Accepted for publication in IEEE ICDH 2023, Chicago, IL, July 2-8, 2023.

Precision Dosing in Critical Care: Application of Machine Learning in Fluid Therapy

Elham Estiri
College of Aeronautics and Engineering
Kent State University
Kent, Ohio, USA
eestiri@kent.edu

Hossein Mirinejad
College of Aeronautics and Engineering
Kent State University
Kent, Ohio, USA
hmiri@kent.edu

***Abstract*—Fluid therapy is a common treatment for hypovolemic scenarios to restore the lost blood volume and stabilize acutely ill patients. Automating fluid therapy can lead to a reduction of delay in care, a decrement in dosing errors, and a reduction of cognitive load on clinicians responsible for patient care resulting in improved patient outcomes. However, this process is highly challenging due to the complexity of patient's physiology and the variability of hemodynamic responses among patients. This work presents a novel machine learning approach based on reinforcement learning (RL) for automated fluid management, where the RL agent is designed to recommend subject-specific infusion dosages without having the knowledge of dose-response models and only by interacting with the environment (virtual subject generator). Compared to the state-of-the-art focusing on the entire population's data, the proposed approach uses individual patient's data to recommend patient-specific fluid dosage adjustment. Simulation results demonstrate that the proposed approach outperforms a proportional-integral-derivative (PID) and a rule-based fluid resuscitation controller previously reported for an animal study.**

***Index Terms*—Machine learning, reinforcement learning, fluid management, automated fluid therapy, mean arterial pressure**

I. Introduction

Fluid therapy is a medical treatment for life-threatening conditions such as hemorrhage, sepsis, and third-degree burn. Automated fluid therapy systems can regulate fluid infusion dosages by targeting a hemodynamic endpoint such as blood volume [1] or mean arterial pressure (MAP) [2]. Such automated systems are capable to significantly reduce the clinician's cognitive overload in critical care and can also be employed in remote or hostile environments. e.g., combat casualty care, to potentially save lives. Additionally, automated fluid calibration has the potential to prevent over- and under-dosing, a common issue in fluid therapy, by providing an optimal infusion dosage schedule [3], [4].

Various control methodologies have been proposed for automated fluid therapy, including adaptive control [4], optimal receding horizon control [5], fuzzy logic control [6], proportional-integral-derivative (PID) control [7] and rule-based (decision table) control [8]. A review of recent advances and open challenges in fluid therapy was presented in [3]. Most of the existing fluid resuscitation controllers are population-based methods that use a one-size-fits-all model and controller that may not hold for a large group of patients. Also, the accuracy of fluid management can be significantly improved using a more accurate dosing adjustment tool.

Machine learning has been successfully applied in the healthcare domain in the development of new diagnosis and treatment methods [9]–[16]. Reinforcement learning (RL) is a machine learning paradigm concerned with training intelligent agents interacting with an environment through a series of actions aimed at maximizing cumulative rewards in the long term. RL-based controllers were previously used in medication dosing for anemia management [9], cancer therapy [10], and blood glucose control [11]. These controllers leverage the power of RL to adjust medication dosages, resulting in a more accurate and personalized treatment plans for patients with complex medical conditions.

In this work, a novel, model-free RL control approach is designed for subject-specific fluid therapy. The proposed algorithm regulates MAP as the hemodynamic endpoint for individual subjects by employing a Q-learning algorithm that enables the RL agent to learn the optimal behavior in a stochastic environment. The RL algorithm does not need a dose-response model, which sets it apart from model-based approaches whose performance depends on the accuracy of the model. Compared to existing model-free controllers, including a rule-based [8] and a PID [7] controller, the proposed approach has a higher performance in adjusting fluid dosages and controlling the physiological variable, as will be demonstrated in Sections III and IV. To the best of Authors' knowledge, this is the first attempt at applying RL to fluid therapy delivery systems.

