Middle East Journal of Pure and Applied Sciences (MEJPAS)
Online ISSN: 3062-343X
Volume 1, Issue 2, 2025, Page No: 16-22
https://mideastjournals.com/index.php/mejpas

Residual Gradient Fuzzy Actor-Critic Learning-Based Control for Two-Tank Level System

Mostafa D. Awheda [1 *], Abdalmajid Almarimi [2]
[1,2] Department of Control Engineering, College of Electronic Technology, Bani Walid, Libya

التحكم القائم على التعلم باستخدام الممثل-الناقد الغامض وتدرج الباقي لنظام الخزانين ذي المستويين

مصطفى ضي أوحيدة [1*]، عبد المجيد المريمي[2]
[1،2] قسم هندسة التحكم الآلي، كلية التقنية الإلكترونية، بني وليد، ليبيا

*Corresponding author: mdawheda@gmail.com

Received: February 28, 2025	Accepted: April 18, 2025	Published: April 24, 2025

Abstract:
The paper presents a novel application of the residual gradient fuzzy actor-critic learning (RGFACL) algorithm to nonlinear level control of a two-tank system. In contrast to traditional fuzzy actor-critic methods using direct gradient methods, which are commonly susceptible to instability and convergence issues, the RGFACL algorithm presented in this work employs a residual gradient formulation to ensure a more precise and stable learning process. Moreover, the algorithm simultaneously adaptively adjusts the premise (input) and consequent (output) parameters of the fuzzy inference systems, increasing the expressiveness and flexibility of the control strategy. To the best of our knowledge, this is the first implementation of the RGFACL approach on a two-tank benchmark system. The simulation results demonstrate that the RGFACL algorithm achieved improved transient response, reduced overshoot, and enhanced robustness compared to the traditional PID controller. The RGFACL algorithm successfully handled abrupt setpoint changes, its ability to adjust within nonlinear, time-varying operating conditions clear. The results confirm the efficacy of employing the RGFACL learning algorithm to the nonlinear and complex environments.

Keywords: Reinforcement Learning, Fuzzy Logic Control, Residual Gradient, Two-Tank Level Control.

الملخص
تقدم هذه الورقة البحثية تطبيقًا جديدًا لخوارزمية تعلم الفاعل الناقد الضبابي بالتدرج المتبقي (*RGFACL*) للتحكم غير الخطي في المستوى لنظام ثنائي الخزانات. وعلى عكس أساليب التعلم الضبابي التقليدية التي تعتمد على الفاعل الناقد باستخدام أساليب التدرج المباشر، والتي غالبًا ما تكون عرضة لمشاكل عدم الاستقرار والتقارب، تستخدم خوارزمية *RGFACL* المعروضة في هذا العمل صيغة تدرج متبقي لضمان عملية تعلم أكثر دقة واستقرارًا. علاوة على ذلك، تضبط الخوارزمية بشكل تكيفي معلمات الفرضية (المدخلات) والنتيجة (المخرجات) لأنظمة الاستدلال الضبابي، مما يزيد من مرونة استراتيجية التحكم وتعبيريتها. وعلى حد علمنا، يُعد هذا أول تطبيق لنهج *RGFACL* على نظام معياري ثنائي الخزانات. وتُظهر نتائج المحاكاة أن خوارزمية *RGFACL* حققت استجابةً مُحسّنة للزمن العابر، وقللت من التجاوز، وعززت المتانة مقارنةً بوحدة التحكم *PID* التقليدية. نجحت خوارزمية *RGFACL* في التعامل مع التغيرات المفاجئة في نقاط الضبط، مع قدرتها الواضحة على التكيف في ظل ظروف تشغيل غير خطية ومتغيرة زمنيًا. وتؤكد النتائج فعالية استخدام خوارزمية التعلم *RGFACL* في البيئات غير الخطية والمعقدة.

