



Enhancing cardiac hemodynamic and pulsatility in heart failure via deep reinforcement learning: An in-silico and in-vitro validation study of percutaneous ventricular assist devices



Yuyang Shi ^a, Zhike Xu ^a, Chenghan Chen ^a, Feng He ^a, Pengfei Hao ^{a,b}, Xiwen Zhang ^{a,*}

^a Department of Engineering Mechanics, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

^b Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Keywords:

Ventricular assist devices
Deep reinforcement learning
Heart failure
Hemodynamics
Pulsatility
Adaptive control

ABSTRACT

Background and objective: Percutaneous ventricular assist devices (pVADs) are critical for bridging heart failure (HF) patients to recovery or transplantation, yet existing control strategies—constant speed control and pre-programmed pulsatile control—lack adaptability to dynamic physiological variations, leading to reduced pulsatility and hemodynamic mismatch.

This study proposes a deep reinforcement learning (DRL)-based adaptive control framework to optimize pVAD performance. The goal is to restore physiological pulsatile hemodynamics while autonomously adjusting to different HF conditions, heart rate fluctuations, and intra-cycle ejection phase variability.

Methods: Following a dual-validation pathway designed to bridge simulation with physical testing, a cardiovascular-pVAD in-silico model was developed and its fidelity confirmed against an in-vitro pulsatile mock circulatory loop. This validated platform was then used to design and test the DRL controller. A modified Deep Deterministic Policy Gradient (DDPG) algorithm with embedded LSTM layers was designed to capture temporal characteristics in aortic pressure (AOP) and aortic flow (AF) waveforms. The reward function integrated hemodynamic recoverability, pulsatile waveform similarity, and control stability and safety penalty.

Results: Comparative simulations and experiments demonstrated the DRL controller's superiority over conventional strategies. Under the moderate HF condition, DRL controller achieved near-physiological AOP (DTW-AOP: 1.17 vs. 16.42 for constant speed control; 2.72 for preprogrammed pulsatile control) and AF (DTW-AF: 21.23 vs. 71.74/48.96), with pulsatility indices (PI: 1.69 vs. 1.05/1.54) and pulse pressures (PP: 34.42 mmHg vs. 3.20/24.90 mmHg) closely matching healthy reference. The framework exhibited robust adaptability to heart rate shifts (75→120 bpm) and ejection phase delays (0.1 s), maintaining stability despite sensor noise and physiological perturbations.

Conclusions: This DRL controller enables real-time synchronization with native cardiac cycles and generalization across pathologies, paving the way for precision pVAD support and future clinical translation.

1. Introduction

1.1. Clinical background and motivation

Heart Failure (HF), affecting over 64 million patients globally, is characterized by impaired myocardial function, ventricular remodeling, and neuroendocrine activation, leading to reduced cardiac output, hypoperfusion, and fluid retention, with a 5-year mortality exceeding 50% [1–3]. While pharmacological therapies (e.g., ARNIs, β-blockers) slow progression in early stages, they offer limited efficacy for end-stage

patients [4,5]. Heart transplantation, though the gold standard, faces severe donor shortages (global supply-to-demand ratio <1:30) [6]. Therefore, effective circulatory support strategies are urgently needed to reduce cardiac load, improve quality of life, and extend survival in these patients [7,8].

Left Ventricular Assist Devices (LVADs), particularly percutaneous ventricular assist devices (pVADs), provide critical hemodynamic support via minimally invasive mechanisms, creating survival windows via bridge-to-transplantation (BTT) or recovery (BTR) strategies [4,9–11]. In this study, the term pVAD specifically refers to catheter-based left

* Corresponding author.

E-mail address: zhangxiw@tsinghua.edu.cn (X. Zhang).

ventricular assist devices, such as the Impella® family (Abiomed). These axial or micro-axial flow pumps are inserted percutaneously, typically via the femoral artery, and advanced retrogradely across the aortic valve into the left ventricle. By actively drawing blood from the left ventricle and expelling it into the ascending aorta, they reduce left ventricular preload and myocardial oxygen demand while increasing systemic perfusion.

Current pVAD control approaches are often limited to constant speed control, which sustains perfusion but causes non-pulsatile flow (reduced pulse pressure <10 mmHg), and preprogrammed pulsatile control, which simulates pulsatility via periodic speed changes but lacks patient-specific adaptability. The performance and limitations of these conventional strategies have been documented in studies on various experimental models and prototypes of axial pumps [12,13]. Experimental validation (Section 4.2) confirms that neither strategy fully restores physiological hemodynamics and pulsatility in HF, underscoring the need for adaptive, feedback-driven control.

1.2. Research gaps and contributions

Recent LVAD control advancements focus on physiological adaptability. To improve safety, some approaches focus on preventing adverse events like suction by adapting to the patient's physiological state, using methods such as supervised fuzzy logic [14] and model-free adaptive control [15]. Other strategies have specifically aimed to improve responsiveness to dynamic conditions like wide variations in heart rate [16]. To further enhance operational safety margins, researchers have explored sensor-integrated designs [17] and multi-objective optimization techniques [18]. A particularly innovative direction involves sensorless algorithms that infer the patient's physiological state from the pump's own operational data, thereby reducing adverse events without requiring additional invasive sensors [19]. In parallel, Frank-Starling-inspired strategies have been developed to optimize pump output across different activity levels [20]. Deep learning (e.g.,

DNNs, CNNs) enables predictive control from flow signals [21–24]. Despite these advances, a common limitation persists: these methods either rely on simplified physiological models and pre-set rules or, in the case of deep learning, require large, labeled datasets of "optimal" control, hindering their ability to generalize across the complex, non-linear dynamics of diverse patient conditions.

Deep Reinforcement Learning (DRL) addresses these limitations through autonomous, environment-driven optimization. Prior applications of DRL to VAD control have shown significant promise. For instance, studies have employed DRL to develop energy-efficient controllers for pulsatile VADs and have utilized algorithms like Soft Actor-Critic (SAC) to create robust physiological control systems [25–27]. However, few studies have explored the application of DRL in pVAD control, particularly in terms of how to adapt to varying heart failure conditions and different cardiac cycles.

Our contribution proposes a DRL-based adaptive strategy enabling real-time synchronization with native cardiac cycles, autonomous adaptation to variable HF conditions (severity, heart rate, ejection phases), and pulsatile hemodynamics comparable to healthy heart. The overall functionality and schematic of the assistive system are illustrated in Fig. 1. To rigorously evaluate this strategy, we employ a dual-validation approach, testing the controller first in a high-fidelity in-silico model and subsequently on a physical in-vitro mock circulatory loop, establishing a practical pathway toward future clinical translation.

Based on these objectives, we formally hypothesise that a DRL controller, trained across a diverse range of simulated cardiac profiles, can safely generalize its adaptive control policy to a physical in-vitro mock circulatory loop. We further hypothesise that this controller can restore near-physiological pulsatility and adapt to dynamic perturbations, all while operating within the strict latency constraints required for real-time embedded deployment.

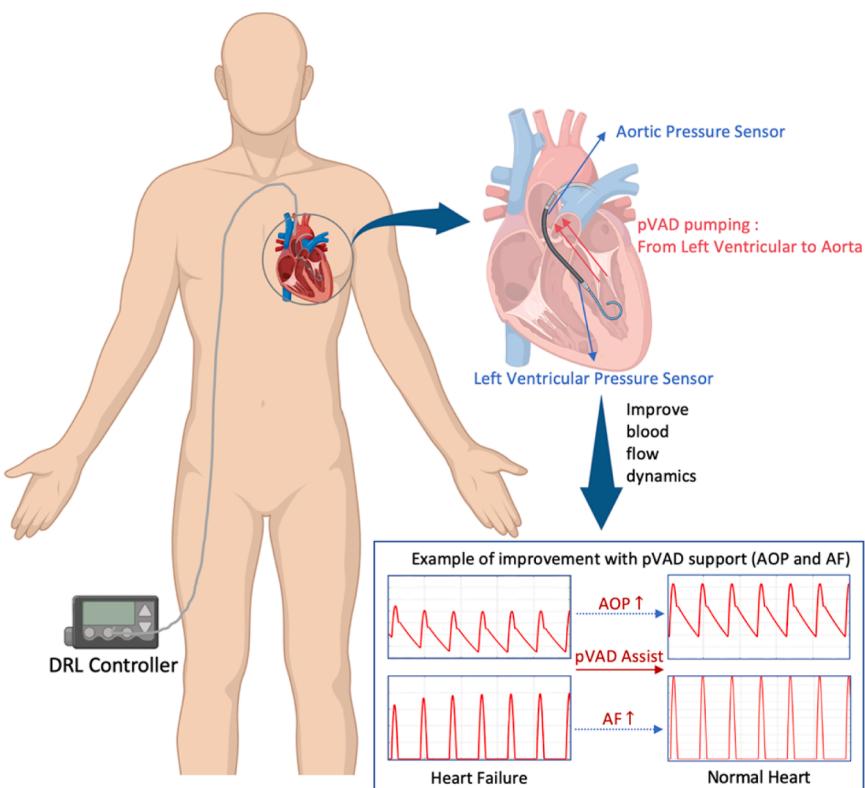


Fig. 1. Schematic diagram of the pVAD assistive system and its overall functionality.

2. Methods

This section primarily introduces the DRL method. DRL algorithms can be categorized along two dimensions [28], the basic structure of the DRL control strategy will be presented in the **Supplementary Material**.

In Section 2.1, the Deep Deterministic Policy Gradient (DDPG) algorithm used in DRL control will be introduced [29]. In our control strategy, several modifications have been made to the classical DDPG structure by incorporating LSTM layers to encode the temporal features of the aortic pressure waveform (AOP) and aortic flow waveform (AF), enabling waveform-based control.

2.1. Improved DDPG network

To address the unique challenge of controlling a pulsatile device based on dynamic, time-series physiological signals, we designed an improved network architecture, hereafter referred to as DDPG-LSTM. Our approach integrates LSTM layers into the standard DDPG framework. This was specifically engineered to overcome the limitations of conventional DRL agents by empowering both the Actor and Critic networks to extract and utilize critical temporal features from sequential inputs, such as the AOP and AF waveforms, to guide control actions. Both the Actor and Critic networks embed LSTM layers to capture the temporal features of AOP waveform and AF waveform.

As a model-free algorithm, DDPG simultaneously learns a deterministic policy (Actor network) and a Q-value approximator (Critic network). The Actor maps physiological states to pump speed adjustment commands, while the Critic evaluates the quality of state-action pairs.

The key difference between the control algorithm in this study and the standard DDPG lies in the state encoding process: the Actor network and Critic network first processes the AOP waveform and AF waveform through LSTM layers, extracting temporally aligned features to guide pulsation synchronization.

To encourage the Actor network to better explore the environment, noise generated by the Ornstein-Uhlenbeck (OU) process is added to the control actions. The structure of DDPG-LSTM is shown in Fig. 2. The detailed algorithmic procedure for our DDPG-LSTM controller, including the initialization, training loop, and update rules, is presented as pseudocode in the **Supplementary Material (Algorithm 1)**. A detailed description of the OU noise generation procedure, including its mathematical formulation and parameterization, is provided in the **Supplementary Material**.

3. Controller design and experimental setup

To meet clinical requirements for perfusion and pulsatility while adapting to diverse patient physiological conditions, pVAD control becomes a complex nonlinear physiological system. Traditional constant speed control and preprogrammed pulsatile control are inadequate to address this challenge. However, the DDPG algorithm within DRL is particularly well-suited, as it efficiently handles high-dimensional continuous action spaces, learns complex system dynamics adaptively, and captures the temporal features of multimodal physiological signals through deep neural networks, such as aortic pressure waveform and aortic flow waveform. This enables automatic speed adjustment to accommodate patient-specific variations (e.g., heart failure severity and heart rate fluctuations) while maintaining hemodynamic stability.

To enable the DRL controller to achieve the above functionalities, extensive offline training is required [25]. Training within a real human physiological environment is impractical, so the most efficient approach is to conduct training in a simulated environment. By establishing a high-fidelity cardiovascular-pump hybrid simulation environment [30], we can train the DRL controller within it. Additionally, tuning the hyperparameters of the DRL controller is relatively straightforward, as multiple DDPG algorithms with different configurations can be trained and tested simultaneously [25].

3.1. Description of the control problem

The pVAD is a sophisticated medical device primarily used to treat patients with HF, serving as a bridge to heart transplantation or a short-term replacement therapy. It assists or replaces the heart's pumping function to meet the patient's perfusion needs. The pVAD used in this study was independently designed by our laboratory [12]. During actual use, it is inserted through the femoral artery and delivered into the left ventricle.

The AOP and AF of heart failure patients are significantly reduced compared to those of healthy individuals. To enhance blood pressure and flow in heart failure patients, the pVAD utilizes a high-speed rotating impeller to generate continuous hydraulic pressure, drawing blood from the left ventricle and pumping it into the aorta, thereby improving systemic blood circulation [12]. The entire system is powered by an external controller and connected to the internal device through a catheter. The pVAD effectively reduces the workload on the failing heart and helps maintain normal hemodynamic conditions. The impeller speed directly determines the pressure and flow generated by the pVAD, making the primary control objective of the entire system to regulate the motor speed driving the impeller.