II. Methodology

RL is a sequential decision-making algorithm with two main components including an environment that represents the system and an agent that learns how to behave optimally in the environment to maximize reward during the training process. The agent interacts with the environment by taking actions, and the environment responds with rewards or penalties, enabling the agent to learn from its actions and refine its behavior over time.

This material is based upon work supported by the National Science Foundation under Grant NO. 2138929.

Accepted for publication in IEEE ICDH 2023, Chicago, IL, July 2-8, 2023.

By formulating the problem as an MDP, we enable the agent to learn an optimal behavior using a Q-learning algorithm [12]. This algorithm allows the agent to estimate the values associated with various actions and improve its action-selection policy during the training process. Equation (1) represents the Q-learning update rule, where the agent updates the action-value function based on the received reward, the estimated future maximum value, and the learning rate and discount factor.

The RL environment can be described as a finite Markov decision process, defined as a 4-tuple, including sets of permissible states (S) and Actions (A) taken by the agent, a reward function (R) that represents reward or penalty received by the agent from the environment, and the probability of transitioning from one state to another by taking a specific action (P) [17]. By formulating the problem as an MDP, we enable the agent to learn an optimal behavior using a Q-learning algorithm [18]. This algorithm allows the agent to estimate the values associated with various actions and improve its action-selection policy during the training process. The Q-learning update rule is shown in (1), where the agent updates the action-value function based on the received reward, the estimated future maximum value, the learning rate, and discount factor.

$$Q_{new}(s_i,a_i) \leftarrow Q(s_i,a_i) + \alpha(r_i + \gamma max(s_{i+1},a_i) - Q(s_i,a_i)) \quad (1)$$

where $i = \{1,2,...,N\}$ denotes the number of current episode, r_i is the reward obtained when transitioning from state s_i to s_{i+1}, $Q(s,a)$ denotes the current value of the action a_i, $max(s_{i+1},a_i)$ provides an estimation of the optimal value in the future, and $Q_{new}(s_i,a_i)$ indicates the accumulated value obtained by selecting action a_i in state s_{i+1}. The learning rate, $\alpha \in [0,1]$ determines how much adjustment of Q-values is used based on new information. The discount factor $\gamma \in [0,1]$ reflects the importance of future rewards versus an immediate reward.

The RL algorithm proposed in this study is model free. In the absence of real subjects, a simulator was employed in this study to create virtual subjects. The virtual subject generator (simulator) has been previously verified as a digital twin of an animal study [19]. The simulator relates the change of hemodynamic response to fluid gain and loss as [19]

$$\Delta\ddot{V}_B(t) + k_p\Delta\ddot{V}_B(t) + k_i\Delta\dot{V}_B(t) = [\ddot{u}(t) - \ddot{v}(t)] + \frac{k_p}{1+\alpha_u}\dot{u} - \frac{k_p}{1+\alpha_v}\dot{v} - \frac{k_i}{1+\alpha_u}u - \frac{k_i}{1+\alpha_v}v \quad (2)$$

$$CO(t) = HR(t)\theta_1 log(\theta_2 CO(t) + \theta_3\Delta V_B(t)\theta_4) \quad (3)$$

$$MAP(t) = CO(t) \times [TPR_0 \times \frac{\Delta TPR}{2} \times \frac{sgn(MAP(t)-MAP_0)\sqrt{|MAP(t)-MAP_0|}}{\sqrt{|MAP(t)-MAP_0|}}] \quad (4)$$

TABLE I: The State Mapping Table

State Number	e(t)(mmHg)
1	$[0,5)$
2	$[5,10)$
3	$[10,15)$
4	$[15,20)$
5	$[20,25)$
6	$[25,30)$
7	$[30,35)$
8	$[35,40)$
9	$[40,45)$
10	$[45,50)$
11	$[50,\infty)$

where α_u and α_v are the steady-state ratio between the changes in intravascular and extravascular volume resulting from fluid gain and loss. The infusion and hemorrhage rates are represented by u and v, respectively, and the change in blood volume is expressed as $\Delta V_B(t)$. CO denotes the cardiac output, and θ_1, θ_2, θ_3, and θ_4 are constant values optimized for each subject. The proportional and integral gains are given by K_p and K_i, respectively. HR is the heart rate, $MAP(t)$ is the current value of MAP, MAP_0 is the nominal MAP, and TPR_0 represents the nominal total peripheral resistance (for more information about the simulator see [19]).

