Dynamic Control Comparison: Tuned-PID vs Trained Reinforcement Learning of the Double Inverted Pendulum on a Two-Wheeled Platform

Sean Audie I. Buscano II
Department of Electronics Engineering, College of Engineering and Geosciences

Re-Ann Cristine O. Calimpusan
Department of Electronics Engineering, College of Engineering and Geosciences

Rudolph Joshua U. Candare
Mine GEARS Caraga State University

Ydron Paul C. Amarga
Department of Electronics Engineering, College of Engineering and Geosciences

https://doi.org/10.5109/7323283

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 10, pp.340-346, 2024-10-17. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Dynamic Control Comparison: Tuned-PID vs Trained Reinforcement Learning of the Double Inverted Pendulum on a Two-Wheeled Platform

Sean Audie I. Buscano II[1], Re-Ann Cristine O. Calimpusan[2], Rudolph Joshua U. Candare[3], Ydron Paul C. Amarga[4]
[1,2,4]Department of Electronics Engineering, College of Engineering and Geosciences, [3]Mine GEARS
Caraga State University, Ampayon, Butuan City 8600, the Philippines
Corresponding author email: seanaudie.buscano@carsu.edu.ph

Abstract: *This study offers a thorough and comprehensive examination of the double-inverted pendulum system, employing both PID control and reinforcement learning (RL) techniques. Through meticulous evaluation, the research compares the efficacy of these methodologies in stabilizing the system's dynamic behaviors. The Stand-up Routine Test meticulously assesses various performance metrics, including response times, overshoot values, and settling times, elucidating RL control's superiority in terms of agility and stability over PID control. Furthermore, the exploration extends into the Circular Routine Test, where RL control demonstrates the potential for achieving faster completion times, highlighting its adaptability in dynamic scenarios. Despite inherent variability, RL control consistently outperforms PID control, indicating its promising applications. The study underscores the dynamic nature of control methodologies, emphasizing the imperative for ongoing experimentation to optimize performance and practical implementation.*

Keywords: Double-Inverted Pendulum, PID control, Reinforcement Learning, Dynamic systems.

1. INTRODUCTION

This study investigates the dynamic control of a Double-Inverted Pendulum on a Two-Wheeled Platform (DIPTWP) using Proportional-Integral-Derivative (PID) and Reinforcement Learning (RL) techniques. The DIPTWP system, chosen for its representation of complex nonlinear dynamics, presents challenges such as instability and under actuation, making it ideal for evaluating control strategies. By comparing linear (PID) and nonlinear (RL) controllers, the study aims to provide insights into the effectiveness of different algorithms for stabilizing highly nonlinear dynamic systems. RL, with its trial-and-error learning approach, is hypothesized to outperform PID in controlling the double-inverted pendulum, showcasing its potential for real-world applications.

The study aims to design an RL-based neural network control, demonstrate its stability and responsiveness compared to PID control, and validate RL's superiority in controlling the DIPTWP. Current robotics technologies will create brand-new control applications based on the double-inverted pendulum stabilization principle. When it is challenging, if not impossible, to explicitly identify the set of behaviors that would cause the system to behave as desired, such as when attempting to make a two-legged robot walk smoothly, applying learning techniques to control problems makes sense. In such situations, it would be ideal if the robot could experiment with various strategies before eventually figuring out the best one.

In this study, the researcher develops a more effective method of controlling the double-inverted pendulum compared to conventional methods. This study also facilitates the creation of a system simulator, which can be used during the learning process.

Table 1. Comparative Performance of Control Techniques in Double Inverted Pendulum System, with Overview of Related Literature Review (RRL)

	Control Technique	Rise Time (s)	Overshoot (%)	Settling Time (s)	Steady-State Error	Simulation (successfully stabilized)	Real-Time Implementation
Banerjee et al. [1]	LQR	-	-	-	-	✓	×
Razzaghi et al. [2]	PID	-	-	-	-	✓	×
Liu et al. [3]	ADRC	-	-	-	-	✓	×
Yi et al. [4]	RL-DQN	-	-	-	-	✓	×
Raj [5]	RL	-	-	0.38	-	✓	×
Brown et al. [6]	RL-PPO	-	-	-	-	✓	×
Simulated Conventional Control	PID	0.04	4.42	0.58	0	✓	×
Conventional Control	PID	0.78	3.55	1.61	-0.22	-	✓
Proposed Method	RL-PPO	0.51	1.16	1.32	0.36	×	✓