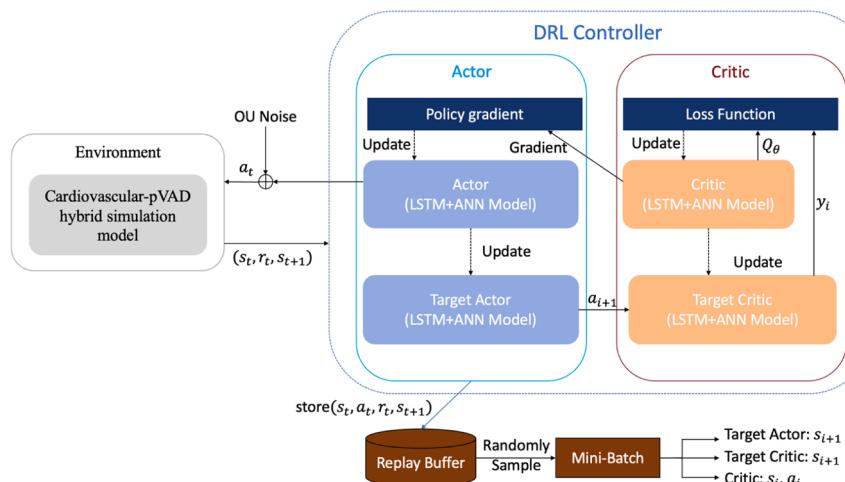


Fig. 2. Structure of the DDPG-LSTM algorithm.

In this study, the motor speed is controlled by the DRL algorithm. The controller inputs include left ventricular pressure, aortic pressure, aortic flow, heart rate, and pVAD flow, which are obtained through sensors and medical devices. The control output (i.e., the action generated by the DRL controller) is the motor speed.

The physical validation of this control problem was performed on our in-vitro pulsatile mock circulatory loop. The key hardware components and signal sources are detailed here for clarity. The pulsatile flow is generated by a programmable linear motor (Model VCAH0445-0300-00A, Suzhou Yuwen Technology Co., Ltd., Suzhou, China). The inputs for the DRL controller are obtained as follows:

Left Ventricular Pressure (LVP) and Aortic Pressure (AOP): Measured using pressure transducers (Model JIBPT-01-UT, Shenzhen Jinchengrui Medical Equipment Co., Ltd., Shenzhen, China).

Instantaneous Aortic Flow (for waveform input): Measured by a custom-made ultrasonic flowmeter (Model IL-100SXB-G2, Shanghai Xunyinke Technology Co., Ltd., Shanghai, China). The time-averaged value of this signal is used to calculate the Cardiac Output (CO) metric reported in L/min.

Heart Rate (HR): This is a set parameter of the linear motor controller, allowing us to precisely simulate different heart rates. pVAD Flow: This value is not directly measured by a sensor but is estimated in real-time from the measured pump head pressure and the pump's empirically-derived hydraulic characteristic surface (Fig. 9).

All sensor data is acquired at a data acquisition rate of 100 Hz.

3.2. Cardiovascular-pump hybrid model

The cardiovascular-pVAD hybrid simulation model in this study was built using MATLAB SIMULINK. The hybrid simulation model is divided into two parts: the cardiovascular model and the pVAD model. The

cardiovascular model can accurately simulate human blood pressure and blood flow, while the pVAD model, based on experimental data from the actual pVAD (to be tested in Section 4.1), simulates the relationship among inlet/outlet pressure difference, rotational speed, and flow rate. These two models are connected to form the cardiovascular-pVAD hybrid simulation model. The structure of the complete cardiovascular-pVAD hybrid simulation model is illustrated in Fig. 3. The construction of each model will be described separately below.

3.2.1. Cardiovascular model

The cardiovascular model was constructed with reference to previously published studies [31,32], effectively simulating the human cardiovascular environment. In this model, current corresponds to blood flow, voltage corresponds to blood pressure, resistance corresponds to vascular resistance, and inductance and capacitance correspond to blood inertia and vascular elasticity, respectively. The construction of the Cardiovascular Model in SIMULINK is presented in the Supplementary Material.

The hemodynamic characteristics of the healthy and failing hearts simulated by the cardiovascular model were compared with those measured from the pulsatile mock circulatory loop platform (to be tested in Section 4.2), as shown in Fig. 4. This comparison demonstrates the high fidelity of the hybrid simulation model and validates the reliability of training the DRL controller within this simulation environment. Figure 4(a) and Figure 4(b) are simulated versus experimental results of AOP and AF of a healthy state heart, and Figure 4(c) and Figure 4(d) are simulated versus experimental results of AOP and AF of a heart in heart failure state, respectively.

3.2.2. pVAD model

In this study, the pVAD is modeled as a controlled current source,

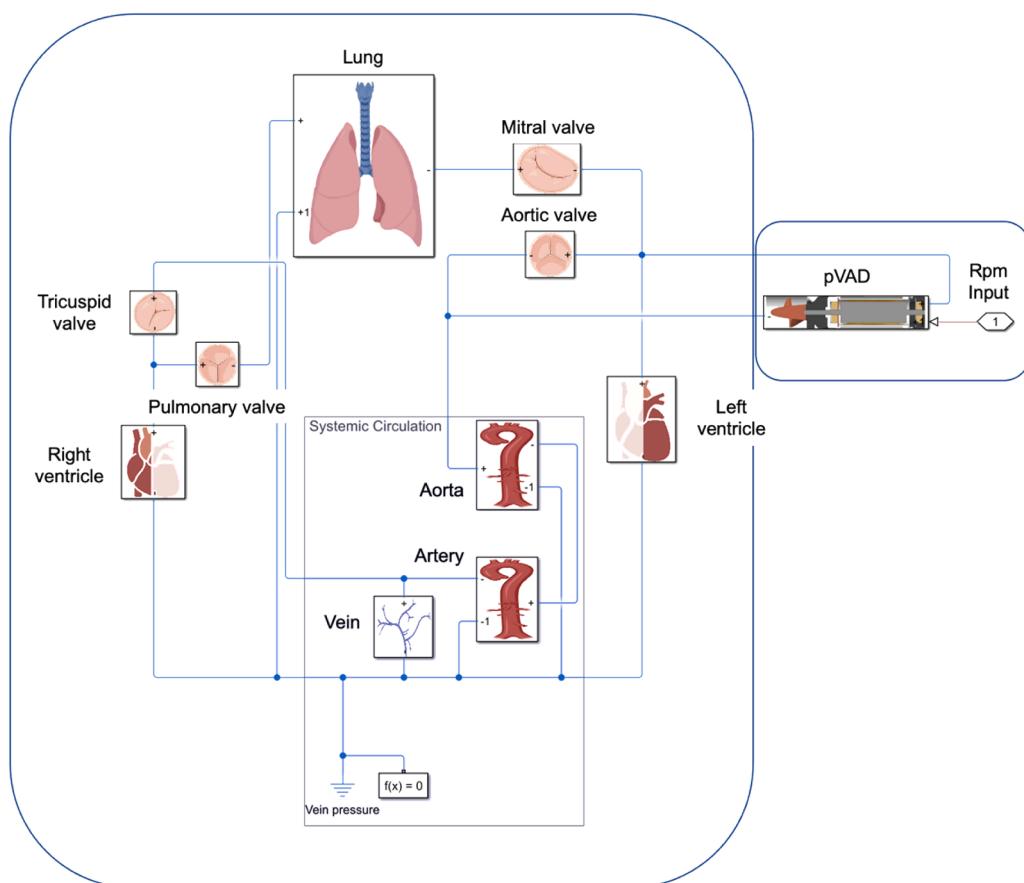


Fig. 3. Structure of the integrated cardiovascular and pVAD hybrid simulation model.

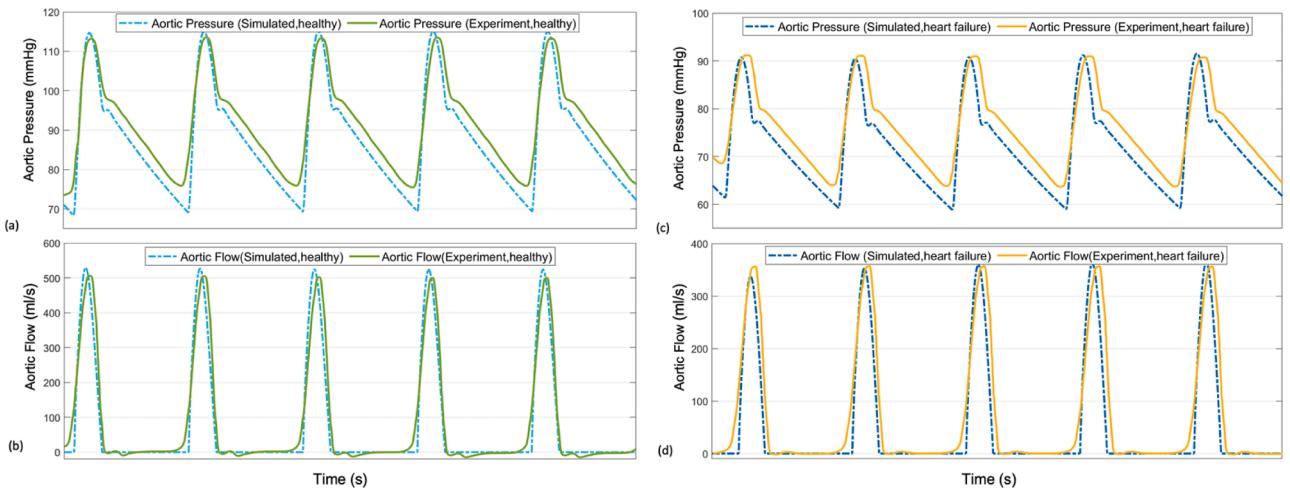


Fig. 4. Comparison of hemodynamic characteristics between simulation and experimental results.

which is connected in parallel to the aortic node of the cardiovascular model, working in coordination with the original heart model to directly compensate for circulatory insufficiency. The pVAD used in this study is an axial flow pump. Under fixed mechanical configuration, there is a specific numerical relationship among pump flow, pressure difference across the pump, and pump speed [33,34]. To obtain this relationship, we conducted hydraulic characteristic experiments using the pulsatile mock circulatory loop platform, measuring data to plot the flow-pressure-speed surface of the pVAD. Once this surface is determined, we can calculate the pump output flow rate based on the pump speed and the pressure at the pump inlet and outlet measured by sensors. This flow rate is then directly fed into the human circulatory loop as a controlled current source in SIMULINK. The detailed experimental procedures are presented in [Section 4.1](#) and [Supplementary Material](#).

3.3. Design of baseline controller

3.3.1. Constant speed control

Based on the set heart failure levels and the experimentally measured hydraulic characteristics of the pVAD, we set the pump speed in the range of 27,000 to 34,000 rpm. Subsequently, we monitor the outputs of cardiovascular modules and compare the improvements achieved at different speeds.

3.3.2. Preprogrammed pulsatile control

Among the various blood pump control strategies, the preprogrammed pulsatile control is effective in maintaining the pulsatility of the assisted cardiovascular system. Therefore, we adopt the preprogrammed pulsatile control strategy and adjust its parameters to improve circulatory requirements and organ perfusion.

Based on the hydraulic characteristics of the pVAD, the pump speed is set to pulse between 20,000 and 34,000 rpm (e.g., increasing the speed to 34,000 rpm during systole and decreasing to 22,000 rpm during diastole).

3.4. Design of the DRL controller

The objective of the DRL framework in this section is to provide a dynamic and adaptive pulsatile control strategy for the pVAD. This strategy aims to ensure that, under varying heart failure severity, heart rate fluctuations, and pathophysiological disturbances, the system can both stabilize core hemodynamic indicators ($\text{MAP} \geq 75 \text{ mmHg}$; $\text{Cardiac Output (CO)} \geq 4.5 \text{ L/min}$) and accurately reproduce physiological pulse waveform characteristics. Additionally, it should exhibit cross-patient generalization (adapting to different heart

failure levels and heart rate variations of 60–150 bpm without additional training).

The following sections will explain how the DRL controller is designed, including the design of the state space, action space, and safety constraints, as well as the design of the reward function and network structure.

3.4.1. State space definition

The state space needs to capture key physiological dynamic information to enable the DRL controller to perceive the current blood flow state. We define the state vector s_t at time t as the fusion of multimodal signals:

$$s_t = [P_{\text{AOP}}(t - \Delta t : t), P_{\text{AF}}(t - \Delta t : t), \text{HR}(t), LVP(t), Q_{\text{pump}}(t)] \quad (1)$$

where $P_{\text{AOP}}(t - \Delta t : t)$ is the AOP values from the past 50 time points, capturing the pulsatile characteristics of AOP. $P_{\text{AF}}(t - \Delta t : t)$ is the AF values from the past 50 time points. $\text{HR}(t)$ denotes the real-time heart rate of the patient. $LVP(t)$ represents the current left ventricular pressure. $Q_{\text{pump}}(t)$ is the current flow rate of the assist pump.

There are several points to explain regarding the observation signal design: The waveform signals P_{AOP} and P_{AF} provide pulsatile phase synchronization information, which helps to a certain extent in avoiding pVAD-heart cycle mismatch. Real-time $\text{HR}(t)$ and $LVP(t)$ help assess cardiac preload and contractile potential, preventing suction events. The instantaneous pump flow value $Q_{\text{pump}}(t)$ reflects the direct effect of pump speed on blood flow.

