Reinforcement Learning-Based Fuzzy Control for E-Bike Power Adaptation

Jin-Shyan Lee and Zheng-Yan Lee

Department of Electrical Engineering, National Taipei University of Technology, Taiwan
*Corresponding author: jslee@mail.ntut.edu.tw

***Abstract*—While electric-assist bicycles (e-bikes) promote sustainable and convenient mobility, current assist mechanisms often ignore the balance between physical exertion and user fatigue. Excessive assistance diminishes exercise benefits, while insufficient power increases injury risk. This paper proposes a fuzzy logic control system that integrates pedal force, riding habits, environmental factors, and real-time physiological data. Unlike traditional heart-rate-only controllers which suffer from latency and instability, our model employs a reinforcement learning (RL) algorithm to optimize assist delivery dynamically. The system maintains the rider's heart rate and speed within optimal, safe ranges across diverse user profiles. Simulation and field test results demonstrate that the proposed method significantly improves riding comfort and physiological stability compared to standard control models.**

***Index Terms*—Electric assist bicycles (e-bikes), Fuzzy logic control, Reinforcement learning, Physiological feedback**

I. Introduction

Traditional electric-assist bicycles (e-bikes) often fail to accommodate riders with physical limitations due to rigid power-assist strategies [1]-[4]. Currently, commercial e-bikes typically rely on either constant-assist or proportional-assist approaches. Neither method accounts for the rider's real-time physiological state, leading to either excessive exertion or insufficient exercise benefits. Rider fatigue is best indicated by metabolic energy expenditure. While direct measurement of Adenosine Triphosphate (ATP) or oxygen consumption is impractical for daily commuting, heart rate (HR) serves as a reliable, non-invasive proxy for physical exertion. Consequently, HR-based regulation has become a focal point for intelligent e-bike control.

Existing literature highlights several control challenges. Methods using Sliding Mode Control (SMC) or Model Predictive Control (MPC) often require pre-established, subject-specific nonlinear heart rate models [5]. These lack generalizability across diverse user profiles. Furthermore, systems focusing on transmission ratios often reduce vehicle speed to lower HR, which is impractical for urban commuting. Early Fuzzy Logic and PID implementations have also shown significant HR fluctuations and high deviation from target ranges, with some experiments showing actual heart rates deviating by over 30 bpm from the target [6]-[8].

To address these gaps, this paper proposes a Three-Tier Fuzzy Logic Heart Rate Controller (FLHRC), further optimized via a Q-Learning reinforcement learning algorithm (Q-FLHRC). This model-free approach adapts to pedaling habits and environmental conditions without requiring prior physiological modeling. The main contributions of this work include:

- Model-Free Universality: Combining Q-Learning with Fuzzy Logic to eliminate the need for individualized HR modeling, ensuring applicability to the general public.
- Multi-Input Integration: A robust controller that synthesizes pedal force, riding habits, environmental factors, and physiological states.
- RL Optimization: Utilizing Q-Learning to dynamically optimize the fuzzy rule base, significantly enhancing system robustness.

Experimental results from both simulations and real-world trials demonstrate that the proposed system maintains HR within target ranges while ensuring stable speed and safe acceleration.

II. Dynamics and Modelling of Bike Systems

A. System Architecture

The proposed e-bike architecture, illustrated in Fig. 1, integrates a centralized controller, a motor driver, a high-torque rear hub motor, and a sensor array (IMU, torque, speed, and electronic braking sensors). A 36V/10Ah LiFePO4 battery provides power, offering a higher energy-to-weight ratio than traditional lead-acid batteries to extend the operational range. At the system's core, the controller executes the proposed regulation algorithms to maintain the rider's heart rate within target zones. For stability and safety, the system synthesizes real-time data from the IMU, torque, and speed sensors to adapt to terrain and intensity, while electronic braking signals ensure immediate power termination during deceleration.

