

Aortic-Flow Reference Shaping via Q-Learning for CF-LVAD Support

Mahmud Karno*, Mert yigit*, Ahmed Alhajyounis[†], Kamuran Kadipasaoglu*

*Biomedical Engineering Department.

Yildiz Technical University, İstanbul, Turkey

mahmud.karno@std.yildiz.edu.tr, Yigitmert1@hotmail.com, kamuran@yildiz.edu.tr

[†]Control and Automation Engineering Department.

Yildiz Technical University, İstanbul, Turkey

ahmed.alhajyounis@std.yildiz.edu.tr

Abstract—Continuous-flow left ventricular assist devices (CF-LVADs) operate at fixed speeds, suppressing pulsatility, prolonging aortic-valve closure, and impairing root washout, thereby elevating thrombotic risk. To address these limitations, we propose a Q-learning-based framework that synthesizes a smooth, bandwidth-limited aortic-flow reference without explicit plant modeling. The reference is parameterized by waveform coefficients optimized offline through interactions with a cardiovascular simulator, guided by a composite reward that penalizes deviations in mean arterial pressure, encourages periodic aortic valve opening, prevents suction events, and maintains sufficient and proper flow. In simulation, the learned waveform maintained mean arterial pressure between 60–90 mmHg, left-ventricular pressure between 5–70 mmHg, pump flow between 2–5 L/min, ventricular volume between 50–140 mL, and a pulsatility index around 2. Importantly, the aortic valve opened approximately 6 times per 20 s, improving washout while avoiding suction. These results underscore the potential of reference-level learning to restore physiological pulsatility and preserve hemodynamic safety in CF-LVAD support.

Keywords—Q-learning, LVAD, Aortic flow reference.

I. INTRODUCTION

Left ventricular assist devices (LVADs) are well-established therapeutic options for patients with advanced heart failure, providing long-term circulatory support either as a bridge to transplantation or as destination therapy [1]. Their ability to restore cardiac output, improve end-organ perfusion, and enhance quality of life has driven widespread clinical adoption. Among the available technologies, continuous-flow LVADs (CF-LVADs) have gained particular prominence due to their compact size, mechanical simplicity, and superior durability compared with pulsatile-flow designs [2]. Nevertheless, the fixed-speed operation of CF-LVADs inherently suppresses physiological pulsatility, which can lead to prolonged aortic valve closure, reduced or irregular valve motion, and diminished aortic root washout [2]. These hemodynamic changes have been associated with adverse events such as aortic insufficiency, thrombosis, and vascular remodeling [3].

In the literature, multiple strategies have been proposed to mitigate these effects by shaping aortic-flow reference signals. For example, Dobarro et al. employed pre-programmed pulsatility patterns to impose periodic variations in pump

speed, thereby promoting intermittent aortic valve opening and improving aortic-root washout [4]. More recently, Fernández-Zapico et al. demonstrated the feasibility of reinforcement learning for physiological modulation in mechanical circulatory support, using it to modulate LVAD motor speed so as to achieve left-ventricular ejection and periodic aortic-valve opening highlighting the promise of data-driven approaches for LVAD reference-signal design [5].

In earlier studies, we utilized a seven-compartment lumped-parameter cardiovascular system (CVS) simulation model [6] in Simulink to evaluate candidate aortic flow reference waveforms for CF-LVAD operation. This platform was employed to characterize the hemodynamic impact of different pulsatility patterns, assess their influence on aortic valve dynamics, and identify key parameters affecting mean arterial pressure and pulsatility under varying preload, afterload, and contractility conditions. These studies provided quantitative evidence of the trade-offs between root washout, pressure stability, and overall pump performance, and revealed that fixed-pattern modulation often failed to maintain optimal hemodynamics across different physiological states.