To ensure the optimal performance of our RL agent, we defined the action set as a predetermined range of fluid infusion dosages at a given time t. To this end, the infusion rate range of [0,100] ml/min [20] was translated to the action set A = {0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100} ml/min. Also, the environment state was represented by MAP, a readily accessible hemodynamic variable displayed on patient monitors. The environment state set was defined based on the absolute MAP error $e(t)$ as

$$e(t) = |MAP(t) - MAP_{target}| \quad (5)$$

where $MAP(t)$ and MAP_{target} denote the current and target (desired) MAP, respectively. Also the state mapping table for different ranges of $e(t)$ is shown in Table I.

The RL agent strengthens its action-selection strategy by evaluating the reward received from the environment by selecting an action a_t. The reward function plays an important role in learning process, as it provides a positive value when the error decreases and a zero value when the error increases. Therefore, the reward function was established based on a_t as

$$r_{i+1} = \begin{cases} \frac{e(i)-e(i+1)}{e(i)} & e(i+1) < e(i) \\ 0 & e(i+1) \geq e(i) \end{cases} \quad (6)$$

III. Results

The resuscitation and hemorrhage scenarios were adopted from [8] where sheep underwent medium and small hemorrhaging events. In the first 15 minutes of the experiment, 25 ml/kg hemorrhage was applied to each animal subject. Then

hemorrhage stopped for the rest of the experiment except for t = 52 min and t = 72 min where a small hemorrhage rate of 5 ml/kg was applied to each subject for 2 minutes. MAP was considered the design endpoint and measured every 5 minutes during the study. The resuscitation started with the lactated Ringer's solution after 30 minutes from the beginning of the study and continued until the end of the study (180 minutes). The simulator [19] used in our work was a digital twin model designed for the aforementioned animal study [8]. The parameter values used in the simulator for each animal subject were identified in [19]. We run our simulation in Python for 180 minutes and recorded 20,000 episodes. The RL controller parameters were set to $\alpha = 0.2$, $\lambda = 0.69$, and $\varepsilon = 0.9$. The selection of the discount factor was based on an exhaustive search and iterative refinement of the RL algorithm. Also, various values of λ were tested to assess their impact on the RL agent's learning and performance. The value of λ was selected as a balance between prioritizing recent rewards and accounting for the importance of future rewards. The RL agent was trained to reach and maintain the desired MAP level during the simulation. The initial value of MAP (MAP_0) was considered the desired MAP (MAP_{target}) for each subject. The Results of the proposed algorithm were compared against two other model-free approaches previously reported for this animal study: a rule-based (decision table) [8] and a PID controller [7]. All controllers used MAP as the hemodynamic endpoint to regulate the infusion rate.

A. The results of MAP Regulation with Different Controllers

Figs. 1 and 2 demonstrate the recommended fluid infusion dosages and MAP responses for 8 subjects using each control algorithm (RL and PID in Fig. 1, RL and rule-based in Fig. 2). The large hemorrhage scenario in the first 15 minutes caused a large drop in the MAP level at the beginning of the study. The MAP slightly recovered after hemorrhage stopped and before resuscitation started, due to the fluid shift mechanism in the body. The controller began at $t = 30$ *min* to compensate for the fluid loss and stabilize the MAP to the target level.