" - " : not " × " : not applied " ✓ " : applied

2. REVIEW OF RELATED LITERATURE

All studies reviewed in Table 1 were conducted exclusively in simulated environments, effectively stabilizing the inverted pendulum system. However, no key parameters such as rise time, overshoot, settling time, or steady-state error were systematically recorded. Only one study observed and compared settling times between their proposed method and conventional controls. In contrast, this study aims to implement and measure these parameters in simulation and real-time hardware implementation.

3. STUDY AREA

The study was conducted at the MINE GEARs office at Caraga State University, Ampayon, Butuan City. This office serves as the focal point for the MINE GEARs project, a dynamic initiative focused on cutting-edge advancements in AI (Artificial Intelligence) and Robotics. Functioning both as a workspace and a laboratory, the MINE GEARs office provides a conducive environment for research and experimentation in the fields of AI and Robotics. Equipped with state-of-the-art facilities and resources, including Arduino and ESP32 microcontrollers, the office offers a robust infrastructure for conducting comprehensive studies on complex systems such as the Double-Inverted Pendulum on a Two-Wheeled Platform (DIPTWP).

3. MATERIALS AND METHODS

This part of the study focuses on the design and control techniques for balancing the double inverted Pendulum on the two-wheel cart platform using two different controls. In addition, the researchers analyze the output (response) of the system in two (2) different tests. Overall, this chapter provides a detailed account of the methodology used to answer the research objectives, demonstrating the validity and reliability of the study's findings.

3.1. General Framework

The figure below provides a visual representation of the flow from design and fabrication to programming, testing, and analysis of the TWDIP system.

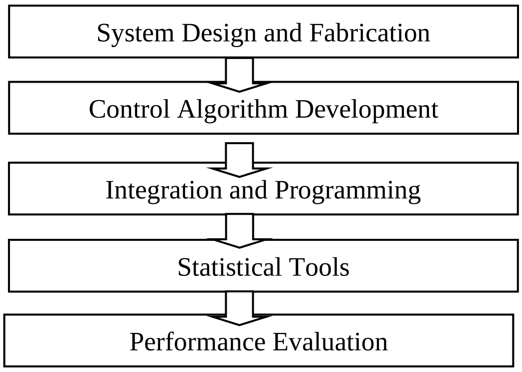


Fig. 1. Overall Framework

3.2. System Circuit Diagram

A top block diagram and circuit diagram illustrate the electrical architecture of the DIPTWP system, highlighting the integration of essential components like sensors, motor drivers, and wireless transceivers.

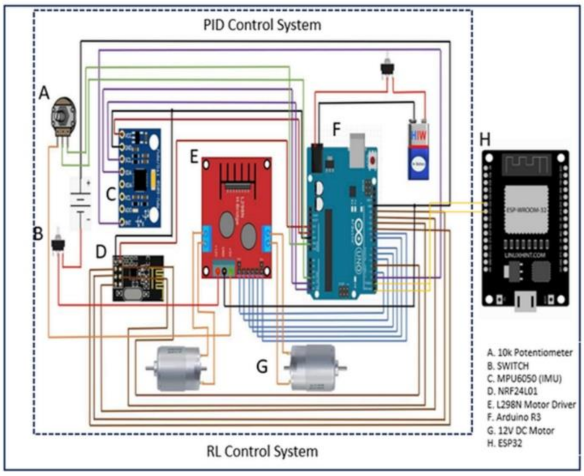


Fig. 2. Top Block Diagram of the DIPTWP System

The circuit design integrates key components, including MPU6050, L298N Motor Driver, NRF24L01 wireless transceiver, 10k Potentiometer, and ESP32 module for stabilizing the Double Inverted Pendulum on a Two-Wheeled Cart. This setup allows for effective motion data gathering, precise motor control, wireless communication, and advanced control algorithms, enabling precise system stabilization in diverse conditions.

Fig 3. Block Diagram of the Hardware PID Algorithm flow.

