Artificial Intelligence In Medicine 128 (2022) 102308

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Intelligent and strong robust CVS-LVAD control based on soft-actor-critic algorithm☆

Te Li [a], Wenbo Cui [b], Nan Xie [b], Heng Li [b], Haibo Liu [a], Xu Li [a], Yongqing Wang [a,*]

[a] Key Laboratory for Precision and Nontraditional Machining Technology of Ministry of Education, Dalian University of Technology, Dalian 116024, China
[b] Graduate School of Mechanical Engineering, Dalian University of Technology, China

ARTICLE INFO

Keywords:
Deep reinforcement learning
Heart failure
Left ventricular assist devices
Physiological control

ABSTRACT

Left ventricular assist device (LVAD) is an effective method to treat ventricular failure. According to the physiological conditions of different patients, the device adaptively adjusts its rotation speed to change LVAD output. In this study, a physiological control system for LVAD based on deep reinforcement learning (DRL) is proposed. The system estimates the amount of blood required by LVAD based on a Starling-like method. The DRL controller regulates LVAD to adjust the speed and quickly approach the target value. The changes of vascular resistance, myocardial contractility, and the transition from rest to exercise were simulated, and the single factor and mixed factor experiments were carried out to compare the effects of DRL controller and proportional integral derivative (PID) controller, which controls the system according to the difference between measured variables and expected values. Two metrics are used to illustrate the regulation effect: the sum of absolute error (SAE) and the response time of the two controllers, where SAE is the difference between the estimated required pumped blood flow $LVADQ_e$ and the actual measured blood flow $LVADQ_m$. The experimental result shows that the SAE of the DRL controller is 47.6% of that of the PID controller, and the response time of the DRL controller is 38.6% of that of the PID controller. This study demonstrates that the LVAD based on the DRL controller can respond more quickly and more effectively to the different physiological needs of a variety of patients than a PID controller.

1. Introduction

HEART failure is a common disease in which the patient's heart is unable to provide sufficient blood flow to the body to meet the normal metabolic conditions of the human body due to impaired systolic or diastolic function of the heart. Heart transplantation cannot be widely used among patients due to a shortage of donor hearts [1]. Therefore, the development of ventricular assist devices has become an important means of treating patients with heart failure [2].

Left ventricular assist device (LVAD) is a promising and effective method for the treatment of left ventricular failure [3]. At present, most of the LVAD used in clinical practice is set by doctors to maintain a constant speed mode, which cannot ensure that the pump speed matches the current physiological state of patients. If the pump speed is too slow, blood may flow back from the aorta to the left ventricle; if the pump speed is too high, the LVAD may attempt to draw too much blood from the left ventricle resulting in ventricular suction, which may lead to myocardial injury, ventricular extrasystoles or other adverse events [4,5]. Ideally, LVAD should be able to meet the blood perfusion needs of patients with different activity levels and degrees of heart failure, and adjust pump speed adaptively, so that patients with heart failure can live and work like healthy humans [3,6]. In order to achieve this goal, LVAD controller is required to be robust.

The estimated or measured physiological signal needs to be fed back to LVAD physiological controller for comparison with the reference signal. Based on the results of the comparison, the controller calculates the amount of variation in LVAD speed required and applies it to the pump in order to simulate the native feedback mechanism of a healthy heart [25]. According to previous research, these physiological feedback signals include: the constant differential pressure [7], the constant pump inlet flow [8], the constant pump pressure [9], the constant left ventricle end-diastolic pressure [10], and the Frank-Starling flow control [11,12].

☆ This work was supported in part by the Dalian High-level Talent Innovation Support Program (no. 2019RQ026), Changjiang Scholar Program of Chinese Ministry of Education (no. T2017030), the Fundamental Research Funds for the Central Universities (grant no. DUT18RC(3)073).

* Corresponding author.
E-mail addresses: teli@dlut.edu.cn (T. Li), 1297691410@mail.dlut.edu.cn (W. Cui), xiedanan@mail.dlut.edu.cn (N. Xie), Liheng0424@mail.dlut.edu.cn (H. Li), hbliu@dlut.edu.cn (H. Liu), imlixu@dlut.edu.cn (X. Li), yqwang@dlut.edu.cn (Y. Wang).

https://doi.org/10.1016/j.artmed.2022.102308
Received 22 August 2021; Received in revised form 26 February 2022; Accepted 16 April 2022
Available online 22 April 2022
0933-3657/

The cardiovascular system (CVS) includes a wide range of nonlinearities, uncertainties, perturbations, and constraints [25]. In order to enable LVAD to automatically control the speed, most scholars in previous studies have used proportional integral derivative (PID) controllers to adjust these feedback signals [12,15]. Different from the constant speed controller, PID controller can automatically regulate the running speed of LVAD according to the error between physiological signal and set value. Although it is a simple method to deploy PID controller on LVAD equipment, it is heavily dependent on the measured or estimated system variables without any knowledge of the underlying interaction between the rotary blood pump and the cardiovascular system (CVS) [25]. Therefore, when the patient's physiological state (e.g. activity level, degree of heart failure) changes, the PID controller may not be able to adjust the speed adaptively and may become unstable, threatening the patient's life [25].

In addition to traditional PID methods, some scholars have used intelligent control methods for adaptive control of LVAD speed. These include the fuzzy control method [13,15], robust sliding mode control [16,17], H-infinite robust control [32], Iterative learning control [33], model-free adaptive control [10] and artificial neural network control [18,19]. Among these methods, fuzzy control requires artificially developed fuzzy rules, and as fuzzy control is mostly combined with PID methods, it cannot take into account the two objectives of response time and stability, both of which are more important indicators in actual clinical practice. The H-infinity robust control method relies on the accurate CVS-LVAD mathematical model. Although the mathematical model of CVS-LVAD has been continuously studied by scholars, there is still no guarantee that it can perfectly reproduce the complex environment in the human body. Iterative learning control is a control method that does not require an accurate CVS-LVAD mathematical model, but the CVS environment does not have strict repeatability and iterative learning control is easily affected by disturbances. Although the model-free adaptive control (MFAC) method can adapt to various physiological changes among patients and shows strong robustness, the experimental result shows that the response time of the method is not satisfactory. Ideally, the neural network controller can overcome the above problems and meet the control requirements of the coupling model of CVS and LVAD. However, it is very difficult and inefficient for training neural network to collect a large number of data sets containing various physiological states for calibration.