B. Performance Assessment

A variety of performance metrics including dynamic performances, measurement errors, and input/output responses were employed for comparison studies. These performance measures followed the IEC 60601-1-10 standard for the evaluation of physiologic closed-loop controllers [21]. The dynamic performances consisted of rise time, the time taken for the MAP response to transition from 10% to 90% of its target value, and settling time, the time required for the MAP to stabilize within a 2% error band of its final value. Measurement errors used for comparison studies were mean absolute performance error (*MAPE*), mean absolute error (*MAE*), and root-mean-square error (*RMSE*) and computed as

$$MAPE = mean(|PE_i|), \quad i = 1, ..., N \tag{7}$$

where,

$$PE_i = \frac{MAP(t) - MAP_{target}}{MAP_{target}} \times 100, i = 1, ..., N, \tag{8}$$

$$MAE = \frac{\Sigma_{i=1}^{N} |MAP(t) - MAP_{target}}{N}, \tag{9}$$

and

$$RMSE = \frac{\sqrt{\Sigma_{i=1}^{N} (MAP(t) - MAP_{target})^2}}{N}. \tag{10}$$

The input/output responses included the total amount of infusion and the final MAP response from each method. Table II demonstrates the comparison of RL, PID, and rule-based fluid resuscitation controllers with respect to the dynamic performances, measurement errors, and input/output responses. The results clearly indicate the higher performance of the RL controller compared to the other two approaches.

IV. Discussion

This study presented a model-free RL-based fluid resuscitation control algorithm for regulating MAP response in hemorrhagic scenarios. The RL framework used a Q-learning algorithm to develop an optimal drug delivery policy for fluid management by interacting with a simulator that represented the digital twin of an animal study. As indicated by different performance measures in Table II, the performance of the RL algorithm was higher than the rule-based and PID fluid resuscitation controllers previously implemented for this animal study.

Figs. 1b and 2b illustrate a large decrease in the value of MAP in the beginning of the simulation due to the large hemorrhage applied to each subject and the lack of a resuscitation strategy to compensate for the blood loss. As the controller started working at t = 30 min, the value of MAP increased significantly due to the large amount of infusion delivered by the controllers. The RL algorithm also demonstrated a faster transition to the target MAP compared to the rule-based and PID controllers in all subjects, most notably in subjects 1, 3, 4, 5, and 7. When two smaller hemorrhages occurred at t = 52 min and t = 72 min, the MAP responses of the rule-based and PID controllers sharply dropped. Conversely, the MAP responses of the RL controller were less negatively affected by these two hemorrhages, as indicated in Figs 1 and 2. This demonstrates the higher robustness of the RL algorithm against the external disturbances, compared to the other two methods.

Comparing dynamic performances of the fluid resuscitation controllers reveals that the RL-based controller provided a faster transition to the target value of MAP (Table II). Both rise time and settling time of the RL controller were smaller than corresponding values from the other two methods, indicating faster convergence of the hemodynamic response (MAP) to the steady-state level in the proposed method. In addition, the measurement errors (*MAPE*, *MAE*, and *RMSE*) of the RL controller were smaller than those from the rule-based and PID, demonstrating the higher accuracy of the proposed algorithm in the control of MAP. Further, while the final MAP