Q-learning, a model-free reinforcement learning method, offers a promising framework for optimizing aortic flow reference signals in CF-LVADs. Unlike conventional algorithms, Q-learning learns optimal state-action mappings through repeated interaction with the environment, maximizing a cumulative reward over time. This enables the simultaneous pursuit of multiple, potentially conflicting objectives [7], such as maintaining mean arterial pressure within 80–100 mmHg, preventing suction by preserving adequate left ventricular volume, sustaining physiological pulsatility, and promoting periodic aortic valve opening, while adapting to changes in preload, afterload, and contractility.

Prior work has largely focused on direct pump-speed modulation or on prescribing aortic-flow references with heuristic, preprogrammed, or model-based schemes; to our knowledge, none has framed the task as synthesizing the aortic-flow reference via Reinforcement Learning (RL). We instead formulate reference design at the flow level, anchoring the objective in clinically salient endpoints (periodic aortic-valve opening, effective aortic-root washout, a target pulsatility index, and

improved ventricular unloading) while leaving low-level actuation to the pump controller. Trained across variations in preload, afterload, and contractility, the approach explicitly negotiates the trade-offs among mean arterial pressure regulation, pulsatility index preservation, valve-opening duty cycle, and root washout, yielding a smooth, bandwidth-limited reference.

In this formulation, a model-free Q-learning agent interacts with a numerical cardiovascular (CVS) simulator to identify the reference waveform from sampled transitions, without requiring an explicit plant model during learning. The resulting reference balances pressure targets with physiologic pulsatility and addresses limitations of fixed-pattern and model-dependent schemes, while remaining fixed at deployment.

II. MATERIALS AND METHODS

A. Simulation environment

All experiments were run in MATLAB/Simulink [R2024b]. The cardiovascular environment is a seven-compartment lumped-parameter model with ideal-diode valves and time-varying ventricular elastance. The plant executes with a fixed step time of 10^{-4} second. MATLAB/Simulink R2024b with MathWorks Reinforcement Learning Toolbox installed; all plant dynamics resided in Simulink, while the toolbox was used only to host the Q-learning agent (details in Section D). The same topology and baseline parameters were used unless otherwise stated. Pressures, flows, and volumes are reported in mmHg, mL/s, and mL, respectively.

B. System Architecture

Overall, the methodology was organized as a closed reinforcement-learning loop centered on reference synthesis. At each decision step, the agent maintained and updated a value function (details in section D.2) over aortic-flow references $Q_{Ao,ref}$ and produced a candidate reference waveform (details in section C). This reference was then supplied to a controller, which generated the corresponding LVAD command summarized as Q_{LVAD} ; the commanded actuation drove the cardiovascular system (CVS) model. The CVS yielded observations from which scalar rewards were computed according to predefined hemodynamic objectives (details in section D.3). Since the agent and plant operated at different sample times, plant measurements were decimated and synchronized through a rate-transition stage before forming the next observation vector. The observations (details in section D.1) and rewards were then fed back to the agent (details in section D.4) which updated its value estimates and issued a new action that modified the aortic-flow reference parameters for the subsequent step. In this way, the cycle iterated continuously to align the learned reference with the desired physiological targets.

C. Parametric Encoding of the Aortic-Flow Reference

We parameterized the aortic-flow reference (1) as a superposition of ten sinusoids with an additional bias:

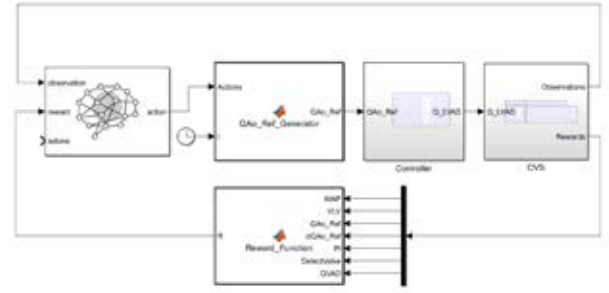


Fig. 1: Block Diagram of the System Architecture.

$$Q_{Ao,ref} = b_0 + \sum_{k=1}^{10} A_k \sin(\omega_k t + \varphi_k). \quad (1)$$

The coefficient set consisted of the amplitudes A_k , angular frequencies ω_k , and phases φ_k , for $k = 1 \dots 10$ together with the DC term b_0 yielding 31 scalar parameters in total. These parameters were treated as the agent's actions (details in section C.4) and were learned offline through interaction with the numerical cardiovascular simulator.