For the past few years, with the rapid development of artificial intelligence techniques and computer hardware devices, deep reinforcement learning (DRL) has been widely used in the field of healthcare. For example, DRL has been applied to areas such as blood glucose control [20,21], and drug sensitivity prediction for cancer treatment [22]. Good results have been achieved with DRL techniques in these tasks. This indicates that DRL performs well in the control tasks under complex and uncertain conditions and shows good robustness [23]. These advantages can precisely solve the difficulties of LVAD physiological controller. In CVS, through continuous interactive training with the CVS system, the agent gradually grasps how to automatically adjust the speed of LVAD to reach the preset reference value under the condition of changes in the human physiological state.

In this study, a physiological control system for LVAD based on deep reinforcement learning is designed. This physiological control system can adapt to the physiological conditions of different patients, allowing the LVAD device to quickly and accurately adjust the CVS to a healthy physiological state. The structure of this paper is as follows. In Section 2, related work on LVAD control is presented. In Section 3, the simulation environment of CVS-LVAD used and the construction method and process of LVAD controller based on DRL are introduced, and the design and indicators of the comparison test are described. Section 4 shows the results of the comparative experiment. The simulation result is discussed, and the superiority of LVAD control method based on DRL is verified in Section 5.

2. Related work

According to one or more controlled variables and their desired set points, the physiological control system replaces the clinician to complete the control task of automatically adjusting LVAD speed. At present, many control strategies and methods for adjusting LVAD speed have been proposed.

The first step is to select the appropriate physiological target to be controlled. The various physiological control systems mentioned in the literature have a variety of control objectives, which can be broadly divided into pressure control, flow control, and control designed to mimic the Frank-Starling mechanism. Giridharan et al. proposed an optimal PI controller that keeps the pressure difference between the left ventricle and the aorta constant. The controller can maintain normal physiological perfusion under various physiological conditions, such as rest, light and strenuous exercise levels [7,30]. In order to improve the preload sensitivity of LV-LVAD system, it is necessary to control the left ventricular preload at a set value, to effectively maintain the constant level of ventricular fluid. Bullister et al. developed a physiological controller targeting left ventricular end-diastolic pressure (LVEDP), which can ensure that perfusion pressure increases with increased patient activity [8]. Control methods to keep the pumping flow of LVAD constant have been proposed for cardiopulmonary bypass, but only when the demand for blood perfusion changes very little. The Frank-Starling mechanism can correlate cardiac output with preload, ensuring that ventricular output matches venous return flow [12]. A physiological controller modeled on the Frank-Starling mechanism was first proposed by Maslen, which could control the artificial heart by varying pump flow and preload [35]. Stevens et al. proposed a physiological control system for LVAD that mimics the Frank-Starling mechanism [12]. With the help of the underlying PID controller, the system can estimate the target flow rate of LVAD pump according to the preload, effectively improving the preload sensitivity of the system.

In addition, appropriate control methods need to be selected. Many different control techniques have been proposed to achieve the regulation of physiological targets. The research on PID control is the most extensive. Moscato et al. used PI controller to regulate LVAD, which was simplified from PID controller [31]. In addition to PID, there are many other control methods applied to LVAD, including fuzzy control [13,15], robust control [32], iterative learning control [33], MFAC [10] and artificial neural network control [18,19]. An intelligent physiological control system for the blood pump was proposed by Fu et al. [13]. They adjusted the input of the pump motor according to the fuzzy logic mechanism so that the pump output can reach the required flow while also preventing the occurrence of ventricular suction. In order to meet the changing needs of patients, Rüschen et al. proposed an H-Infinite robust physiological control system, which adjusted LVAD rotational speed according to the load sharing ratio between LVAD and native left ventricle, and verified the robust stability of the system [32]. Ketelhut et al. proposed a norm optimal iterative learning controller with variable cycle duration without precise model knowledge to control LVAD, which overcame the influence of heart rate changes on traditional iterative learning control by utilizing the repetition characteristics of CVS and realized the adaptive control of LVAD [33]. Similar to iterative learning control is model-free adaptive control, which is a data-driven controller without knowledge of dynamic and structural information of the controlled system. In [10], a physiological controller using MFAC is proposed to maintain LVEDP within the normal range of 3–15 mm Hg. The controller simulated 100 different patient conditions in six different patient scenarios, showing superior experimental results than PID control, and reducing the risk of suction and congestion events in patients. In order to maintain pulmonary and systemic circulation volume balance, Ng et al. integrated artificial neural network into the control system of biventricular assist device [18]. The control system combines a multi-objective neural predictive controller with a Starling-like controller based on preload and solves the problems of pulmonary artery congestion and ventricular aspiration in addition to meeting the appropriate blood flow for organ perfusion. The scheme was verified to be superior to outperform the Dual Independent Starling-like controller and Constant speed control modes.

3. Method

3.1. CVS-LVAD numerical coupling model

Firstly, the simulation model used in this experiment and its specific experimental parameters are introduced. Combined with the practical application of LVAD in clinical practice, a common method to verify the LVAD control algorithm is to use Kirchhoff's law to simulate hemodynamics and establish a CVS-LVAD mathematical coupling model. The model proposed in [24] is used to test the physiological control system, and it is implemented in MATLAB. The CVS-LVAD numerical coupling model is shown in Fig. 1.

In this physiological system, an appropriate physiological feedback signal is needed, which can be used as a control target to reflect the physiological state of the patient. According to the previous studies, the Starling-Like controller can well correlate cardiac output with preload. This method can be used to estimate the blood flow required by LVAD in the current state of the human body, and the difference between the actual measured LVAD output flow and the estimated blood flow required by LVAD can be used as the controller target for adjustment [12].

The meanings and parameters represented by the state variables in the diagram are shown in Table 1:

3.2. Artificial heart control system based on DRL

Considering the high complexity of CVS-LVAD, the LVAD controller proposed in this paper is designed based on Soft-Actor-Critic (SAC) algorithm [26], which is a model-free DRL algorithm based on maximum entropy theory. The SAC algorithm framework inherits the Deep Deterministic Strategy Gradient (DDPG) algorithm. Compared with other DRL algorithms, SAC has stronger environment detection ability and stable convergence ability. DRL focuses on how the agents take different actions in the environment to maximize cumulative rewards. LVAD control problem can be regarded as a Markov decision process, which is represented by tuple {S, P, A, R, Υ}. State S represents the physiological state of the patient, and the transition probability P is based on the CVS-LVAD numerical model in the experiment. Action A represents the change of LVAD speed, and reward R represents the regulation effect of LVAD under the current physiological state of the patient. Discount factor Υ reflects the importance of the regulation effect