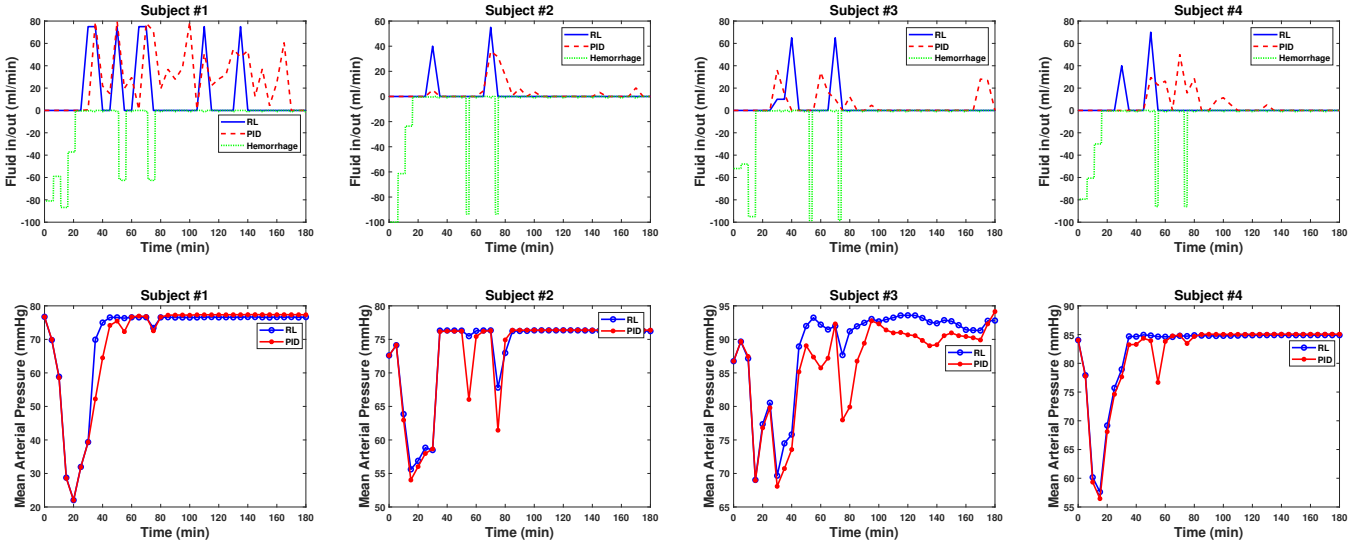

Fig. 1: (a) Fluid dose adjustments and (b) achieved mean arterial pressure levels from the RL and PID controller

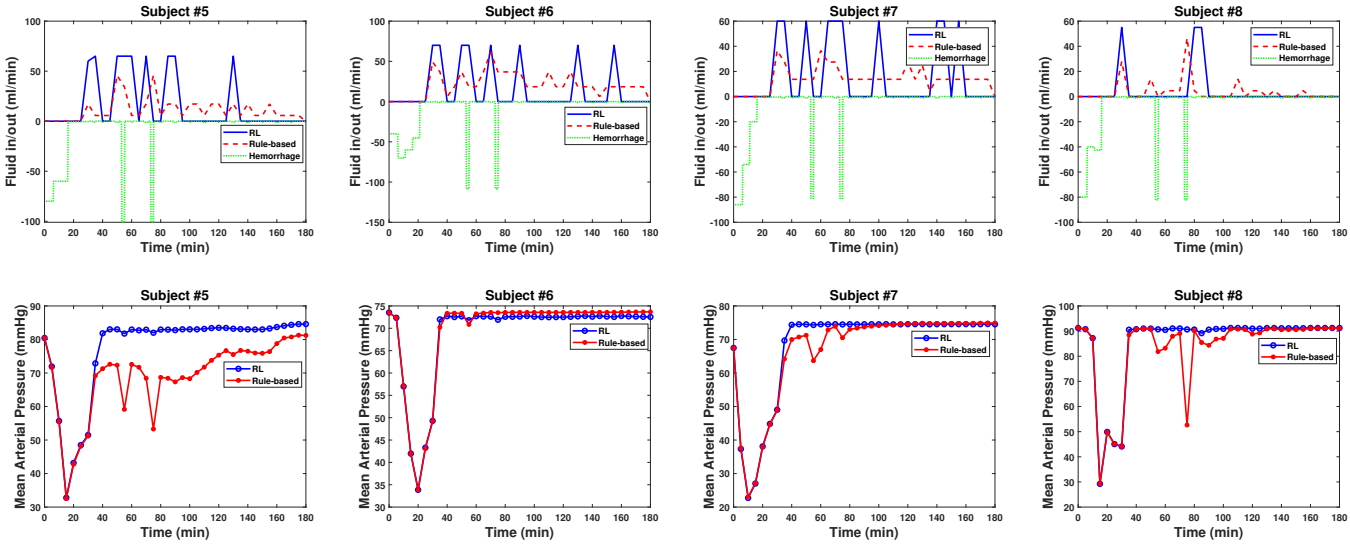

Fig. 2: (a) Fluid dose adjustments and (b) achieved mean arterial pressure levels from the RL and rule-based controller

responses of all three algorithms were similar and close to the target level, the RL controller used a lower amount of total infusion than the rule-based and PID controllers. This may suggest that the proposed approach was more effective in preventing fluid overdoing, a common issue in fluid therapy delivery.