The Arduino R3 processes data from IMU sensors to determine the orientation and angular velocity of the Double Inverted Pendulum (DIP) system. This data generates a control signal for balancing the DIP, which is transmitted to DC motors through PWM signals. The core of this strategy is the PID controller, which continuously evaluates the error signal between desired and actual orientation, ensuring precise and stable control, even in the presence of disturbances. This integration enhances the system's capability for agile and dependable control of dynamic systems like DIP.

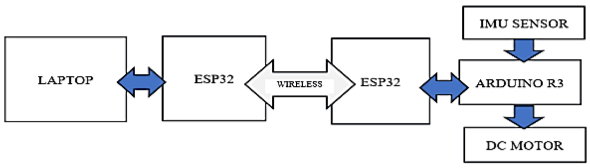


Fig. 4. Block Diagram of the Hardware RL Algorithm Flow.

Two ESP32 modules enable robust bidirectional communication. One connects to a laptop for real-time monitoring and remote parameter adjustments, while the other interfaces with Arduino, facilitating seamless data exchange with the physical system. This setup enhances versatility by enabling remote monitoring, real-time adjustments, and prompt responses to changing conditions without direct physical interaction.

3.3. Overall System Design

The overall system design of the double inverted Pendulum on a two-wheeled cart includes the system's mechanical, electrical, and control aspects. It involves the creation of the physical structure, including the cart and Pendulum, and selecting and integrating electrical components, such as motors, sensors, and controllers, to achieve the desired system behavior. The system design also includes the implementation of control algorithms, such as PID or reinforcement learning, to regulate the motion of the Pendulum and cart. The procedure takes into consideration the system requirements, such as stability, accuracy, and performance, and may involve iterative testing and optimization to achieve the desired system performance.

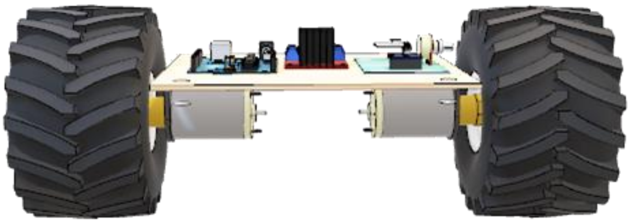


Fig. 5 Two-wheeled Platform of the TWDIP system

The figure above depicts the two-wheeled Platform or cart, a highly dynamic and underactuated system. The Platform comprises a rigid frame with two wheels attached to opposite sides, propelled by electric motors. The first layer consists of the leading electronics components, such as the Arduino microcontroller, motor driver, Inertial Measurement Unit (IMU) sensor, and potentiometer.

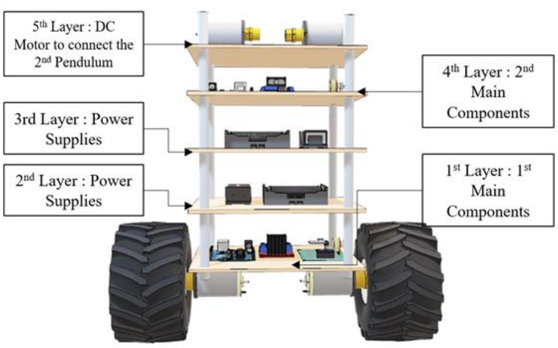


Fig. 6. Inverted Pendulum on a Cart

The cart and single inverted pendulum system, as shown in Fig. 6, constitute a dynamic and nonlinear system. It consists of a cart that traverses along a track and a pendulum attached to the cart, carrying all the electronics within each platform responsible for controlling the system. The Pendulum, actuated by gravity and internal forces, exhibits inherently unstable behavior. The system's objective is to stabilize the Pendulum upright while simultaneously controlling the cart's motion along the track.

The system is organized into several layers for efficient functionality. The first layer includes essential electronic components such as the Arduino microcontroller, motor driver, Inertial Measurement Unit (IMU) sensor, and potentiometer. Layers two and three are dedicated to power supplies, ensuring stable operation throughout. Layer four replicates the first, housing additional main electronics components. Finally, the fifth layer incorporates a DC motor designed to interface with the secondary pendulum for enhanced system dynamics.

