, under the sensor fault, the blood glucose level will have higher variations about the desired value. This may have dangerous outcomes for diabetic patients. Also, insulin variations in the desired value is higher than two other methods, which may result in the death of the diabetic patient. According to Fig.9, under the actuator fault in the system and using the ANNs method, although a higher dosage is injected, blood glucose and insulin have higher variations about their desired values. However, when the actor-critic method is used, such variations are low. The eligibility traces algorithm can significantly control blood glucose and insulin levels and with lower variations, under the actuator fault in the system.

Fig.10 represents the blood glucose and insulin variations, for three mentioned methods, in the case both sensor and actuator faults are present in the system.

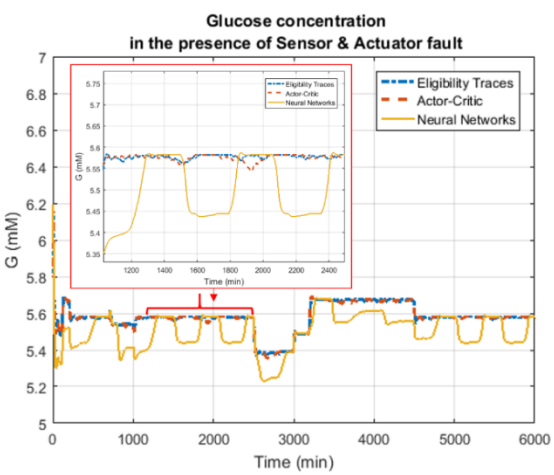

Fig. 10. (a)

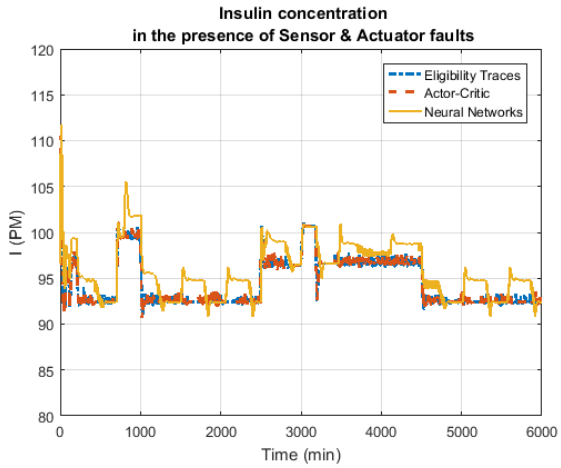

Fig. 10 (b)

Fig. 10. Simulation results in the presence of sensor-actuator fault; (a) blood glucose concentration; (b) insulin concentration

As shown in Fig.10, if both sensor and actuator faults are simultaneously applied to the system, the eligibility traces algorithm can control the blood glucose and insulin levels with a lower variations and drug dosage and keep them about the desired values, when compared with two other methods. Table II represents the injected dosage determined by the three considered methods, in the presence of sensor and actuator faults.

TABLE II

TOTAL DOSAGE INJECTED TO THE DIABETIC PATIENT, USING THE ELIGIBILITY TRACES ALGORITHM, ACTOR-CRITIC METHOD, AND NEURAL NETWORK

Control Input	Method	Applied Fault Type	Total Injected Dosage
u^f	Eligibility Traces	Sensor	19.01 mg
		Actuator	18 mg
		Sensor-Actuator	18.01 mg
	Actor-Critic	Sensor	20.48mg
		Actuator	34.05 mg
		Sensor-Actuator	25.64mg
	Neural Network	Sensor	357.8235mg
		Actuator	325.1007mg
		Sensor-Actuator	368.1947mg

As shown in Table 2, when the ANNs method is used and there is a fault in the system, the injected dosage determined by this method is higher than two other methods. This will lead to higher side effects.

Table 3, represents changes in the blood glucose and insulin concentration control using the eligibility traces algorithm as well as the actor-critic method and neural networks, in the case

V. CONCLUSIONS

In this paper, the eligibility traces algorithm is used as one of the best methods in the RL framework. On fault control, this method is compared with two previous works[11], i.e. actor-critic and ANNs. Simulation results revealed that the eligibility traces algorithm due to the fact that a backward view for updating the value of the state-action pair at the current time step can control the blood glucose level with an increased speed, lower dosage and reduced changes about the desired value and bring it to the desired value, in the case the sensor or actuator fault occurs in the system. However, the ANNs method has a lower speed and uses a higher dosage. Therefore, in the presence of faults, the proposed method shows a better performance in terms of accuracy and the drug dosage use, which implies the mitigated side effects for diabetic patients.