الكلمات المفتاحية: التعلم التعزيزي، التحكم المنطقي الضبابي، التدرج المتبقي، التحكم في مستوى الخزانين.

Introduction

Fuzzy logic controllers (FLCs) have become popular for controlling nonlinear, uncertain, and ill-defined systems [3]-[6], [9], [22], [23]. A lot of supervised learning techniques including genetic algorithms, gradient descent and clustering have been employed for the tuning of FLCs by the input-output data [10]. However, these methods tend to depend on expert knowledge and can potentially have high cost in terms of expensive or difficult-to-obtain data. In contrast, reinforcement learning provides a model-free, reward-based paradigm that demands no a priori knowledge or expert guidance [17].

Reinforcement learning (RL) enables an agent to learn how to perform best actions from trial-and-error interaction in its world through only evaluative feedback, rather than through expert demonstrations [18], [19]. RL has been

successfully applied to robotic control [12], [14], [20] and nonlinear optimal control issues [21], [26], [27]. Traditional RL approaches utilize lookup tables as value functions, which suffer from the curse of dimensionality when the state size increases and cannot handle continuous differential-game worlds without discretization [11], [18]. To address these limitations, function approximation systems (FASs), namely gradient-descent–based approximators, are utilized in order to make RL scale up to large or continuous state–action spaces as well as enable online learning [18], [28]. Several fuzzy reinforcement learning algorithms have been introduced in the literature to control systems with continuous state-action spaces through gradient-descent-based FASs [12], [13], [15], [16]. The author in [15] proposed a fuzzy actor critic learning (FACL) algorithm that uses fuzzy inference systems (FISs), or FASs, to represent the continuous state-action spaces. The FACL algorithm used the temporal difference (TD) error calculated by the state value function to tune the parameters of the FASs. However, the FACL algorithm only tuned the output parameters of the FISs, while the input parameters of the FISs were kept fixed. The authors in [16] proposed the Q-learning fuzzy inference system (QLFIS) algorithm. Unlike the FACL algorithm, the QLFIS algorithm tuned both the input and output parameters of its FISs. However, both the FACL and QLFIS algorithms relied on what commonly referred to as "direct algorithms described in [29]" to tune their FASs parameters. In spite of the fact that the direct algorithms have been widely applied in the tunning process for the FISs' parameters, the direct algorithms may lead FISs to indeterminate solutions and to divergence in others [29]-[31]. The authors in [12], [13] proposed the residual gradient fuzzy actor critic learning (RGFACL) algorithm. The RGFACL used the TD error of the state-action value functions of the two successive states in the state transition to tune the input and output parameters of its FASs. Unlike the FACL and QLFIS algorithms, the RGFACL algorithm relied on the residual gradient algorithms to tune the input and out of its FISs, where the residual gradient algorithms are always guaranteed to converge to a local minimum compared to the direct methods [29]-[31].

This paper represents the novelty of applying a residual gradient-based fuzzy actor-critic learning (RGFACL) algorithm for the nonlinear two-tank level control problem. As compared to traditional fuzzy actor-critic schemes using direct gradient methods that are highly sensitive to instability and convergence issues, the proposed solution implements a residual gradient formulation to yield a more accurate and stable learning trajectory. In addition, the algorithm also dynamically changes both the premise (input) and consequent (output) parameters of the fuzzy inference systems in parallel, enhancing the flexibility and expressiveness of the control policy. To our knowledge, it is the initial application of the RGFACL algorithm to the two-tank benchmark system and has demonstrated superior control performance in terms of tracking accuracy, convergence rate, and disturbance rejection, and thereby its potential in practical applications for industrial process control.