Fig. 7. Double Inverted Pendulum on a Cart

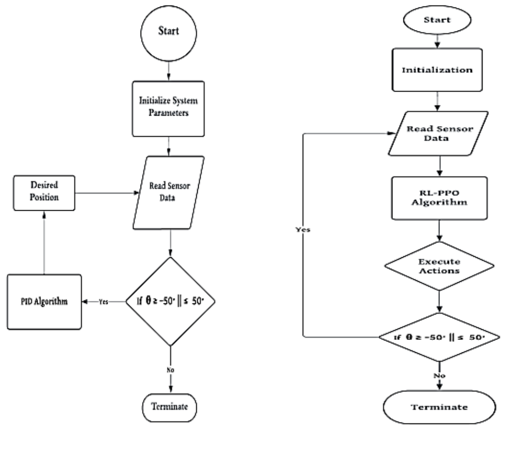
Fig. 7 shows the overall system design, depicting the double-inverted pendulum mounted on a cart. Notably, the system includes the integration of a second pendulum, which is connected to the DC motor located on the 5th layer. Alongside the pendulum, the 2nd IMU sensor is also incorporated into the system.

Despite the system's apparent simplicity, the double-inverted pendulum system exemplifies a typical nonlinear dynamic. The model exhibits significant nonlinearity, particularly concerning the pendulum angle during the transition from a pendulum position to an upright stance. The real challenge lies in developing an effective control algorithm to manage its complex behavior.

Fig. 8. Actual Double Inverted Pendulum on a Cart

3.4. Flow Chart of the System

The control process begins by setting system parameters and then measuring crucial variables like pitch angle and velocity to evaluate the system's condition. Next, the control algorithm calculates the error by comparing the desired state (setpoint) with the measured state, determining the system's deviation from the target.



(a) (b)

Fig. 9. Flow Chart of the System. (a) PID Control. (b) RL Control

Figure 9(a) showcases the implementation of the PID algorithm. It generates a control signal to guide the system towards its target state, with parameters continuously evolving for improved performance. Termination occurs when objectives are met or if the pitch angle deviates beyond defined limits.

In contrast, Figure 9(b) illustrates the RL approach, employing the Proximal Policy Optimization (PPO) algorithm, known for its stability and effectiveness. It gathers sensor data from an Arduino via an ESP32 module, transmitting wirelessly to a laptop for real-time processing. The RL-PPO algorithm then maps this data to numerical commands, refining decision-making through iterative learning for enhanced stability and performance.

3.5 Performance Evaluation

This phase involves a comprehensive analysis of the double-inverted pendulum system's dynamic response through step response metrics, including rise time, settling time, overshoot, and steady-state error. Comparative assessments between Proportional-Integral-Derivative (PID) and Reinforcement Learning (RL) control strategies unveil nuanced insights into their respective efficacy.

•Rise Time: The time taken for the response to reach from 10% to 90% of its final value.

•Overshoot: The maximum percentage or degrees by which the response exceeds its final steady-state value

•Settling Time: The time required for the response to settle within a specified error band around its final value.

•Steady-State Error: The difference between the desired and actual values once the response has stabilized.

3.6 Statistical Analysis

To quantify and ascertain any statistically significant differences between the control methods, a two-sample independent t-test was conducted. This test is commonly employed to compare the means of two independent groups, making it suitable for evaluating the efficacy of distinct control algorithms. The t-test formula utilized for this comparison is presented below [13]:

$$t = \frac{(\overline{x_1} - \overline{x_2})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

Where $\overline{x_1}$ *and* $\overline{x_2}$ represent the sample mean of the PID and RL control groups respectively.

Here, s_1 and s_2 represent the standard deviations of the PID and RL control groups, while n_1 and n_2 denote the sample sizes of the two groups.