In order to detect the fault, a residual calculation method was used. For calculating the residual, system states were supposed to be estimated at each time step. To this aim, the RBF neural network was employed and the estimated values from the RBF neural network were compared with the real outputs of the diabetic system, so the level of residual was calculated. Results obtained from the proposed method and ANNs method [11] were compared. Based on the results, if any fault (either sensor or actuator fault) occurs in the system, the ANNs method fails to control the blood glucose and insulin levels and there will be severe variations in the desired level. Also, the injection dosage determined by the ANNs method is high, which will cause side effects for diabetic patients and in some cases, this may result in the patients' death.

In addition, determining an individual drug plan for each patient and the drug personalization concept are important issues, which have received much attention in recent years. One of the principal advantages of the Reinforcement Learning method is the fact that it takes this issue into consideration for each individual patient in the real world. In this sense, after the learning has occurred for the nominal model of the patient, the Q-table, containing the value of each state-action pair, will be used and another learning phase will be performed for each real patient. In this case, less trial and error will take place and the optimal drug plan will be specified for each person individually. In this paper, since there was no access to real patients, all simulations were performed on the nominal model. However, it is possible to determine the drug plan for a real diabetic patient, when learning is terminated and the Q-table is obtained. The proposed algorithm can be used for real patients. In future works this algorithm can be examined on other types of faults. Also, for further works, Continues RL can be implemented on this model.

TABLE III
GLUCOSE AND INSULIN VALUES AFTER EACH 1000 MINUTES

Glucose Concentration (Sensor fault)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	6.98(mM)	5.67(mM)	5.57(mM)	5.56(mM)	5.662(mM)	5.571(mM)	5.57(mM)
Actor-Critic	6.981(mM)	5.67(mM)	5.571(mM)	5.58(mM)	5.677(mM)	5.58(mM)	5.58(mM)
Neural Network	6.954(mM)	5.57(mM)	5.6(mM)	5.75(mM)	5.63(mM)	5.43(mM)	5.44(mM)
Insulin Concentration (Sensor fault)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	105(PM)	96.2(PM)	92(PM)	96.3(PM)	96(PM)	93.1(PM)	93(PM)
Actor-Critic	105(PM)	97(PM)	92.3(PM)	96.3(PM)	96.3(PM)	92.2(PM)	93(PM)
Neural Network	112(PM)	93(PM)	95(PM)	92(PM)	98(PM)	93(PM)	92(PM)
Glucose Concentration (Actuator fault)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	6.198(mM)	5.535(mM)	5.58(mM)	5.39(mM)	5.58mM)	5.58(mM)	5.57(mM)
Actor-Critic	6.199(mM)	5.536(mM)	5.581(mM)	5.38mM)	5.577(mM)	5.58(mM)	5.58(mM)
Neural Network	6.1(mM)	5.376(mM)	5.485(mM)	5.44(mM)	5.611(mM)	5.43(mM)	5.423(mM)
Insulin Concentration (Actuator fault)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	109(PM)	89(PM)	93(PM)	96(PM)	93(PM)	93.3(PM)	93(PM)
Actor-Critic	109(PM)	90(PM)	93.2(PM)	97(PM)	94(PM)	92.2(PM)	92.8(PM)
Neural Network	111(PM)	94(PM)	95(PM)	96.7(PM)	92.5(PM)	92.5(PM)	95(PM)
Glucose Concentration (Sensor and Actuator faults)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	6.198(mM)	5.5(mM)	5.58(mM)	5.39(mM)	5.68mM)	5.585(mM)	5.59(mM)
Actor-Critic	6.199(mM)	5.52(mM)	5.581(mM)	5.38mM)	5.681(mM)	5.588(mM)	5.587(mM)
Neural Network	6.198(mM)	5.325(mM)	5.6(mM)	5.4(mM)	5.65(mM)	5.6(mM)	5.6(mM)
Insulin Concentration (Actuator and Sensor faults)							
Time / Method	1 min	1000 min	2000 min	3000 min	4000 min	5000 min	6000 min
Eligibility trace	110(PM)	98(PM)	93(PM)	101(PM)	97(PM)	93.3(PM)	93(PM)
Actor-Critic	220(PM)	100(PM)	93.3(PM)	101.1(PM)	97.8(PM)	93.8(PM)	92.8(PM)
Neural Network	111(PM)	102(PM)	92.5(PM)	96.5(PM)	97(PM)	92.5(PM)	92.5(PM)

REFERENCES

[1] N. C. Van Der Ven *et al.*, "The confidence in diabetes self-care scale: psychometric properties of a new measure of diabetes-specific self-efficacy in Dutch and US patients with type 1 diabetes," *Diabetes care*, vol. 26, no. 3, pp. 713-718, 2003.

