IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 May 2024, accepted 10 June 2024, date of publication 12 June 2024, date of current version 21 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3413572

RESEARCH ARTICLE

Personalized Blood Pressure Control by Machine Learning for Remote Patient Monitoring

V. T. MAI[1], KHALID A. ALATTAS[2], (Member, IEEE), YASSINE BOUTERAA[3], (Member, IEEE), EBRAHIM GHADERPOUR[4], AND ARDASHIR MOHAMMADZADEH[5]

[1]Faculty of Engineering, Dong Nai Technology University, Bien Hoa City, Vietnam
[2]Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia
[3]Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-kharj 11942, Saudi Arabia
[4]Department of Earth Sciences, Sapienza University of Rome, 00185 Rome, Italy
[5]Multidisciplinary Center for Infrastructure Engineering, Shenyang University of Technology, Shenyang 110870, China

Corresponding authors: Ebrahim Ghaderpour (ebrahim.ghaderpour@uniroma1.it), V.T. Mai (vmai_99@dntu.edu.vn), and Ardashir Mohammadzadeh (a.mzadeh@sut.edu.cn)

This work was supported in part by the Sapienza University of Rome.

ABSTRACT In the midst of a global health crisis, it is of utmost importance for healthcare technologies to possess the capability to regulate and monitor the physiological variables of patients remotely and automatically. The effective control of mean arterial pressure (MAP) in a closed-loop manner is particularly critical for individuals who are critically ill or in the process of recovering from surgical procedures. Within the framework of the present research, an adaptive closed-loop structure has been formulated with the objective of controlling a patient's MAP through governed administration of the medication sodium nitroprusside (SNP), to attain the desired MAP levels under varying conditions. The proposed closed-loop technique incorporates an intelligent controller known as the active disturbance rejection control (ADRC) with the intention of tracking the desired MAP value, alongside the utilization of continuous action policy gradient (CAPG) for the optimization of the controller's coefficients. Under the DRL strategy, an actor is responsible for generating policy requests, while a critic assesses the efficacy of the actor's policy directives. This approach uses gradient descent to train the weight values of both actor and critic networks, and it is dependent on the reward return linked to the MAP fault. Upon comparing the outcomes of the recommended structure with conventional models, numerical simulation results demonstrate the superiority of the proposed system in coping with varying working conditions, key-value fluctuations, and uncertainties, while effectively maintaining the desired mean arterial pressure and drug administration rate.

INDEX TERMS Drug infusion pump, mean arterial pressure (MAP), adaptive closed-loop strategy, active disturbance rejection control (ADRC), deep reinforcement learning (DRL).

I. INTRODUCTION

Recently, there has been notable progress in closed-loop physiological variable regulation. This century demands that medical equipment be automatically controlled. Researchers now understand how important it is to automate medical equipment in order to save time and labor in the pandemic age. Given the impracticality of individualized patient care by physicians, patients in urgent situations must have access to an autonomously managed medicine delivery system [1].

The mean arterial pressure (MAP) is an indicator of the average blood pressure (BP) during a cardiac cycle, determined by the amalgamation of cardiac output, central venous pressure, and systemic vascular resistance [2]. One of the most crucial hemodynamic factors is that it must be managed properly and within stable bounds in a variety of acute, life-threatening situations, including cardiac arrest, the delivery of anesthetic, and the recuperation period following surgery. Complications from post-operative hypertension include subendocardial ischemia, hemorrhage,

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

disruption of vascular suture lines, and cerebrovascular issues.

A fast-acting drug called sodium nitroprusside (SNP) has long been used to treat individuals with post-operatively high BP. However, because of the medication's rapid onset of effect, a patient needs a caregiver to adjust the SNP drug's infusion based on the patient's needs. To keep physiological variables within clinical limits, it is important to continuously monitor the administration of SNP drugs during and after surgery [3]. However, because minor variations can have detrimental effects on a patient receiving medical care, it is imperative to employ automated means to manage the administration of SNP medications to modulate MAP. Thus, for any patient receiving post-surgical medication infusion, such as SNP, maintaining MAP within an ideal range is a difficult challenge. In this case, maintaining MAP at the correct level may be accomplished conventionally and simply by the medical professional using manual control [4], [5], [6].

However, this practice is not recommended due to certain severe drawbacks such as time consumption, inaccurate dose management, and MAP level fluctuation. As the BP must be monitored frequently to adjust the drug infusion rate, proper feedback is necessary to keep the MAP within the intended ranges. This is because external disturbances can alter the MAP level, as can frequent changes in the patient's condition, variations in the other patient's response characteristics, and intra-patient variability. Therefore, automated drug administration management is an intriguing way to keep an eye on the patient and will drastically cut down on human labor under pressure [7], [8], [9]. The literature states that clinical experiment findings show that automated control is far safer, more effective, and superior to manual control. This topic has received increased attention to improve closed-loop medicine distribution and regulation during critical scenarios like as surgery, accidental instances, etc. [10], [11]. Figure 1 shows a schematic illustration of the closed-loop strategy for controlling MAP by administering SNP medication infusion.

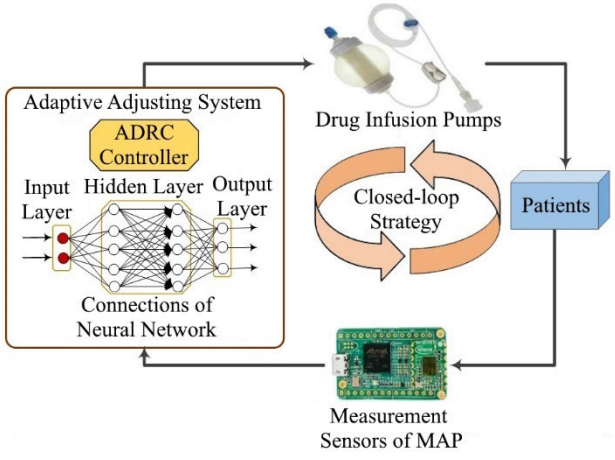


FIGURE 1. Outline of the closed-loop strategy to regulate MAP by drug infusion SNP.

Nowadays, it is simple to regulate physiological variables in real-time, such as MAP, since compact actuators and sensors are readily available to set up the closed-loop control's measurement and feedback units. The automated modulation of the MAP by SNP medication administration has been the focus of several research projects, and this might represent a bright future for the healthcare industry. Adjustable control [12], fuzzy logic control [1], [5], [13], model predictive control (MPC) [14], interval type-2 fuzzy logic control [5], fractional order control [15], robust multi-model adaptive control [16], and reinforcement learning (RL) control [17] are some of the control approaches that have been discussed. Modern methods can be more efficient, even if traditional fuzzy features or PID controllers have the benefit of being easily tunable. To find the ideal settings, these methods, however, are more intricate and demand a lot more processing power.

Recently, researchers have introduced an alternative approach called active disturbance rejection control (ADRC). The ADRC falls in between new intelligence techniques and traditional PID systems. It retains the durability and stability to compete with modern methods even though it inherits the simplicity of PID technique. Additionally, such un-modeled strategy endows the system with enhanced resilience against disturbances. By using ADRC, the system is guided toward a reinforced version that includes all unpredictability, nonlinearities, and external disruptions as enhanced states [18], [19].