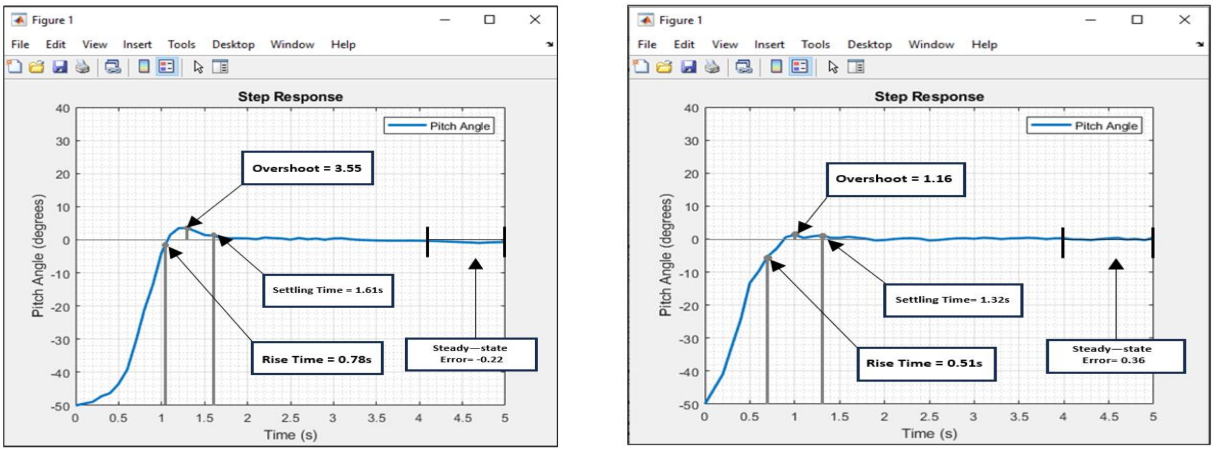
The obtained t-value is compared to the critical value of the t-distribution with degrees of freedom equal to $n_1+n_2-2n_1+n_2-2$ at a predetermined significance level (e.g., 0.05). A statistically significant t-value indicates a notable performance disparity between the PID and RL control methods. The statistical analysis, conducted using MATLAB, employed a two-sample independent t-test to compare the means of the PID and RL control groups, which was then compared to the critical t-value to determine statistical significance, with a significant result indicating a significant difference in performance between the control methods.

4. RESULTS AND DISCUSSION

This section demonstrates the efficacy of PID and RL control techniques in stabilizing a double-inverted pendulum on a self-balancing robot. It entails implementing both techniques on a physical robot platform and evaluating their performance in terms of stability, accuracy, and response time.

4.1. Stand-Up Routine Test

The Step Response of PID control, depicted in Figures 6 a and b, unveils how the system reacts to a step input, revealing nuances in its behavior and performance. Analyzing this response provides valuable insights into the effectiveness of the employed PID control method in regulating system dynamics, aiding researchers in understanding its intricacies and performance.



(a) (b)

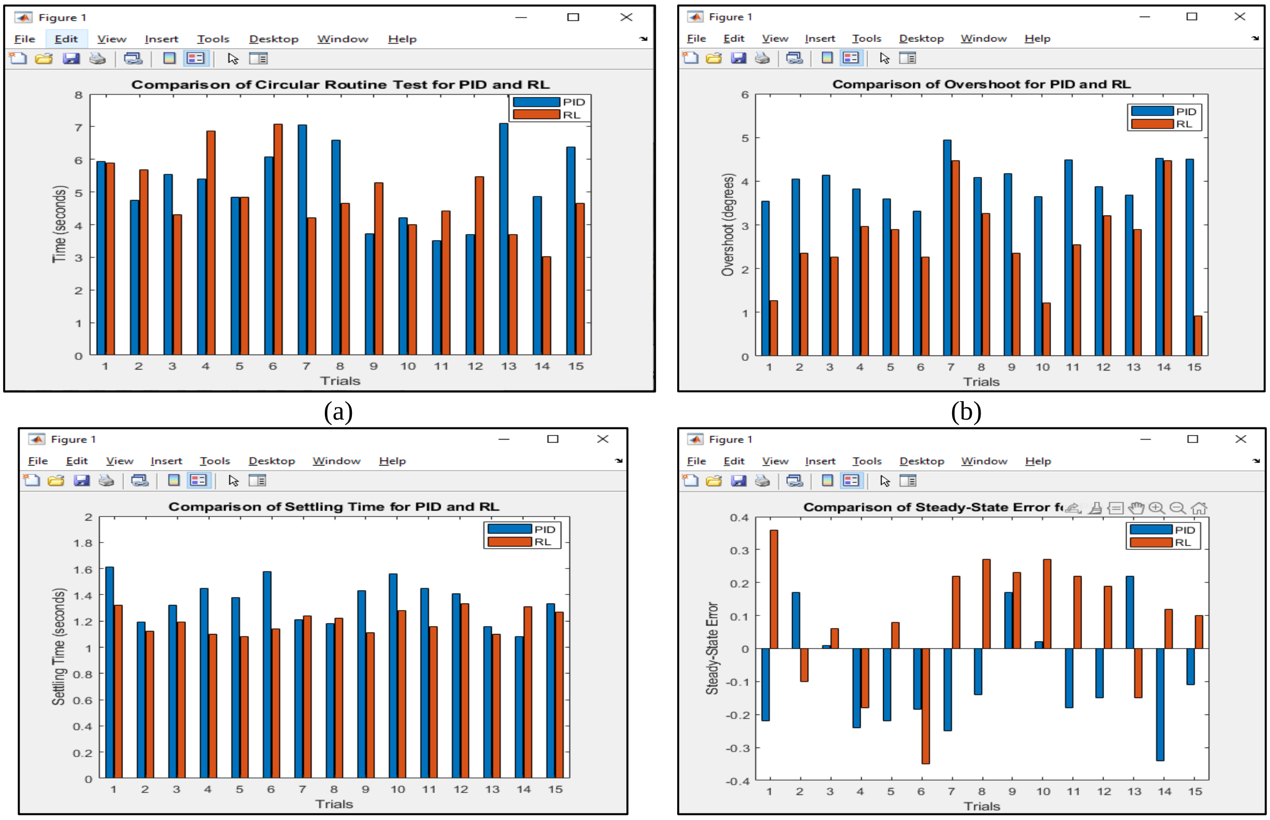
Figure 10. Step Response. (a) PID Control (b) RL Control