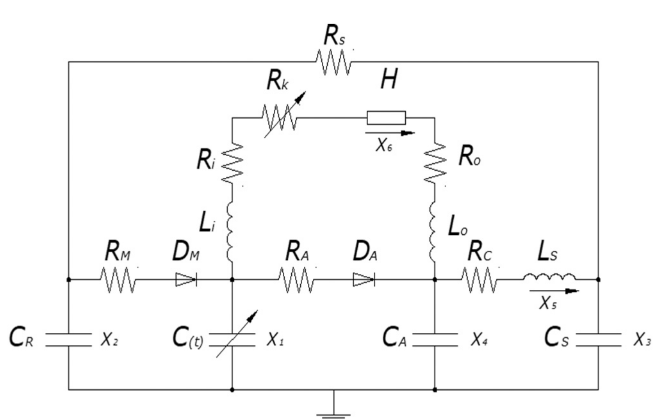

Fig. 1. CVS-LVAD numerical coupling model. Description of related variables in the model: x1, left ventricular pressure; x2, left atrial pressure; x3, Arterial pressure; x4, aortic pressure; x5, total flow, and x6, LVAD output flow.

Table 1
Cardiovascular system model parameters.

Parameter	Normal value	Physiological meaning
R_S	1.0000 mm Hg·s/ml	Systemic vascular resistance
R_M	0.0050 mm Hg·s/ml	Mitral valve resistance
R_A	0.0010 mm Hg·s/ml	Aortic valve resistance
R_C	0.0398 mm Hg·s/ml	Characteristic resistance
R_i	0.0677 mm Hg·s/ml	Inlet resistance of cannulae
R_o	0.0677 mm Hg·s/ml	Outlet resistance of cannulae
R_k	0 ($x_1 > 1$) $-3.5(x_1 - 1)$ ($x_1 \leq 1$)	Suction resistance with parameters
C_t	Time varying	Left ventricular compliance
C_R	4.4000 ml/mm Hg	Left atrial compliance
C_S	1.3300 ml/mm Hg	Systemic compliance
C_A	0.0800 ml/mm Hg	Aortic compliance
L_S	0.0005 ml/mm Hg	Inertance of blood in aorta
L_i	0.0127 ml/mm Hg	Inlet inertance of cannulae
L_O	0.0127 ml/mm Hg	Outlet inertance of cannulae
DM		Mitral valve
DA		Aortic valve
H		Pump pressure difference

of LVAD in the future. The DRL controller is designed to adopt a fully connected neural network structure with two hidden layers. Each hidden layer has 512 neurons to analyze the measured signals and enhance the understanding of the CVS system by the agent. In order to facilitate the agent to perceive and learn CVS-LVAD, this task needs to be artificially modeled, and the state representation, action space, and reward function need to be designed for this task.

1) *State Representation:* In the LVAD control task, a suitable physiological signal is needed, which can well reflect the current physiological condition of the patient. Left atrial pressure (LAP) can be used as a measure of preload in response to the simulation system. LAP was measured and the pumped blood flow required for the current physiological state of the human was estimated according to Eq. (1) [12].

$$LVADQ_e = 7 \times \left(\frac{-5.259 + LAP}{7.124 + 0.2665 LAP^2} + 0.8698 \right) \tag{1}$$

It is worth mentioning that since both the estimated required pumped blood flow $LVADQ_e$ and the actual measured blood flow $LVADQ_m$ fluctuate periodically, these two values are averaged in the interval of 3 s. Finally, the obtained ΔQ is stored in the DRL experience pool as state s_t after normalization processing. s_t is defined as:

$$s_t = \Delta Q = \overline{LVADQ_e} - \overline{LVADQ_m} \tag{2}$$

where $\overline{LVADQ_e}$ is the estimated average blood flow required by LVAD, and $\overline{LVADQ_m}$ is the average blood flow actually measured by LVAD.

2) *Action Space:* In the LVAD system, it is hoped that the LVAD can adaptively adjust the speed of the LVAD after feature extraction according to the sensor information. Therefore, the speed change of the LVAD in a time step $\Delta\omega$ is taken as the final output action of the DRL controller. $\Delta\omega$ is obtained by the transformation of the output action a_t by the neural network through Eq. (3), where the value range of a_t is $[-1, 1]$. $\Delta\omega$ is defined as:

$$\Delta\omega = \begin{cases} a_t \times 300 & |\Delta Q| > 10 \\ a_t \times 300 \times \lambda & |\Delta Q| < 10 \end{cases} \tag{3}$$

where the stability factor $\lambda = 0.01$ is used to enable the DRL controller to stabilize quickly and avoid frequent local oscillations. There are two modes of this action. When the estimated blood flow required for LVAD output differs too far from the actual measured

LVAD output, the patient may experience adverse symptoms such as insufficient perfusion and ventricular suction, which can further damage the patient's heart. Therefore, the output speed of the LVAD is not limited at this stage because it is important to bring the ΔQ back into the normal range as quickly as possible. When the estimated blood flow required for LVAD output does not differ significantly from the actual measured LVAD output, but still requires further adjustment, a limit is imposed on action A to quickly stabilize the DRL controller to a value of 0.

3) *Reward Function:* For CVS-LVAD, a reward-intensive reward function is designed to reflect the performance of the DRL controller under the current physiological state of the human body. The DRL controller is rewarded based on the ΔQ mentioned above. The reward function is designed as:

$$r(s_t, a_t) = |\Delta Q| \times \beta + r_{target} \tag{4}$$

$$r_{target} = \begin{cases} 0.5 & |\Delta Q| < 3 \\ 0 & |\Delta Q| > 3 \end{cases} \tag{5}$$

where $\beta = -0.02$ is the reward normalization factor, r_{target} is a coefficient introduced to assist DRL algorithm in judging whether the target value is close to 0. The reward function is divided into two parts. In the first part, the DRL controller is punished appropriately according to the degree of ΔQ approaching 0 to encourage the DRL controller to eliminate the error as soon as possible. In the second part, a piecewise function r_{target} is designed. In the interval of $\Delta Q \in [-3, 3]$, the DRL controller will get a reward of 0.5, which is beneficial to accelerate the convergence of the DRL controller.