3.4.2. Action space and safety constraints

After designing the state space, we need to design the action space, which defines the physical control commands that the DRL controller can output. The DRL controller outputs the pump speed every 10 ms, specifically controlling the pump's rotational speed $\text{rpm} \in [20000, 34000]$. To ensure clinical safety, constraints must be set: when the left ventricular pressure drops below the threshold (left ventricular end-diastolic pressure ($\text{LVEDP} < 5 \text{ mmHg}$)), the pump speed is forcibly reduced to avoid ventricular collapse [35,36]. Moreover, during the training process, a gradient penalty term should be introduced to suppress severe speed fluctuations, ensuring smooth and stable control.

3.4.3. Reward function design

In reinforcement learning, the reward function is the only direct feedback guiding the agent's behavior and determines the optimization goal of reinforcement learning [37]. The ultimate goal of the agent is to maximize the expected cumulative reward, so the design of the reward function directly influences the agent's behavior direction. In this study,

the reward function needs to balance clinical objectives and engineering feasibility. We propose a hybrid multi-objective reward structure that addresses the following core clinical goals:

Maintaining hemodynamic stability: Mean arterial pressure (MAP) and cardiac output (CO) must meet tissue perfusion needs.

Restoring physiological pulsatile waveforms: The AOP waveform and AF waveform should closely match healthy reference waveforms in terms of temporal features. **Adapting to different patient conditions:** The system should be robust to variations in heart failure severity and heart rate.

To achieve these goals, the reward function consists of the following components: Hemodynamic stability reward term R_1 , Pulsatile waveform similarity reward term R_2 , Control stability and safety penalty term R_3 .

Hemodynamic Stability Reward R_1 :

$$R_1 = w_1 \cdot e^{-\alpha |MAP - MAP_{target}|} + w_2 \cdot e^{-\beta |CO - CO_{target}|} \quad (2)$$

Here, MAP_{target} and CO_{target} represent the target values of MAP and CO. An exponential decay form is used as the reward function to ensure high gradient density near the target values, accelerating convergence. Compared with squared error ($R = -(error)^2$), the exponential form imposes a stronger gradient when far from the target, preventing the strategy from falling into local optima. The decay coefficients α and β are chosen as $\alpha = \beta = 0.15$.

Pulsatile Waveform Similarity Reward R_2 :

$$R_2 = w_3 \cdot \frac{1}{1 + \sum_{(i,j) \in \pi} |AOP_{current}(i) - AOP_{ref}(j)|} + w_4 \cdot \frac{1}{1 + \sum_{(i,j) \in \pi} |AF_{current}(i) - AF_{ref}(j)|} \quad (3)$$

To make the AOP waveform and AF waveform resemble the physiological waveform of a healthy individual, we use Dynamic Time Warping (DTW) to calculate the reward. DTW measures the similarity between two time series waveforms, solving the issue of phase delay that traditional Mean Squared Error cannot handle. To accelerate computation, we employ the FastDTW algorithm, which reduces the time complexity from $O(n^2)$ to approximately $O(n)$ while maintaining high alignment accuracy.

Control Stability and Safety Penalty R_3 :

$$R_3 = R_{rpm} + w_6 \cdot R_{LVEDP} \quad (4)$$

Here, R_{rpm} limits excessive pump speed fluctuations, and R_{LVEDP} penalizes LVEDP falling below the critical value of 5 mmHg to prevent ventricular suction.

The speed fluctuation penalty is defined as:

$$R_{rpm} = \begin{cases} 0 & |\Delta rpm| < 500 \\ -w_5 \cdot |\Delta rpm| & |\Delta rpm| > 500 \end{cases} \quad (5)$$

The LVEDP suction penalty is set as:

$$R_{LVEDP} = -50 \text{ LVEDP} < 5 \text{ mmHg} \quad (6)$$

This multi-objective reward function comprehensively integrates hemodynamic stability, waveform similarity, and safety considerations, guiding the agent to learn clinically safe and effective control strategies.

3.4.4. Network architecture overview

Fig. 5(a) illustrates the structures of the Actor network in the DDPG-LSTM algorithm used in this study and **Fig. 5(b)** shows the structures of the Critic network. The Actor network consists of two input branches: a **temporal input branch** and a **scalar input branch**. The **temporal input branch** handles sequential data, including historical AOP values and AF values over the past 50 time steps. An LSTM layer with 64 neurons is utilized to capture the temporal features in the physiological

signals, followed by layer normalization to enhance stability and prevent gradient vanishing or explosion. The **scalar input branch** processes scalar signals, including the current HR, LVP, and Q_{pump} , through a fully connected (FC) layer containing 16 neurons, followed by layer normalization to maintain numerical stability. After feature extraction from both branches, the outputs are concatenated and further processed through two FC layers, each comprising 256 neurons. The final output layer generates the action, specifically the pump speed, which lies within the defined range of [20,000, 34,000] rpm.

The Critic network follows a similar architectural design and also

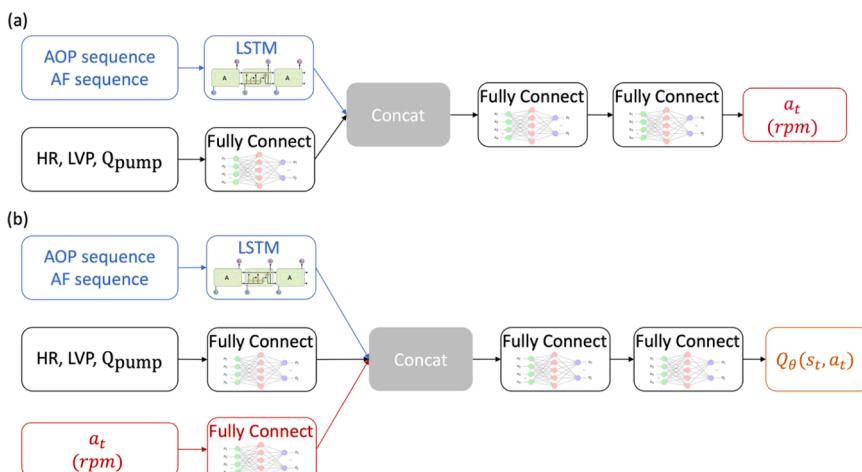


Fig. 5. Architecture of the Actor network and Critic network in DDPG-LSTM.

consists of two input branches: a state input branch (s_t) and an action input branch (a_t). The state input branch mirrors the structure of the Actor network to maintain consistency in feature extraction, where temporal inputs (AOP and AF) are processed through an LSTM layer with 64 neurons and layer normalization, while scalar inputs (HR, LVP, and Q_{pump}) are processed through an FC layer with 16 neurons. The action input branch takes the action signal (pump speed) as input and processes it through an FC layer with 16 neurons. Subsequently, the outputs from the state and action branches are concatenated and passed through two additional FC layers, each containing 256 neurons. The final output layer produces a scalar value representing the estimated Q-value, which is utilized to evaluate the quality of the current action given the observed state. The integration of LSTM layers to capture temporal features and FC layers to handle scalar data allows the Actor and Critic networks to accurately model the complex dynamics of heart pump speed control.

3.4.5. Training process description

The training of the DRL controller in this study was conducted within the SIMULINK environment. Both the Critic and Actor networks were trained using GPU-based parallel processing over 20,000 episodes, which took approximately 2 days. For each training episode, the cardiovascular system parameters were randomly initialized to simulate diverse pathological conditions such as varying degrees of heart failure and heart rates.

Notably, imitation learning was incorporated into the training process. Since DDPG relies on random exploration, which can result in excessively long training durations if started from scratch, imitation learning was introduced at the early stage to pre-train the Actor network. Specifically, the preprogrammed pulsatile control was used as the expert strategy. The preprogrammed pulsatile control strategy was applied to control the pVAD within the cardiovascular-pump hybrid simulation model, during which the pump speed a and sensor signals $s = [\text{AOP}, \text{AF}, \text{LVP}, \text{HR}, Q_{\text{pump}}]$ were recorded. These state-action pairs (s, a) formed the expert dataset, which was used to supervise the training of the Actor network by minimizing the action error. This pre-training effectively reduced the exploration burden during the initial training phase of DDPG. A comprehensive list of all cardiovascular model parameters, physiological constants, and DRL controller hyperparameters used in this study is provided in the **Supplementary Material** (Table 2).

After the pre-training phase, the DDPG algorithm was allowed to interact with the cardiovascular-pump hybrid simulation model, ultimately resulting in a deployable DRL controller for subsequent simulation and experimental testing.

To ensure the DRL controller learns a truly adaptive policy rather than merely memorizing specific states, we employed a domain randomization strategy during training. For each training episode, key cardiovascular system parameters were randomly sampled from wide, continuous ranges defined in **Supplementary Table 2(a)**. This method forced the agent to experience a diverse spectrum of pathological conditions, covering a continuum from near-healthy to severe heart failure, thereby fostering a robust and generalizable control policy.

The specific cardiac profiles used for evaluation in **Section 4** (i.e., 'Moderate HF' and 'Severe HF') are discrete, fixed parameter sets. While their parameter values fall within the training randomization ranges, these exact configurations were held out and not explicitly used during the training phase. This experimental design constitutes a valid test of the controller's ability to interpolate and generalize to specific, unseen points within the learned pathological distribution. Furthermore, the subsequent tests involving abrupt heart rate and ejection phase shifts serve as additional out-of-distribution assessments of the controller's temporal adaptability.

To validate the stability and reproducibility of this training methodology, the entire process was repeated five times ($N = 5$), each initiated with a different random seed. The resulting learning convergence is presented in **Fig. 6**. The curve shows a consistent upward trend, starting

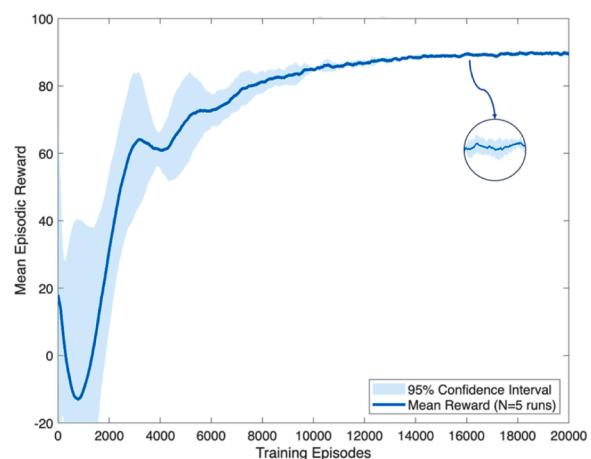


Fig. 6. Learning curve of the DRL controller over 20,000 training episodes.

from a competent baseline established by imitation learning and steadily improving as the agent explores. The tight 95 % confidence interval, which narrows as the training progresses, demonstrates the high stability and low variance of our training approach, confirming that it reliably leads to an effective controller.

3.5. Statistical analysis and reproducibility

To rigorously evaluate the performance and consistency of the DRL controller and to ensure the reproducibility of our findings, we conducted statistical repeat experiments in both the simulation and experimental platform environments.

Simulation Environment: In the SIMULINK environment, each evaluation scenario was repeated five times. To ensure statistical independence while maintaining reproducibility, each of the five runs was initiated with a unique random seed from a predefined set. For the representative time-series plots shown in the figures in the next Section, a fixed random seed of 42 was used. This strategy allows for the assessment of the DRL algorithm's average performance and its variance due to the inherent stochasticity in training and execution (e.g., from OU-noise and network initializations).

Experimental Platform: For the experiments on the pulsatile mock circulatory loop, a five-repeat protocol was also employed for each scenario. To ensure consistent initial conditions, key platform parameters defining each physiological state—including the linear motor's operating amplitude, the flow resistor's value, and the compliance chamber's air pressure—were first calibrated, recorded, and saved. Each of the five trials for a given scenario was then initiated using these identical pre-set parameters. This method allows for the evaluation of the controller's robustness and performance consistency against the inherent noise and variability of a real-world physical system.

All quantitative results from both simulation and experiments reported in this study are presented as the mean \pm standard deviation (\pm SD) calculated from these five independent runs.

3.6. Criteria for performance evaluation

To establish a clear and rigorous framework for evaluating the DRL controller against the baseline strategies, we predefined the following criteria for declaring clinical superiority, non-inferiority, and safety within the context of this pre-clinical study.

Superiority in Restoring Physiological Pulsatility: The primary objective of the DRL controller is to overcome the limitations of conventional methods by restoring natural pulsatility. Therefore, the DRL controller was deemed superior if it demonstrated statistically significant improvements in key pulsatility metrics compared to both constant

speed and preprogrammed pulsatile controls. The specific criteria were:

A significantly higher Pulse Pressure (PP) and Pulsatility Index (PI).

A significantly lower Dynamic Time Warping (DTW) distance for both AOP and AF waveforms when compared against the healthy reference waveforms.

Non-inferiority in Hemodynamic Perfusion: While enhancing pulsatility, the DRL controller must ensure that fundamental perfusion is not compromised. The controller was considered non-inferior if its ability to maintain core hemodynamic targets was not clinically significantly worse than the baseline controllers. The criteria were:

Mean Arterial Pressure (MAP): More closely aligns with the healthy physiological reference.

Cardiac Output (CO): More closely aligns with the healthy physiological reference.