Control engineers are now paying more attention to the ADRC technique because of these benefits, and they are investigating how it might be used to address a range of problems in this field. For instance, Wu and Zheng [19] recommends improving the artificial blood pumping used by individuals with end-stage congestive heart failure by applying an ADRC method. A treatment called ADRC has also been proposed to help reduce hand tremors in Parkinson's sufferers [20]. Furthermore, in [21], an approach known as ADRC was introduced to control the flow of a rotary blood pump when there are significant fluctuations in pathological state and activity. Designing a durable optimum controller is a fundamental need for autonomous control of systems. This makes it even more crucial in the healthcare business, where physiological parameter fluctuations may occur. To get a superior closed-loop response, an efficient control system needs ideal control parameters. Numerous writers have created optimization algorithms based on heuristics and metaheuristics because of advances in computing. When using such tactics, the best outcomes may be obtained for a brief amount of time, but the learning capacity from observed signals and the generalization capacity is restricted.

One type of intelligence algorithm is called deep reinforcement learning (DRL). RL and deep learning (DL) are combined in DRL. Continuous action policy gradient (CAPG) has gained popularity recently as a subset of artificial intelligence (AI) and machine learning (ML) for resolving high-dimensional problems. Through interaction and

exploration between RL and the environment, CAPG achieves the optimal objective during the training phase. DL is then used to suit the control strategy [22], [23]. The primary use of CAPG is in Markov decision process (MDP) problem-solving. The agent can provide the best strategy for choosing actions in the environment, with the ultimate objective being the maximization of reward value. CAPG requires less computing since it does not require sophisticated models, unlike traditional approaches that call for fixed structures and parameters. When the environment changes (MAP model), CAPG performs better than other approaches in making quick adjustments. Reward learning offers no chattering and a greater accuracy rate than other smart techniques [24].

The fundamental accomplishment of the current study is the creation of a regulated closed-loop medication infusion system that keeps a patient's MAP stable during treatment. This system may serve as a solid basis for the development of an affordable automatic drug administration control system. The most important thing for a critically ill, post-surgery patient undergoing heart surgery is to keep all significant physiological parameters within the specified ranges. This means that a controlled amount of medication should be infused into the patient to prevent any negative side effects. To achieve this, CAPG-based ADRC, an intelligent adaptive technique, has been used to regulate the medication infusion. The main contributions of the current study are summarized below.

1) An intelligent control strategy has been implemented to regulate MAP through the administration of a controlled infusion of SNP drugs, as a means of addressing the previously mentioned concerns.

2) An intelligence adaptive framework was employed in the development of an ADRC controller to effectively manage MAP through modulated SNP medication infusion in the present investigation. To estimate the complex and elusive dynamics of the ADRC, the ADRC controller makes use of an extended state observer (ESO). By incorporating the I/O of the MAP model, the design of the ADRC controller was accomplished without necessitating the identification of the model pertaining to the MAP system.

3) The learning capacity of CAPG is employed for the purpose of regulating MAP through engagement with the agent via the dynamic model. The Actor and Critic components of the learning control mechanism are trained in a manner devoid of any model constraints by means of incorporating a reward function as the objective for optimization.

4) The performance of the ADRC approach that has been proposed is compared to the ADRC counterpart that is based on particle swarm optimization (PSO), as well as on ADRC and traditional PID controllers. Robustness, noise rejection, and the reference model are compared at the control results and feedback points.

The current work is organized as follows: In Part II, the MAP model is expressed. Part III discusses the recommended controller's design for use with the MAP system. The stimulation findings and a comparison analysis are presented in Part IV. Parts V and VI are the discussion and conclusions, respectively.

II. FORMULATING OF MEAN ARTERIAL PRESSURE

To control the patient's desired MAP through the SNP medication infusion, this part gives the MAP model. The control goals limitations for the model under study are also discussed in the following. The overall structure of the studied model is indicated in Fig. 2. The implemented model shows how the MAP fluctuation for drug administration and SNP medicine infusion volumes relate to one another, as seen in Fig. 2 [1], [5]. The dynamic system is as follows:

$$G_p = \frac{Y_p(s)}{I_p(s)} = \frac{[S_p\left(1 + L_{p3}s\right)e^{-(\theta_p)s}]}{[\left(1 + L_{p3}s\right)\left(1 + L_{p2}s\right) - \delta_p](1 + L_{p1}s)} \quad (1)$$

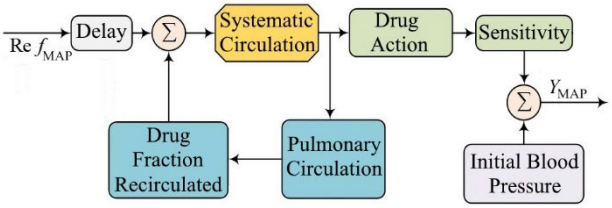


FIGURE 2. Schematic of the MAP system.

In the dynamic model, the BP changes caused by the infusion of the SNP drug are represented by $Y_p(s)$, whereas the rate at which the drug is infused is denoted by I_p. The drug infusion time constants, L_{p1}, L_{p2}, and L_{p3}, determine the speed at which the drug is absorbed into the patient's system. The parameter δ_p signifies a fraction of the recirculated SNP drug, while θ_p refers to the time-delay between the drug infusion and its impact on BP. Lastly, S_p illustrates the patient's sensitivity to the drug and its influence on their BP. Finally, the actual MAP output is formulated as [5] and [14]:

$$\text{MAP}_p\,(t) = Y_p\,(t) + I_{\text{bp}}(0) \quad (2)$$

where, $I_{\text{bp}}(0)$ is the initial BP. The dynamic model of each block and nominal values are presented in Table 1.

TABLE 1. The configurations of the dynamic system.

Coefficients	Mathematical Model	Value
Systematic Circulation	$\frac{1}{1 + L_{p2}s}$	$L_{p2} = 30s$
Drug Action	$\frac{1}{1 + L_{p1}s}$	$L_{p1} = 50s$
Pulmonary Circulation	$\frac{1}{1 + L_{p3}s}$	$L_{p3} = 10s$
Drug Fraction Recirculated	δ_p	0.5
Delay	$e^{-s\theta_p}$	$\theta_p = 60s$
Sensitivity	S_p	-0.25 mmHg
Initial Blood Pressure	I_{bp_0}	150

As was mentioned, the main goal of this experiment is to manage MAP by carefully administering the medication SNP. In the context of this investigation, the starting input value, which starts at 150 mmHg and subsequently drops to 100 mmHg must be efficiently controlled by the right dosage of SNP medication to produce a MAP that falls within the permissible range of 100 mmHg. Because of this divergence from the initial elevated MAP value of 150 mmHg, the restrictions related to the SNP medication rate of administration must be carefully considered.

To achieve the optimal outcome, it is imperative to note the constraints imposed by the administration of SNP drugs. The limitations below are enumerated [3], [5], [14]:

- **i)** At no point should oscillations occur in the final closed-loop strategy.
- **ii)** The patient's MAP should never go below the critical level or 70 mmHg.
- **iii)** The administration of a substantial quantity of SNP medication may pose a potential risk to the patient undergoing treatment or recuperation. Therefore, it is essential that the patient's medication dosage be between 0 and 180 milliliters each hour.
- **iv)** The suggested closed-loop system's settling time ought to be lower than 10 minutes and definitively below 15 minutes.
- **v)** After settling, the intended MAP should fall between [70, 120] mmHg values.
- **vi)** The ultimate steady-state MAP rate must remain within a range of ± 5 mmHg from the intended MAP level of 100 mmHg.