Preliminary concepts and notations

There are several types of fuzzy inference systems (FIS) in literature. The ones used in this work are the zero-order Takagi-Sugeno-Kang (TSK) FISs with constant consequences which are represented in [32], [33]. In each FIS, there are L rules, and there are n states and one constant consequent in each rule. Each rule ($l = 1, \ldots, L$) is represented as follows:

$$R_l\text{: IF } s_1 \text{ is } f_1^l, \ \ldots \ , \text{and } s_n \text{ is } f_n^l \quad \text{THEN} \quad z_l = k_l \qquad (1)$$

where s_i, ($i = 1, \ldots, n$), is the ith input state to the FIS, n is the number of input states, and f_i^l is the linguistic value of the input state s_i at the rule l. Each input state s_i has h membership functions (MFs). The variable z_l is the output of the rule l, and k_l is a constant that describes the consequent parameter of the rule l. We used Gaussian membership functions (MFs) for the inputs to the FISs, where each MF is defined as follows,

$$\mu^{F_i^l}(s_i) = \exp\left(-\left(\frac{s_i - m}{\sigma}\right)^2\right) \qquad (2)$$

where σ and m represent the standard deviation and the mean of the MF, respectively.

In each FIS, there are H of standard deviations and H of means of all MFs, where $H = n \times h$.

The residual gradient fuzzy actor critic learning (RGFACL) algorithm

The RGFACL algorithm uses three FISs to represent the continuous state-action spaces; one for the actor (fuzzy logic controller, FLC), and two for the critics [12], [13]. The critic estimates the value functions $V_t(s_t)$ and $V_t(s_{t+1})$ of the learning agent at two different states, s_t and s_{t+1} respectively. The actor, on the other hand, is responsible for providing a continuous action u_t at each continuous state st. The RGFACL algorithm is shown in Fig (1). The critic has input and output parameters. The input parameters are the parameters of the membership functions (MFs) of the critic's inputs: σ_j and m_j (where $j = 1, \ldots, H$); while the output parameters are the

consequent parameters of its rules, k_l. We refer to the input and output parameters of the critic as ψ^C. Similarly, the actor has input and output parameters. The input parameters of the actor are σ_j and m_j, the parameters of the MFs of the critic's input. On the other hand, the critic's outputs are k_l, the consequent parameters of the critic's rules k_l. We refer to the input and the output parameters of the actor as ψ^A. The temporal difference error, Δ_t, and its mean square error, E, are defined as follows,

$$\Delta_t = r_t + \gamma V_t(s_{t+1}) - V_t(s_t) \quad (3)$$

$$E = \frac{1}{2}\Delta_t^2 \quad (4)$$

where r_t is the immediate reward of the learning agent, and γ is a discount factor.

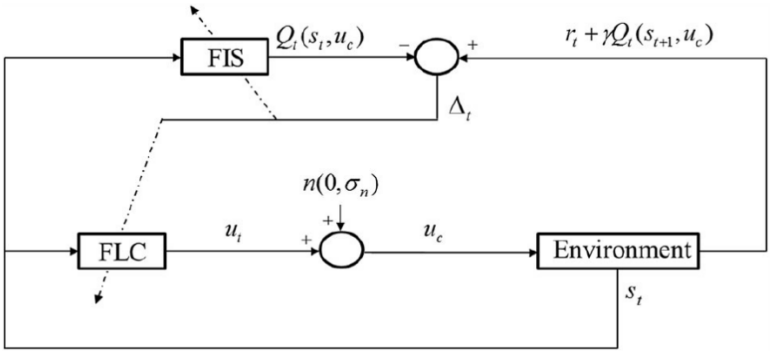

Figure 1. The RGFACL algorithm [12], [13], [16].

The RGFACL algorithm updates the input and output parameters of its actor and critics based on the residual gradient value iteration algorithm described in [29]. The RGFACL algorithm uses the following rules to update the input and output parameters of its actor and critics [12], [13]:

$$\psi_{t+1}^C = \psi_t^C - \alpha\frac{\partial E}{\partial \psi_t^C} \quad (5)$$

$$\psi_{t+1}^A = \psi_t^A + \beta\Delta_t\frac{\partial u_t}{\partial \psi_t^A}\left[\frac{u_c-u_t}{\sigma_n}\right] \quad (6)$$

where α and β are learning rates, u_t is the output of the actor, and u_c is the output of the actor with a random Gaussian noise.