Safety Criterion (Suction Avoidance): A critical safety requirement for any VAD controller is the prevention of ventricular suction. The controller's safety performance was evaluated based on its adherence to the suction avoidance rule defined in our model:

Suction Index: The controller must maintain the Left Ventricular End-Diastolic Pressure (LVEDP) consistently above the safety threshold of 5 mmHg. Any instance of LVEDP dropping below this threshold would be registered as a suction event.

These predefined criteria form the basis for the comparative analysis and conclusions presented in [Section 4.4](#).

4. Experimental validation and results

This section systematically presents the validation results for our DRL controller, with each subsection designed to address a key research question and demonstrate the controller's capabilities in a logical progression. The workflow is as follows: First, we present the **experimental hydraulic characterization** of the pVAD itself ([Section 4.1](#)), which provides the foundational model for our subsequent simulations and experiment. Next, we evaluate the controller's performance against the baseline strategies in the **in-silico simulation environment** ([Section 4.2](#)) [12] and then validate these findings on the **in-vitro experimental platform** ([Section 4.3](#)). A detailed quantitative comparison across all key metrics is subsequently provided to objectively assess the controllers ([Section 4.4](#)). Finally, we analyze the DRL controller's **robustness and adaptability** under various dynamic perturbations to demonstrate its resilience ([Section 4.5](#)).

Before presenting these results, it is important to define the key conditions tested. In our study, the **severity of heart failure (HF)** is quantitatively defined by a set of hemodynamic parameters, benchmarked against standard clinical classifications. 'Moderate HF' approximates NYHA Class III (e.g., MAP \approx 75 mmHg, CO \approx 3.5 L/min), while 'Severe HF' approximates NYHA Class IV (e.g., MAP $<$ 70 mmHg, CO \leq 2.5 L/min). The term '**ejection phase**' used in our perturbation tests refers specifically to the **systolic phase** of the cardiac cycle.

4.1. Pulsatile mock circulatory loop platform and experimental methods

To validate and compare the performance of the different control strategies in a physically realistic environment, we constructed a comprehensive pulsatile mock circulatory loop platform, as shown in [Fig. 7](#). This platform is designed to simulate the hemodynamics of the human left heart circulatory system under various conditions, including both a healthy state and different severities of heart failure.

The platform is a closed-loop system with its primary components assembled as follows. A programmable linear motor (Model VCAH0445-0300-00A, Suzhou Yuwen Technology Co., Ltd., Suzhou, China) drives a piston within the left ventricle chamber to generate pulsatile flow, simulating cardiac contraction and relaxation. Key hemodynamic signals are continuously measured using high-precision sensors. Specifically, aortic and left ventricular pressures are monitored by pressure transducers (Model JIBPT-01-UT, Shenzhen Jinchen-gui Medical Equipment Co., Ltd., Shenzhen, China), which feature a working pressure range of -50 to +300 mmHg and a frequency response exceeding 200 Hz. Aortic flow is measured by a custom-made ultrasonic flowmeter (Model IL-100SX-B-G2, Shanghai Xunyinke Technology Co., Ltd., Shanghai, China). The systemic compliance and peripheral resistance of the circulatory system are simulated using a sealed air chamber and an adjustable flow resistor, respectively. During experiments, the pVAD is positioned across the aortic valve region to provide active circulatory support. [Fig. 7\(a\)](#) shows the complete cardiovascular loop system, [Fig. 7\(b\)](#) illustrates the transvalvular transport driven by the pVAD, and [Fig. 7\(c\)](#) presents the physical view of the pVAD used in this study. The hemodynamic characteristics of the simulated healthy and failing hearts using the platform are also illustrated in the [Fig. 4](#). To ensure the physiological relevance of the hemodynamic conditions, a 40 % glycerol-water mixture by weight was used as the working fluid. At the operating temperature of our laboratory (approximately 37 °C), this

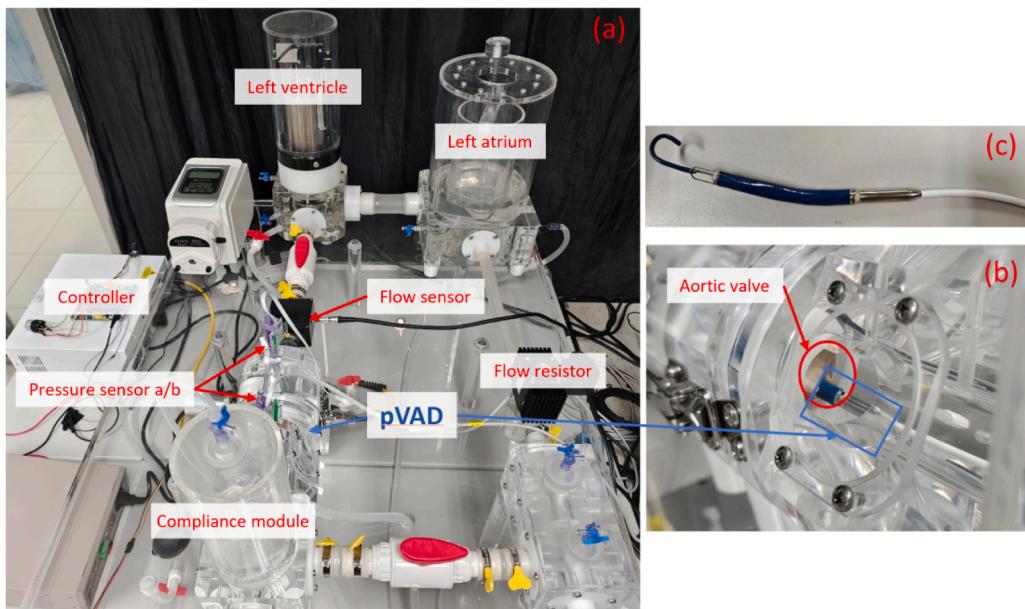


Fig. 7. The experimental platform and the pVAD.

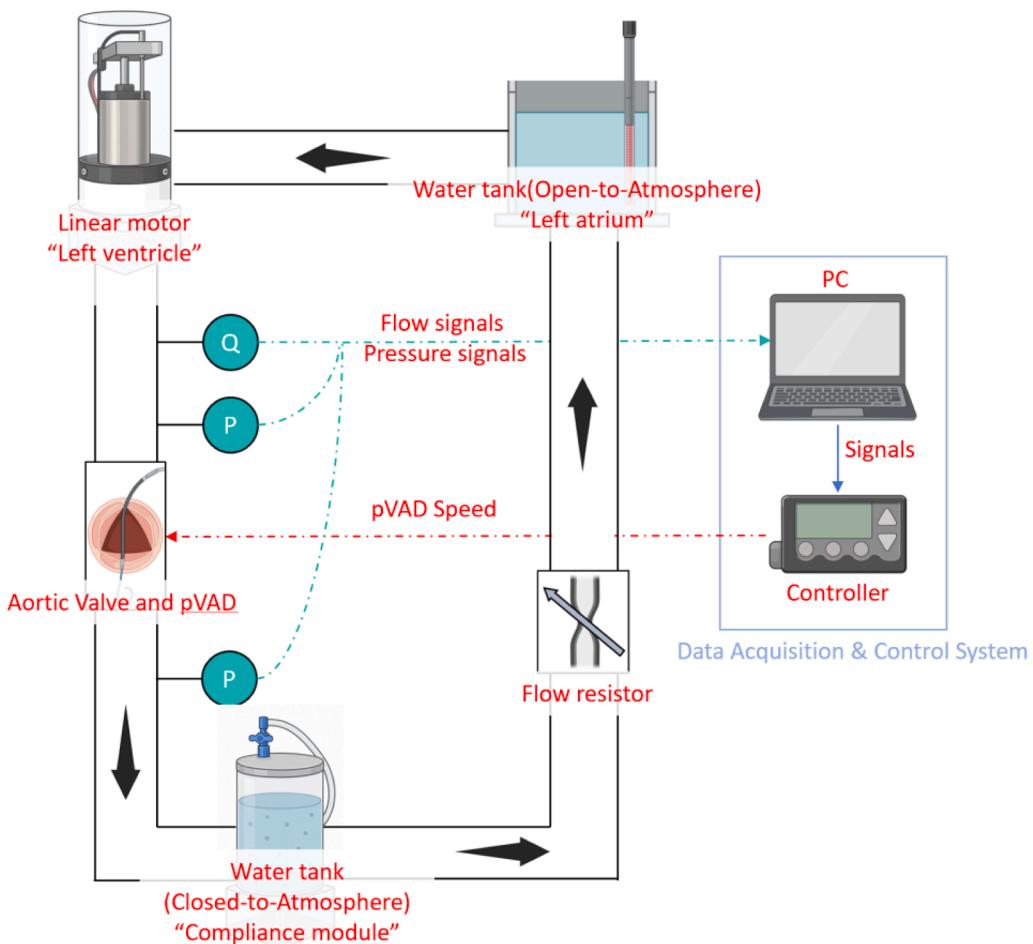


Fig. 8. Schematic of the experimental circulatory platform.

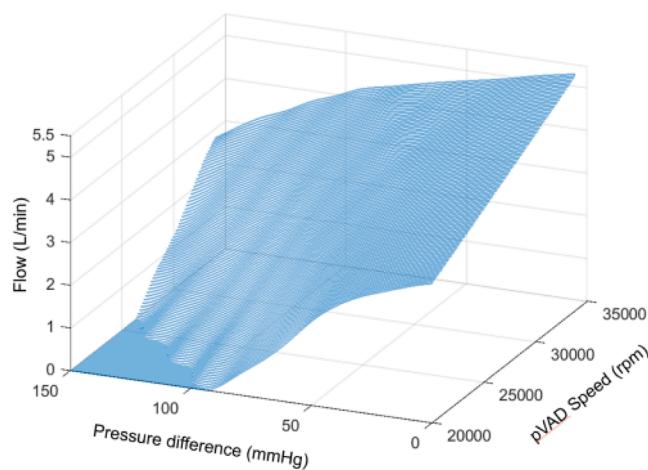


Fig. 9. Hydraulic characteristics surface of the pVAD.

solution has a dynamic viscosity of approximately 3.5 mPa·s, which closely mimics the viscosity of human blood at body temperature (typically in the range of 3–4 mPa·s), thus validating the fluid-dynamic aspects of our experimental setup.

The data acquisition and control flow of the system is illustrated in Fig. 8. Real-time hemodynamic signals from the pressure and flow sensors are acquired at a sampling frequency of 100 Hz via a National Instruments data acquisition board and transmitted to a host PC. The PC processes these signals and sends the state information to the external DRL controller. The controller then computes the appropriate pVAD speed command based on its learned policy and drives the pump motor accordingly, thus closing the real-time control loop.

The detailed experimental protocol for generating the pVAD's hydraulic characteristic surface (Fig. 9), including the test bench setup and measurement steps, is presented in the **Supplementary Material**. The hydraulic performance surface of the pVAD is shown in the Fig. 9. This surface allows quick estimation of the flow rate Q based on a given pressure difference Δp and pump speed n , and can then be directly used as the input current to the SIMULINK simulation.

4.2. Test results in the simulation environment

The baseline controllers referred to in this section are the constant speed controller and preprogrammed pulsatile controller. In this section, we first configure the cardiovascular model with parameters representing a healthy individual to simulate a healthy heart. We then adjust the model to represent a moderate HF condition and a severe HF condition. Finally, we modify the heart rate and the phase of the cardiac

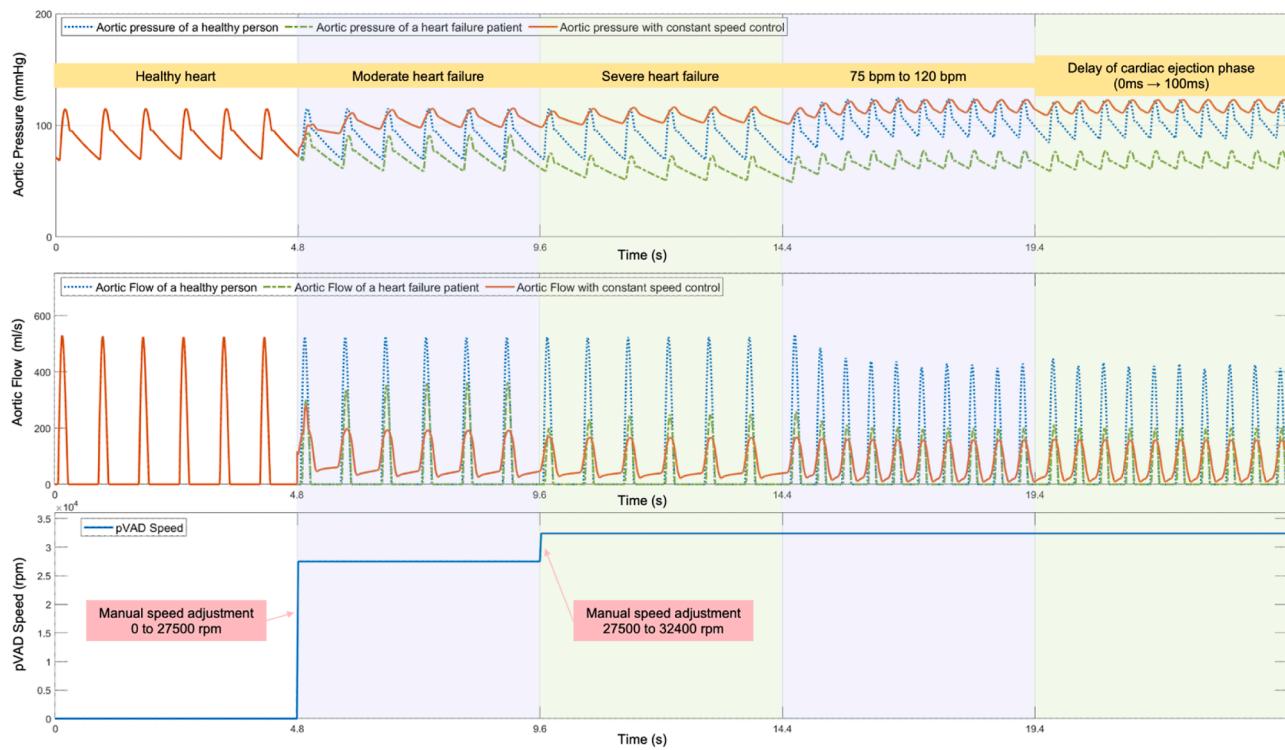


Fig. 10. Simulation results (AOP and AF) and pump speed variation under constant speed control.

ejection period.