III. ADAPTIVE DESIGN OF DRL-BASED ADRC

A. CONCEPT OF THE ADRC CONTROLLER

The Advanced ADRC introduces a novel and intrinsically reliable controller-building component for controlling systems, which necessitates minimal knowledge about the plant. The control mechanism constantly assesses and adjusts for the entire chaos, including both external and internal dynamics, in real-time by use of an ESO. The estimated info is then eliminated in the control rule, which causes the system to behave like an ongoing integrator.

This method gets rid of the present design's predominant model need. In other words, the controller gets the data required to regulate the plant from the ESO rather than relying on the plant's model. The framework of ADRC control design is illustrated in Fig. 3, and its description is provided below. The ADRC controller, as seen in Fig. 3, consists of three components, each of which is explained in more detail below [20], [25]:

1) TRACKING DIFFERENTIAL (TD)

In MAP, the intended result is contingent upon the SNF medication infusion, which is continuously modified. Like a proportional integrator, TD can set up the perfect error transition process between the reference signal and y. This allows for easy tracking of the anticipated production.

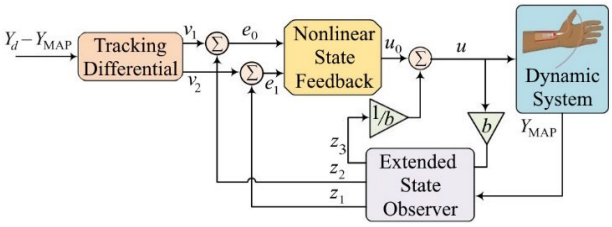


FIGURE 3. The structure of the ADRC controller.

Additionally, the supplied signal can be filtered by TD to extract an effective differential signal, either first order or higher order.

$$TD = \begin{cases} v_1 = v_1 + hv_2 \\ v_2 = v_2 + h * r * fa \end{cases} \tag{3}$$

where the tracking signal is denoted by v_1 and the deferential signal by v_2. The simulation step and the transitional trend's time are indicated by the symbols h and r, respectively, and fa is a specific function.

2) NONLINEAR STATE FEEDBACK (NLSEF)

Based on the nonlinear combination of total disturbance and predicted output error, the nonlinear feedback control rule is created. The nonlinear feedback control rule is composed similarly to differential and proportional gain. On the other hand, the control law's convergence speed is quicker than the PID controller. Effective compensation for the unknown model and disturbance may be achieved by this nonlinear combination. Furthermore, a big gain is used to boost speed when the mistake is minor, while a small gain is utilized to decrease overshoot when the error is significant. Formulating the nonlinear function is as follows:

$$\text{NLSEF} = \begin{cases} e_1 = v_1 - z_1 \\ e_2 = v_2 - z_2 \\ u_0 = \xi_1 e_1 + \xi_2 e_2 \\ u = u_0 - z_3/b \end{cases} \tag{4}$$

where the controller's parameters represented by ξ_1 and ξ_2.

3) EXTENDED STATE OBSERVER (ESO)

The ESO is used for monitoring system status and calculating system disturbance [26], [27], [28], [29].

$$\text{ESO} = \begin{cases} e = z_1 - y \\ z_1 = z_1 + h\,(z_2 - \beta_{01} e) \\ z_2 = z_2 + h\,(z_3 - \beta_{02} e + b_0 u) \\ z_3 = z_3 + h\,(-\beta_{03} e) \end{cases} \tag{5}$$

In Equation (5), the nonlinear parameters of the controller are denoted as β_{01}, β_{02}, β_{03}, and b_0. The controller yields the signal output, represented as y, while u signifies the input signal of the control objective. Both y and u serve as inputs for ESO, where z_1, z_2, and z_3 are the resulting outputs of ESO.

B. ESTIMATING AND TUNING ONLINE PARAMETERS BY DRL

A DRL algorithm is used in the study's suggested parameter estimation technique to actively estimate and modify the parameters in the compensator and ADRC controllers. An RL agent has been built in the Simulink platform with the goal of developing an adaption topology for actively predicting the adjusting coefficients of the designated controllers. The RL algorithm first engaged with the environment (MAP model) to ascertain the appropriate tuning rules. While the drug infusion is running, the controllers' gain values are continually changed using the training policy. This study uses CAPG as its RL algorithm.

The deep Q-learning network (DQN) and the DPG concept are combined in the CAPG method to provide an off-policy algorithm that can operate in continuous action space. Its goal is to provide the best possible action strategy so that the agent may achieve its goals and maximize rewards. The CAPG algorithm's capacity to function across continuous action spaces presents a significant challenge to more traditional RL techniques, such as Q-learning. The policy gradient and the value function are both integrated into the hybrid CAPG algorithm. The actor-critic framework serves as the foundation for the CAPG architecture. The policy function is referred to by the actor in the algorithm, while the value Q-function is referred to by the critic. Based on the rewards and the state that follows from the environment, the critic network assesses the actor's activities. The critic's job is to modify the actor network's weights to have the actor's future actions provide the greatest possible cumulative benefit.

The primary aim of the CAPG algorithm is to train a parameterized deterministic policy $\mu_\theta(s)$ so that, across all states accessible to the policy, the optimal policy maximizes the expected reward [27], [28], [29]. The architecture of the actor-critic network is illustrated in Fig. 4.

$$J(\theta) = \mathbb{E}_{s\sim\rho\mu}[R(s, \mu_\theta(s))] \tag{6}$$

where the distribution of all states that the policy can reach is denoted by ρ_μ. It has been demonstrated that the best course of action is always the one that maximizes returns or genuine Q-value. The initial presentation of this concept took place within the framework of dynamic programming, wherein the process of policy evaluation entails the determination of the true Q-value associated with each state-action combination prior to the selection of the action (s) possessing the highest Q-value, which subsequently leads to the modification of said policy.

$$a_t^* = \text{argmax}_a Q_\theta(s_t, a) \tag{7}$$

The action(s) having the highest Q-value are denoted by a_t^*. The activity is anticipated to produce the highest possible predicted benefit. The gradient of the goal function (a) and the gradient of the Q-value are equivalent in continuous space. A policy change $\mu_\theta(s)$ in the direction of $\nabla_\theta Q^\mu(s, a)$ results in an action with a greater Q-value and associated return if there is an estimate $Q^\mu(s, a)$ of all the values of any action (a).

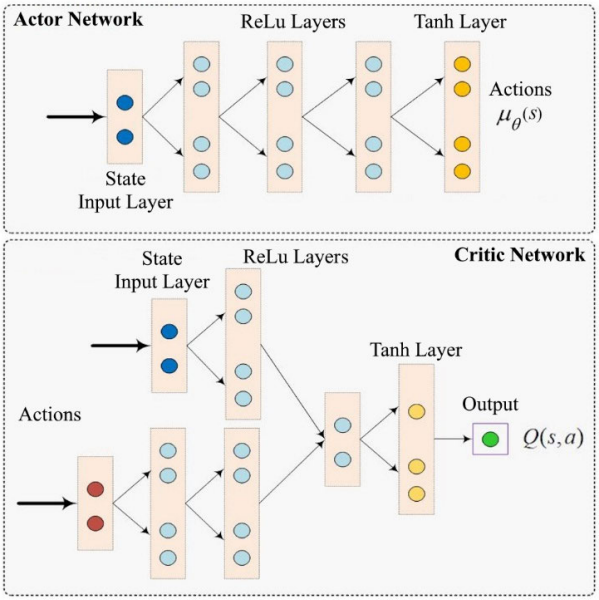


FIGURE 4. The present study's actor-critical network topology.

A gradient in relation to the Q-value's action (a) is made.