The RL-based control algorithm was designed based on the error between the instantaneous MAP and MAP_{target} . While the desired level of MAP was considered 90 mmHg in the original animal study [8], MAP_{target} varied from subject to subject depending on the initial value of MAP (at t=0) in training of the RL method. Our evaluation of the digital twin simulator [19] showed that the final value of MAP computed by this simulator cannot exceed the initial MAP for each subject. As a result, If MAP_{target} was set to 90 mmHg for all subjects, the RL agent would have not been able to observe the desired target state for some of these subjects and would have trapped into an infinite loop preventing its training. Due to this limitation of the simulator, MAP_{target} was set to the initial MAP of each subject in training of the RL agent.

This study used MAP as the design endpoint. A multivariable control approach that considers multiple hemodynamic endpoints such as cardiac output, HR, and MAP can be examined in the future to consider the influence of multiple factors that affect the physiological state of the patient. The proposed algorithm can also be extended to assess the effect of multiple medication infusions (e.g., fluid and vasopressors) in critical care. In addition, further assessments are needed

TABLE II
COMPARISON OF RL, PID, AND RLUE-BASED FLUID RESUSCITATION CONTROLLERS

		PID	Rule-based	RL
Dynamic Performances	Rise Time (min)	16.65 ± 3.23	20.77 ± 7.45	13.06 ± 16.51
	Settling Time (min)	63.53 ± 25	82.63 ± 50.17	62.22 ± 47.86
Measurement Errors	MAPE (%)	4.61 ± 2.46	7.37 ± 5.88	4.27 ± 1.57
	MAE (%)	3.98 ± 2.37	6.25 ± 5.19	2.81 ± 1.12
	RMSE (%)	6.69 ± 3.15	9.49 ± 5.07	5.29 ± 1.79
Input/Output Responses	Total infusion (ml)	1801 ± 1502	2373 ± 1202	1612 ± 1115
	MAP Response (mmHg)	82.23 ± 6.55	80.12 ± 8.05	81.68 ± 7.77

to evaluate the robustness of the proposed algorithm against clinical disturbances.

V. CONCLUSION

The study presented a machine learning algorithm based upon a model-free Q-learning RL for fluid management in hypovolemia. The method was tested on eight animal subjects using their digital twin models. Simulation Results indicated that the proposed approach outperformed PID and rule-based fluid resuscitation controllers previously implemented for this animal study. Evaluating the proposed method in presence of clinical disturbances and extending it a multivariable approach for simultaneous control of multiple hemodynamic endpoints will be investigated in future research directions.

REFERENCES

[1] R. Bighamian, A. T. Reisner, and J. Hahn, "A lumped-parameter subject-specific model of blood volume response to fluid infusion," *Front. Physiol.*, vol. 7, p. 390, Aug. 2016.

[2] J. Rinehart et al., "Review article: Closed-loop systems in anesthesia: Is there a potential for closed-loop fluid management and hemodynamic optimization?" *Anesthesia Analgesia.*, vol. 114, pp. 130–143, July. 2012.

[3] B. Gholami, W. M. Haddad, J. M. Bailey, and W. W. Muir, 2021. "Closed-Loop Control for Fluid Resuscitation: Recent Advances and Future Challenges," *Frontiers in Veterinary Science.*, vol. 8, p. 2297-1769, Feb. 2021.