3.3. Controller algorithm flow

As shown in Fig. 2 the controller consists of three components: the CVS-LVAD model system, the signal processing module, and an agent based on SAC algorithm. The DRL controller selects a suitable pump speed change at each time step and inputs this into the LVAD, and after the CVS-LVAD interaction process, the sensor data is fed into the Starling-like controller in the data processing module and the error ΔQ is calculated. This ΔQ is normalized as a state s_t, which is fed into the DRL controller together with the reward R = $r(s_t,a_t)$ calculated by the reward function. In this process, the hope is to maximize the cumulative amount of reward $r(s_t,a_t)$. The Q-function $Q_{soft}(s_t,a_t)$ is used in the DRL problem to represent this cumulative reward. $Q_{soft}(s_t,a_t)$ Q-function can be solved by Bellman equation. Different from DDPG, SAC algorithm adds the entropy regularization term $\log_\pi(s_{t+1},a_{t+1})$ into the Bellman equation. The effect of introducing this term is to make the agent more capable of exploring the environment. The Bellman equation of SAC algorithm is defined as:

$$\begin{aligned} Q_{soft}(s_t, a_t) &= r(s_t, a_t) \\ &+ \gamma E_{s_{t+1}, a_{t+1}}\left[Q_{soft}(s_{t+1}, a_{t+1}) - \alpha log_\pi(s_{t+1}, a_{t+1})\right] \end{aligned} \tag{6}$$

The soft value function is trained to minimize the squared residual error [26]:

$$J_Q(\theta_i) = E_{(s_t, a_t, s_{t+1})}\left[\frac{1}{2}Q_\theta(s_t, a_t) - \left(r(s_t, a_t) + \gamma V_{\bar{\theta}}(s_{t+1})\right)^2\right] \tag{7}$$

where $Q_\theta(s_t,a_t)$ represents the Q-function fitted by the neural network, and θ represents the network parameters. The corresponding pseudo-code and training process are represented in Algorithm 1.

Algorithm 1. DRL-LVAD control.

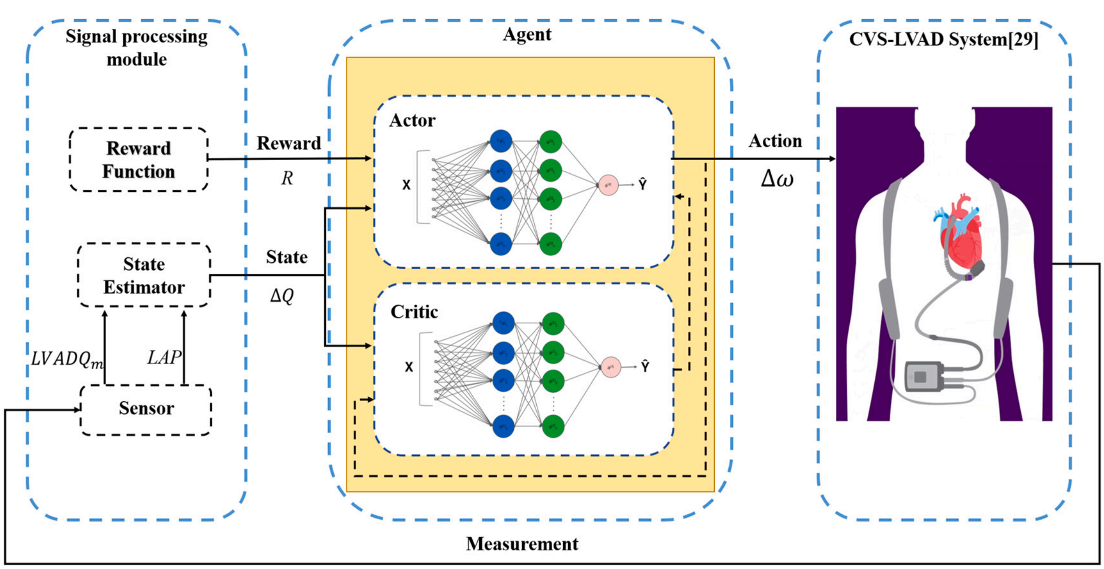

Fig. 2. Block diagram of a physiological controller based on SAC.

Input : initial parameters θ_1, θ_2, ϕ_2
Initial target network weights $\overline{\theta_1} \leftarrow \theta_1$, $\overline{\theta_2} \leftarrow \theta_2$
Initial an empty reply pool $D \leftarrow \varnothing$
Initial pump speed ω
For each iteration **do**
Initialize the simulation environment s'
Initialize random patient and physiological conditions
For each environmental step **do**
If $t > t_l$ then Change the CVS parameters end if
Sample action from the policy $a_t \sim \pi_\phi(a_t \mid s_t)$, and apply it to the LVAD.
Get the state $s_{t+1} \sim p(s_{t+1} \mid s_t, a_t)$ from the CVS-LVAD environment and calculate the rewards $r(s_t, a_t)$
Store the transition in the replay pool $D \leftarrow D \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$
End for
For each gradient step **do**
Update the Q-function parameters
$\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} J_Q(\theta_i)$ for $i \in \{1,2\}$
Update policy weights
$\phi \leftarrow \phi - \lambda_\pi \nabla_{\theta_i} J_\pi(\phi)$
Adjust temperature α
$\alpha \leftarrow \alpha - \lambda \nabla_\alpha J(\alpha)$
Update target network weights
$\overline{\theta_i} \leftarrow \tau \overline{\theta_i} + (1-\tau)\overline{\theta_i}$ for $i \in \{1,2\}$
End for
End for

The entire controller system operation can be described as follows:

1) At each time step, the system measures LAP and feeds it into the state estimator.
2) Estimated the blood demand flow of the LVAD by the human under the current physiological state through Eq. (1) and measured the actual flow at the outlet of the LVAD. The average measured flow rate of LVAD and average estimated LVAD flow needed to output are calculated, and the ΔQ is obtained by difference. After normalization, it is input into SAC system as state s_t.
3) According to the current state s_t, DRL controller judges and gives the action a_t that the next time step should execute.
4) a_t is processed by Eq. (3) and fed into the CVS-LVAD coupled model system for simulation and the states measured by the SAC controller in the next time step are obtained and rewarded according to Eqs. (4) and (5).

3.4. Experimental design and indicators

In order to fully compare the performance of DRL physiological controller (DRL controller) and PID physiological controller (PID controller), the changes of myocardial contractile force (Emax), systemic vascular resistance (Rs), and heart rate (HR) are simulated by changing the parameters.

In each set of experiments, in order to simulate the differences in physiological values between patients, a ± 20% numerical fluctuation is made for the physiological parameters mentioned in Table 1. During the test, single factor experiments are conducted to compare each of these three cases, and finally a comparison test is conducted through a mixed factor experiment. In order to ensure the experimental effect, the PID controller in all experiments adopts the same set of P-I-D parameters, and the DRL controller also selects only one trained agent as the DRL controller. Variable HR is used to simulate the transition from rest to exercise.