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim\rho\mu}[\nabla_\theta Q^\mu(s, a)|_{a=\mu_\theta(s)}] \tag{8}$$

Using the chain rule,

$$\frac{\partial Q_{(s,a)}}{\partial\theta} = \frac{\partial Q_{(s,a)}}{\partial\alpha} \times \frac{\partial\alpha}{\partial\theta}$$
$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim\rho\mu}\left[\nabla_\theta\mu_\theta(s) \times \nabla_a Q^\mu(s, a)|_{a=\mu_\theta(s)}\right] \tag{9}$$

As long as the function approximator $Q_\emptyset(s, a)$ is compatible, it is feasible to construct a fair estimate of the Q-value of any action, compute its gradient, and minimize the quadratic error with the true Q-values:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim\rho\mu}\left[\nabla_\theta\mu_\theta(s) \times \nabla_a Q_\emptyset(s, a)|_{a=\mu_\theta(s)}\right]$$
$$J(\emptyset) = \mathbb{E}_{s\sim\rho\mu} \tag{10}$$

The CAPG was formulated with the intention of incorporating non-linear function approximators into the DPG framework. By merging the concepts of DQN and DPG, an algorithm was developed that is capable of operating within continuous spaces and accomplishing the desired objective. In the following, a memory for retaining previous transitions and acquiring knowledge off-policy, known as an experience reply memory and target networks to stabilize learning, was added to the original DPG. CAPG uses a soft update technique, i.e., after every update of the taught network, the target networks coefficients for both networks (actor/critic) are changed [30], [31]:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{11}$$

The hyperparameter, denoted as τ, varies within the interval $[0, 1)$. The acquisition of Q-values exhibits enhanced stability due to the implementation of an update rule that guarantees the target networks consistently trail behind the

trained networks. The fundamental notion of policy gradient for the actor is derived from the deep policy gradient (DPG) approach. Target networks, alongside ordinary Q-learning, are harnessed to facilitate the learning process of the critic.

$$J(\theta) = \mathbb{E}_{s\sim\rho\mu}\left[\left(r\left(s,a,\acute{s}\right)+\gamma Q_{target}\left(\acute{s},\mu_{\acute{\theta}}\left(\acute{s}\right)\right)-Q_{\emptyset}(s,a)\right)^2\right] \tag{12}$$

Thus, based on all conceivable actions that might be taken in the next state, $Q_{\acute{\emptyset}}\left(\acute{s},\mu_{\acute{\theta}}\left(\acute{s}\right)\right)$ is the value of the action that is expected to deliver the highest total future reward. The discount factor, γ, is found in Eq. (12). The addition of noise enhances exploration.

$$a_t = \mu\left(s_t \mid \theta^{\mu}\right) + \mathcal{N} \tag{13}$$

An Ornstein–Uhlenbeck process that produces temporally correlated noise with zero mean is the source of the additive noise. The target network may be used to calculate the target value.

$$y_{RLi} = r_i + \gamma Q_{target}\left(s_{i+1}, \mu_{target}\left(s_{i+1} \mid \theta^{\mu_{target}}\right) \mid \theta^{Q_{target}}\right) \tag{14}$$

The critic is updated as a consequence of minimizing the following loss function. To update the actor, a gradient sampling policy may also be used. The framework of the CAPG strategy is indicated in Fig. 5.

$$L = \left(\frac{1}{N}\right)\sum_i \left(y_{\mathrm{RLi}} - Q(s_i, a_i|\theta^Q)\right)^2 \tag{15}$$

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q\left(s, a \mid \theta^Q\right)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}\nabla_a \tag{16}$$

Since the primary NLSEF coefficients (ξ_1 and ξ_2) are crucial for estimating the disturbance and plant condition, and these coefficients have a significant impact on the effectiveness of ADRC's control measures. The parameters are thought of as control objective parameters that need to be appropriately modified. To do this, the NLSEF parameters are adaptively constructed by:

$$\xi_1 = \xi_{01} + d\xi_1 \tag{17}$$

$$\xi_2 = \xi_{02} + d\xi_2 \tag{18}$$

where ξ_{01} and ξ_{02} are the primary values of the NLSEF coefficients and $d\xi_1$ and $d\xi_2$ are the regulatory signals that are created by the CAPG agent. This is accomplished by using the CAPG with actor-critic architecture. A reward must be established for the NNs of the CAPG tuner to guarantee that the SNP drug infusion system meets the required requirements from the perspective of biomedical engineering. Since managing the MAP is the dynamic system's control aim, this term is used in the reward signal specification, which is as follows:

$$R_t = 1/\left(Y_{\mathrm{MAP}}\right)^2 \tag{19}$$

For each instant reward of R_t, the Critic component evaluates the effectiveness of the Actor's control policy in response to immediate rewards, while the Actor component introduces corrective actions to compensate for tracking errors observed by the system. With regard to the particular BG system, the actor NN generates two control signals $\{d\xi_1, d\xi_2\}$ to modify the NLSEF parameters after receiving Y_{MAP} and $\int Y_{\mathrm{MAP}} \cdot dt$ as the system states, $s_t = \left\{Y_{\mathrm{MAP}}, \left(\frac{dY_{\mathrm{MAP}}}{d_t}\right)\right\}$. During the iteration of the CAPG agent with the MAP system, the critic NN receives $s_t = \left\{Y_{\mathrm{MAP}}, \left(\frac{dY_{\mathrm{MAP}}}{d_t}\right)\right\}$ and the regulatory signals $\{d\xi_1, d\xi_2\}$. A reward, R_t, is then issued. The MAP value will be decreased by training the CAPG agent NNs' parameter weights in this arrangement of the actor and critic networks. The structure of the tuning ADRC coefficients based on the environment (MAP model) is represented in Fig. 6.

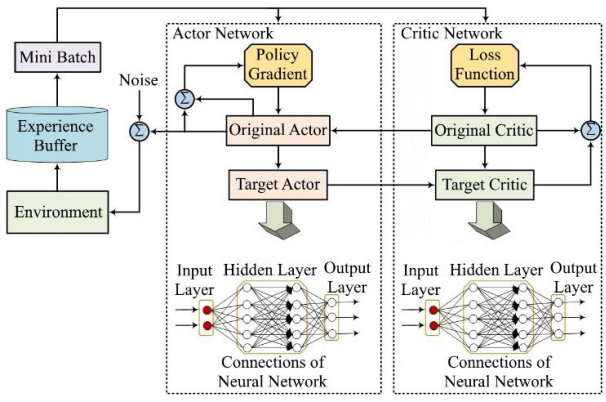


FIGURE 5. The framework of CAPG.

IV. NUMERICAL SIMULATION RESULTS

In this section, to evaluate the effectiveness of the proposed control architecture in managing the medication infusion SNP rate to regulate MAP, a dynamics model of MAP was selected (Section II) and implemented in a simulated environment as a controlled system. In the Simulink domain, two experiments were conducted to assess the efficacy of the suggested control framework, as delineated below:

1) The efficacy of the proposed closed-loop approach in a predetermined state (Reference signal),

2) The examination of the strategy's robustness in confronting changes in parameters is conducted in two distinct sections:

- Changes in the sensitivity parameter (S_p).
- Changes in the input signal.