At $t = 0$, the cardiovascular system is set to a healthy heart condition (initial heart rate: 75 bpm, initial blood pressure: 115/70 mmHg, initial CO: 4.73 L/min, initial stroke volume: 63 mL).

During $t \in [0, 4.8]$, the cardiovascular system operates under healthy conditions.

During $t \in (4.8, 9.6]$, the system is adjusted to a moderate heart

failure condition (blood pressure: 92/60 mmHg, CO: 3.29 L/min, stroke volume: 44 mL).

During $t \in (9.6, 14.4]$, the system is adjusted to a severe heart failure condition (blood pressure: 72/50 mmHg, CO: 2.19 L/min, stroke volume: 29 mL).

During $t \in (14.4, 19.4]$, the heart rate is increased from 75 bpm to 120 bpm.

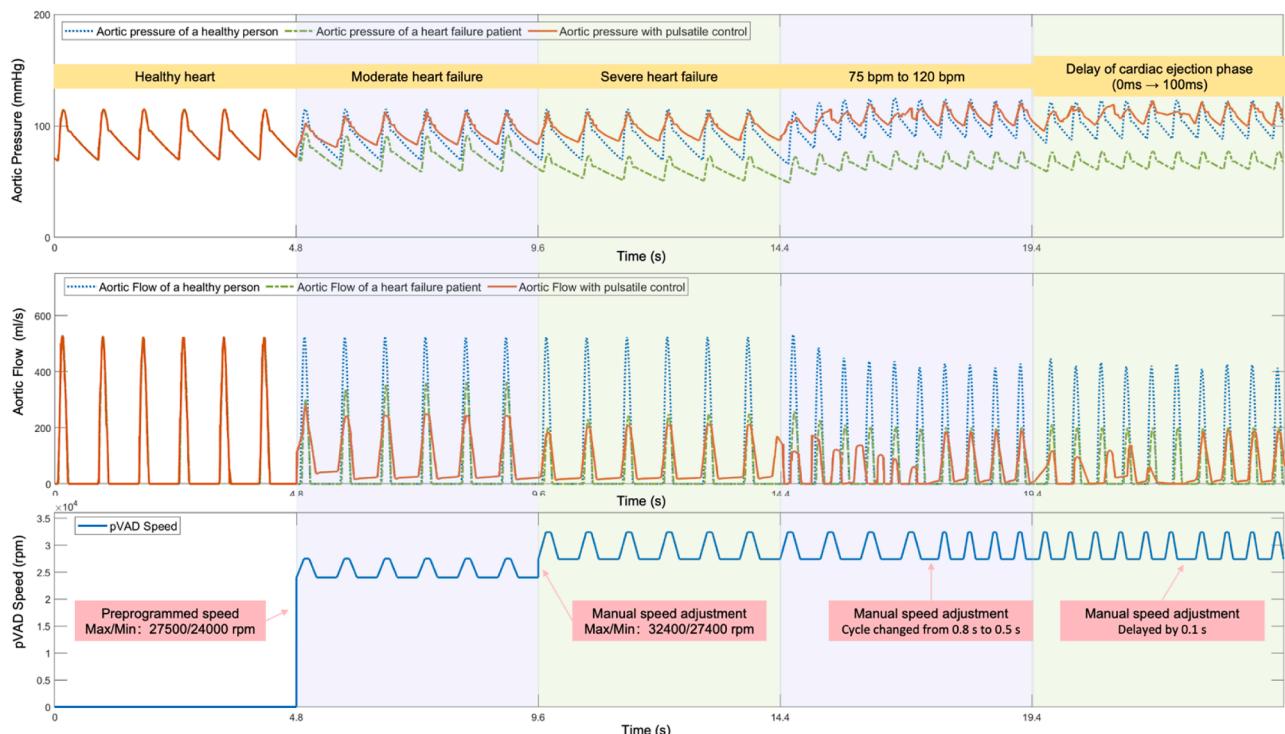


Fig. 11. Simulation results (AOP and AF) and pump speed variation under pulsatile control.

During $t \in (19.4, 24.4]$, the cardiac ejection timing is delayed by 0.1 s.

4.2.1. Constant speed control results

We first present the results of constant speed control. The AOP waveform, AF waveform and pump speed variations are shown in the Fig. 10. At the beginning, under healthy heart conditions, the pump speed was manually set to 0 rpm, and both AOP and AF met healthy perfusion demands. When the condition was changed to moderate heart failure, the pVAD speed was manually set to 27,500 rpm. It can be observed that under constant speed control, the pulsatility of the blood pressure is significantly reduced. When further adjusted to severe heart failure, the pump speed was manually increased to 32,400 rpm. After increasing the heart rate from 75 bpm to 120 bpm, the pump speed remained unchanged. Similarly, after manually delaying the cardiac ejection time by 100 ms in each cycle, the pump speed was still kept constant. The AOP and AF results indicated that constant speed control can restore the hemodynamic indicators of cardiac perfusion. However, the pulsatility was significantly reduced, and the controller lacked adaptability to varying degrees of heart failure, requiring manual adjustment of pump speed.

4.2.2. Preprogrammed pulsatile control results

The results of preprogrammed pulsatile control are shown in the Fig. 11. Under the initial healthy heart condition, the pump speed was manually set to 0, and both AOP and AF met the healthy perfusion demand. When the heart was changed to a moderate heart failure state, the pVAD speed was preprogrammed to vary between 24,000 rpm and 27,500 rpm in synchronization with the heartbeat frequency and the systolic phase within each cardiac cycle. Compared to constant speed control, pulsatile control improved the pulsatility of blood pressure, although it was still somewhat diminished. In the severe heart failure state, the pVAD speed was manually adjusted to vary between 27,400 rpm and 32,400 rpm.

When the heart rate increased from 75 bpm to 120 bpm, the pulsatile control settings initially remained unchanged. As a result, a significant mismatch occurred between the pVAD's pulsatile cycle and the cardiac

cycle, leading to loss of synchrony and disordered hemodynamics—AOP and AF lost their periodic pattern. After manually adjusting the pulsation frequency of the control to match the new heart rate, the periodicity of AOP and AF was restored.

When the cardiac ejection phase was delayed by 100 ms per cycle, the pVAD's pulsatile pattern remained unchanged at first. Similar to the previous case, this caused a mismatch and disrupted the hemodynamic performance. After manually delaying the pulsatile output by 100 ms to re-align with the new ejection timing, the hemodynamic performance returned to healthy.

These results show that pulsatile control can effectively restore the heart's hemodynamic characteristics and enhance pulsatile characteristics. However, it cannot adapt to different degrees of heart failure without manual adjustment of the pump speed, and more critically, it fails to adapt to changes in heart rate and ejection timing without manual tuning of the pulsation frequency and phase.

4.2.3. DRL control results

The results of the DRL control are shown in the Fig. 12. Under the initial healthy heart condition, the pump speed was manually set to 0, and both AOP and AF met the healthy perfusion demand. When the system transitioned to a moderate heart failure state, the DRL controller automatically adjusted the pVAD speed based on sensor feedback, modulating the speed within a single cardiac cycle between 20,000 rpm and 27,000 rpm.

In the severe heart failure state, the DRL controller detected the change in heart condition and autonomously adjusted the pump speed within two cardiac cycles, varying it between 22,000 rpm and 32,740 rpm. When the heart rate increased from 75 bpm to 120 bpm, the DRL controller automatically updated the pVAD speed within three cardiac cycles, achieving synchronization with the patient's heart rate.

When the cardiac ejection phase was manually delayed by 100 ms per cycle, the DRL controller adapted within two cycles, automatically modifying the pump speed to match the delayed ejection timing.

From the results of AOP and AF, it is evident that the DRL-controlled heart under heart failure conditions achieves superior outcomes in both hemodynamic characteristics and pulsatility characteristics compared to

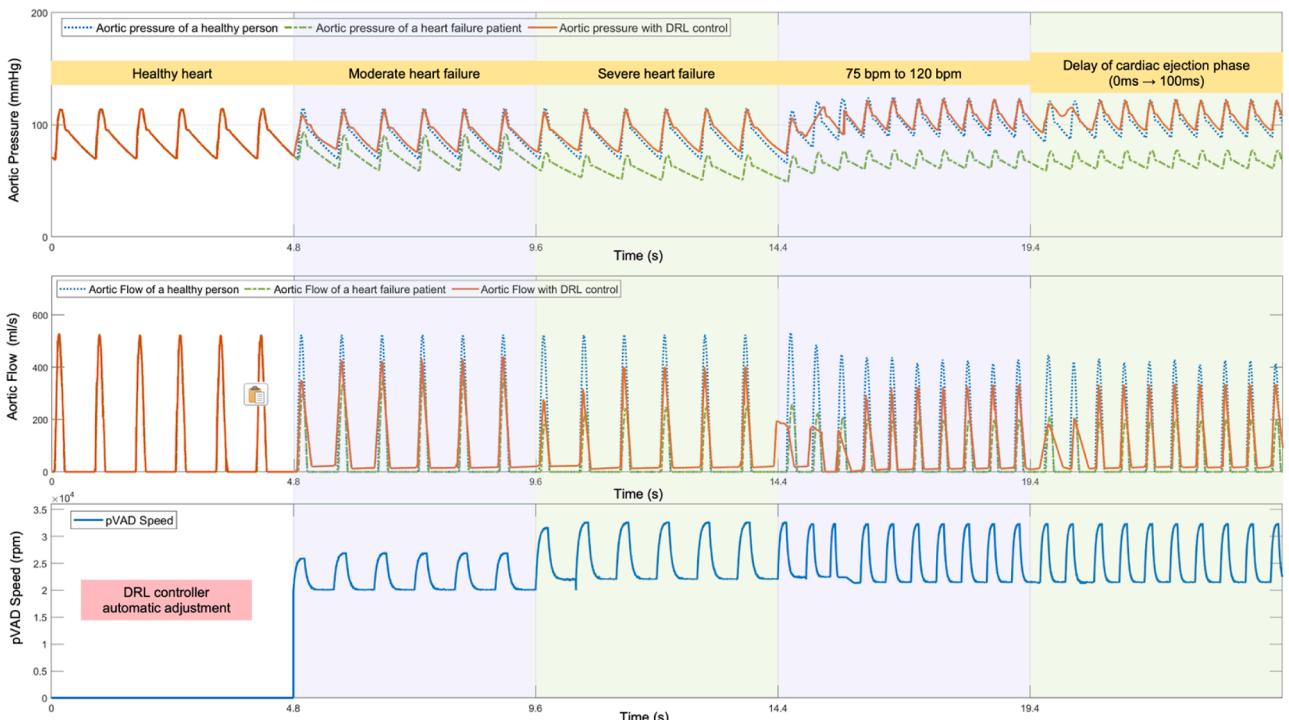


Fig. 12. Simulation results (AOP and AF) and pump speed variation under DRL control.

baseline controllers. The outputs closely resemble that of a healthy heart.

4.3. Experimental platform test results

Using the experimental platform as shown in the Fig. 7, we evaluated the performance of the pVAD under different controllers, examining the restoration of hemodynamic and pulsatility characteristics. During the experiments, the controller and motor driver were placed outside the experimental platform, while the pVAD was installed inside the platform to facilitate transvalvular fluid transport. A glycerol-water mixture was used as the working fluid to simulate the viscosity of blood. The controller was connected to the pump via wires, and both the baseline control strategy and the DRL control algorithm were preloaded into the controller for comparison. During the experiments, the DRL control algorithm was deployed on an external embedded controller to drive the pVAD in real-time. A detailed description of the embedded deployment toolchain, hardware specifications are provided in the **Supplementary Material**.

Flow rate and pressure data, along with heart rate and other signals collected by sensors on the platform, were transmitted in real-time to the controller via a computer to enable adaptive pump speed control.

Following a procedure similar to the simulation tests, we initially set the experimental platform to simulate a healthy heart at $t = 0$ (initial heart rate = 75 bpm, initial blood pressure = 112/76 mmHg, initial CO = 5.0 L/min, initial stroke volume = 69 mL). The heart remained in the healthy condition during $t \in [0, 4.8]$. At $t \in (4.8, 9.6]$, the heart was set to a moderate heart failure state (blood pressure = 91/64 mmHg, CO = 3.5 L/min, stroke volume = 47 mL). From $t \in (9.6, 14.4]$, the heart was adjusted to a severe heart failure condition (blood pressure = 73/54 mmHg, CO = 2.6 L/min, stroke volume = 36 mL). At $t \in (14.4, 19.4]$, the heart rate was increased from 75 bpm to 120 bpm. Finally, at $t \in (19.4, 24.4]$, the cardiac ejection phase was manually delayed by 0.1 s.