In order to clearly evaluate the performance difference between DRL controller and PID controller numerically, the evaluation method proposed in [10] is adopted. Use the sum of the absolute error between the estimated required flow rate of the LVAD and the measured actual flow rate as the evaluation quantity. The definition is as follows:

$$SAE = \sum_{i=1}^{n} \left| \overline{LVADQ_e} - \overline{LVADQ_m} \right| \tag{8}$$

SAE and response time are used as experimental indexes to compare the DRL controller to the PID controller.

During the experiment, the total simulation time is set as 100 s for these situations, and the disturbance is applied when the simulation is carried out to 50 s respectively. In each group of experiments, when the physiological state of each patient is about to mutate, the physiological state of the patients is stable, that is, $\Delta Q = 0$. The function of the two physiological controllers is to adjust the pump speed and pull the ΔQ back to 0 as soon as possible. There is a step-change in the system parameters because if the controller can respond appropriately to this extreme case, then it is assumed that the controller is also able to respond appropriately to small changes [10,29]. Since ΔQ is a periodic cyclic signal caused by heartbeat, in order to facilitate observation and analysis, this paper presents ΔQ as an average in Figs. 3 and 4. The SAE and the response time between the sudden change in physiological conditions and the completion of system regulation are used to compare the two controllers.

4. Results

4.1. Single factor perturbation

The experiment begins with a comparison of three conditions: change in systemic vascular resistance, change in myocardial contractility, and change in heart rate (rest to exercise). In the myocardial contractility change and systemic vascular resistance change experimental group, the total duration of the simulation is 100 s. As the simulation progresses to the 50th second, the system is in a stable state, and the CVS system parameters are changed by adjusting systemic vascular resistance and myocardial contractile force, thereby simulating the changes of the patient's cardiovascular system.

Fig. 3(a) and (b) respectively shows the control effect of the two controllers on ΔQ after the patient's systemic vascular resistance changes and myocardial contractile force changes; Fig. 3(d) and (e) respectively shows the changes in the speed of the two LVAD in these two experiments. In the experimental group with changes in systemic vascular resistance, the DRL controller can complete the control task in 10s and quickly adjust ΔQ to 0. In contrast, the PID controller needs 50 s to complete the task. In the experimental group with changes in myocardial contraction force, the DRL controller can still eliminate ΔQ quickly (5–8 s), while the PID controller needs about 20 s to complete the task.

In the heart rate variation group, the total duration of the simulation is 100 s. The system is already in a stable state at 50 s of simulation, and the heart rate parameter HR changed at this time. Fig. 3(c) and (f) respectively shows the regulation effect of the two controllers on ΔQ and the rotation speed change of the two LVADs after the patient's heart rate changes. At 50 s of simulation, the heart rate changes. It takes 11 s for the DRL controller to eliminate the flow error, while it takes 100 s for the PID to eliminate the flow error.

4.2. Mixed factor disturbance

Fig. 4 shows the regulation effect of DRL controller and PID controller on three simulated patients. The total duration of the simulation is 100 s. At the 50th second of the simulation, the system is in a stable state, and the myocardial contractility, systemic vascular resistance and heart rate changed simultaneously. As shown in Fig. 4, the error increases suddenly under the influence of mixed factors, and the DRL controller and the PID controller react to the disturbance almost at the same time. It takes about 8–15 s for the DRL controller to eliminate the error, while it takes 25–35 s for the PID controller to achieve the same effect.

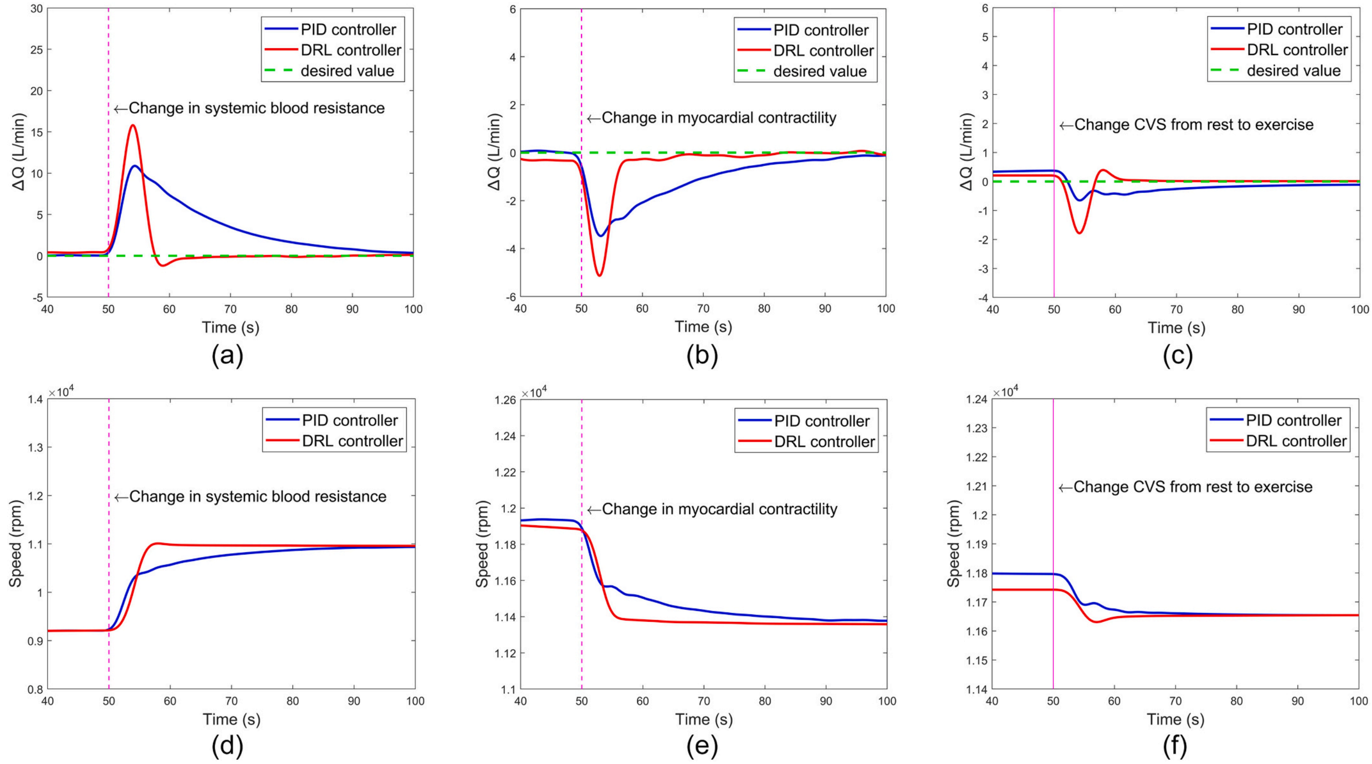

Fig. 3. Control effects of DRL controllers and PID controllers in single factor perturbation experiments. (a), (b), and (c): The ΔQ-Time curve of blood vessel resistance test, myocardial contractility test, and heart rate test. Where the pink dotted line represents the time point perturbation applied to CVS, the green dotted line represents the expected value of ΔQ, the ΔQ in the system based on the DRL controller is represented by a solid red line, and the ΔQ in the system based on the PID controller is represented by a solid blue line. (d), (e), and (f): Speed-Time curve of vascular resistance test, myocardial contractile force test, and heart rate test. Where the pink dotted line represents the time point at which the disturbance acts on CVS, and the LVAD speed based on the DRL controller is represented by the solid red line, while the LVAD speed based on the PID controller is represented by the solid blue line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Discussion