Table 2: Data collection from the step response of the DIP system in one (1) sample.

Controller Type	Rise Time (s)	Settling Time (s)	Overshoot (degrees)	Steady-state Error
PID	0.78	3.55	1.61	-0.22
RL	0.51	1.16	1.32	0.36

Figure 10 shows the Step Response of the PID-based method, while Figure 10b illustrates the Step Response of the RL-based method. Analyzing this response provides valuable insights into the effectiveness of the employed control method.

Table 2 summarizes data from analyzing the step response of the Double Inverted Pendulum (DIP) system. Comparing the PID and RL controllers, the RL controller outperforms PID with the faster rise and settling times (0.51s and 1.16s, respectively), lower overshoot (1.32 degrees), but a slightly higher steady-state error (0.36 compared to PID's -0.22).



(a) (b) (c) (d)

Fig. 11. Comparison of each Key Parameters for 15 trials. (a) Rise Time. (b) Overshoot. (c) Settling Time. (d) Steady-State Error.

The comparative analysis of 15 trials between PID and RL control methods for the Double Inverted Pendulum (DIP) system is shown in Figure 11, revealing significant differences in performance metrics. RL consistently demonstrated faster rise times (0.762s average) and lower overshoot (2.944 degrees average) compared to PID (0.886s average rise time and 4.25 degrees average overshoot). Additionally, RL exhibited more consistent settling times and slightly higher steady-state error values compared to PID. These findings emphasize the potential superiority of RL-based control strategies in managing the DIP system, offering faster responses and better control over oscillations compared to traditional PID approaches.

Table 3: Data collection from the step response of the DIP system across 15 trials

Controller Type	Avg Rise Time (s)	Avg Settling Time (s)	Avg Overshoot (degrees)	Avg Steady-state Error
PID	0.886	1.397	4.25	0.2053
RL	0.762	1.323	2.944	0.2253

Table 3 presents a summary of the performance metrics for the Double Inverted Pendulum (DIP) system across 15 trials, comparing the average values of key parameters between the PID and RL controllers. The metrics include Average Rise Time, Average Settling Time, Average Overshoot (in degrees), and Average Steady-state Error. The data indicates that the RL-based system demonstrates superior performance compared to the PID controller across all evaluated parameters.

4.2. Circular Routine Test

A circular routine test assessed the performance of a double-inverted pendulum on a two-wheeled cart platform. The test involved driving the cart in a circular path with a 0.350-meter radius while keeping the pendulums upright for 15 trials. The time taken for a 360-degree rotation was recorded to evaluate the system's control capabilities and stability.