D. RL design algorithm

1) *Observation space*: The policy received a seven-dimensional observation vector at each decision step: mean arterial pressure (MAP), left-ventricular volume V_{LV} , aortic-valve-open flag (0/1), pulsatility index (PI), aortic-flow ($Q_{Ao,ref}$), aortic-flow derivative ($dQ_{Ao,ref}/dt$), LVAD-flow (Q_{LVAD}). MAP and V_{LV} were read directly from the CVS. $Q_{Ao,ref}$ and $dQ_{Ao,ref}/dt$ were fed back from the agent's current reference to provide context for subsequent decisions. The valve-open flag was generated by evaluating the physical condition $P_{LV} > P_{Ao}$, and $Q_{LV} > 0$. It was emitted as a unit pulse on each rising event. PI was computed over a sliding window as the moving (max-min) of Q_{LVAD} divided by the moving mean. Q_{LVAD} was taken from the controller output. Before being provided to the policy normalized to the interval $[-1, 1]$.

2) *Action space*: The policy emitted a continuous action vector $a_t \in \mathbb{R}^k$ that parameterized an aortic-flow reference generator. Elements of a_t set (i) a DC bias (mean flow level) and (ii) waveform-shape parameters (e.g., amplitudes/phases). Each element were constrained to fixed bounds to preserve physiological plausibility and maintain training stability. The baseline coefficient b_0 and all phase terms φ_k used global limits in every experiment: $b_0 \in [40100]$ and $\varphi_k \in [-\pi, \pi]$. The remaining coefficients followed scenario-specific intervals that matched the operating regime, as summarized in Table 1. These bounds were enforced at every update step. After convergence, the learned coefficient vector defined the aortic-flow reference waveform used in subsequent evaluations.

3) *Reward function*: At each decision step the agent received a scalar reward composed of seven additive terms, $R_T = R_1 + \dots + R_7$, designed to (i) keep the learned aortic-flow reference within physiological bounds, (ii) regulate arterial pressure, (iii) avoid suction by preserving left-ventricular filling, (iv) keep pump flow in a safe operating window, (v)

TABLE I: Scenario-specific bounds for reference-waveform coefficients

k	$A_k^{\min}$	$A_k^{\max}$	$\omega_k^{\min}$	$\omega_k^{\max}$
1	-2	+2	0.5	10
2	-2	+2	0.5	10
3	-10	+10	0.1	5
4	-10	+10	0.1	5
5	-3	+3	0.1	3
6	-3	+3	0.1	3
7	-8	+8	0.5	8
8	-8	+8	0.5	8
9	-4	+4	0.1	3
10	-4	+4	0.1	3

promote periodic aortic-valve opening, (vi) pulsatility index, and (vii) limit aggressive upstrokes in the reference.

i. Aortic-flow reference shaping, $R_1(Q_{Ao,ref})$: This term (2) encouraged a strictly positive, physiologic mean aortic flow with a soft target around 5 L/min and penalties for non-physiologic values:

$$R_1 = \begin{cases} 5Q_{Ao,ref}, & Q_{Ao,ref} < 0, \\ -5, & Q_{Ao,ref} = 0, \\ Q_{Ao,ref}/5, & 0 < Q_{Ao,ref} < 5, \\ 5/Q_{Ao,ref}, & 5 < Q_{Ao,ref} < 8, \\ -Q_{Ao,ref}, & Q_{Ao,ref} \geq 8. \end{cases} \quad (2)$$