[2] A. A. Sharief and A. Sheta, "Developing a mathematical model to detect diabetes using multigene genetic programming," *IJARAI) International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 10, 2014.

[3] M. O. M. Javad, S. Agboola, K. Jethwani, I. Zeid, and S. Kamarthi, "Reinforcement Learning Algorithm for Blood Glucose Control in Diabetic Patients," in *ASME 2015 International Mechanical Engineering Congress and Exposition*, 2015: American Society of Mechanical Engineers, pp. V014T06A009-V014T06A009.

[4] I. Hochberg, G. Feraru, M. Kozdoba, S. Mannor, M. Tennenholtz, and E. Yom-Tov, "A reinforcement learning system to encourage physical activity in diabetes patients," *arXiv preprint arXiv:1605.04070*, 2016.

[5] A. Noori and M. A. Sadrnia, "Glucose level control using Temporal Difference methods," in *2017 Iranian Conference on Electrical Engineering (ICEE)*, 2017: IEEE, pp. 895-900.

[6] W.-H. Weng, M. Gao, Z. He, S. Yan, and P. Szolovits, "Representation and reinforcement learning for personalized glycemic control in septic patients," *arXiv preprint arXiv:1712.00654*, 2017.

[7] P. D. Ngo, S. Wei, A. Holubová, J. Muzik, and F. Godtliebsen, "Control of Blood Glucose for Type-1 Diabetes by Using Reinforcement Learning with Feedforward Algorithm," *Computational and mathematical methods in medicine*, vol. 2018, 2018.

[8] J. Skach, I. Punčochář, and F. L. Lewis, "Temporal-difference Q-learning in active fault diagnosis," in *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, 2016: IEEE, pp. 287-292.

[9] J. Cao, "Using reinforcement learning for agent-based network fault diagnosis system," in *2011 IEEE International Conference on Information and Automation*, 2011: IEEE, pp. 750-754.

[10] J. Škach and I. Punčochář, "Input design for fault detection using extended kalman filter and reinforcement learning," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7302-7307, 2017.

[11] F. Farivar and M. N. Ahmadabadi, "Continuous reinforcement learning to robust fault tolerant control for a class of unknown nonlinear systems," *Applied Soft Computing*, vol. 37, pp. 702-714, 2015.

[12] K.-Z. Han, J. Feng, and X. Cui, "Fault-tolerant optimised tracking control for unknown discrete-time linear systems using a combined reinforcement learning and residual compensation methodology," *International Journal of Systems Science*, vol. 48, no. 13, pp. 2811-2825, 2017.

[13] D. Zhang, Z. Lin, and Z. Gao, "Reinforcement-learning based fault-tolerant control," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, 2017: IEEE, pp. 671-676.

[14] H. H. Afshari, D. Al-Ani, and S. Habibi, "Fault Prognosis of Roller Bearings Using the Adaptive Auto-Step Reinforcement Learning Technique," in *ASME 2014 Dynamic Systems and Control Conference*, 2014: American Society of Mechanical Engineers Digital Collection.

[15] P. Herrero *et al.*, "Robust fault detection system for insulin pump therapy using continuous glucose monitoring," *Journal of diabetes science and technology*, vol. 6, no. 5, pp. 1131-1141, 2012.

[16] Z. Mahmoudi, K. Nørgaard, N. K. Poulsen, H. Madsen, and J. B. Jørgensen, "Fault and meal detection by redundant continuous glucose monitors and the unscented Kalman filter," *Biomedical Signal Processing and Control*, vol. 38, pp. 86-99, 2017.

[17] K. Turksoy, I. Hajizadeh, E. Littlejohn, and A. Cinar, "Multivariate statistical monitoring of sensor faults of a multivariable artificial pancreas," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10998-11004, 2017.

[18] K. Kölle, A. L. Fougner, K. A. F. Unstad, and Ø. Stavdahl, "Fault detection in glucose control: Is it time to move beyond CGM data?," *IFAC-PapersOnLine*, vol. 51, no. 27, pp. 180-185, 2018.

[19] X. Yu *et al.*, "Fault Detection in Continuous Glucose Monitoring Sensors for Artificial Pancreas Systems," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 714-719, 2018.

[20] I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: literature review," *Journal of medical Internet research*, vol. 20, no. 5, p. e10775, 2018.

[21] A. Khajeh and Z. Shabani, "Adaptive Gain Scheduling Control of Doubly Fed Induction Generator Based Wind Turbines to Improve Fault Ride Through Performance," *International Journal of Industrial Electronics, Control and Optimization*, vol. 1, no. 1, pp. 61-70, 2018.