[4] X. Jin, R. Bighamian, and J. Hahn, "Development and in silico evaluation of a model-based closed-loop fluid resuscitation control algorithm," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7 pp. 1905-1914, July. 2019.

[5] J. Grant and H. Mirinejad, "An optimal control approach for automated fluid resuscitation," Presented at IEEE Intern. Conf. Biomed. Health Inform. (BHI), Virtual Conference, Jul. 2021.

[6] N. R. Marques et al., "Automated closed-loop resuscitation of multiple hemorrhages: A comparison between fuzzy logic and decision table controllers in a sheep model," *Disaster Mil. Med.*, vol. 3, no. 1, pp. 1- 10, Jan. 2017.

[7] G. C. Kramer et al., "Closed-loop control of fluid therapy for treatment of hypovolemia," *J. Trauma Acute Care Surg.*, vol. 64, no. 4, pp. S333-S341, Apr. 2008.

[8] A. D. Rafie, P. A. Rath, M. W. Michell, R. A. Kirschner, D. J. Deyo, D. S. Prough, J. J. Grady, G. C. Kramer, "Hypotensive resuscitation of multiple hemorrhages using crystalloid and colloids," *Shock.*, vol. 22, pp. 262-269, Sep. 2004.

[9] J. D. Martín-Guerrero et al., "A reinforcement learning approach for individualizing erythropoietin dosages in hemodialysis patients," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 9737-9742, Aug. 2009.

[10] R. Padmanabhan, N. Meskin, W. M. Haddad, "Reinforcement learning-based control of drug dosing with applications to anesthesia and cancer therapy," *Control Appl. Biomed. Eng. Sys.*, pp. 251-97, Jan. 2020.

[11] M. K. Bothe et al., "The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas," *Expert Rev. Med. Dev.*, vol. 10, no. 5, pp. 661-673, Jan. 2014.

[12] H. Mirinejad, A. E. Gaweda, M. E. Brier, J. M. Zurada, and T. Inanc, "Individualized drug dosing using RBF-Galerkin method: Case of anemia management in chronic kidney disease," *Comput. methods programs Biomed.*, vol. 148, pp. 45-53, 2017.

[13] H. Mirinejad, "A radial basis function method for solving optimal control problems," Ph.D. Dissertation, University of Louisville, Louisville, KY, 2016.

[14] H. Mirinejad and T. Inanc, "Individualized anemia management using a radial basis function method," in *Proc IEEE Great Lakes Biomed. Conf. (GLBC)*, Milwaukee, WI, pp. 1-4, 2015.

[15] H. Mirinejad and T. Inanc, "RBF method for optimal control of drug administration in the anemia of hemodialysis patients," in *41st Annual Northeast Biomed. Eng. Conf. (NEBEC)*, Troy, NY, pp. 1-2, 2015.

[16] S. I. Ali Meerza, A. Affan, H. Mirinejad, M. E. Brier, J. M. Zurada, and T. Inanc, "Precise warfarin management through personalized modeling and control with limited clinical data," in *43rd Annual Intern. Conf. IEEE Eng. Med. Biol. Society (EMBC)*, Mexico, 2021.

[17] M. L. Puterman, "Markov decision processes," *Handbooks in operations research and management science*, vol. 2, pp. 331–434, 1990

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

[19] R. Bighamian, B. Parvinian, C. G. Scully, G. C. Kramer, and J. Hahn, "Control-oriented physiological modeling of hemodynamic responses to blood volume perturbation," *Control Eng. Pract.*, vol. 73, pp. 149- 160, March. 2018.

[20] H. Mirinejad et al., "Evaluation of fluid resuscitation control algorithms via a hardware-in-the-loop test bed," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 471-481, Feb. 2020.

[21] International Electrotechnical Commission: IEC 60601-1-10 Medical electrical equipment - Part 1-10: General requirements for basic safety and essential performance - Collateral Standard: Requirements for the development of physiologic closed-loop controllers, 2013.