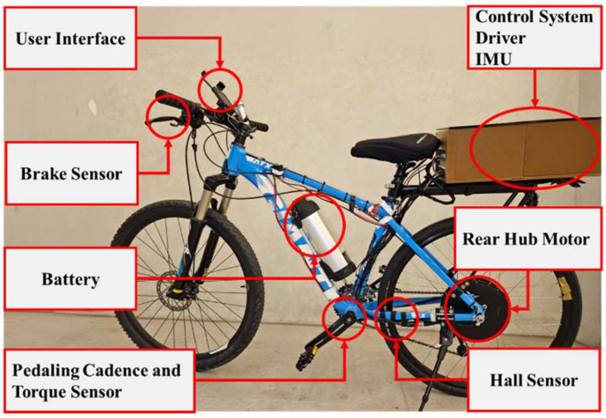

Fig. 1. Developed electric-assist bike.

B. Dynamics and Modelling of Bike Systems

During operation, an electric-assist bicycle is subject to external resistive forces, primarily aerodynamic drag, gradient resistance, and rolling friction. The integration of specialized hardware—including the controller, battery pack, and hub motor—substantially increases the vehicle's gross weight compared to conventional bicycles. This added mass exacerbates the gravitational force components acting parallel and perpendicular to the road surface, thereby intensifying slope resistance and rolling friction. Consequently, the rider must exert higher torque to maintain velocity, which can negatively impact pedaling efficiency [2]–[3]. The comprehensive force distribution is characterized in the free-body diagram shown in Fig. 2.

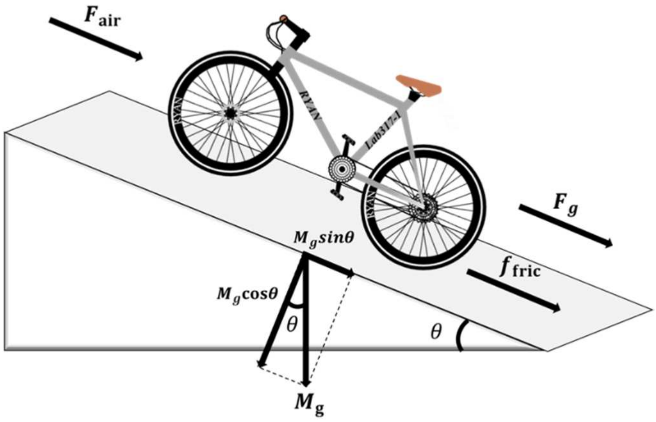

Fig. 2. . Resisting forces acting on a bike in the upward motion on a slope.

The riding resistance includes air resistance F_{air}, friction resistanceF_{fric}, and slope resistance F_{g}, as shown in (1) to (3).

$$F_{\text{air}} = \frac{C_{\text{x}} A \rho v^2}{2} \tag{1}$$

$$F_{\text{fric}} = \mu M g \cos\theta \tag{2}$$

$$F_{\text{g}} = M g \sin\theta \tag{3}$$

The power sources for an electric assist bicycle under ideal conditions are mainly divided into two types: one is the torque T_{pedal} generated by human pedaling, which produces an angular velocity ω_{pedal} during the pedal stroke. This angular velocity is transmitted to the rear sprocket via the chain, driving the tire to move the bicycle forward. The human output here is referred to as pedaling power P_{pedal}, calculated as shown in (4); the other source is the torque output by the hub motor under the controller's command, acting on the rear wheel to generate an angular velocity. The motor output here is referred to as motor power P_{motor}, calculated as shown in (5).

$$P_{\text{pedal}} = \omega_{\text{pedal}} \times T_{\text{pedal}} \tag{4}$$

$$P_{\text{total}} = P_{\text{pedal}} + P_{\text{motor}} \tag{5}$$

Therefore, when the total power P_{total} of the power sources exceeds the total power P_{resist} of the environmental resistance, the bicycle can achieve an instantaneous speed V_{instant} moving forward. Additionally, this paper does not consider friction or heat loss in the transmission system or energy conversion losses. The calculation method for P_{resist} is shown in (6).

$$P_{\text{resist}} = (F_{\text{g}} + F_{\text{fric}} + F_{\text{air}}) \times V_{\text{instant}} \tag{6}$$

The power required for forward movement of the bicycle, obtained by subtracting the total power of environmental resistance P_{resist} from the total power P_{total} of the power sources, divided by V_{instant}, yields the thrust force F_{thrust} needed for riding, calculated as shown in (7).

$$F_{\text{thrust}} = (P_{\text{total}} - P_{\text{resist}})/V_{\text{instant}} \tag{7}$$

C. Cyclist Heart Rate Model

The cardiovascular response in humans varies dynamically with exercise intensity. Precise regulation of intensity is critical for both athletic performance optimization and clinical rehabilitation, as tailored exertion levels facilitate specific physiological adaptations. Consequently, HR has emerged as a standard auxiliary metric for monitoring physical load, leading to a surge in HR-targeted control systems across various domains. For instance, recent studies have demonstrated successful HR regulation by adjusting treadmill velocity via automated controllers or by establishing subject-specific heart rate response models to guide running intensity.