Furthermore, to indicate the effectiveness of the designed strategy compared to the conventional controllers and meta-heuristic algorithms like ADRC-based PSO, ADRC, SMC, and PI. For the fair comparative analysis, the initial coefficients of the controllers are chosen the same. Additionally, to facilitate the training of the neural networks, various

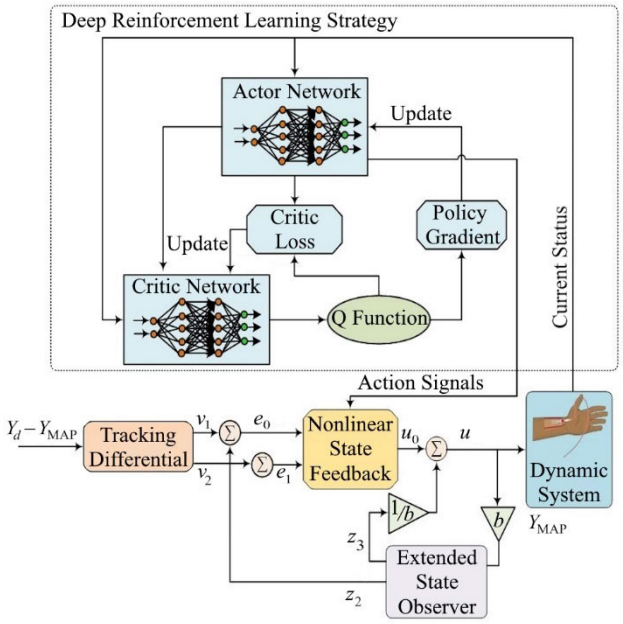


FIGURE 6. The process of tuning controllers' coefficients by CAPG.

numerical settings for hyperparameters were utilized with the intention of determining the most optimal values. The configurations of the controllers and hyperparameter of the neural network are furnished in Table 2.

TABLE 2. The configurations of controllers and CAPG hyper parameters.

ADRC Coefficients			
h	0.01	β_{02}	23
β_{01}	85	β_{03}	10
ξ_{01}	0.1	ξ_{02}	0.1
b	105	r	100
SMC Coefficients			
δ_1	0.06	δ_2	0.006
PI Coefficients			
P	0.06	I	0.0065
CAPG hyperparameters			
Learning rate, λ	0.0001	Discount factor, γ	0.99
Minibatch Size	64	Target smooth factor	0.001
Soft target update, τ	0.01	Replay buffer size	20000

A. FIRST EXPERIMENT

In the first test, the reference signal input (beginning value = 150 mmHg, time delay = 60 s, and ending value = 100 mmHg) is used to assess the effectiveness of the closed-loop method. Figure 7 shows the strategy's results in comparison to the traditional approaches for controlling MAP rate, drug infusion SNP rate, and error (error = Y_{Ref} - Y_{MAP}). The suggested closed-loop system (ADRC-based CAPG) has better output than other conventional controllers and optimization algorithms, as can be shown in Fig. 7.

Furthermore, it is evident that the suggested approach, which has no overshot (overshot = 0), reaches the

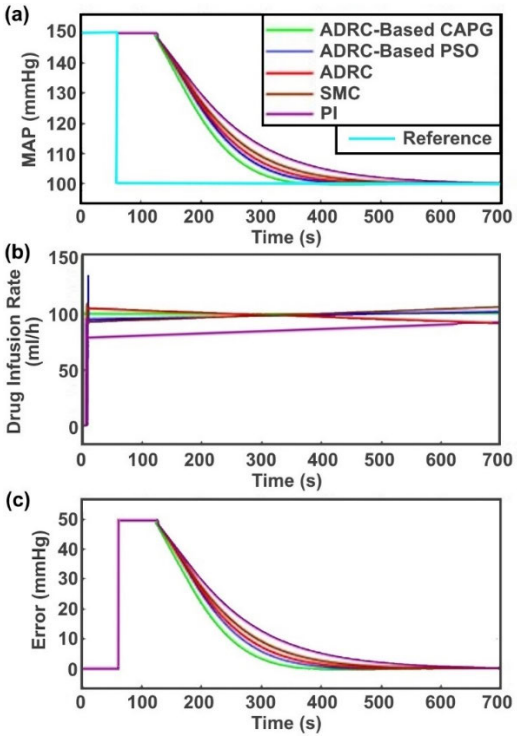


FIGURE 7. Outcomes of the controllers, a) MAP, b) drug infusion SNP rate, c) Error.

MAP's acceptable zone after around 268 seconds (settling time = 268). It is evident from Fig. 7(b) that, over the course of 700 seconds, the medication infusion SNP rate is steadier than with the other approaches. It should be noted that while the other approaches are able to regulate the MAP rate with longer settling times, the SNP rate is not stable while the MAP is being controlled. Lastly, Fig. 7(c) shows the controllers' mistakes during the MAP regulation. The suggested method, as shown in Fig. 7(c), has a smaller error in the early stages, reaches zero error faster, and is stable during the controlled period.

B. SECOND EXPERIMENT

Because every patient is unique, the MAP's dynamic system may vary, and some parameters like sensitivity and reference signal are set in accordance with that fact. Furthermore, as the recommended approach is made to function remotely, the robustness test is required to demonstrate how well the system handles fluctuations in parameters. To achieve this, the sensitivity parameter (S_p) is changed in six parts: 0.2, 0.21, 0.22, 0.23, 0.24, 0.25, and Fig. 8 shows the corresponding system results.

Figure 8 illustrates the reciprocal relationship between the sensitivity gain and MAP. This indicates that the MAP's settling time increases as sensitivity decreases. Also, the reference input signal is modified in five parts: 130, 140, 160, 170, 180 in the context of robustness. The effectiveness of the method in the face of fluctuating reference input shows that the recommended structure tends to the desired rate instantly

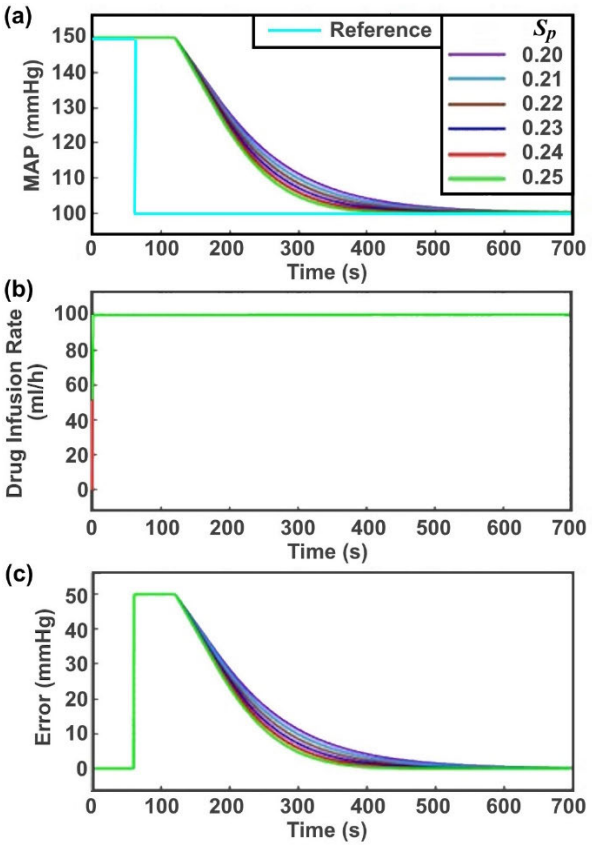


FIGURE 8. The results of the ADRC-based CAPG in dealing with changing of S_p, a) MAP, b) Drug infusion rate, c) Error.