In this study, the proposed control method of LVAD based on deep reinforcement learning algorithm is tested by computer simulation. In the experiment, the tracking goal of the control system is to estimate the blood flow demand of LVAD, which is based on [12], and then use the DRL technology to control the actual blood flow in the system to approximate the estimated value. The performance of the control method is verified by simulating different patient conditions (Changes in systemic vascular resistance, myocardial contractility, and heart rate) in a comparative experiment. The result shows that this control method not only has good robustness among patients with different physiological conditions but also can track the preset value more speedily than the PID controller.

Table 2 shows the control effects of PID controller and DRL controller in 120 groups of patients with different physiological conditions, with SAE and the response time as the comparison indexes. In each case, 20 simulations were conducted for each case. In single factor perturbation experiments, the SAE of DRL controller is 40–80% of PID controller and the response time is only 20–45% of PID controller. In the mixed factor experimental group, the DRL controller reduces SAE by approximately 50% compared to the PID controller and the response time is only 38% of that of the PID controller.

The ideal control system of LVAD should be able to provide an appropriate and safe blood flow for each patient, which requires the controller to adapt to different patients and human activity levels between patients, and requires the controller to be strongly robust to the CVS-LVAD. The controller based on PID method only adjusts the system input (pump speed) according to the measured or estimated system variables, which leads to the neglect of the relationship between LVAD and CVS [27]. However, the neural network of DRL controller stores the previous experience, passing on the control experience. The operation mechanism of CVS-LVAD is briefly modeled, and the appropriate speed is finally output. This method may also show good control performance if the P-I-D parameters are adjusted for a single patient or in patients with similar physiological conditions. In practice, however, each person's physiology is different and the degree of heart failure and the cause of heart failure varies from patient to patient. Therefore, for traditional PID, its fixed parameters do not perform well in all cases. For fuzzy PID, the establishment of fuzzy rules gives the method stronger adaptive ability, but because the formulation of fuzzy rules is also man-made, it naturally brings the prior knowledge of human beings. A neural network based multi-objective adaptive controller proposed in [18] relies on a trained artificial neural network to control the cardiovascular system, which provides greater robustness and faster response to the physiological controller. But this method requires collecting and calibrating a large number of data sets that can simulate a variety of patients before training, a process that is complex and time-consuming. [32] propose an H-infinite robust control method. This approach requires knowing the exact CVS-LVAD mathematical model, which is very difficult for CVS. In addition, iterative learning control is also proposed in [33]. Although this control method does not need to know the precise CVS-LVAD mathematical model in advance, it requires strict repeatability in the control environment and is prone to disturbance by interference. In addition, the author only simulated four patients with different conditions, which could not fully explain the applicability of this method to various conditions. [10] propose a physiological controller based on model-free adaptive control, in which the method shows strong robustness in simulations with 600 patients, but the results showed that the method is too unresponsive, taking 100 s to stabilize the LVAD