To evaluate the proposed control strategy, this study integrates the heart rate (HR) model from [7] for simulation. This model simulates real-time physiological responses across diverse environmental conditions. The resulting HR error-the deviation between the measured and target HR-serves as the primary input for the fuzzy controller, which modulates motor assistance to maintain exertion within the desired range. The cardiovascular dynamics are governed by the first-order nonlinear state-space equations in (8)-(12). Specifically, x_1 captures the instantaneous HR response to pedaling intensity, while x_2 represents the progressive cardiovascular drift during prolonged exercise. Driven by the rider's instantaneous pedaling power (normalized relative to maximum aerobic power), the model outputs a normalized HR response, which is then converted to the absolute heart rate via the transformation in (11).

$$\dot{x}_1 = -a_1 \times x_1 + a_2 \times x_2 + \frac{a_6}{P_{\text{max}}} \times P_{\text{pedal}} \tag{8}$$

$$\dot{x}_2 = \frac{a_4 \times x_1}{1 + \exp(-(x_1 - a_5))} - a_3 \times x_2 \tag{9}$$

$$y = x_1 \tag{10}$$

$$x_1 = (HR - HR_{\text{rest}})/HR_{\text{max}} \tag{11}$$

$$P_{\text{pedal}} = T_{\text{pedal}} \times \omega_{\text{pedal}} \tag{12}$$

III. THE PROPOSED POWER-ASSISTED CONTROL SYSTEM

A. Fuzzy Logic Heart Rate Controller (FLHRC)

The architecture of the proposed Fuzzy Logic Heart Rate Controller (FLHRC) is illustrated in Fig. 3. The control process begins with user initialization via a mobile application, where demographic data—including gender, height, weight, and age—are inputted. The system utilizes these parameters to calculate the user's theoretical maximum heart rate, subsequently deriving an optimal target exercise heart rate range. Alternatively, riders may manually define a custom target heart rate based on specific training requirements. Once the target heart rate is established and transmitted to the control panel, the real-time operational phase commences. The controller continuously monitors a comprehensive set of inputs: braking status, battery state-ofcharge, pedaling power, road slope, velocity error, and heart rate error. These variables are processed through the three-tier fuzzy logic controller to determine the optimal motor assistance level. This hierarchical approach aims to regulate the rider's heart rate within the target zone while simultaneously optimizing for riding comfort and operational safety. Furthermore, the application provides real-time feedback on recommended pedaling power, assisting the rider in maintaining their physiological targets through a human-in-the-loop interface.

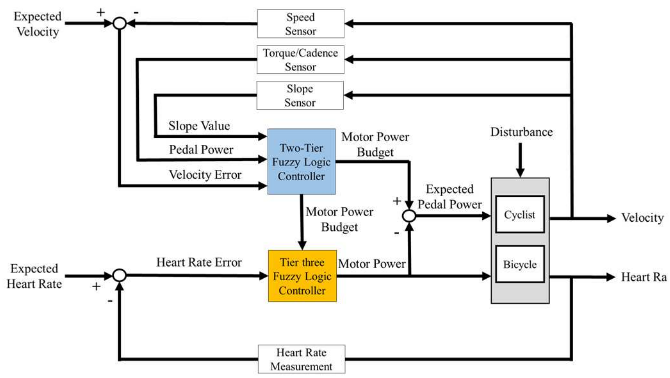

Fig. 3. Overall control system architecture of FLHRC.