Two-tank level system

A. System description:

The two-tank level system consists of two vertically placed tanks. Tank 1 receives input from the exterior and output to Tank 2. Tank 2's output goes to the environment. The two tanks should be cylindrical with a constant cross-sectional area and incompressible fluid.

The following notations can be employed:

- A_1, A_2: Cross-sectional area of Tank 1 and Tank 2 [m2].
- $h_1(t)$, $h_2(t)$: Liquid heights in Tank 1 and Tank 2 at time t [m].
- $q_{in(t)}$: Rate of inflow into Tank 1 $[m^3/s]$.
- $q_{12}(t)$: Rate of flow from Tank 1 into Tank 2 $[m^3/s]$.
- $q_{out}(t)$: Rate of outflow from Tank 2 to the environment $[m^3/s]$.
- C_1, C_2: Flow coefficients for Tank 1 and Tank 2 discharge outlets.

Assuming flow rates follow Torricelli's Law for free outflow under gravity [1], the inter-tank and output flow rates are given by:

$$q_{12}(t) = C_1\sqrt{h_1(t) - h_2(t)} \quad (7)$$

$$q_{out}(t) = C_2\sqrt{h_2(t)} \quad (8)$$

B. Continuous-time model:

Imposing a mass balance on the two tanks results in the following first-order nonlinear differential equations [1]:

Tank 1:

$$A_1 \frac{dh_1(t)}{dt} = q_{in}(t) - C_1\sqrt{h_1(t) - h_2(t)} \quad (9)$$

Tank 2:

$$A_2 \frac{dh_2(t)}{dt} = C_1\sqrt{h_1(t) - h_2(t)} - C_2\sqrt{h_2(t)} \quad (10)$$

These equations capture the nonlinear behavior of the tank system due to the square-root flow relationships.

C. Discrete-time model:

Using the forward Euler method for discretization with a sampling time T_s, the discrete-time model becomes:

$$h_1(k+1) = h_1(k) + \frac{T_s}{A_1}\left(q_{in}(k) - C_1\sqrt{h_1(k) - h_2(k)}\right) \quad (11)$$

$$h_2(k+1) = h_2(k) + \frac{T_s}{A_2}\left(C_1\sqrt{h_1(k) - h_2(k)} - C_2\sqrt{h_2(k)}\right) \quad (12)$$

This discrete model is suitable for digital control design and simulation applications.

Reward function for the RGFACL algorithm in a two-tank level system

The RGFACL-based learning agent tunes the parameters of both the actor and the critics so that the liquid level of the second tank, $h_2(t)$, converges to a reference level h_{ref} . The reward function at time t is formulated as:

$$r(t) = -\alpha_1 e(t)^2 - \alpha_2 \Delta e(t)^2 - \alpha_3 \Delta u(t)^2 \quad (13)$$

where, α_1, α_2, $\alpha_3 > 0$ are weighting coefficients,

$$e(t) = h_{\text{ref}} - h_2(t)$$

is the error at time$t$,

$$\Delta e(t) = e(t) - e(t-1)$$

is the change in error, and

$$\Delta u(t) = u(t) - u(t-1)$$

is the change in the actor's output.

- The term $-\alpha_1 e(t)^2$ encourages accuracy, and penalizes large deviations from the reference level.
- The term $-\alpha_2 \Delta e(t)^2$ penalizes rapid changes in error, encouraging stability and discouraging oscillations.

The term $-\alpha_3 \Delta u(t)^2$ penalizes aggressive control actions, promoting smooth control signals.