The control results of the HF under different controllers are shown in Fig. 13, Fig. 14, Fig. 15. The data collected from the experimental

platform demonstrate clear differences in the improvements of hemodynamic and pulsatile performance between the baseline controllers and DRL controller.

Consistent with the results observed in the simulation environment, under the constant speed control, the pump speed was manually adjusted to 29,000 rpm during moderate heart failure and to 32,700 rpm during severe heart failure. The experimental results show that this controller eliminates most of the pulsatility in the AOP. Although the total AF improved significantly, the pulsatile nature of the output was largely diminished.

For the preprogrammed pulsatile control, the AOP and AF of the failing heart were better restored, and pulsatility was well preserved. However, when there were significant changes in heart rate or ejection timing, phase mismatches between the native heart and the pump output occurred, leading to disordered AOP and AF. Manual adjustments were required to restore proper synchronization.

In contrast, under the DRL control, both the AOP and AF of the failing heart were effectively restored in terms of hemodynamic and pulsatility. Moreover, when sudden changes occurred in heart rate or ejection timing, the DRL-controlled pVAD was able to adaptively adjust pump speed based on real-time pressure and flow measurements, thereby restoring hemodynamic and pulsatile performance without the need for manual intervention.

4.4. Comparison of different controllers

Through both simulation testing and experimental platform validation, we found that the pVAD exhibits significantly different performance in improving heart failure conditions under the baseline controllers versus the DRL controller. The previous sections have shown the AOP and AF waveforms for the healthy heart, the failing heart, and the heart assisted by the pVAD under different controllers.

To further compare the performance of these controllers, this section analyzes the results from two key dimensions: **hemodynamic characteristics** and **pulsatility characteristics**.

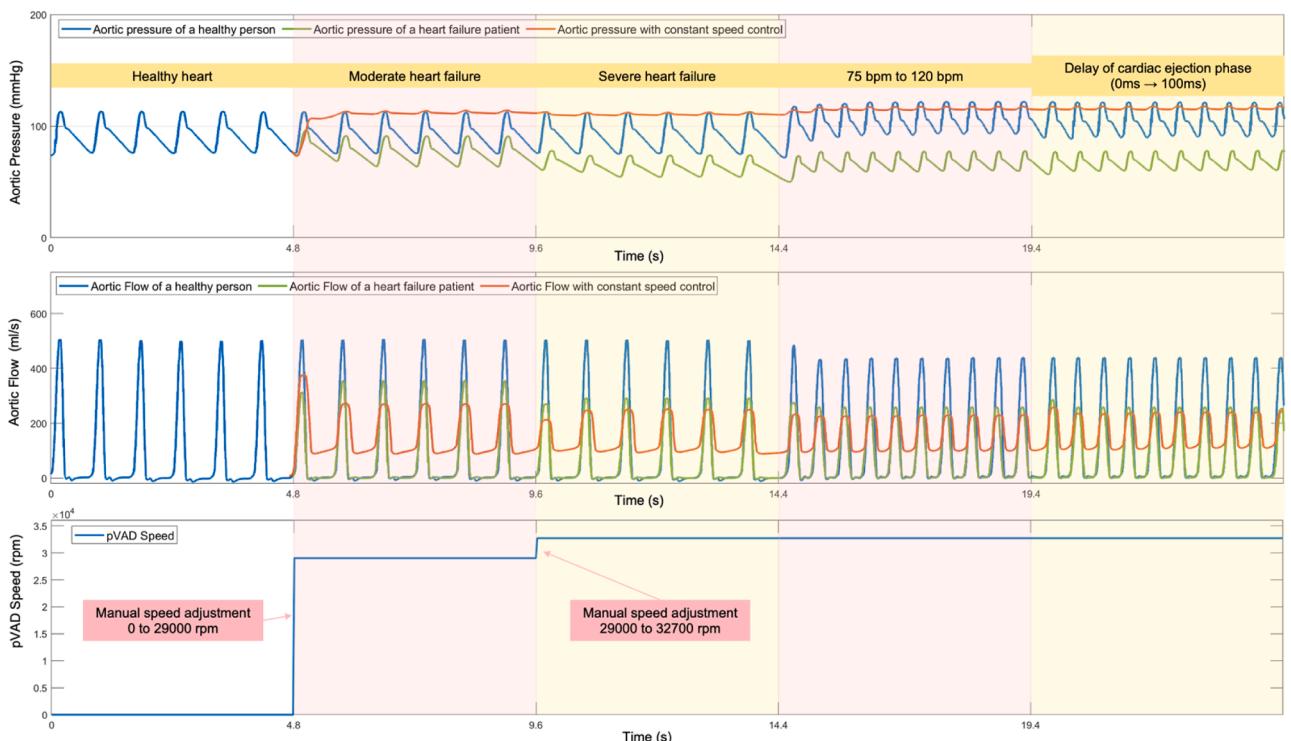


Fig. 13. Experimental results (AOP and AF) and pump speed variation under constant speed control.

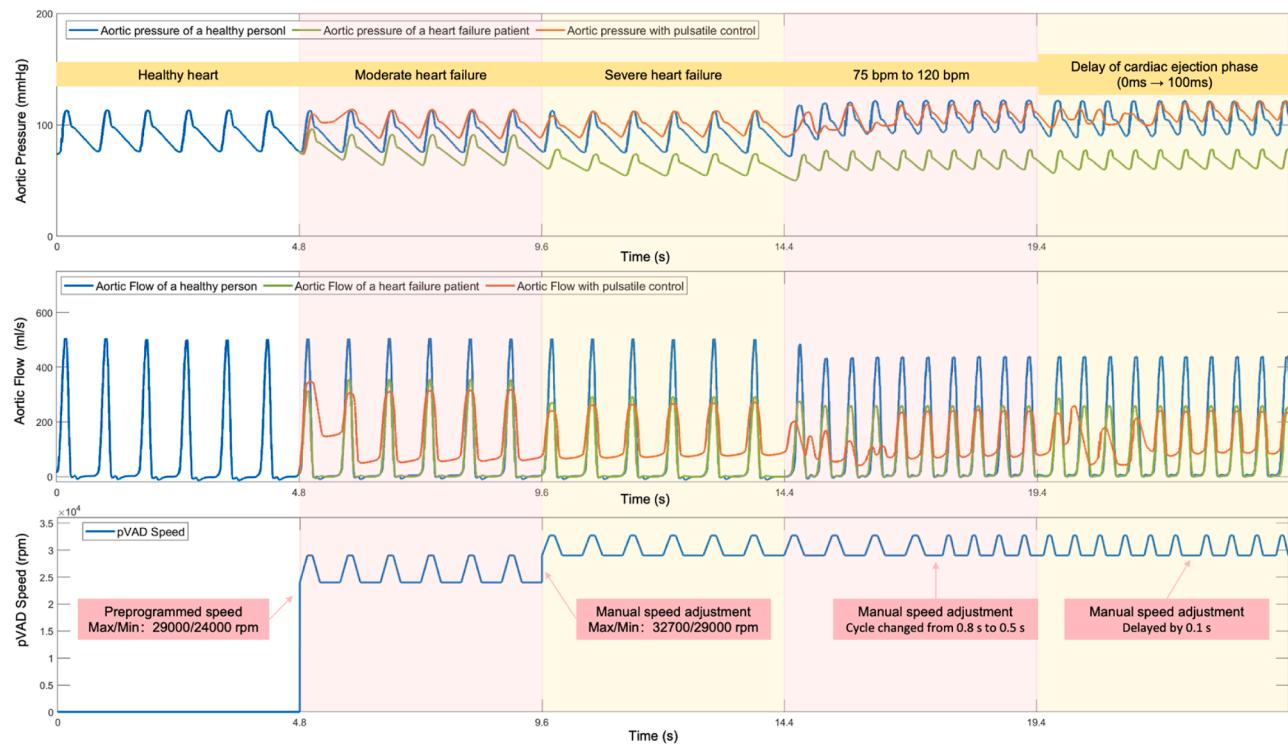


Fig. 14. Experimental results (AOP and AF) and pump speed variation under pulsatile control.

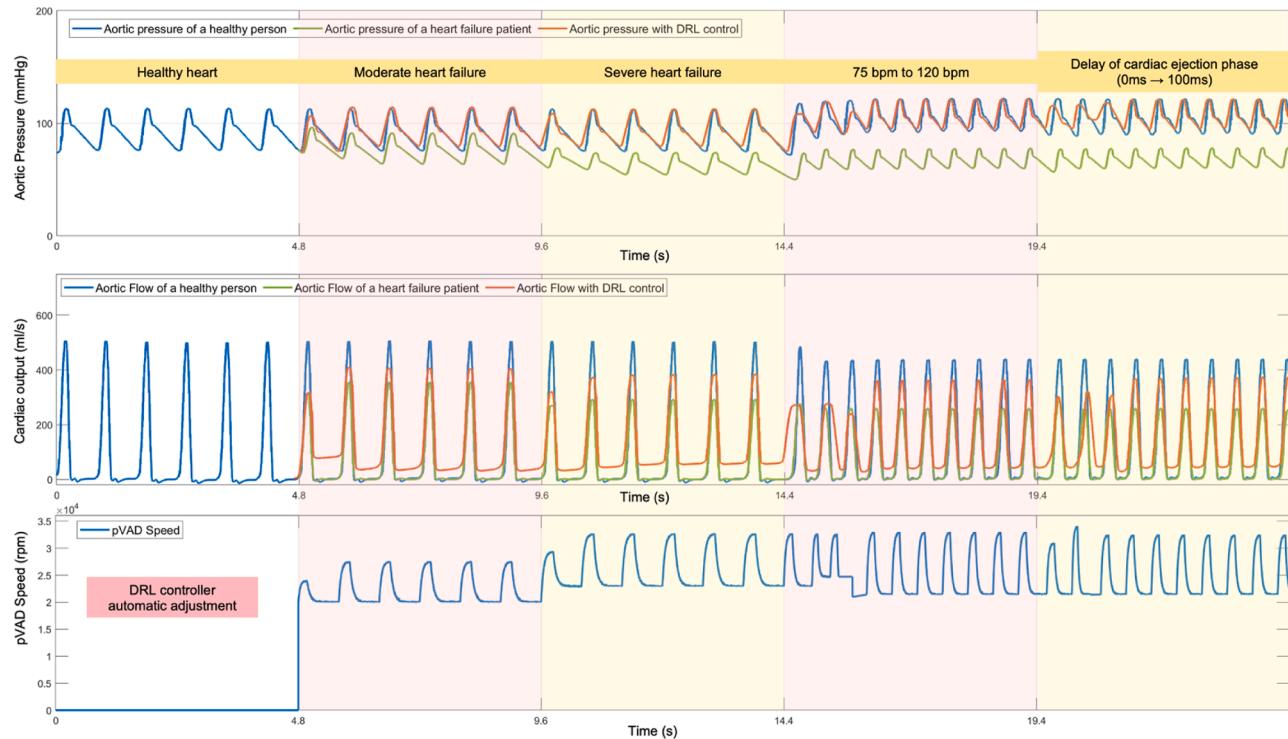


Fig. 15. Experimental results (AOP and AF) and pump speed variation under DRL control.

In terms of hemodynamic performance, we evaluated the recovery of MAP and CO when assisted by the pVAD. These metrics reflect how effectively each controller improves perfusion capability. MAP refers to the time-averaged AOP over one cardiac cycle and is a critical indicator

for assessing the perfusion capability of the heart. CO refers to the average AF per unit time, and it serves as an important indicator of the heart's or artificial heart's ability to maintain systemic circulation.

$$MAP = \frac{1}{T} \int_0^T AOP(t) dt \quad (7)$$

$$CO = \frac{1}{T} \int_0^T AF(t) dt \quad (8)$$

where $AOP(t)$ represents the time-varying AOP, and $AF(t)$ denotes the time-varying AF. T is the duration of one complete cardiac cycle.

In terms of pulsatility performance, we employed two well-established indicators—**Pulsatility Index (PI)** and **Pulse Pressure (PP)**—to evaluate the pulsatility of AOP and AF [14,18]. Additionally, the DTW algorithm is employed to assess the similarity between the AOP waveforms, AF waveforms and the physiological reference waveforms thereby quantifying the controller's ability to generate pulsatile outputs. PI is the ratio of the fluctuation amplitude of the AF to its mean value, used to describe the strength of pulsatility. PP is the difference between the maximum and minimum AOP within a cardiac cycle, directly reflecting the pulsatility and rhythmicity of the AOP.

$$PI = \frac{AF_{max} - AF_{min}}{\bar{AF}} \quad (9)$$

$$PP = P_{max} - P_{min} \quad (10)$$

where AF_{max} and AF_{min} represent the maximum and minimum flow rates within one cardiac cycle, respectively, and \bar{AF} denotes the average flow rate. P_{max} and P_{min} represent the maximum and minimum aortic pressures within one cardiac cycle, respectively.