B. *Q-FLHRC: Integration of Q-Learning and Fuzzy Logic*

This section details the integration of the FLHRC with a Q-Learning algorithm, forming the proposed Q-FLHRC framework illustrated in Fig. 4. In this architecture, the inputs to the FLHRC function as the state space within the environment. As the system operates, the RL agent observes the resulting state transitions and evaluates the performance through a predefined reward function. The agent then updates the Q-table—a state-action value mapping—based on these transitions, rewards, and the controller's outputs. Through an iterative optimization process guided by the chosen policy (e.g., ε-greedy), the agent refines the rule-base of the Tier-Three Fuzzy Logic controller. This continuous learning loop allows the system to converge toward an optimized rule-set, ensuring the assist power remains robust under varying riding conditions.

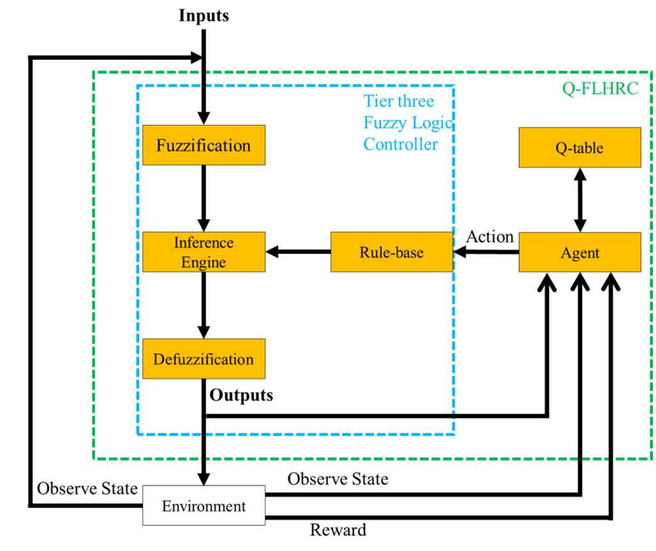

Fig. 4. Q-FLHRC schematic diagram.

The system parameters of Q-FLHRC are designed as follows:

- *State Space and Actions*

Because the traditional Q-learning algorithm operates over a finite, discrete state space, the continuous input variables must be discretized to facilitate the learning process. In this study, the Q-learning agent is integrated into the third-tier fuzzy control layer—which specifically governs heart rate regulation—to optimize the underlying rule base. Consequently, the inputs for the third-tier rules are mapped into a multi-dimensional state space. The state variable is defined by a triplet of discretized variables: heart rate error, estimated motor power, and road slope. Experimental results indicated that the inclusion of the slope variable significantly enhances the system's predictive performance and adaptive stability across varying terrains.

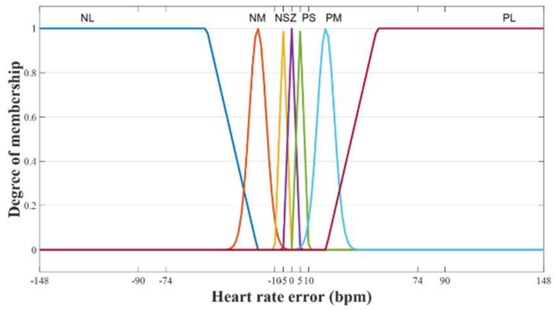

Fig. 5. The heart rate error membership function.

Fig. 5 illustrates the membership functions employed for discretizing the heart rate error. The state space is defined by 297 discrete states, spanning from -148 to 148 bpm. In this configuration, each 1 bpm interval is mapped to one specific state. Similarly, the membership functions for the motor power budget are established as shown in Fig. 6, comprising 251 states. For the slope membership functions shown in Fig. 7, the interval from -3% to 3% is divided into 7 states. Based on this composition, the total Q-learning state space is 521,829 states. The action a_{ij} represents the output linguistic variables from the third-tier fuzzy logic rule base. Consequently, the state space and action sets used in this study are defined in (13) and (14), respectively.

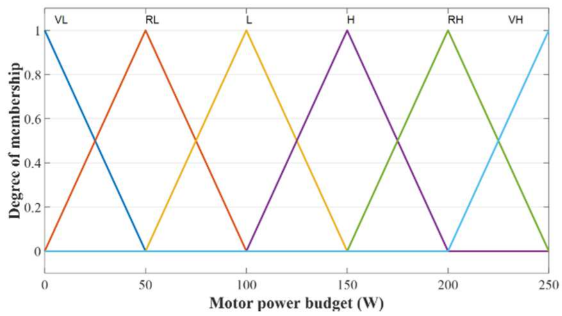

Fig. 6. The motor power budget membership function.