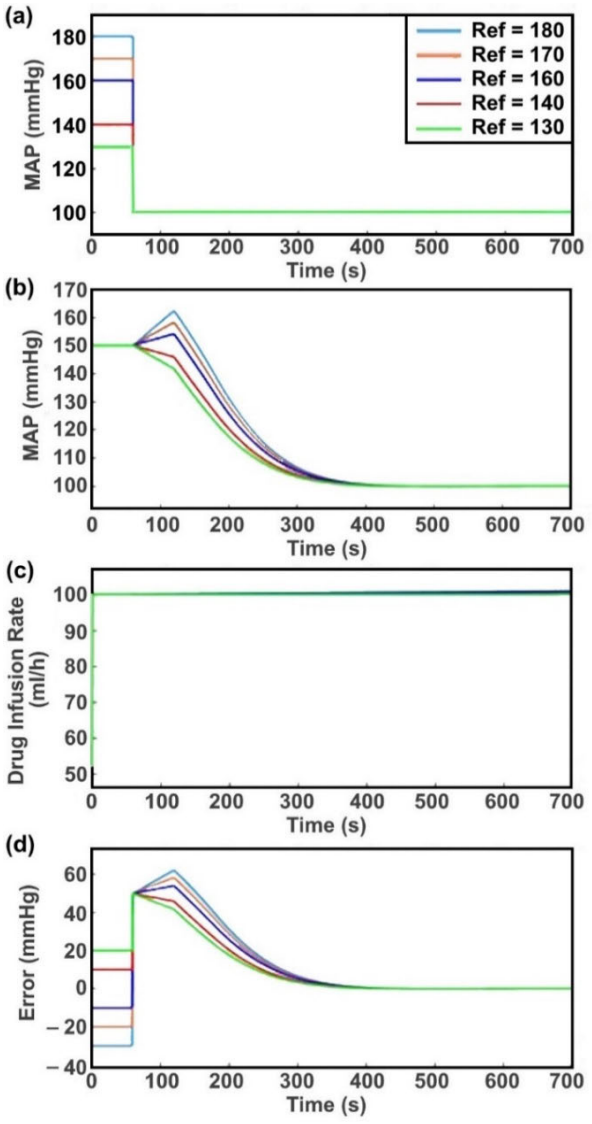


FIGURE 9. The results of ADRC-based CAPG in dealing with changing of reference, a) Input signals, b) MAP, c) Drug infusion rate, d) Error.

and without variation. The system's outputs demonstrate how stable and resilient the system is to changes in its many parameters. Figure 9 shows the result of the technique against variations in the reference input signal. After proving the efficiency of the system in dealing with coefficients variations, now, the performance of the purposed architecture based on $\int_0^\infty t \cdot \text{MAP}^2 dt$ compared to the other methodologies is presented in Fig. 10 and Table 3.

V. DISCUSSION

Amidst the ongoing worldwide health crisis, there exists a heightened necessity for cutting-edge healthcare technologies capable of remotely monitoring and managing patients' physiological parameters. The precise modulation of MAP in a closed-loop fashion holds particular significance for individuals in critical condition or those recuperating from surgical procedures. Within this investigation, a responsive closed-loop mechanism was suggested to oversee the regulation of a patient's MAP by means of the controlled dispensation of SNP, thus attaining the specified MAP levels across varying circumstances.

An ADRC, which functions as an intelligent mechanism to achieve the target MAP value, is included in the closed-loop design of the proposed system. Additionally, CAPG is employed in the optimization of the controller's coefficients. The foundation of this methodology is the use of a DRL technique, which involves continuously adjusting a policy gradient to improve system performance. During training, gradient descent is used to update the actor and critic networks' weight parameters based on the reward signals connected to the MAP error. The key component is the reward function, linked to the MAP fault. The aim is minimizing the MAP error, and the reward signal guides the optimization of the controller parameters to achieve the desired MAP levels.

To mitigate the limitation of overreliance on programming frameworks and mathematical models for verifying the accuracy and robustness of adaptive techniques, it is crucial to conduct thorough experimental validation initially, followed by implementation in real clinical environments. Furthermore, performing rigorous assessment and sensitivity analysis can assist in identifying potential weaknesses and optimizing the operational performance of the system across various scenarios. Given the theoretical nature of the present investigation, robustness analysis serves as a validation method to demonstrate how the system operates

TABLE 3. The performance index results of the methodologies against parameters variations.

Variations	Controllers														
	ADRC-CAPG			ADRC-PSO			ADRC			SMC			PI		
	MAP	Error	Drug	MAP	Error	Drug	MAP	Error	Drug	MAP	Error	Drug	MAP	Error	Drug
$S_p = 0.2$	2.82	4.03	2.44	2.88	4.64	2.32	2.91	4.87	2.28	2.96	5.33	2.17	3.2	6.09	1.85
$S_p = 0.21$	2.79	3.5	2.43	2.85	4.37	2.38	2.87	4.59	2.21	2.92	5	2.26	3.02	5.87	1.68
$S_p = 0.22$	2.76	3.31	2.42	2.82	4.14	2.43	2.84	4.33	2.39	2.89	4.71	2.55	2.98	5.51	1.52
$S_p = 0.23$	2.74	3.05	2.44	2.79	3.94	2.52	2.81	4.12	2.06	2.86	4.46	1.96	2.95	5.2	2.1
$S_p = 0.24$	2.72	2.91	2.43	2.77	3.76	2.23	2.79	3.92	1.96	2.83	4.24	2.08	2.92	4.92	2
$S_p = 0.25$	2.7	2.28	2.44	2.74	3.6	2.05	2.77	3.75	1.86	2.8	4.05	1.95	2.89	4.68	2.2
Ref = 130	2.65	2.35	2.44	2.69	2.65	2.4	2.71	2.79	2.26	2.75	3.06	2.59	2.82	3.64	1.87
Ref = 140	2.67	2.77	2.43	2.72	3.08	2.29	2.74	3.23	2.17	2.78	3.51	2.38	2.85	4.12	1.86
Ref = 160	2.73	3.88	2.44	2.77	4.2	2.08	2.79	4.36	2.29	2.83	4.66	2.46	2.92	5.3	1.95
Ref = 170	2.75	4.58	2.42	2.8	4.9	1.65	2.82	5.05	2.01	2.87	5.35	1.95	2.95	6.01	1.73
Ref = 180	2.78	5.36	2.45	2.82	5.67	2.38	2.85	5.82	1.82	2.9	6.13	2.14	2.98	6.79	1.62

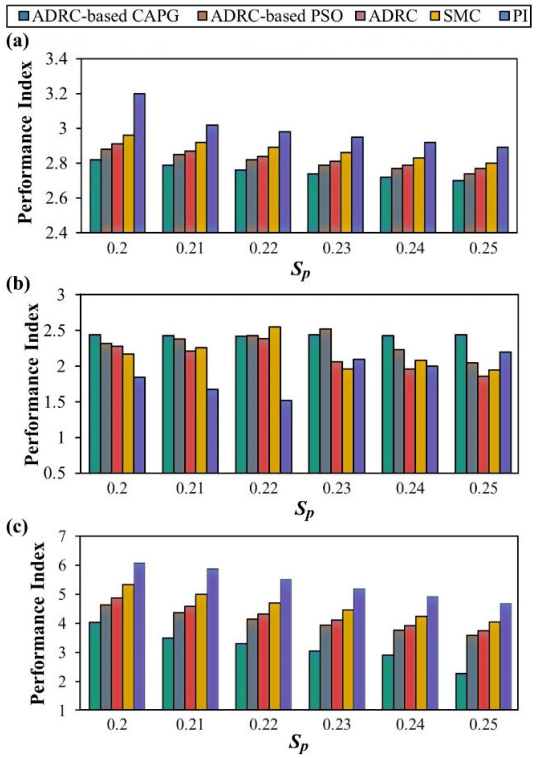


FIGURE 10. Performance Index of the various methods against changing S_p, a) MAP, b) Drug Infusion rate, c) Error.

in enhancing output and self-updating functionality across diverse conditions.