Simulation and results

A. Simulation Setup:

The RGFACL algorithm is implemented to control the level of the second tank in a nonlinear two-tank liquid level system. The RGFACL algorithm tunes the parameters of its actor and critics so that the error between the second tank's level h_2 and a reference level is minimized (i.e. to maximize the reward function $r(t)$).
The system dynamics are described by:

$$Q_{\text{out1}} = 0.3\sqrt{h_1}$$

$$Q_{\text{out2}} = 0.3\sqrt{h_2}$$

where h_1 and h_2 are the fluid heights in tanks 1 and 2, respectively. The tanks cross-sectional areas and timestep used are A_1 = 1.0, A_2 = 0.8, Δ_t = 0.05 seconds. The reference level h_{ref} was randomly initialized during training within the range 5 to 10 meters. The state s_t is defined as the error between the level of the second tank and the

target level. Three Gaussian membership functions (MFs) are used to define the fuzzy sets of each input of the three FISs. The learning rates of the RGFACL algorithm are chosen as those presented in [13].

B. Results:

The training was conducted over 50 episodes with 5000 steps in each episode. During the test, the level reference was initially set at h_{ref} =10.0; the simulation took place over 10,000 steps, with the change of level reference applied at t = 3000 (down to 5.0m) and at t = 6000 (back up to 10.0m), to measure the tracking performance.

Figure 2 shows the tank level response over time. The RGFACL algorithm successfully tuned the input and output parameters of its actor and critics so that the level of the second tank reaches the target level with fast convergence and minimal overshoot, despite sudden changes in the target level.

C. Discussion:

The simulation experiments evaluated the performance of the RGFACL algorithm against a conventional PID controller in a two-tank level control system. The dynamics of the two-tank level system were governed by nonlinear flow equations, with Tank 2's level (h_2) regulated to track reference setpoints (h_{ref}) under abrupt changes at t = 3000s (10m → 5m) and t = 6000s (5m → 10m). The RGFACL algorithm was more adaptable than the PID controller and took less time to settle and had lower overshoot, as seen in Figure 2. The reward function successfully balanced transient and steady-state performance through penalizing the squared error, control effort, and input deviations. The PID controller, although stable, exhibited steady oscillations and small steady-state errors, highlighting the limitation of fixed gains on nonlinear systems. The RGFACL algorithm provided rapid convergence of the level of the second tank to its desired level, even under setpoint changes. The results confirm the efficacy of employing the RGFACL algorithm to the nonlinear and complex environments.

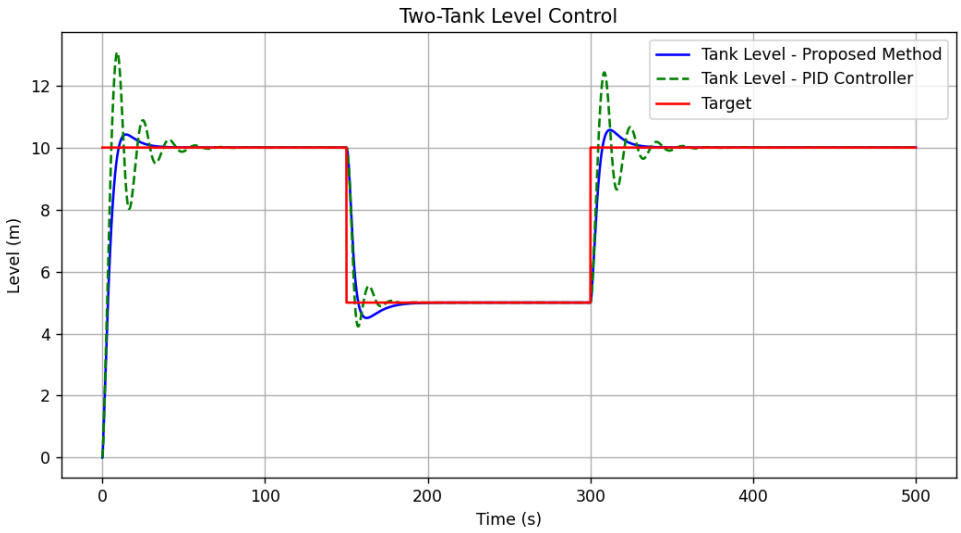

Figure 2. Tank level h_2 response over time under the RGFACL algorithm with varying reference levels h_{ref}.