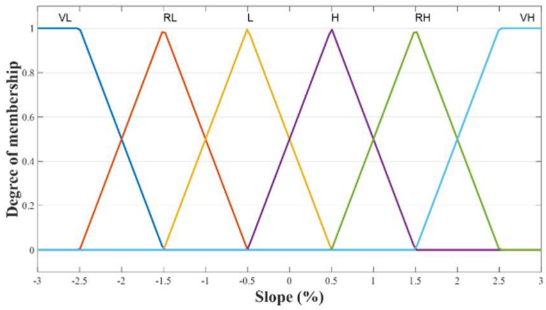

Fig. 7. The slope membership function.

$$\mathcal{S} = \{s_i | i = 1, \dots, 521{,}829\} \tag{13}$$

$$\mathcal{A} = \{a_{ij} | i = 1, \dots, 521{,}829; j = 1, \dots, 6\} \tag{14}$$

- *Action policy*

The choice of strategy in the Q-Learning algorithm will greatly affect the performance of the final system. Therefore, we adopt the ε-greedy strategy, as shown in equation (15).

$$\pi = \begin{cases} arg \max_{a_{ij}} Q(s_i, a_{ij}), & \text{with probability } 1-\varepsilon \\ random(\mathcal{A}(s_i)), & \text{otherwise} \end{cases} \tag{15}$$

The parameter ε affects the probability of selecting the action with the highest Q-value. A larger ε results in a lower probability of selecting the action with the maximum Q-value. Therefore, during the initial stages, agents typically set ε to a larger value to facilitate exploration and accumulation of Q-table values by making it easier to select other available actions in the same state. As the number of iterations increases, the agent gradually decreases ε, eventually setting it to 0, allowing the system to consistently choose the action with the highest Q-value in future iterations.

- *Rewards*

After the first two points and setting up the environment, the agent will select an action based on the current state s_t according to its behavioral policy. After executing the action and interacting with the environment, it transitions to the next state s_{t+1}. At this point, the agent receives an immediate reward r from the reward mechanism to evaluate the effectiveness of the action. This paper defines two types of rewards: range reward and trend reward. The range reward is based on whether the next state falls within the target range, while the trend reward is based on whether the next state is progressively closer to the target range.

IV. Experimental Evaluation

To prevent excessive fatigue or insufficient exercise intensity, this paper evaluates the controller's ability to maintain the rider's heart rate within the target zone. Based on the intensity and safety intervals defined in [7,9], we introduce Heart Rate Regulation Efficiency (HRRE) as a performance indicator. This metric assesses whether the heart rate remains within the target zone with a maximum error of ±10 bpm.

A. *Simulation Results*

Considering that the main application is urban riding, the heart rate reference values in the simulation phase of this paper are set to a low-intensity range and fixed at 110 bpm. This reference value is relatively easy to maintain while riding and can achieve a training effect. We apply normal pedaling force scenarios in our simulation.

The gradient of the slope significantly affects the force exerted by the rider, which can lead to instability in the rider's heart rate. The simulation scenarios are set to typical road gradients in urban areas, with the gradient range set from 0% to 3%. Through these simulation scenarios, the paper verifies whether the proposed method can maintain stable heart rates across different gradient changes. The gradient is defined as the elevation gain per 100 meters of horizontal travel; for example, a 3% gradient means an elevation gain of 3 meters for every 100 meters traveled. The simulation scenario is defined as shown in Fig. 8.

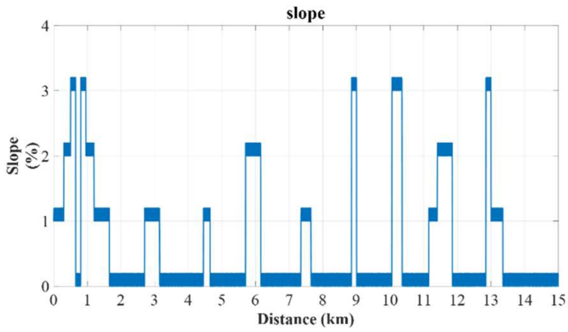

Fig. 8. Simulation scenario gradient settings.