The findings of the numerical simulations indicate that the adaptive closed-loop system exhibits superior performance compared to traditional models in handling dynamic variations in working environments, fluctuations in key values, and uncertainties, while ensuring the maintenance of the desired MAP and rate of drug administration. A comparison is presented in Table 4, showing the key differences between the proposed system and previous studies in the domain of closed-loop MAP regulation.

TABLE 4. Comparison results of the previous research.

Paper	Control Strategy	Optimization Method	Setling time	Over shot
Su et al. [14]	Model predictive control (MPC)	Particle swarm optimization (PSO)	-	-
Ahmed & Özbay [3]	Robust control	Hysteresis switching	-	-
Sharma et al. [32]	IT2-FLC-PID	Cuckoo search algorithm	350	0
Sharma & Kumar [1]	Interval Type-2 Fuzzy Logic Control	Grey wolf optimization (GWO)	277	0
Kumar & Raj [5]	Fractional order two layers fuzzy logic controller	Grey wolf optimization (GWO)	274	0
Current research	ADRC	Deep Reinforcement Learning	268	0

VI. CONCLUSION

In this scholarly article, the functioning of an active disturbance rejection control (ADRC) is thoroughly investigated to effectively regulate the mean arterial pressure (MAP) by employing sodium nitroprusside (SNP) drug infusion in a

closed-loop fashion. The distinguishing characteristic of this method lies in its utilization of a pre-compensator, known as the extended state observer (ESO), to counterbalance the impact of external uncertainties and disturbances. By carefully providing the prescribed dosage of the SNP medication, the suggested framework can ensure that the MAP stays at its predetermined rate of 100 mmHg throughout surgical operations, post-surgery recovery, or anesthetic delivery. The control parameters necessary for designing an efficient control system are derived using the highly effective and relatively novel method known as continuous action policy gradient (CAPG). Two experiments are used to assess how well the suggested technique performs: **i)** utilizing a reference input signal, and **ii)** conducting robustness analysis by varying key parameters such as S_p and the reference input signal. The merits of the presented closed-loop system can be listed as follows: (1) Unlike traditional controllers like PID, which require the system model to be designed, the ADRC controller does not. Additionally, ADRC employs an ESO to manage external uncertainties and disruptions; (2) Unlike heuristic-based controllers like GA and PSO, which are only optimal for a specific operative state, CAPG-developed controllers are adaptive throughout a broad variety of closed-loop procedures; (3) Since they need q-values for every possible state and action pair, typical RL algorithms are ill-suited to handle systems with uncertainties or continuous state space. However, continuous problems may be efficiently addressed by merging RL with deep neural networks (DNN). Finally, the efficacy of the closed-loop structure is compared with that of traditional controllers and meta-heuristic approaches, including ADRC-based particle swarm optimization (PSO), ADRC, sliding mode control (SMC), and proportional-integral (PI) controller. The proposed ADRC-based CAPG method is found to surpass its conventional counterparts. Consequently, this research offers an appropriate alternative for medical professionals to save time and manpower, while simultaneously providing timely and relevant assistance to critically ill patients. It is important to consider the testing platform's constraints for flexible closed-loop techniques. The selection of computational models for mean arterial BP has been conducted with great care to ensure alignment with the characteristics of measured physiological data. It is crucial to recognize that these models might not accurately represent the dynamics of a real patient.

Recognizing the limitations of these models is crucial. These models enable a more realistic portrayal of the variations in amplitude and drug rate concentration over time that are seen in actual patients by incorporating a dynamic input signal into the closed-loop mechanism. These models are useful for evaluating the accuracy and robustness of adaptive algorithms, but maybe they are not very useful for proving how beneficial therapeutic interventions are. However, the initial stage of verifying and certifying the precision and robustness of adaptive methods heavily relies on the programming framework and related mathematical models. It will take more work to build on this foundational study. More advanced models may eventually be able to include real patient data (physiological and sensitivity data) and make more accurate predictions about how the patient will react to treatment. The incorporation of a significant amount of virtual patient models can accelerate the procedure of authenticating and ascertaining adaptable closed-loop structures and treatments.

REFERENCES

[1] R. Sharma and A. Kumar, "Optimal interval type-2 fuzzy logic control based closed-loop regulation of mean arterial blood pressure using the controlled drug administration," *IEEE Sensors J.*, vol. 22, no. 7, pp. 7195–7207, Apr. 2022.

[2] S. J. Silva, T. A. Scardovelli, H. N. Martucci, S. R. M. S. Boschi, and A. P. Silva, "Mean arterial pressure-embedded model for real-time simulation and controller validation," *Electron. Lett.*, vol. 53, no. 7, pp. 450–452, Mar. 2017.

[3] S. Ahmed and H. Özbay, "Design of a switched robust control scheme for drug delivery in blood pressure regulation," *IFAC-PapersOnLine*, vol. 49, no. 10, pp. 252–257, 2016.

[4] C. W. Frei, M. Derighetti, M. Morari, A. H. Glattfelder, and A. M. Zbinden, "Improving regulation of mean arterial blood pressure during anesthesia through estimates of surgery effects," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 11, pp. 1456–1464, 2000.

[5] A. Kumar and R. Raj, "Design of a fractional order two layer fuzzy logic controller for drug delivery to regulate blood pressure," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 104024.

[6] D. Liu, X. Liu, Z. Chen, Z. Zuo, X. Tang, Q. Huang, and T. Arai, "Magnetically driven soft continuum microrobot for intravascular operations in microscale," *Cyborg Bionic Syst.*, vol. 2022, Jan. 2022, Art. no. 9850832.

[7] S. Isaka and A. V. Sebald, "Control strategies for arterial blood pressure regulation," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 4, pp. 353–363, Apr. 1993.

[8] P. L. Mage, B. S. Ferguson, D. Maliniak, K. L. Ploense, T. E. Kippin, and H. T. Soh, "Closed-loop control of circulating drug levels in live animals," *Nature Biomed. Eng.*, vol. 1, no. 5, p. 0070, May 2017.

[9] H.-W. Huang, J. Chen, P. R. Chai, C. Ehmke, P. Rupp, F. Z. Dadabhoy, A. Feng, C. Li, A. J. Thomas, M. da Silva, E. W. Boyer, and G. Traverso, "Mobile robotic platform for contactless vital sign monitoring," *Cyborg Bionic Syst.*, vol. 2022, Jan. 2022, Art. no. 9780497, doi: 10.34133/2022/9780497.

[10] Y. Gao and M. J. Er, "An intelligent adaptive control scheme for postsurgical blood pressure regulation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 475–483, Mar. 2005.

[11] Y. Kuroda, Y. Yamanoi, S. Togo, Y. Jiang, and H. Yokoi, "Coevolution of myoelectric hand control under the tactile interaction among fingers and objects," *Cyborg Bionic Syst.*, vol. 2022, Jan. 2022, Art. no. 9861875, doi: 10.34133/2022/9861875.

[12] J. M. Arnsparger, B. C. Mcinnis, J. R. Glover, and N. A. Normann, "Adaptive control of blood pressure," *IEEE Trans. Biomed. Eng.*, vol. BME-30, no. 3, pp. 168–176, Mar. 1983.