DTW is a distance metric that measures the similarity between two time-series waveforms; a smaller DTW value indicates a closer match to the reference waveform. To visually illustrate this concept, representative DTW alignment plots for the AOP waveforms under each control strategy are provided in the **Supplementary Material (Figure S1)**. These plots visually confirm that the DRL controller's waveform aligns much more closely with the healthy reference compared to the baseline strategies. The reference waveform here is the AOP and AF waveform of a healthy heart, and normalization is performed based on sequence length. Specifically, **DTW-AOP** represents the DTW value between the AOP waveform of the heart failure condition after pVAD support and reference AOP waveform, while **DTW- AF** denotes the DTW value between the AF waveform of the heart failure condition after pVAD support and reference AF waveform.

The calculated evaluation metrics based on the data obtained from the aforementioned experiments as shown in **Table 1**.

From the metric data in the **Table 1**, it is evident that the DRL controller demonstrates superior hemodynamic maintenance and pulsatile output performance across various abnormal cardiac conditions.

In terms of hemodynamic characteristics, the MAP and CO values generated by the DRL controller are generally closer to the healthy reference values and more reasonable compared to other control strategies. For example, in the case of moderate heart failure, the DRL-controlled MAP is 96.4 mmHg and CO is 5.10 L/min, which closely approximate the healthy references (94.9 mmHg and 5.00 L/min), outperforming constant speed control (MAP 111.9 mmHg) and the preprogrammed pulsatile control (MAP 101.1 mmHg). This trend is similarly observed in other scenarios such as severe heart failure, heart rate variation, and ejection time variation, indicating the superior adaptability and regulatory capability of the DRL controller.

In terms of pulsatility characteristics, the DRL controller significantly outperforms other controllers in both PI and PP, reflecting a better restoration of physiological pulsatile patterns. For example, in the case of moderate heart failure, the PI and PP under DRL control are 1.69 and 34.42 mmHg respectively—markedly higher than those under constant speed control (PI 1.05, PP 3.20 mmHg) and the preset pulsatile strategy (PI 1.54, PP 24.90 mmHg).

Furthermore, in the DTW similarity analysis, the AOP and AF waveform produced by the DRL controller show a higher resemblance to healthy physiological waveforms, with significantly lower DTW values than those of the other controllers. For example, in the moderate heart failure case, the DRL controller achieves a DTW-AOP of 1.17 and DTW- AF of 21.23, far outperforming the constant speed (16.42 and 71.74) and preprogrammed pulsatile strategies (2.72 and 48.96). This consistent difference across various scenarios indicates that the DRL controller not only restores hemodynamic performance but also more effectively reconstructs the natural pulsatility of the heart.

While **Table 1** provides a summary of each controller's absolute performance, a direct quantification of the DRL controller's superiority is essential. To this end, we calculated the effect sizes of the improvements, presented in **Table 2**. This table reports the mean difference in key hemodynamic and pulsatility endpoints between the DRL controller and the two baseline strategies. This analysis clearly presents both the magnitude and the statistical significance of the observed improvements.

Table 1

Comparison of hemodynamic and pulsatility characteristics under different controllers. data are presented as mean \pm standard deviation from $n = 5$ independent runs for the DRL controller.

Scenario	Control Strategy	Hemodynamic Characteristics		Pulsatility Characteristics			
		MAP (mmHg)	CO (L/min)	PI	PP (mmHg)	DTW-AOP (vs. Healthy)	DTW- AF* (vs. Healthy)
Moderate Heart Failure	Constant Speed	111.9	5.29	1.05	3.20	16.42	71.74
	Preprogrammed Pulsatile	101.1	5.26	1.54	24.90	2.72	48.96
	DRL	96.4 \pm 1.2	5.10 \pm 0.07	1.69 \pm 0.05	34.42 \pm 0.7	1.17 \pm 0.57	21.23 \pm 5.07
	Ref (Healthy)	94.9	5.00	2.37	36.87		
Severe Heart Failure	Constant Speed	110.6	5.26	0.96	2.99	15.56	79.81
	Preprogrammed Pulsatile	100.7	5.20	1.24	22.40	3.52	68.45
	DRL	96.1 \pm 1.6	5.07 \pm 0.06	1.67 \pm 0.11	33.05 \pm 1.2	1.29 \pm 0.41	29.21 \pm 5.41
	Ref (Healthy)	94.9	5.00	2.37	36.87		
Heart Rate Variation	Constant Speed	115.4	5.49	0.80	3.36	8.92	65.26
	Preprogrammed Pulsatile	110.1	5.41	1.16	18.74	1.87	53.77
	DRL	108.1 \pm 1.1	5.19 \pm 0.09	1.62 \pm 0.12	26.05 \pm 1.2	0.89 \pm 0.27	17.20 \pm 2.11
	Ref (Healthy)	107.1	5.21	2.17	29.69		
Ejection Time Variation	Constant Speed	116.2	5.47	0.79	3.52	10.27	64.61
	Preprogrammed Pulsatile	111.2	5.40	0.99	17.01	3.14	58.28
	DRL	107.7 \pm 2.1	5.19 \pm 0.05	1.59 \pm 0.19	26.36 \pm 2.7	0.92 \pm 0.29	16.49 \pm 2.66
	Ref (Healthy)	105.4	5.21	2.15	30.05		

* Due to the large difference between the peak values of the AF waveform after using the pVAD and those of the reference waveform, the overall DTW- AF values are relatively high. However, it can still be observed that the DTW- AF value for the DRL controller is significantly lower than that of the baseline controller.

Table 2

Effect size analysis: Improvement in closeness to healthy in hemodynamic and pulsatility characteristics for DRL controller compared to baseline controllers.

Scenario	Comparison	Improvement in Closeness to Healthy*					
		MAP (mmHg)	CO (L/min)	PI	PP (mmHg)	DTW-AOP (vs. Healthy)	DTW- AF (vs. Healthy)
Moderate Heart Failure	DRL vs. Constant Speed	+15.5	+0.19	+0.64	+31.22	+15.25	+50.51
	DRL vs. Preprogrammed Pulsatile	+4.7	+0.16	+0.15	+9.52	+1.55	+27.73
Severe Heart Failure	DRL vs. Constant Speed	+14.5	+0.19	+0.71	+30.06	+14.27	+50.60
	DRL vs. Preprogrammed Pulsatile	+4.6	+0.13	+0.43	+10.65	+2.23	+39.24
Heart Rate Variation	DRL vs. Constant Speed	+7.3	+0.3	+0.82	+22.69	+8.03	+48.06
	DRL vs. Preprogrammed Pulsatile	+2.0	+0.22	+0.46	+7.31	+0.98	+36.57
Ejection Time Variation	DRL vs. Constant Speed	+8.5	+0.28	+0.80	+22.84	+9.35	+48.12
	DRL vs. Preprogrammed Pulsatile	+3.5	+0.21	+0.60	+9.35	+2.22	+41.79

Improvement in Closeness to Healthy.

* For example, for Pulse Pressure (PP), the value is calculated as ($|PP_{Healthy} - PP_{Baseline}| - |PP_{Healthy} - PP_{DRL}|$). A positive value consistently indicates a greater improvement towards the healthy reference state by the DRL controller.

4.5. Robustness and adaptability of the DRL controller

To ensure that the DRL controller possesses strong reliability and generalization capability in real clinical scenarios, this study places particular emphasis on the training and evaluation of its robustness and adaptability. Compared with traditional constant speed control or pre-programmed pulsatile control, the DRL controller is capable of dynamically adjusting based on real-time state information, thereby demonstrating superior adaptability and responsiveness to physiological changes.

During the training phase, OU noise was introduced into the system to simulate transient disturbances in pressure and flow signals, accounting for possible fluctuations and errors in sensor signals encountered in practice. This design enables the DRL to learn to operate under a certain range of signal variability. In simulation tests, the controller's performance was further evaluated under artificially added random noise interference. The results showed that the DRL controller maintained stable control performance despite feedback signal perturbations, effectively stabilizing both aortic pressure and cardiac output.

Moreover, in our experimental platform, we observed some degree of signal drift in the pressure and flow sensors. Nonetheless, the DRL

controller was still able to respond quickly and make appropriate adjustments, demonstrating its robustness in real-world physical systems.

To further probe the controller's adaptability beyond its response to a single, large perturbation, we conducted an additional experiment evaluating its performance under more gradual, step-wise physiological fluctuations in heart rate. In this test, the heart rate was increased sequentially from 75 bpm to 90 bpm, then to 105 bpm, and finally to 120 bpm. A detailed visualization of the controller's response, is provided in the **Supplementary Material (Figure S2)**. This confirms that the DRL controller is robust not only to large, abrupt disturbances but also to sequential, incremental changes, highlighting its potential for reliable operation under dynamically varying patient conditions.

To address the critical question of long-term controller stability, we conducted a dedicated experiment to validate that the DRL controller's performance is sustained over an extended duration. This test was designed to ensure the system is not susceptible to performance degradation, where minor, uncorrected errors could accumulate over time, leading to system drift or instability—a phenomenon sometimes masked by short-term evaluations. For this experiment, the pulsatile mock circulatory loop was configured to maintain Moderate Heart Failure, and the DRL controller was run continuously for a duration of 30 min.

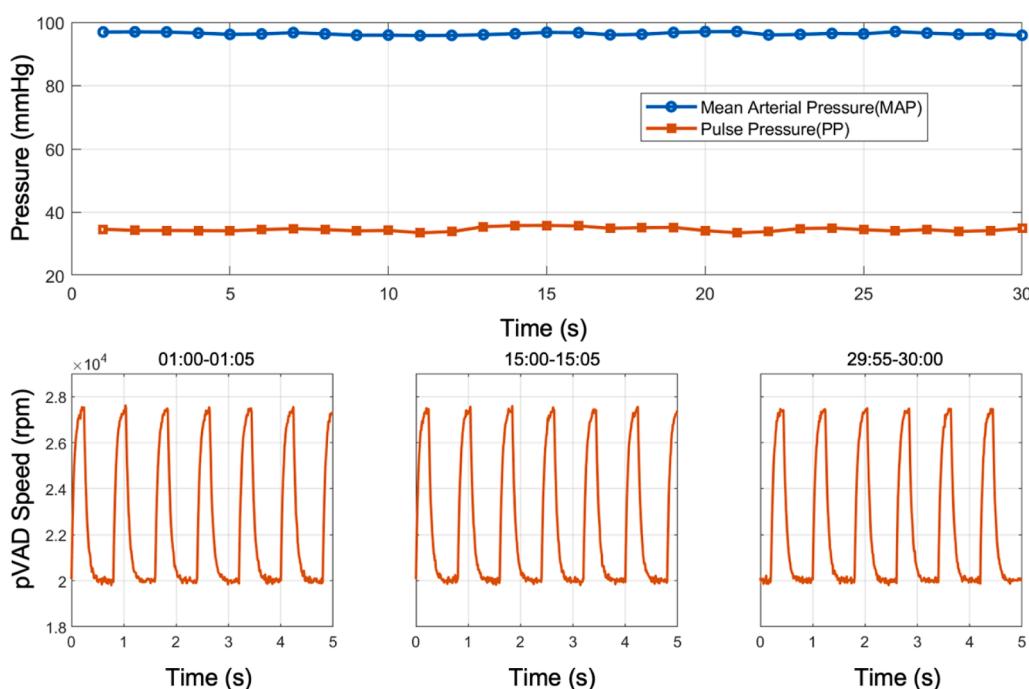


Fig. 16. Long-horizon stability test of the DRL controller over 30 min under a stable Moderate HF condition. (a) MAP and PP, averaged over each minute, remain stable with no significant drift. (b) Comparison of detailed pVAD speed waveforms from the start (01:00–01:05), middle (15:00–15:05), and end (29:55–30:00) of the run, illustrating consistent and stable pulsatile control action over time.

The results of this long-duration test are summarized in Fig. 16. As shown in Fig. 16(a), both the MAP and PP remained remarkably stable throughout the entire 30-minute period, exhibiting no significant drift or decay. This demonstrates the controller's ability to maintain consistent hemodynamic support over time. Furthermore, Fig. 16(b) compares detailed snippets of the pVAD speed waveform from the beginning, middle, and end of the run. The control pattern remains consistent, confirming that the learned policy is stable and does not degrade.

This long-horizon experiment provides strong evidence that the DRL controller is not only adaptive to dynamic physiological changes but is also robust and stable for sustained periods of operation, a critical requirement for its potential clinical application.

In summary, the DRL controller exhibits not only robustness against noise and sensor errors, but also strong adaptability, continuously delivering appropriate control strategies under dynamically changing physiological conditions. This provides a solid foundation for its application in clinical environments.

5. Conclusions

This paper designs a DRL controller based on the Deep Deterministic Policy Gradient (DDPG) architecture for the control of pVAD. The controller adapts the pVAD's rotational speed in real time by sensing key physiological signals such as the patient's aortic pressure, aortic flow, and left ventricular pressure. Specifically, the DRL controller designed in this paper can adapt to varying degrees of heart failure in patients, adjust to changes in heart rate, and accommodate variations in the ejection time within a heartbeat cycle, demonstrating strong resistance to interference. By building a cardiovascular-pVAD simulation model in MATLAB SIMULINK, we can simulate different heart failure conditions, and by comparing the simulation model with experimental data from a pulsatile experimental platform, the high fidelity of the model can be verified. Compared with constant speed control or preprogrammed pulsatile control, the proposed DRL controller shows significant advantages in restoring hemodynamic stability (e.g., mean arterial pressure and cardiac output) and physiological pulsatile characteristics (e.g., pulse index and waveform similarity). Both simulation and experiments indicate that the DRL controller can adapt to a wide range of heart failure degrees (e.g., it can effectively improve heart failure conditions with a blood pressure of 91/64 mmHg and cardiac output of 3.5 L/min, as well as those with a blood pressure of 73/54 mmHg and cardiac output of 2.6 L/min) and dynamic heart rate variations (60–150 bpm), while maintaining stable physiological output even under noisy and interfered conditions, demonstrating its robustness and clinical potential.