This structure rewarded flows near 5 L/min, discouraged zeros/negatives, and penalized excessive outputs.

ii. Arterial-pressure regulation, $R_2(MAP)$: MAP was rewarded within 80–100 mmHg and penalized outside (3):

$$R_2 = \begin{cases} -0.3, & MAP = 0, \\ -0.3/MAP, & 0 < MAP < 80, \\ MAP/240, & 80 \leq MAP < 100, \\ MAP, & MAP \geq 100. \end{cases} \quad (3)$$

iii. LV-volume safeguard, $R_3(V_{LV})$: To avoid suction, small ventricular volumes were penalized (4) while physiologic volumes (≈ 30 –150 ml) were modestly rewarded:

$$R_3 = \begin{cases} 0.2V_{LV}, & V_{LV} < 0, \\ -0.2, & V_{LV} = 0, \\ -0.2/V_{LV}, & 0 < V_{LV} < 30, \\ V_{LV}/1500, & 30 \leq V_{LV} < 150, \\ 0, & V_{LV} \geq 150. \end{cases} \quad (4)$$

iv. LVAD-flow window, $R_4(Q_{LVAD})$: Pump flow was encouraged to remain within 20–150 ml/s (≈ 1 –9 L/min) and penalized otherwise (5).

$$R_4 = \begin{cases} -0.1, & Q_{LVAD} = 0, \\ -0.1/Q_{LVAD}, & 0 < Q_{LVAD} < 40, \\ Q_{LVAD}/4000, & 40 \leq Q_{LVAD} < 140, \\ -Q_{LVAD}/1000, & Q_{LVAD} > 140. \end{cases} \quad (5)$$

v. Aortic-valve opening incentive, $R_5(ValveOpen)$: To promote periodic washout and physiologic pulsatility as specified (6);

$$R_5 = \begin{cases} 0.1, & ValveOpen = 1, \\ -0.1, & ValveOpen = 0. \end{cases} \quad (6)$$

vi. Pulsatility index shaping, R_6 : PI was computed as (7) [8]:

$$\frac{\max_{\tau \in [t-10, t]} Q_{LVAD}(\tau) - \min_{\tau \in [t-10, t]} Q_{LVAD}(\tau)}{\frac{1}{10} \int_{t-10}^t Q_{LVAD}(\tau) d\tau} \quad (7)$$

This term promoted physiologic pulsatility (8) while discouraging both near-aperiodic and overly large oscillations:

$$R_6 = \begin{cases} -PI10^{-3}, & PI < 0 \text{ OR } PI > 8, \\ -10^{-3}, & 0 \leq PI < 2, \\ PI10^{-3}, & 2 \leq PI \leq 8, \end{cases} \quad (8)$$

vii. Slope penalty, $R_7(dQ_{Ao,ref}/dt)$: To enforce a bandwidth-limited, smooth aortic-flow reference and suppress chatter, we penalized the instantaneous slope symmetrically for up- and down-ramps (9):

$$R_7 = -0.01|dQ_{Ao,ref}/dt| \quad (9)$$

The total reward was the unweighted sum of all components. Each term's scale (constants and piecewise bounds) was selected to keep individual contributions in typical operating ranges, ensuring that pressure regulation, suction avoidance, safe pump operation, and valve opening were simultaneously reflected in the learning signal without any single metric dominating.

4) *Agent & training*: Training proceeded in episodes of 10 s. The actor and critic were multilayer perceptrons with two hidden layers of 64 units and ReLU activations. Actor outputs were squashed and then affinely rescaled to the action bounds listed in Table 1.

Parameters were optimized with Adam using learning rates of 10^{-4} for the actor and 10^{-3} for the critic, with rate-transition handling and no weight decay. The discount factor was $\gamma = 0.99$.