Conclusion

In this research study, the residual gradient fuzzy actor-critic learning (RGFACL) algorithm was evaluated and implemented for controlling a nonlinear two-tank level system. The RGFACL algorithm succeeded in maintaining the level of the second tank in the nonlinear two-tank level system at its desired level. The RGFACL algorithm successfully handled abrupt setpoint changes, its ability to adjust within nonlinear, time-varying operating conditions clear. Compared to traditional PID control strategies, the RGFACL method demonstrated greater robustness and flexibility with faster convergence, less overshoot, and improved tracking precision. These results highlight the engineering usability of the RGFACL algorithm in process control in industries, particularly in applications where nonlinearity and disturbance rejection are of greatest significance.

References

1. K. Ogata, Modern Control Engineering, 5th ed. Upper Saddle River, NJ: Prentice Hall, 2010.
2. K. J. A˚ stro¨m and R. M. Murray, Feedback Systems: An Introduction for Scientists and Engineers. Princeton, NJ: Princeton University Press, 2010.
3. N. Bhuvaneswary, B. G. Reddy, K. P. Kumar Reddy, M. V. Vardhan, K. Jagadeesh Kumar and M. A. Venkat Reddy, "Quadcopter with Fuzzy Logic based Self-Tuning PID Controller," 2025 5th