However, the DRL controller proposed in this paper also has some limitations. First, the controller's performance was primarily developed and evaluated on a lumped-parameter Simulink model. While this model is computationally efficient for DRL training and was validated against our mock loop data, it has inherent limitations as a simplified representation of the cardiovascular system. It does not capture complex 3D spatial flow dynamics, arterial wave propagation and reflection phenomena, or the detailed non-linear mechanics of heart valves. Recognizing this limitation was a primary motivation for validating our controller on the physical mock circulatory loop, which introduces a higher degree of real-world non-linearity, hardware delays, and sensor noise. Furthermore, the DRL agent was trained and validated on a finite set of cardiac profiles that, while covering varying degrees of heart failure, were all based on regular cardiac rhythms. The controller's performance and safety in the presence of more complex and clinically common pathologies, particularly arrhythmias (e.g., atrial fibrillation or ventricular tachycardia), remain untested and represent a critical area for future investigation. Additionally, the experiments only validated short-term hemodynamic stability, and the potential long-term complications (e.g., hemolysis risk and thrombosis tendency) of using the DRL controller have not been assessed. In the future, animal experiments

could be conducted to evaluate the long-term effects of DRL control on the pVAD, with comparative analysis of blood damage indicators and vascular endothelial response. It is crucial to emphasize that despite the promising results from our in-silico and in-vitro platforms, the generalization of these findings to complex in vivo human physiology remains putative. Our current study should therefore be viewed as a successful pre-clinical proof-of-concept that establishes a strong foundation for these necessary next steps. Future work could also benefit from a formal global sensitivity analysis (GSA). The simulation model and experimental methods used in this study could also be used to compare the performance of different DRL algorithms (e.g., DDPG, PPO, SAC, etc.) to determine the most suitable algorithm for application in implanted artificial heart control systems.

Ethics statement

This study did not involve any human participants or animal subjects. All simulations were conducted in a computational environment using a high-fidelity cardiovascular-pump hybrid model developed in MATLAB SIMULINK. Experimental validations were performed exclusively on a pulsatile mock circulatory loop platform using a glycerol-water mixture to simulate blood viscosity. No biological tissues, patient data, or live models were used at any stage of this research.

CRediT authorship contribution statement

Yuyang Shi: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zhike Xu:** Validation, Data curation. **Chenghan Chen:** Methodology, Investigation. **Feng He:** Software, Resources. **Pengfei Hao:** Methodology. **Xiwen Zhang:** Validation, Resources, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No 2022YFC2402600), Tsinghua-Toyota Joint Research Fund, the National Natural Science Foundation of China (Grant No 12388101), and Cross-strait Tsinghua Research Institute (Xiamen) (Grant No.HXY-Med-20240627-02).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2025.108975.

References

- [1] J.M. Beattie, Palliative care for heart failure: challenges and opportunities, Eur. J. Cardiovasc. Nurs. 13 (2) (2014) 102–104, <https://doi.org/10.1177/1474515114521697>.
- [2] A. Groenewegen, F.H. Rutten, A. Mosterd, A.W. Hoes, Epidemiology of heart failure, Eur. J. Heart Fail. 22 (8) (2020) 1342–1356, <https://doi.org/10.1002/ejhf.1858>.
- [3] G. Savarese, L.H. Lund, Global Public Health burden of Heart failure | CFR Journal, Accessed: Apr 13, 2025. [Online]. Available, <https://www.cfrijournal.com/articles/global-public-health-burden-heart-failure>, 2016.
- [4] S. Alishetti, L. Braghieri, D.L. Jennings, N. Uriel, P.C. Colombo, M. Yuzefpolskaya, Angiotensin receptor neprilysin inhibitor use in patients with left ventricular assist devices: a single-center experience, Int. J. Artif. Organs 45 (1) (2022) 118–120, <https://doi.org/10.1177/0391398821989066>.
- [5] S.J. Greene, et al., Abstract 8872: comparative effectiveness of dosing of medical therapy for heart failure: from the CHAMP-HF registry, Circulation (2021), https://doi.org/10.1161/circ.144.suppl_1.8872.

- [6] S.A. DeVries, J. Smith, Increasing donor heart use in cardiac transplantation, *JAA* 36 (11) (2023) 1, <https://doi.org/10.1097/01.JAA.0000979512.09945.c6>.
- [7] J.L. Vieira, H.O. Ventura, M.R. Mehra, Mechanical circulatory support devices in advanced heart failure: 2020 and beyond, *Prog. Cardiovasc. Dis.* 63 (5) (2020) 630–639, <https://doi.org/10.1016/j.pcad.2020.09.003>.
- [8] G. Tarantini, A. Panza, G. Lorenzoni, D. Gregori, G. Masiero, Breaking down cardiogenic shock: an analytical reflection on the DanGer-SHOCK and ECLS-SHOCK trials, *Am. J. Cardiol.* 236 (2025) 30–33, <https://doi.org/10.1016/j.amjcard.2024.10.032>.
- [9] C.A. den Uil, et al., Short-term mechanical circulatory support as a bridge to durable left ventricular assist device implantation in refractory cardiogenic shock: a systematic review and meta-analysis, *Eur. J. Cardiothorac. Surg.* 52 (1) (2017) 14–25, <https://doi.org/10.1093/ejcts/exz088>.
- [10] P. Shah, et al., Twelfth Interagency Registry for Mechanically Assisted Circulatory Support Report: readmissions after left ventricular assist device, *Ann. Thorac. Surg.* 113 (3) (2022) 722–737, <https://doi.org/10.1016/j.athoracsur.2021.12.011>.
- [11] J.G. Rogers, et al., Intrapericardial left ventricular assist device for advanced heart failure, *N. Engl. J. Med.* 376 (5) (2017) 451–460, <https://doi.org/10.1056/NEJMoa1602954>.
- [12] C. Chen, M. Zhang, P. Hao, F. He, X. Zhang, An *in silico* analysis of unsteady flow structures in a microaxial blood pump under a pulsating rotation speed, *Comput. Methods Programs Biomed.* 243 (2024) 107919, <https://doi.org/10.1016/j.cmpb.2023.107919>.
- [13] H. Wang, Y. Zhang, X. Wu, Simulation of axial flow left ventricular Assist Device's control algorithm used to assist different levels of heart failure based on the circuit model, *J. Phys.: Conf. Ser.* 2005 (1) (Aug. 2021) 012192, <https://doi.org/10.1088/1742-6596/2005/1/012192>.
- [14] M. Azizkhani, Y. Chen, Supervised adaptive fuzzy control of LVAD with pulsatility ratio modulation, in: 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE), Aug. 2022, pp. 2429–2434, <https://doi.org/10.1109/CASE49997.2022.9926444>.
- [15] M. Fetanat, M. Stevens, C. Hayward, N.H. Lovell, A physiological control system for an implantable heart pump that accommodates for interpatient and intrapatient variations, *IEEE Trans. Biomed. Eng.* 67 (4) (2020) 1167–1175, <https://doi.org/10.1109/TBME.2019.2932233>.
- [16] S. Ravanshadi, M. Jahed, Robust control of LVAD based on the sub-regional modeling of the heart, *Scientia Iranica* 23 (6) (Dec. 2016) 2934–2943, <https://doi.org/10.24200/sci.2016.4003>.
- [17] A. Verbeni, et al., An innovative adaptive control strategy for sensorized left Ventricular assist devices, *IEEE Trans. Biomed. Circuits Syst.* 8 (5) (2014) 660–668, <https://doi.org/10.1109/TBCAS.2014.2346015>.
- [18] H. Liu, S. Liu, A multi-objective physiological control system for continuous-flow left ventricular assist device with non-invasive physiological feedback, *Biomed. Signal Process. Control* 100 (2025) 107020, <https://doi.org/10.1016/j.bspc.2024.107020>.
- [19] M. Fetanat, M. Stevens, C. Hayward, N.H. Lovell, A sensorless control system for an implantable heart pump using a real-time deep convolutional neural network, *IEEE Trans. Biomed. Eng.* 68 (10) (2021) 3029–3038, <https://doi.org/10.1109/TBME.2021.3061405>.
- [20] M. Bakouri et al., “A feasible method to control left ventricular assist devices for heart failure patients: a numerical study,” 2022.
- [21] B. Rim, N.-J. Sung, S. Min, M. Hong, Deep Learning in physiological signal data: a survey, *Sensors* 20 (4) (2020), <https://doi.org/10.3390/s20040969>. Art. no. 4.
- [22] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, *Comput. Methods Programs Biomed.* 161 (2018) 1–13, <https://doi.org/10.1016/j.cmpb.2018.04.005>.
- [23] Marwen Belkacem, Farah Jemili, Omar Ellouze, et al., Optimizing Left Ventricular Assist Device Therapy: A Machine Learning Approach for Predicting Cardiac Output, 18 August 2023, PREPRINT (Version 1) available at Research Square. <http://doi.org/10.21203/rs.3.rs-3262331/v1>.
- [24] M. Fetanat, M. Stevens, C. Hayward, N.H. Lovell, A sensorless control system for an implantable heart pump using a real-time deep convolutional neural network, *IEEE Trans. Biomed. Eng.* 68 (10) (2021) 3029–3038, <https://doi.org/10.1109/TBME.2021.3061405>.
- [25] T. Li, W. Cui, X. Liu, X. Li, N. Xie, Y. Wang, A physiological control system for pulsatile ventricular assist device using an energy-efficient deep reinforcement learning method, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–10, <https://doi.org/10.1109/TIM.2023.3277993>.
- [26] T. Li, et al., Intelligent and strong robust CVS-LVAD control based on soft-actor-critic algorithm, *Artif. Intell. Med.* 128 (Jun. 2022) 102308, <https://doi.org/10.1016/j.artmed.2022.102308>.
- [27] D. Fernández-Zapico, T. Peirelinck, G. Deconinck, D.W. Donker, L. Fresiello, Physiological control for left ventricular assist devices based on deep reinforcement learning, *Artif. Organs* (2024), <https://doi.org/10.1111/aor.14845> p. aor.14845.
- [28] Z. Nagy, et al., Ten questions concerning reinforcement learning for building energy management, *Build. Environ.* 241 (Aug. 2023) 110435, <https://doi.org/10.1016/j.buildenv.2023.110435>.
- [29] T.P. Lillicrap et al., “Continuous control with deep reinforcement learning,” Jul. 05, 2019, *arXiv: arXiv:1509.02971*. doi: [10.48550/arXiv.1509.02971](https://doi.org/10.48550/arXiv.1509.02971).
- [30] X. Zuo, et al., Co-simulation of hypertensive left ventricle based on computational fluid dynamics and a closed-loop network model, *Comput. Methods Programs Biomed.* 216 (2022), <https://doi.org/10.1016/j.cmpb.2022.106649>.
- [31] J.F. De Canete, P.D. Saz-Orozco, D. Moreno-Boza, E. Duran-Venegas, Object-oriented modeling and simulation of the closed loop cardiovascular system by using SIMSCAPE, *Comput. Biol. Med.* 43 (4) (2013) 323–333, <https://doi.org/10.1016/j.combiomed.2013.01.007>.
- [32] L. Rosalia, C. Ozturk, D. Van Story, M.A. Horvath, E.T. Roche, Object-oriented lumped-parameter modeling of the cardiovascular system for physiological and pathophysiological conditions, *Advcd. Theory Sims.* 4 (3) (2021) 2000216, <https://doi.org/10.1002/adts.202000216>.
- [33] Z. Zuo, L. Tan, W. Shi, C.-K. Chen, J. Ye, E.M. Francis, Transient characteristic analysis of variable frequency speed regulation of axial flow pump, *Sustainability* (2022), <https://doi.org/10.3390/su141811143>.
- [34] L. Shi, et al., Numerical investigation and experiment on pressure pulsation characteristics in a full tubular pump, *Renew. Energy* 163 (2021) 987–1000, <https://doi.org/10.1016/j.renene.2020.09.003>.
- [35] S. Sadatieh, M. Dehghani, M. Mohammadi, R. Boostani, Extremum-seeking control of left ventricular assist device to maximize the cardiac output and prevent suction, *Chaos Solitons Fractals* 148 (2021) 111013, <https://doi.org/10.1016/j.chaos.2021.111013>.
- [36] M. Mekki, Y. Wang, P. Sethu, M. Ghazal, A. El-Baz, G. Giridharan, A sensorless rotational speed-based control system for continuous flow left ventricular assist devices, *IEEE Trans. Biomed. Eng.* 67 (4) (2020) 1050–1060, <https://doi.org/10.1109/TBME.2019.2928826>.
- [37] R.T. Icarte, T.Q. Klassen, R. Valenzano, S.A. McIlraith, Reward machines: exploiting Reward function structure in reinforcement learning, *J. Artif. Intell. Res.* 73 (2022) 173–208, <https://doi.org/10.1613/jair.1.12440>.