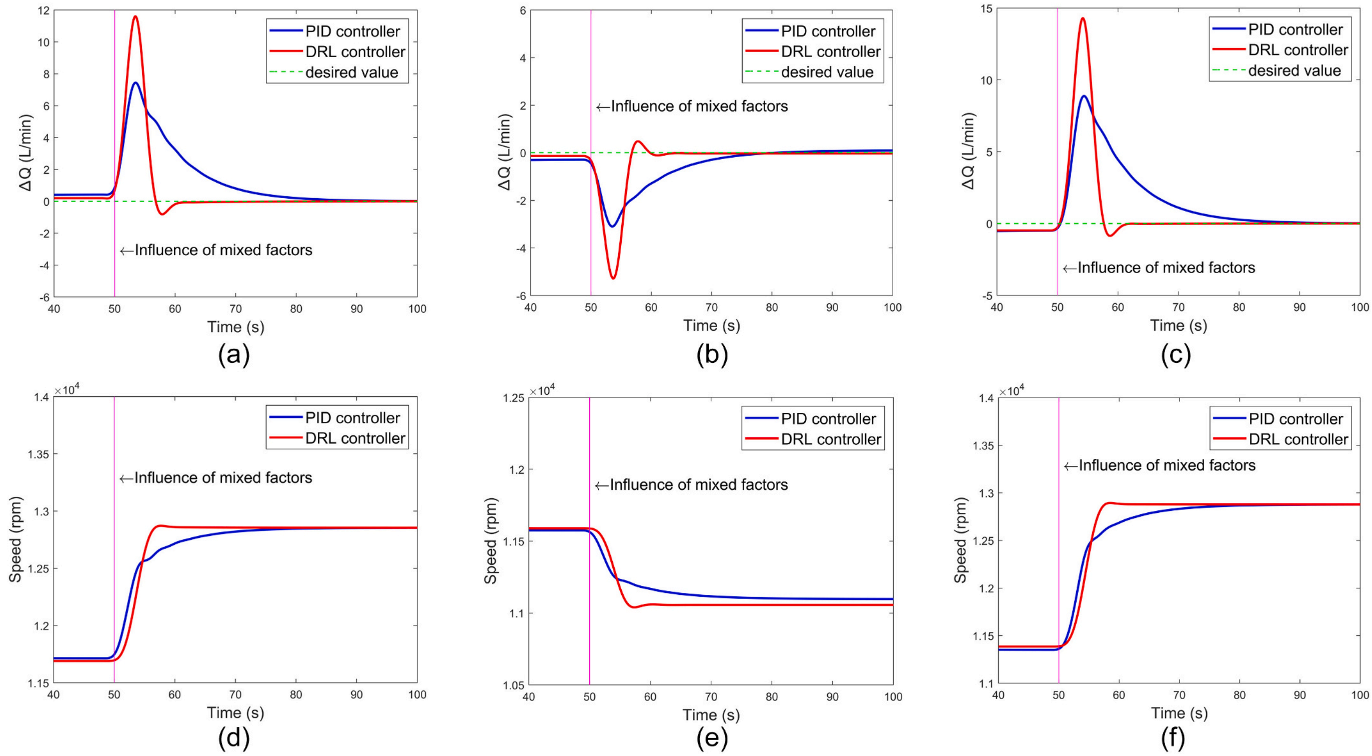

Fig. 4. The control effect of DRL controller and PID controller in mixed-factor disturbance experiment is studied. (a), (b), and (c): ΔQ-Time curve of three groups of mixed factor perturbation experiments. Where the pink dotted line represents the time point perturbation applied to CVS, the green dotted line represents the expected value of ΔQ, the ΔQ in the system based on the DRL controller is represented by a solid red line, and the ΔQ in the system based on the PID controller is represented by a solid blue line. (d), (e), and (f): Speed-Time curves of three groups of mixed factors perturbation experiments. Where the pink dotted line represents the time point at which the disturbance acts on CVS, and the LVAD speed based on the DRL controller is represented by the solid red line, while the LVAD speed based on the PID controller is represented by the solid blue line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
The control results of DRL controller and PID controller.

Scenarios	Controller	SAE	The response time
Rising Rs	DRL	758 ± 119	10.19 ± 2.02
	PID	1992 ± 651	43.01 ± 6.01
Falling Rs	DRL	793 ± 385	11.69 ± 2.99
	PID	1704 ± 798	40.76 ± 5.20
Rising Emax	DRL	207 ± 34	5.67 ± 1.38
	PID	253 ± 126	17.74 ± 6.39
Falling Emax	DRL	216 ± 48	6.26 ± 1.07
	PID	341 ± 125	21.77 ± 5.07
Rest to Exercise	DRL	734 ± 214	10.85 ± 2.3016
	PID	848 ± 313	23.46 ± 12.80
Action of mixed factors	DRL	514 ± 126	12.30 ± 2.69
	PID	1080 ± 435	31.89 ± 3.61

speed. Different from them we control the LVAD using the SAC algorithm, a type of DRL algorithm that directly replaces the PID in the regulation of the LVAD controller. Since DRL is a method suitable for sequential decision problems, the well-trained DRL agent can show strong robustness in the test environment if the neural network can fit the policy network well. The results shown in Figs. 3–4 show that, regardless of single factor or multi-factor mixed action, compared with conventional PID, well-trained DRL agents can track ΔQ to zero quickly and accurately.

When a patient's physiological condition or activity level changes, the controller needs to adjust the physiological signal to a preset value or interval. For a patient, the course of heart failure may be slowly changing, when it may not be necessary for a controller to consider the speed of its corresponding time. But when the patient is switching between rest-exercise in daily life, the indicator of response speed becomes very important, as it may lead to short periods of dyspnea. During the switch from rest to exercise, the body's heart rate and systemic vascular resistance, and the body require more blood flow than under normal conditions. It has been mentioned in previous studies that most physiological controller experiments are carried out for specific physiological conditions of specific patients [10], which may be due to the fact that conventional PIDs only regulate according to a specific set of parameters and do not guarantee the robustness of this controller. In addition, the PID controller is often unable to quickly return the tracked values to normal levels in the event of a sudden change in physiological parameters [10,28,29]. In our experiment, the controller using conventional PID control needed about 30–50 s to make LVAD pump the right amount of blood flow to reach the current blood flow required by the Starling-like controller, in previous studies [10,13,28], the conventional PID controller or fuzzy PID controller used by researchers takes about 50–150 s to make the system stable. Correspondingly, the time can be compressed to 5–20 s using a DRL-based LVAD control method. This is a consequence of the fact that the PID-based method only observes ΔQ as an indicator and does not provide a clear and effective evaluation of the whole system, so the closer ΔQ is to zero, the more careful the LVAD regulation action becomes. This point can be more clearly seen in the LVAD rotational speed change curve of each group of experiments shown in Fig. 4(d)–(f). During the control process, the DRL controller adjusts the left ventricular assisted rotation at a nearly constant rate change. In contrast, the acceleration of the PID controller is getting smaller and smaller as the steady speed approaches. By comparison, the control process of DRL controller is more determined, which is also the reason why the response time of DRL controller is shorter than that of PID controller. In addition, as the LAP signal is a periodic fluctuating signal in the CVS-LVAD, the signal input to the PID controller in the experiment is processed by averaging the measured data over a period of time. Therefore, the effect of the PID controller on the LVAD speed is not immediately reflected in the PID input signal, which also slows down the response of the PID controller. The goal of an agent in DRL is to optimize in the direction of a strategy that maximizes the expected reward match. For the DRL controller, if the CVS is not physiologically healthy, the agent will receive a negative reward for each time step. This reward encourages DRL controller to learn how to quickly respond to disturbances in the system and eliminate the errors caused by such disturbances. With this reward and training, the DRL controller gradually learned how to control the speed of the left ventricular assist by relying on LAP signals with delayed feedback properties.

The proposed LVAD physiology controller based on DRL adaptively and rapidly adjusts the pump speed to provide adequate blood flow to the patient for different physiological conditions (heart failure, rest-exercise transition). In addition to showing good robustness for different patients and different physiological conditions, DRL controller can respond to perturbations more rapidly than traditional PID controllers, eliminating errors to near zero within 5–50s, more closely resembling the adaptive capacity of a normal heart in a healthy population.

There are still some limitations to this study. First of all, the ΔQ of DRL controller is greater than that of PID controller during the initial phase of disturbance in the simulation process. This is because the DRL controller is considered to limit the speed of the CVS system after a sudden change, so the PID controller can adjust the LVAD at a faster rate in the initial stage of the change. Moreover, the experiment in this paper only simulated patients with left heart failure and did not involve the condition of right heart failure. Third, different from traditional control methods, the DRL controller involving neural network is difficult to prove the robustness and stability of the controller at the mathematical level, which is also one of the limitations of this paper. In addition, only computer simulation experiments are carried out. Since this is an ideal numerical simulation of the complex task environment of the cardiovascular system, some information is omitted. In order to better validate the control method proposed in this paper, the next stage the trained agent should be transferred to a simulated circulation simulation environment for evaluation.

6. Conclusion

In this study, a physiological control system for the left ventricle based on deep reinforcement learning techniques is presented. The control system uses a method based on the Frank-Starling principle to estimate the current blood flow to be pumped by the patient, and the DRL controller regulates the speed of the LVAD according to the difference between the currently measured LVAD output blood flow and the estimated flow that requires LVAD output. The robustness of the DRL controller is verified by single factor and mixed factor comparison tests, and the proposed DRL controller has a shorter response time and a lower sum of accumulated errors than the PID controller. It suggests that the DRL control approach has great potential in regulating LVAD-CVS tasks.