[22] R. Sedaghati and M. R. Shakarami, "A New Sliding Mode-based Power Sharing Control Method for Multiple Energy Sources in the Microgrid under Different Conditions," *International Journal of Industrial Electronics, Control and Optimization*, vol. 2, no. 1, pp. 25-38, 2019.

[23] S. M. E. Oliaee, "Fault Detection and Identification of High Dimension System by GLOLIMOT," *International Journal of Industrial Electronics, Control and Optimization*, vol. 2, no. 4, pp. 331-342, 2019.

[24] S. Baniardalani, "Fault Diagnosis of Discrete-Time Linear Systems Using Continuous Time Delay Petri Nets," *International Journal of Industrial Electronics, Control and Optimization*, vol. 3, no. 1, pp. 81-90, 2020.

[25] A. Roy and R. S. Parker, "Dynamic modeling of exercise effects on plasma glucose and insulin levels," ed: SAGE Publications, 2007.

[26] P. Magni and R. Bellazzi, "A stochastic model to assess the variability of blood glucose time series in diabetic patients self-monitoring," *IEEE Transactions on biomedical engineering*, vol. 53, no. 6, pp. 977-985, 2006.

[27] A. Makroglou, J. Li, and Y. Kuang, "Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview," *Applied numerical mathematics*, vol. 56, no. 3-4, pp. 559-573, 2006.

[28] Y. C. Kueh, T. Morris, E. Borkoles, and H. Shee, "Modelling of diabetes knowledge, attitudes, self-management, and quality of life: a cross-sectional study with an Australian sample," *Health and quality of life outcomes*, vol. 13, no. 1, p. 129, 2015.

[29] F. Nani and M. Jin, "Mathematical modeling and simulations of the pathophysiology of Type-2 Diabetes Mellitus," in *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2015: IEEE, pp. 296-300.

[30] J. R. Moore and F. Adler, "Mathematical modeling of type 1 diabetes in the NOD mouse: separating incidence and age of onset," *arXiv preprint arXiv:1412.6566*, 2014.

[31] A. Mahata, S. P. Mondal, S. Alam, and B. Roy, "Mathematical model of glucose-insulin regulatory system on diabetes mellitus in fuzzy and crisp environment," *Ecological Genetics and Genomics*, vol. 2, pp. 25-34, 2017.

[32] E. Lehmann and T. Deutsch, "A physiological model of glucose-insulin interaction in type 1 diabetes mellitus," *Journal of biomedical engineering*, vol. 14, no. 3, pp. 235-242, 1992.

[33] A. Onvlee, H. Blauw, N. Middelhuis, and H. Zwart, "In silico modeling of patients with type 1 diabetes mellitus," MS thesis, Dep. Tech. Med., Univ. Twente, Enschede, Netherlands, 2016.

[34] P. Palumbo, S. Panunzi, and A. De Gaetano, "Qualitative behavior of a family of delay-differential models of the glucose-insulin system," *Discrete and Continuous Dynamical Systems Series B*, vol. 7, no. 2, p. 399, 2007.

[35] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[36] J.-J. E. Slotine and W. Li, *Applied nonlinear control* (no. 1). Prentice hall Englewood Cliffs, NJ, 1991.

Amin Noori was born in Mashhad, Iran. He received the B.SC. degree in electric engineering form the Sadjad University of Technology, Mashhad, Iran, the M.S. degree in control engineering from Ferdowsi University of Mashhad, Mashhad, Iran. Currently, He is Ph.D. candidate of control engineering at Shahrood University of Technology, Shahrood, Iran. He is a lecturer with the Faculty of electric and biomedical engineering, Sadjad University of Technology, Mashhad, Iran. He is also ahead of machine learning and Artificial Intelligence laboratory. It is research interest include reinforcement learning, Neural Networks, Fault Tolerant Control, and applications of Artificial Intelligence in biomedicine and biomedical.

Mohammad Ali Sadrnia was born in mashhad, Iran. He received his B.S. degree in Electronics Engineering from Ferdowsi University of Mashhad. He Received the M.S. and Ph.D. degree from University of Hull, United Kingdom in Control Engineering. His research interests are Robust Control, Fault Diagnosis, Fault Tolerant Control Systems, and, Flight Control.

Mohammad Bagher Naghibi-Sistani received the B.Sc. and M.Sc. (Hons.) degrees in control engineering from the University of Tehran, Tehran, Iran, in 1991 and 1995, respectively, and the Ph.D. degree from the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran, in 2005. He was a Lecturer with the Ferdowsi University of Mashhad from 2001 to 2005, where he is currently an Associate Professor. His current research interests include artificial intelligence, reinforcement learning, and control systems.