- *Fixed heart rate reference value*

In this case, the Q-FLHRC method proposed in this paper is used, and the experimental results are shown in Fig. 9. The fixed heart rate reference value is set to 105.66 bpm. This value is used for low-intensity exercise heart rate simulation scenarios. The results show good performance on low-gradient sections. Although there are fluctuations in heart rate on high-gradient sections, it remains effectively controlled within the low-intensity exercise heart rate range, as indicated by the black dashed line. In the evaluation indicators, the HRRE value proposed in this paper is 99.27%, and the average heart rate is 106.35 bpm, which preliminarily verifies the Q-FLHRC method proposed in this paper.

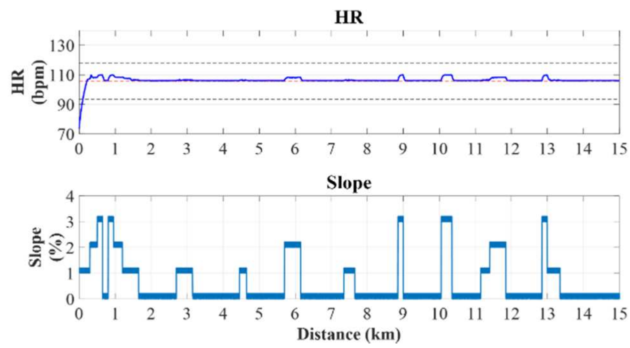

Fig. 9. Simulation results of the fixed heart rate reference value for Q-FLHRC

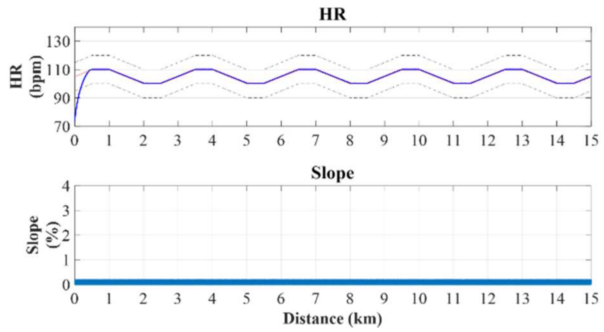

Fig. 10. Simulation results of the changing heart rate reference value for Q-FLHRC

- *Changing heart rate reference value*

To effectively verify the performance of the heart rate controller, reference [6] used a slow-changing heart rate reference value and set the simulation scenario on flat terrain to evaluate the controller's performance. To test whether the Q-FLHRC method proposed in this paper can also achieve good tracking performance with continuously varying heart rate reference values, simulation experiments were included. The simulation results are shown in Fig. 10.

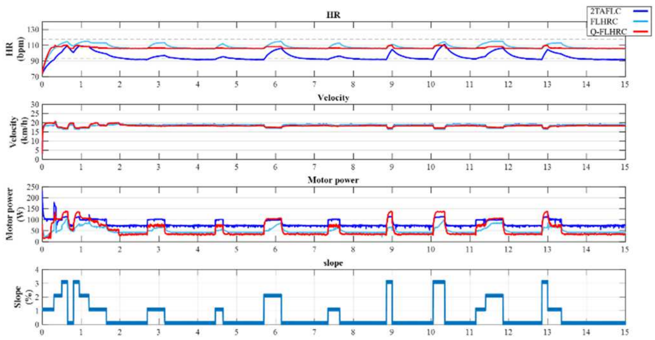

Fig. 11. Simulation results of the low-intensity exercise heart rate reference value for Q-FLHRC

The heart rate reference value is set to vary between 100 bpm and 110 bpm with fixed intervals, and the simulation scenario is on flat terrain. The results show that even with varying heart rate reference values, the controller can still effectively maintain the heart rate within a ±10 bpm range. In the evaluation indicators, the *HRRE* value proposed in this paper is 98.8%, and the root mean square error of heart rate is 2.38 bpm, which preliminarily verifies the Q-FLHRC method proposed in this paper, demonstrating its performance even with changing heart rate reference values.

- *Low-intensity exercise heart rate zone*

This section focuses on simulating and comparing the low-intensity exercise heart rate zone. The comparison methods include the two-tier adaptive fuzzy-logic control (2TAFLC) [10], FLHRC, and Q-FLHRC. The low-intensity exercise heart rate zone ranges from 93.5 bpm to 117.8 bpm, with the heart rate reference value set at the average of this range, 105.66 bpm. The speed reference value is set to the average of the comfort zone, 18 km/h, which ranges from 16 km/h to 20 km/h.