Declaration of competing interest

The authors have no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.artmed.2022.102308.

References

[1] Gustafsson Finn, Rogers Joseph G. Left ventricular assist device therapy in advanced heart failure: patient selection and outcomes. Eur J Heart Fail 2017;19 (5):595–602. https://doi.org/10.1002/ejhf.779.

[2] Drakos Stavros G, Mehra Mandeep R. Clinical myocardial recovery during long-term mechanical support in advanced heart failure: insights into moving the field forward. J Heart Lung Transplant 2016;35(4):413–20. https://doi.org/10.1016/j.healun.2016.01.001.

[3] Wang Y, Faragallah G, Divo E, Simaan MA. Detection of ventricular suction in an implantable rotary blood pump using support vector machines. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011. p. 3318–21.

[4] Wang Yu, Koenig Steven C, Slaughter Mark S, Giridharan Guruprasad A. Rotary blood pump control strategy for preventing left ventricular suction. ASAIO J 2015; 61(1):21–30.

[5] Vollkron Michael, et al. Development of a suction detection system for axial blood pumps. Artif Organs 2004;28(8):709–16.

[6] Verbeni A, et al. An innovative adaptive control strategy for sensorized left ventricular assist devices. IEEE TransBiomedCircSyst Oct. 2014;8(5):660–8.

[7] Giridharan GA, Skliar M. Control strategy for maintaining physiological perfusion with rotary blood pumps. Artif Organs 2003;27(7):639–48.

[8] Bullister E, Reich S, Sluetz J. Physiologic control algorithms for rotary blood pumps using pressure sensor input. Artif Organs 2002;26(11):931–8.

[9] Casas Fernando, Ahmed Nisar, Reeves Andrew. Minimal sensor count approach to fuzzy logic rotary blood pump flow control. ASAIO J 2007;53(2):140–6.

[10] Fetanat M, Stevens M, Hayward C, Lovell NH. A physiological control system for an implantable heart pump that accommodates for interpatient and intrapatient variations. IEEE TransBiomedEng April 2020;67(4):1167–75.

[11] Francis Salamonsen Robert, et al. Theoretical foundations of a Starling-like controller for rotary blood pumps. Artif Organs 2012;36(36):787–96.

[12] Stevens MC, et al. Frank-starling control of a left ventricular assist device. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2011. p. 1335–8.

[13] Fu M, Xu L. Computer simulation of sensorless fuzzy control of a rotary blood pump to assure normal physiology. ASAIO J 2000;46(3):273–8.

[15] Andreas, et al. Physiological control of a rotary blood pump with selectable therapeutic options: control of pulsatility gradient. ArtifOrgans 2008;32(10): 761–71.

[16] Bakouri MA, Salamonsen RF, Savkin AV, Alomari AH, Lim E, Lovell NH. Physiological control of implantable rotary blood pumps for heart failure patients. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013. p. 675–8.

[17] Gwak KW, et al. Sliding mode control for the Frank-Starling response of a piston pump mock ventricle. JProcess Control 2015;25:70–7.

[18] Ng BC, et al. Application of multiobjective neural predictive control to biventricular assistance using dual rotary blood pumps. BiomedSignal ProcessControl 2018;39:81–93.

[19] Kim Sanghyun, Kim Hunmo, Ryu Jungwoo, Chung Sungtaek. Control of left ventricular assist device using artificial neural network. In: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136). 3; 1997. p. 1363–6.

[20] Tejedor Miguel, Woldaregay Ashenafi Zebene, Godtliebsen Fred. Reinforcement learning application in diabetes blood glucose control: a systematic review. ArtifIntell 2020;104:101836.

[21] Lee S, Kim J, Park SW, Jin S-M, Park S-M. Toward a fully automated artificial pancreas system using a bioinspired reinforcement learning design:in silico validation. IEEE J Biomed Health Inform Feb. 2021;25(2):536–46.

[22] Daoud S, Mdhaffar A, Jmaiel M, Freisleben B. Q-rank: reinforcement learning for recommending algorithms to predict drug sensitivity to cancer therapy. IEEE J Biomed Health Inform Nov. 2020;24(11):3154–61.

[23] Al-Nima Raid Rafi Omar, Han Tingting, Al-Sumaidaee Saadoon Awad Mohammed, Chen Taolue, Woo Wai Lok. Robustness and performance of deep reinforcement learning. Appl Soft Comput 2021;<journal-title>Appl Soft Comput</journal-title>:107295.

[24] Simaan MA, Ferreira A, Chen S, Antaki JF, Galati DG. A dynamical state space representation and performance analysis of a feedback-controlled rotary left ventricular assist device. IEEE TransControl SystTechnol Jan. 2009;17(1):15–28.

[25] Daners Marianne Schmid, Dual Seraina Anne, AnneDual Seraina. Pathophysiological determinants relevant in blood pump control. In: Mechanical support for heart failure. Cham: Springer; 2020. p. 253–77.

[26] Haarnoja Tuomas, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. Aug. 2018. arXiv: 1801.01290.

[27] Qu, et al. In: Percutaneous and transcutaneous connections: mechanical circulatory and respiratory support. Elsevier; 2018. p. 627–57.

[28] Mansouri Mahdi, et al. Preload-based Starling-like control of rotary blood pumps: an in-vitro evaluation. PloS one 2017;12(2):e0172393.

[29] Pauls JP, et al. Evaluation of physiological control systems for rotary left ventricular assist devices: an in-vitro study. Ann Biomed Eng 2016;44(8):2377–87.

[30] AlOmari Abdul-Hakeem H, et al. Developments in control systems for rotary left ventricular assist devices for heart failure patients: a review. Physiol Meas 2012;34 (1):R1.

[31] Moscato Francesco, et al. Left ventricle afterload impedance control by an axial flow ventricular assist device: a potential tool for ventricular recovery. Artif Organs 2010;34(9):736–44.

[32] Rüschen Daniel, et al. Robust physiological control of rotary blood pumps for heart failure therapy. at-Automatisierungstechnik 2018;66(9):767–79.
[33] Ketelhut Maike, et al. Iterative learning control of ventricular assist devices with variable cycle durations. Control EngPract 2019;83:33–44.
[35] Maslen Eric H, et al. Feedback control applications in artificial hearts. IEEE Control SystMag 1998;18(6):26–34.