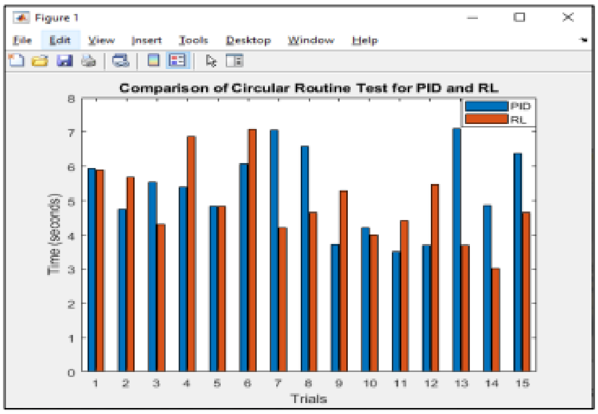


Fig. 12. Comparison of PID and RL Circular Routine Test.

The circular routine test revealed variability in completion times for both PID and RL control methods. PID control ranged from 3.51s to 7.10s, while RL control ranged from 3.01s to 7.88s across trials. Figure 12 illustrates that, on average, RL tended to complete faster than PID. These results underscore the adaptability of both methods in regulating dynamic systems, with RL showing a slight edge in average completion time.

5. CONCLUSION

This chapter provides a concise and comprehensive overview of the key findings, insights, and implications derived from the research on the double-inverted pendulum system using PID control and reinforcement learning (RL) techniques. It includes a summary of the research findings and results, a discussion of implications and contributions, an acknowledgment of limitations, suggestions for future research, and a final reflection on the research.

5.1 Evaluation of Control Methodologies: Stand-up Routine and Circular Routine Tests

Various performance metrics were assessed after the comprehensive analysis conducted in the Stand-up Routine Test, which evaluated the performance of both PID and RL control methodologies in stabilizing the dynamic behavior of the Double Inverted Pendulum (DIP) system. The results highlighted the effectiveness of RL control in achieving faster responses, lower overshoot values, and more consistent settling times compared to PID control. With RL control showcasing advantages in terms of agility and stability, the subsequent Circular Routine Test aimed to further explore the capabilities of both control methods in a dynamic environment. This test evaluated the completion times of the DIP system across 15 trials as it navigated a circular path, providing insights into the adaptability and responsiveness of each control methodology. Despite variability observed in completion times for both PID and RL control, RL control tended to exhibit faster average completion times, underscoring its potential advantages in dynamic scenarios. These findings emphasize the dynamic nature of control methodologies and the need for further experimentation to optimize their performance for practical applications.

5.1.1 Stand-up Routine Test

Various performance metrics were evaluated in the Stand-up Routine Test, where both PID and RL control methodologies were critically assessed for their performance in stabilizing the dynamic behavior of the Double Inverted Pendulum (DIP) system. The step response analysis provided, as shown in Table 3, insights into their behaviors, revealing that RL control consistently exhibited faster responses compared to PID control. Specifically, the rise time, indicating the time taken for the response to reach from 10% to 90% of its final value, was noticeably shorter for RL control across various trials, averaging approximately 0.762s compared to approximately 0.886s for PID control.

Additionally, RL control demonstrated lower overshoot values, indicating better control over oscillations and deviations from the target pitch or reference signal, with an average overshoot of approximately 2.944° compared to approximately 4.25° for PID control. While both methods exhibited similarities in settling time, RL control showcased more consistent response times and stability, as indicated by narrower ranges of overshoot and settling time values.

Furthermore, the analysis of steady-state error revealed that RL control exhibited a narrower range of steady-state error values and generally maintained steadier responses closer to zero compared to PID control. Although PID control showed a slightly lower average steady-state error of approximately 0.2053 compared to approximately 0.2253 for RL control, RL offered greater consistency in achieving desired steady-state conditions.

5.1.2 Circular Routine Test

In the Circular Routine Test, both PID and RL control methodologies were evaluated based on their completion times across 15 trials. This test aimed to assess the performance of each control method in stabilizing the double-inverted pendulum (DIP) system while navigating a circular path.