In the case of general pedaling force, with pedaling force set between 42 and 98 W, the simulation results are shown in Fig. 11. It can be observed that although the 2TAFLC method generally maintains the speed within the comfort zone, the heart rate is consistently lower throughout the entire route. This is due to the higher motor output power of 2TAFLC, making the ride easier for the rider. The FLHRC and Q-FLHRC methods proposed in this paper are able to maintain both speed within the comfort zone and heart rate within the low-intensity exercise heart rate zone. While FLHRC performs well in both speed and heart rate, noticeable fluctuations occur on high-gradient sections, and the heart rate on low-gradient sections is lower than the reference value.

TABLE I

THE LOW-INTENSITY EXERCISE HEART RATE ZONE UNDER GENERAL PEDALING FORCE SCENARIOS

Parameters / Methods	Average HR (bpm)	HRRE (%)	HR Maximum (bpm)	HR RMSE (bpm)	HR Standard Deviation (bpm)	Average Velocity (km/h)	Comfortability (%)	Safety (%)
2TAFLC	95.70	20.64	**110.98**	15.21	5.20	18.62	99.64	99.91
Proposed-FLHRC	106.98	98.46	112.91	4.48	3.31	**18.12**	**99.78**	**99.92**
Proposed-Q-FLHRC	**110.18**	**98.55**	114.89	**3.00**	**3.00**	18.56	99.64	99.91

Based on Table I, it can be seen that although the Q-FLHRC does not perform the best in terms of vehicle average speed, comfort, and safety, it does achieve good results in heart rate-related indicators. In this study, it is observed that the FLHRC shows the best performance in vehicle average speed, comfort, and safety. However, in heart rate-related evaluation indicators, the overall average heart rate is relatively low, leading to a reduced maximum heart rate. The 2TAFLC demonstrates good performance in speed, comfort, and safety based on various indicators, but there is a significant gap in heart rate-related evaluation indicators compared to FLHRC and Q-FLHRC.

From the above observations, Q-FLHRC still shows the best performance. In the average heart rate, the error for Q-FLHR is only 0.16%, FLHRC is 2.8%, and 2TAFLC is 5.7%. In this scenario, it can be seen that Q-FLHRC performs better in heart rate-related evaluation indicators. Regarding the maximum heart rate, Q-FLHRC has the highest value, followed by FLHRC. The reason is that the 2TAFLC provides more motor assistance compared to FLHRC and Q-FLHRC, resulting in a lower heart rate during the process and consequently a lower maximum heart rate.

B. Real-World Experiments

This section describes the field trial environment. A 850-meter course within a Forest Park, featuring typical urban gradients (up to 3%) and minimal traffic interference, was selected to evaluate the controllers' performance. While long-distance endurance was tested via simulation, these field trials focused on real-world feasibility.

As illustrated in Fig. 12, the baseline 2TAFLC method operates on an open-loop basis relative to the rider's physiology, focusing exclusively on physical variables. This lack of feedback results in significant heart rate (HR) fluctuations and an inability to maintain the target training intensity, which may lead to excessive fatigue or insufficient exercise efficacy. In contrast, FLHRC and Q-FLHRC exhibit superior physiological synchronization. By detecting early HR elevations and dynamically modulating motor assistance, both proposed controllers effectively stabilize the rider within the prescribed exercise zone. The metrics summarized in

Table II confirm that incorporating physiological feedback optimizes the assistance logic, significantly enhancing both safety and riding comfort.

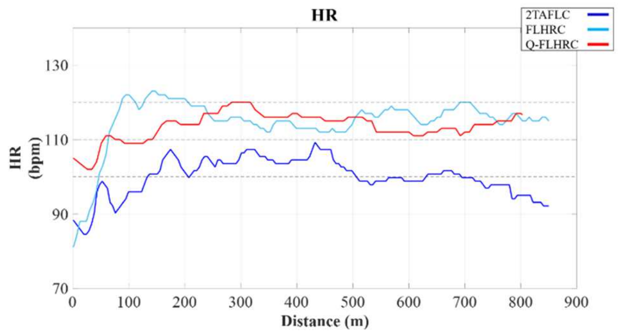

Fig. 12. . The real-world experiments.