III. RESULTS

The raw aortic-flow reference exhibited high-frequency transients. We therefore applied a low-pass filter followed by a smoothing operator, which suppressed spurious spikes and yielded a continuous, bandwidth-limited waveform suitable for actuator tracking (Fig. 2).

Simulation with the filtered reference yielded physiologically acceptable dynamics: aortic flow (Q_{Ao}) tracked the prescribed waveform ($Q_{Ao,ref}$) within a 2–5 L/min band. Q_{LVAD} supplied the predominant share of cardiac output, while native ejection reappeared whenever LVAD flow transiently approached 1L/min, indicating coordinated interaction between the assist device and the ventricle (Fig. 3).

At 60 bpm, hemodynamics was stable and aortic pressure remained pulsatile. Ventricular volumes (V_{LV} , V_{RV}) exhibited

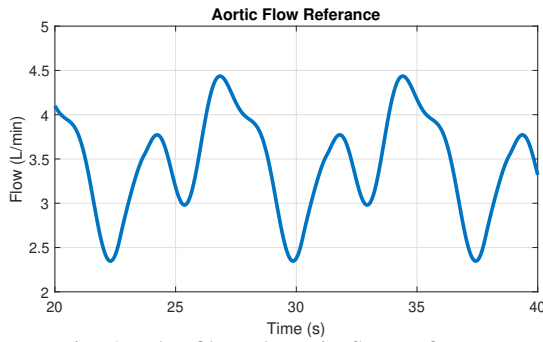


Fig. 2: The filtered aortic-flow reference.

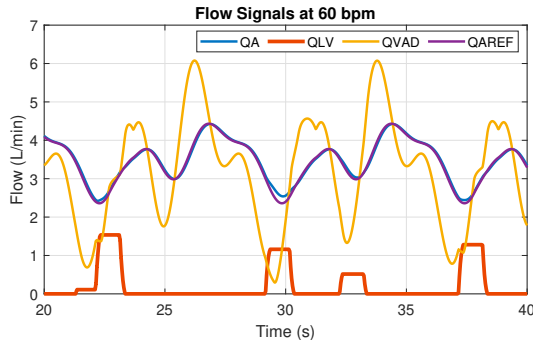


Fig. 3: Flow Responses to the Filtered Reference.

regular filling–ejection cycles: V_{LV} oscillated roughly 70–140 mL with no approach to collapse. No suction events or runaway pressures were observed as shown (Fig. 4).

IV. DISCUSSION & CONCLUSIONS

This study shows that casting aortic-flow reference as a reinforcement-learning problem restored pulsatility while maintaining pressure targets in a numerical CVS model. Optimizing at the reference layer kept the approach controller-agnostic and compatible with existing speed controllers. In simulation, the learned waveform promoted aortic-valve opening and improved root-washout surrogates without suction. These gains resulted from an explicit reward trade-off between pressure regulation and pulsatility/washout rather than ad-hoc speed patterns. Pulsatility was quantified using an LVAD-flow PI rather than an aortic-flow PI. Generally, episodes in which the measured LVAD flow (Q_{LVAD}) approaches zero are a reliable marker of native ventricular ejection, whereas a near-zero aortic-flow reference ($Q_{Ao,ref}$) is merely a command and does not confirm that the pump actually reduced flow at that instant; we therefore considered the LVAD-based PI the more appropriate metric. Baseline speed controller tracked the aortic-flow reference with $\sim 90\%$ fidelity across operating points, sufficient to reveal benefits and limits of reference-level learning. Tighter tracking would likely sharpen predefined hemodynamic objectives but lies outside our scope.