Upon analyzing the completion times, it becomes evident that both PID and RL control exhibited variability in their performance across trials. The completion times for PID control ranged from 3.51 seconds to 7.10 seconds,

showcasing a considerable spread in the time taken for the DIP system to complete a 360-degree rotation around the center point. Similarly, RL control also demonstrated variability in completion times, ranging from 3.01 seconds to 7.88 seconds across trials.
Despite the variability observed, it is noteworthy that RL control tended to exhibit faster completion times on average compared to PID control. While certain trials yielded faster completion times for PID control, others showcased faster times for RL control. This observation suggests that RL-based control strategies offer advantages in terms of agility and responsiveness in dynamic environments, allowing the DIP system to navigate the circular path more efficiently.

The variability in completion times across trials for both control methodologies could be attributed to several factors, including the specific configuration of the control algorithms, environmental conditions, and inherent uncertainties in the system dynamics. Additionally, variations in the initial conditions and perturbations during the experiment may have influenced the performance of the control methods.
Overall, the results of the Circular Routine Test highlight the dynamic nature of control methodologies and their adaptability to varying conditions. While both PID and RL control demonstrated capabilities in stabilizing the DIP system during circular motion, RL control showed a slight advantage in terms of average completion times. However, further analysis and experimentation may be necessary to elucidate the underlying factors contributing to the observed differences and to optimize the performance of both control methodologies for practical applications.

6. REFERENCES

[1] Banerjee, R., Dey, N., Mondal, U., & Hazra, B. (2018). Stabilization of double link inverted pendulum using LQR. 2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT). https://doi.org/10.1109/icctct.2018.8550915

[2] Razzaghi, K., & Jalali, A. A. (2011). A new approach on stabilization control of an inverted pendulum, using PID controller. Advanced Materials Research, 403-408, 4674–4680. https://doi.org/10.4028/www.scientific.net/amr.403-408.4674

[3] Liu, B., Hong, J., & Wang, L. (2019). Linear inverted pendulum control based on improved ADRC. Systems Science & Control Engineering, 7(3), 1–12. https://doi.org/10.1080/21642583.2019.1625081

[4] Yi, J., & Yubazaki, N. (2000). Stabilization fuzzy control of inverted pendulum systems. Artificial Intelligence in Engineering, 14(2), 153–163. https://doi.org/10.1016/s0954-1810(00)00007-8

[5] Raj, S. (2016). Reinforcement learning based controller for stabilization of double inverted pendulum. 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES). https://doi.org/10.1109/icpeices.2016.7853147

[6] Brown, D., & Strube, M. (2020). Design of a neural controller using reinforcement learning to control a rotational inverted pendulum. 2020 21st International Conference on Research and Education in Mechatronics (REM). https://doi.org/10.1109/rem49740.2020.9313887

[7] Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2021). Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. Machine Learning, 110(9), 2419–2468. https://doi.org/10.1007/s10994-021-05961-4

[8] N. Mellatshahi, S. Mozaffari, M. Saif and S. Alirezaee, "Inverted Pendulum Control with a Robotic Arm using Deep Reinforcement Learning," 2021 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 2021, pp. 1-6, doi: 10.1109/ISSCS52333.2021.9497411.

[9] S. Ryoo and S. Han, "Unveiling the Potential and Limitations of Visual Reinforcement Learning: An Experimental Analysis on the Cart-Pole Inverted Pendulum System," 2023 23rd International Conference on Control, Automation and Systems (ICCAS), Yeosu, Korea, Republic of, 2023, pp. 1348-1353, doi: 10.23919/ICCAS59377.2023.10317047.

[10] A. Zeynivand and H. Moodi, "Swing-up Control of a Double Inverted Pendulum by Combination of Q-Learning and PID Algorithms," 2022 8th International Conference on Control, Instrumentation and Automation (ICCIA), Tehran, Iran, Islamic Republic of, 2022, pp. 1-5, doi: https://ieeexplore.ieee.org/document/9737201#:~:text=10.1109/ICCIA54998.2022.9737201

[11] A. Khalil, "Transforming Foundation Industries: Challenges and Effective Resource Efficient Strategies," Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES), 2023,vol 9 pp 22-23, doi: https://doi.org/10.5109/7157930

[12]Zhou, Jingyu, "The RFID Electronic Tag Defect Detection System Based On YOLOv5 For Industrial Scenarios" Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES),2022, vol 8, pp 44-51, doi: https://doi.org/10.5109/5909061

[13] Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). "Introduction to the Practice of Statistics" (7th ed.). W.H. Freeman and Company.ISBN: 978-1-4292-7601-1