TABLE II
QUANTITATIVE ANALYSIS OF THE VALIDATION RESULTS

Parameters / Methods	Average HR (bpm)	HRRE (%)	HR Maximum (bpm)	HR RMSE (bpm)	HR Standard Deviation (bpm)	Average Velocity (km/h)	Comfortability (%)	Safety (%)
2TAFLC	99.96	51.05	109	11.22	5.04	17.12	83.41	71.76
Proposed-FLHRC	114.78	82.24	123	8.44	6.96	18.87	**97.76**	87.76
Proposed-Q-FLHRC	**113.70**	**100**	**120**	**5.22**	**3.67**	**18.03**	89.65	**90.82**

V. CONCLUSIONS

This study developed the FLHRC and self-optimizing Q-FLHRC systems to address the limitations of conventional e-bike control. By synthesizing fuzzy logic with Q-learning, the proposed controller adaptively modulates assistance across diverse terrains and individual rider profiles. Experimental results demonstrate that Q-FLHRC significantly outperforms existing benchmarks, minimizing average heart rate error in both simulations and real-world trials while ensuring optimal safety and comfort. Ultimately, this research validates that the synergy of reinforcement learning and fuzzy logic provides a robust, feasible, and high-performance framework for personalized e-bike adaptation, effectively bridging the gap between mechanized assistance and human physiological health.

ACKNOWLEDGMENTS

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant 113-2221-E-027-070.

REFERENCES

[1] J. S. Lee and R. You, “A fuzzy logic-based automatic gear-shifting system for electric bicycles in urban mobility solutions for smart cities,” *Systems*, vol. 13, no. 4, Article 228, Mar. 2025.

[2] J. S. Lee, Z. H. Chen, and Y. Hong, “Enhancing urban mobility with self-tuning fuzzy logic controllers for power-assisted bicycles in smart cities,” *Sensors*, vol. 24, no. 5, Article 1552, Feb. 2024

[3] J. S. Lee and J. W. Jiang, “Enhanced fuzzy-logic-based power-assisted control with user-adaptive systems for human-electric bikes,” *IET Intelligent Transport Systems*, vol. 13, no. 10, pp. 1492-1498, Oct. 2019

[4] J. H. Dai and J. S. Lee, “Development of fuzzy-logic-based controllers for energy conservation in human-electric bikes,” *IEEE Conf. Consumer Electronics-Taiwan (ICCE-TW)*, Penghu, Taiwan, September 2021.

[5] D. Meyer, W. Zhang, M. Tomizuka, and V. Senner, “Heart rate regulation with different heart rate reference profiles for electric bicycle riders,” *Procedia Manuf.*, vol. 3, pp. 4213–4220, Jan. 2015.

[6] M. Corno, P. Giani, M. Tanelli, and S. M. Savaresi, “Human-in-the-loop bicycle control via active heart rate regulation,” *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 1029-1040, May 2015.

[7] D. Meyer, M. Körber, V. Senner, and M. Tomizuka, “Regulating the heart rate of human–electric hybrid vehicle riders under energy consumption constraints using an optimal control approach,” *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 2125-2138, Sept. 2019.

[8] R. G. Cedillo, D. M. Lopez, E. M. Quintana, A. P. Guerra, J. P. S. H. Moreno, J. M. V. Hernández, E. Moya-Albor, H. Ponce, and J. Brieva “Development of an electric powered assisted cycle with a heart rate sensor control system,” in *Proc. IEEE Int. Conf. Symp. Auton. Decentralized Syst. (ISADS)*, Mexico, Mar. 2023, pp. 1-7.

[9] R. C. Hsu, C. T. Liu, and D. Y. Chan, “A reinforcement-learning-based assisted power management with Qor provisioning for human–electric hybrid bicycle,” *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3350-3359, Aug. 2012.

[10] Jhen-Hong Dai, “Applications of Two-Tier Adaptive Fuzzy Control Systems to Electricity-Assisted Bikes,” *M.S. Thesis, Dept. Elect. Eng.*, National Taipei Univ. of Tech. (TaipeiTech), Taipei, Taiwan, 2021. Available: https://hdl.handle.net/11296/43s65q