International Conference on Trends in Material Science and Inventive Materials (ICTMIM), Kanyakumari, India, pp. 559-563, 2025.
4. S. Sahoo, N. K. Jena, A. Anurag, A. K. Naik, B. K. Sahu and P. C. Sahu, ”Frequency Regulation of Microgrid using Fuzzy-logic based controllers,” 2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT), Bhubaneswar, India, pp. 120-123, 2023.
5. N. W. Madebo, ”Enhancing Intelligent Control Strategies for UAVs: A Comparative Analysis of Fuzzy Logic, Fuzzy PID, and GA-Optimized Fuzzy PID Controllers,” in IEEE Access, vol. 13, pp. 16548-16563, 2025.
6. Singh, S. Yadav, S. Sarangi, S. De, A. K. Singh and R. K. Singh, ”Optimizing BLDC Motor Performance: A Study of PID, Artificial Neural Network and Fuzzy Logic Controllers,” 2025 International Conference on Innovation in Computing and Engineering (ICE), Greater Noida, India, pp. 1-6, 2025.
7. Jahnavi, S. R. Gummadi, U. Gunteiah, J. Tulluri and D. Niharika, ”Enhancing matrix converter output voltage control through fuzzy logic controller,” 2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST), Vijayawada, India, pp. 1-5, 2024.
8. Rafiee and F. Baghbani, ”Control of a Four-Wheeled Mobile Robot Using Fuzzy Logic Controller,” 2024 19th Iranian Conference on Intelligent Systems (ICIS), Sirjan, Iran, Islamic Republic of, pp. 163-168, 2024.
9. H. K. Lam and FH F. Leung, Fuzzy controller with stability and performance rules for nonlinear systems, Fuzzy Sets and Systems 158.2, pp. 147-163, 2007.
10. S. F. Desouky and H. M. Schwartz, Self-learning fuzzy logic controllers for pursuit-evasion differential games, Robotics and Autonomous Systems, 59, pp. 22-33, 2011.
11. T. Jaakkola, M. I. Jordan, and P. S. Singh, On the convergence of stochastic iterative dynamic programming algorithms, Neural Computation, 6.6, pp. 1185-1201, 1994.
12. M. D. Awheda and H. M. Schwartz, The residual gradient FACL algorithm for differential games, Electrical and Computer Engineering (CCECE), IEEE 28th Canadian Conference on, pp. 1006-1011, 2015.
13. M. D. Awheda and H. M. Schwartz, A residual gradient fuzzy reinforcement learning algorithm for differential games, International Journal of Fuzzy Systems, Accepted for publication, Oct. 2016.
14. H. M. Schwartz, Multi-agent machine learning: A reinforcement approach, John Wiley and Sons, Inc., New York, 2014.
15. L. Jouffe, Fuzzy inference system learning by reinforcement methods, IEEE Trans. Syst., Man, Cybern. C, 28.3, pp. 338-355, 1998.
16. S. F. Desouky and H. M. Schwartz, Q (λ)-learning adaptive fuzzy logic controllers for pursuit-evasion differential games, International Journal of Adaptive Control and Signal Processing, 25.10, pp. 910-927, 2011.
17. G. Barto, R. S. Sutton and C. W. Anderson, Neuronlike adaptive elements that can solve difficult learning control problems, Systems, Man and Cybernetics, IEEE Transactions on, 5, pp. 834-846, 1983.
18. R. S. Sutton and A. G. Barto, Reinforcement learning: an introduction, The MIT Press, Cambridge, Massachusetts, 1998.
19. L. P. Kaelbling, M. L. Littman and A. W. Moore, Reinforcement learning: A survey, arXiv preprint cs/9605103, 1996.
20. M. Rodr´ıguez, R. Iglesias, C. V. Regueiro and J. Correa and S. Barro, Autonomous and fast robot learning through motivation, Robotics and Autonomous Systems, 55.9, pp. 735-740, 2007.
21. Luo, H.N. Wu and T. Huang, Off-policy reinforcement learning for H∞ control design, Cybernetics, IEEE Transactions on, 45.1, pp. 65-76, 2015.
22. Jahnavi, S. R. Gummadi, U. Gunteiah, J. Tulluri and D. Niharika, ”Enhancing matrix converter output voltage control through fuzzy logic controller,” 2024 International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST), Vijayawada, India, pp. 1-5, 2024.
23. Rafiee and F. Baghbani, ”Control of a Four-Wheeled Mobile Robot Using Fuzzy Logic Controller,” 2024 19th Iranian Conference on Intelligent Systems (ICIS), Sirjan, Iran, Islamic Republic of, pp. 163-168, 2024.
24. M. V. Kumar and M. Kudadala, ”A Novel Fuzzy Logic Controller Topology for Grid Connected PV System by DC Voltage Droop Controller,” 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC), Mysore, India, pp. 1-5, 2023.
25. K. -H. Choi and H. -J. Chang, ”Improvement of Adaptive Cruise Control System Performance on Sloped Roads Based on Adaptive Neuro-Fuzzy Inference System,” in IEEE Access, vol. 13, pp. 60519-60531, 2025.
26. H. Modares, F. L. Lewis and M.B. Naghibi-Sistani, Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems, Automatica, 50.1, pp. 193-202, 2014.

27. W. Dixon, Optimal adaptive control and differential games by reinforcement learning principles, Journal of Guidance, Control, and Dynamics, 37.3, pp. 1048-1049, American Institute of Aeronautics and Astronautics, 2014.
28. Bonarini, A. Lazaric, F. Montrone, and M. Restelli, Reinforcement distribution in fuzzy Q-learning, Fuzzy sets and systems, 160.10, pp. 1420-1443, 2009.
29. L. Baird, Residual algorithms: Reinforcement learning with function approximation, In ICML, pp. 30-37, 1995.
30. G. J. Gordon, Reinforcement learning with function approximation converges to a region, Advances in neural information processing systems, 2001.
31. R. Schoknecht and A. Merke, TD(0) converges provably faster than the residual gradient algorithm, ICML, 2003.
32. T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modelling and control, IEEE Transactions on Systems, Man and Cybernetics, SMC, 15.1, pp. 116-132, 1985.
33. M. Sugeno and G. Kang, Structure identification of fuzzy model, Fuzzy Sets Syst., 28, pp. 15-33, 1988.