[13] C.-T. Chen, W.-L. Lin, T.-S. Kuo, and C.-Y. Wang, "Adaptive control of arterial blood pressure with a learning controller based on multilayer neural networks," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, pp. 601–609, Jul. 1997.

[14] T.-J. Su, S.-M. Wang, H.-Q. Vu, J.-J. Jou, and C.-K. Sun, "Mean arterial pressure control system using model predictive control and particle swarm optimization," *Microsyst. Technol.*, vol. 24, no. 1, pp. 147–153, Jan. 2018.

[15] S. Sondhi and Y. V. Hote, "Fractional-order PI controller with specific gain–phase margin for MABP control," *IETE J. Res.*, vol. 61, no. 2, pp. 142–153, Mar. 2015.

[16] N. Malagutti, A. Dehghani, and R. A. Kennedy, "Robust control design for automatic regulation of blood pressure," *IET Control Theory Appl.*, vol. 7, no. 3, pp. 387–396, Feb. 2013.

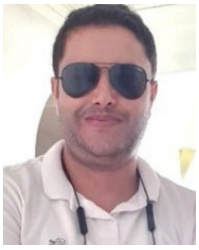
[17] R. Padmanabhan, N. Meskin, and W. M. Haddad, "Closed-loop control of anesthesia and mean arterial pressure using reinforcement learning," *Biomed. Signal Process. Control*, vol. 22, pp. 54–64, Sep. 2015.

[18] M. Sun, W. Zhang, Y. Zhang, T. Luan, X. Yuan, and X. Li, "An anti-rolling control method of rudder fin system based on ADRC decoupling and DDPG parameter adjustment," *Ocean Eng.*, vol. 278, Jun. 2023, Art. no. 114306.

[19] Y. Wu and Q. Zheng, "ADRC or adaptive controller—A simulation study on artificial blood pump," *Comput. Biol. Med.*, vol. 66, pp. 135–143, Nov. 2015.
[20] B. Faraji, M. Gheisarnejad, M. Yalsavar, and M.-H. Khooban, "An adaptive ADRC control for Parkinson's patients using machine learning," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8670–8678, Mar. 2021.
[21] Y. Wu, Q. Zheng, and E. Lim, "Preload based ADRC physiological controller for rotary blood pump," in *Proc. 35th Chin. Control Conf. (CCC)*, Jul. 2016, pp. 9428–9433.
[22] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1329–1338.
[23] B. Faraji, K. Rouhollahi, S. Mollahoseini Paghaleh, M. Gheisarnejad, and M.-H. Khooban, "Adaptive multi symptoms control of Parkinson's disease by deep reinforcement learning," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104410.
[24] M. Gheisarnejad, H. Farsizadeh, M.-R. Tavana, and M. H. Khooban, "A novel deep learning controller for DC–DC buck–boost converters in wireless power transfer feeding CPLs," *IEEE Trans. Ind. Electron.*, vol. 68, no. 7, pp. 6379–6384, Jul. 2021.
[25] J. Heidary, M. Gheisarnejad, and M. H. Khooban, "Stability enhancement and energy management of AC-DC microgrid based on active disturbance rejection control," *Electric Power Syst. Res.*, vol. 217, Apr. 2023, Art. no. 109105.
[26] Y. Zheng, Z. Chen, Z. Huang, M. Sun, and Q. Sun, "Active disturbance rejection controller for multi-area interconnected power system based on reinforcement learning," *Neurocomputing*, vol. 425, pp. 149–159, Feb. 2021.
[27] B. Faraji, M. Gheisarnejad, K. Rouhollahi, Z. Esfahani, and M. H. Khooban, "Machine learning approach based on ultra-local model control for treating cancer pain," *IEEE Sensors J.*, vol. 21, no. 6, pp. 8245–8252, Mar. 2021.
[28] M. H. Khooban and M. Gheisarnejad, "A novel deep reinforcement learning controller based type-II fuzzy system: Frequency regulation in microgrids," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 4, pp. 689–699, Aug. 2021.
[29] A. Barzegar and D.-J. Lee, "Deep reinforcement learning-based adaptive controller for trajectory tracking and altitude control of an aerial robot," *Appl. Sci.*, vol. 12, no. 9, p. 4764, May 2022.
[30] M. Gheisarnejad and M. H. Khooban, "IoT-based DC/DC deep learning power converter control: Real-time implementation," *IEEE Trans. Power Electron.*, vol. 35, no. 12, pp. 13621–13630, Dec. 2020.
[31] B. Faraji and D. Khezri, "Ultra-local model control of Parkinson's patients based on machine learning," *J. Adv. Sport Technol.*, vol. 5, no. 1, pp. 1–16, 2021.
[32] R. Sharma, K. K. Deepak, P. Gaur, and D. Joshi, "An optimal interval type-2 fuzzy logic control based closed-loop drug administration to regulate the mean arterial blood pressure," *Comput. Methods Programs Biomed.*, vol. 185, Mar. 2020, Art. no. 105167.

V. T. MAI received the Ph.D. degree from Korea Maritime University. He is currently a Lecturer with the Department of Engineering, Dong Nai Technology University, Vietnam. His research interests include robotics and control, modeling systems, artificial intelligence, and computer science. He serves as a reviewer for many international journals.

KHALID A. ALATTAS (Member, IEEE) received the B.Sc. degree in computer science from King Abdulaziz University, Saudi Arabia, the M.Sc. degree in telecommunication networks from New York University, New York, NY, USA, and the M.Sc. and Ph.D. degrees in computer science from the University of Louisiana, Lafayette, LA, USA. He is currently an Assistant Professor with the College of Computer Science and Engineering, University of Jeddah, Saudi Arabia. His research interests include networks, machine learning, data analytics, robotics, and unmanned vehicles. He serves as a reviewer for many international journals.



YASSINE BOUTERAA (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Orléans, France, in February 2012. He is currently a Full Professor with Prince Sattam Bin Abdulaziz University, Saudi Arabia. He is the author/coauthor of more than 80 scientific papers and five patents published by INNORPI (WIPO). His research interests include concerns with robotics and control systems.

EBRAHIM GHADERPOUR received the first Ph.D. degree in theoretical and computational and the second Ph.D. degree in remote sensing from Canada, in 2013 and 2018, respectively. He is currently an Assistant Professor with the Department of Earth Sciences, Sapienza University of Rome, Italy. He is also the CEO of Earth and Space Inc., Calgary, AB, Canada. His research interests include big data analytics and artificial intelligence with their applications in remote sensing, geology, geosciences, robotics, and medicine. He is an academic editor and a reviewer of many journals and publishers.

ARDASHIR MOHAMMADZADEH received the B.Sc. degree from the Sahand University of Technology, Tabriz, Iran, in July 2011, the M.Sc. degree from the K. N Toosi University of Technology, Tehran, Iran, in September 2013, and the Ph.D. degree from the University of Tabriz, Tabriz, in November 2016. Since December 2017, he has been an Assistant/Associate Professor with the University of Bonab, Bonab, Iran. In 2023, he joined Shenyang University of Technology, Shenyang, China, as a Professor. He has published many papers in many reputed journals. His research interests include control theory, fuzzy logic systems, machine learning, neural networks, intelligent control, electric vehicles, power systems control, chaotic systems, and medical systems. He is an academic editor and a reviewer of several journals.

Open Access funding provided by 'Università degli Studi di Roma "La Sapienza"' within the CRUI CARE Agreement