Limitations: (i) learning and evaluation were performed only at 60 bpm; higher rates and rhythm variability may require different optima and reward weights. (ii) The CVS included ideal-diode valves and lumped elastance, omitting valve inertia, coronary coupling, and baroreflex effects. (iii) Evaluation was in silico. Learning the aortic-flow reference is a viable route to reintroduce pulsatility and valve opening while ro-

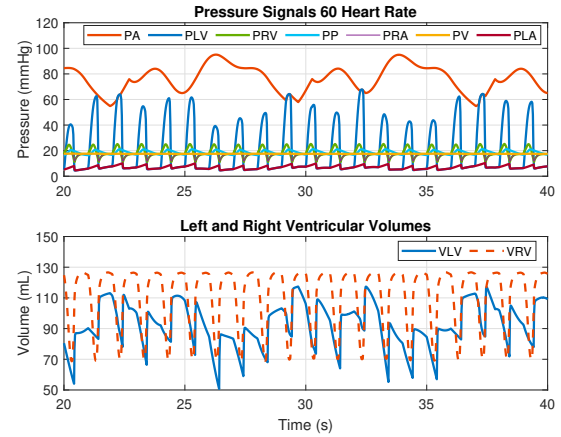


Fig. 4: Pressures and Ventricular Volumes at 60 bpm. Where VLV and VRV are left and right ventricular volumes and PLV, PA, PV, PRA, PRV PP and PLA are left ventricular, aortic, venous, right atrial, right ventricular, pulmonary, and left atrial pressures respectively

bustly preserving pressure. Future work will broaden training, incorporate aortic-based metrics, explore controller–reference co-design, and validate on hardware and in vivo.

ACKNOWLEDGMENT

This work was initiated and completed at the Physiological Control Laboratory with funding from the Scientific and Technological Research Council of Turkey (TUBITAK) 1003 Project (No. 118S098) and Yildiz Technical University Scientific Research Projects Coordination Unit (BAP) (No. FBA-2024-6491)

REFERENCES

- [1] National Center for Health Statistics, “Multiple cause of death 2018–2021 on cdc wonder database,” <https://wonder.cdc.gov/mcd.html?utm=Opt>, 2018.
- [2] L. W. Miller and J. G. Rogers, “Evolution of left ventricular assist device therapy for advanced heart failure: a review,” *JAMA cardiology*, vol. 3, no. 7, pp. 650–658, 2018.
- [3] S. A. Rabi and D. A. D’Alessandro, “Suction event after lvad placement,” in *Near Misses in Cardiac Surgery*. Springer, 2022, pp. 179–181.
- [4] D. Dobarro, M. Urban, K. Booth, N. Wrightson, J. Castrodeza, J. Jungschleger, N. Robinson-Smith, A. Woods, G. Parry, S. Schueler *et al.*, “Impact of aortic valve closure on adverse events and outcomes with the heartware ventricular assist device,” *The Journal of Heart and Lung Transplantation*, vol. 36, no. 1, pp. 42–49, 2017.
- [5] D. Fernández-Zapico, T. Peirelinck, G. Deconinck, D. W. Donker, and L. Fresiello, “Physiological control for left ventricular assist devices based on deep reinforcement learning,” *Artificial organs*, vol. 48, no. 12, pp. 1418–1429, 2024.
- [6] A. Alhajyounis, Y. Aydoğan, A. Canbaş, E. Kulakoğlu, U. Kurnaz, L. Soysal, M. Yiğit, D. Arısoy, and K. Kadıpaşaoğlu, “Nonlinear control of a hybrid pneumo-hydraulic mock circuit of the cardiovascular system,” 2025.
- [7] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, “A systematic review on supervised and unsupervised machine learning algorithms for data science,” *Supervised and unsupervised learning for data science*, pp. 3–21, 2020.
- [8] J. E. Dodge, A. L. Covens, C. Lacchetti, L. M. Elit, T. Le, M. Devries-Aboud, M. Fung-Kee-Fung, G. C. D. S. Group *et al.*, “Preoperative identification of a suspicious adnexal mass: a systematic review and meta-analysis,” *Gynecologic oncology*, vol. 126, no. 1, pp. 157–166, 2012.