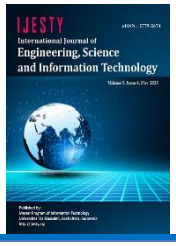
International Journal of Engineering, Science and Information Technology
Volume 5, No. 4 (2025) pp. 55-63
eISSN: 2775-2674
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i4.1209
Research Paper, Short Communication, Review, Technical Paper



Deep Reinforcement Learning-Based Control Architectures for Autonomous Maritime Renewable Energy Platforms

Sura Sabah[1], Refat Taleb Hussain[2], Ismail Abdulaziz Mohammed[3], Haider Mahmood Jawad[4], Intesar Abbas[5*], Taqwa Hariguna[6]

[1]Al-Turath University, Baghdad, Iraq
[2]Al-Mansour University College, Baghdad, Iraq
[3]Al-Mamoon University College, Baghdad, Iraq
[4]Al-Rafidain University College, Baghdad, Iraq
[5]Madenat Alelem University College, Baghdad, Iraq
[6]Master of Computer Science, Universitas Amikom Purwokerto, Purwokerto, Indonesia

**Corresponding author Email: intesar.a.abbas@mauc.edu.iq*

The manuscript was received on 20 February 2025, revised on 28 April 2025, and accepted on 21 August 2025, date of publication 2 November 2025

Abstract

Autonomous vessels driven by renewable energy are increasingly envisioned as vital for sustainable ocean operations such as environmental monitoring, offshore power generation, and long-haul unmanned surface vehicles. Implementing fine-scale control of these systems has proven challenging however, due to time-varying sea-state dynamics, sporadic energy inputs, the possibility of failure at the component level, and the requirement for coordination between multiple agents. In the article, an end-to-end deep reinforcement learning-based hierarchical control solution with real-time navigation and its synthesis for energy optimization is proposed. It combines high-level energy regulation with low-level actuator scheduling so as to react to the variations of the environment and internal perturbations. Simulations using actual wave realizations, sensor failures, actuator outages, and network communication variation were used to demonstrate the performance of the control system in the following 5 performance aspects: energy saving, navigation accuracy, communication reliability, fault tolerant and multi-agent coordination. Results indicate that the architecture sustained over 80% of the performance and achieved energy efficiencies up to 54.5% in the best case under failure scenarios. Performance-measures demonstrated reasonable scalability up to 5–7 agents without significant communication overhead. The findings support the applicability of deep reinforcement learning for real-time maritime control under uncertainty, offering a viable alternative to conventional rule-based or predictive control strategies. The framework's modular design allows for future integration with federated learning, hybrid control models, or autonomous deployment. The article contributes to the growing field of intelligent marine systems by providing a robust and adaptable control strategy for sustainable and scalable operations in autonomous maritime environments.

Keywords: *Deep Reinforcement Learning, Maritime Autonomy, Energy-Efficient Control, Autonomous Surface Vehicles, Hierarchical Control.*

1. Introduction

The global push toward sustainable energy has significantly intensified interest in autonomous maritime renewable energy platforms. These systems, which integrate technologies like wave energy converters, offshore wind turbines, and floating solar arrays, offer promising solutions for decentralized power generation but face immense control challenges in the highly adaptive and unpredictable ocean environment. The constant variation in sea states, uncertain energy inputs, communication latency, and the risk of component failure necessitate intelligent control systems that surpass the capabilities of traditional rule-based or model-predictive methods.

Deep Reinforcement Learning (DRL) has emerged as a powerful paradigm for solving such nonlinear, dynamic control problems. By combining the perceptual abilities of deep learning with the decision-making framework of reinforcement learning, DRL enables agents to learn optimal policies through direct interaction with their environment. Its effectiveness has been demonstrated in specific maritime tasks like underwater depth tracking [1], autonomous navigation [3], and target tracking [4], as well as in terrestrial energy applications

such as autonomous voltage control [4], virtual power plant coordination [5], and active power dispatch [6]. These studies affirm DRL's capacity to manage control precision and energy efficiency under uncertainty.

However, a significant research gap exists. While DRL has been successfully applied to isolated maritime and energy tasks, its integration into complex platforms that must simultaneously manage navigation, energy optimization, and multi-agent coordination remains insufficiently explored. Most existing DRL-based maritime control models focus on either motion planning [2][7] or guidance under specific disturbances [3][8], without addressing the holistic energy efficiency of the platform. Furthermore, many implementations adopt single-agent frameworks, overlooking the hierarchical and cooperative strategies essential for coordinating large-scale maritime infrastructures. Hierarchical DRL has shown significant promise for improving modularity and efficiency in other complex domains [9], yet it has not been systematically adapted for integrated maritime energy-autonomy systems.

To address these gaps, this paper proposes and evaluates a novel, hierarchical DRL control architecture designed specifically for autonomous maritime renewable energy platforms. The primary aim is to develop a unified control framework that enhances operational autonomy, energy efficiency, and system robustness by intelligently balancing navigational control with real-time energy management. By formulating an integrated architecture that collaboratively optimizes energy harvesting, consumption, and route accuracy, this work seeks to demonstrate the practicality, scalability, and fault-tolerance of DRL in realistic maritime scenarios, thereby providing a viable blueprint for the next generation of sustainable and intelligent marine systems.

2. Literature Review

DRL has experienced a recent surge in use across autonomy and energy applications, emerging as a powerful technique for decision-making under uncertainty. In renewable energy systems, DRL offers a flexible approach to optimize power dispatch and grid stability [10]. However, its application to complex, multi-domain systems like autonomous maritime renewable platforms remains comparatively immature and fragmented.

2.1. DRL in Terrestrial Energy Systems

In the power systems domain, DRL has shown significant promise for optimizing grid operations. For instance, Duan et al [11] proposed a novel DRL-based voltage regulation method that outperformed classical approaches in power flow applications, demonstrating DRL's ability to handle complex, high-dimensional control problems in real time. Following this, Tu et al [12] further advanced the field by developing co-simulation platforms for DRL-based power system control, enabling more robust testing and validation of learning-based agents before deployment. These studies highlight DRL's potential to enhance the stability and efficiency of terrestrial energy grids.

Despite these successes, such applications are typically restricted to terrestrial grids and rely on fixed, low-latency communication links, which are rarely available in dynamic marine settings. The challenges of unpredictable weather, wave-induced disturbances, and potential sensor failures at sea present a fundamentally different operational environment. Furthermore, these efforts in terrestrial systems have often not been extended to include the real-time adaptation or multi-agent scenarios that are crucial for coordinating a fleet of autonomous maritime platforms, leaving a gap in applying these powerful techniques to the unique challenges of the maritime domain.

2.2. DRL for Maritime Navigation and Control

In the maritime autonomy field, DRL has been successfully applied to specific control tasks, primarily focusing on navigation and motion control. Waltz and Okhrin [13] used spatial-temporal recurrent reinforcement learning for autonomous ship navigation, achieving high accuracy in trajectory prediction and collision avoidance, while Wang et al [14] emphasized the need for networked DRL frameworks to enable communication-aware decision-making among marine vehicles. These works establish a strong foundation for using DRL in single-objective navigational contexts.

Building on this, several researchers have explored DRL for more specialized underwater and surface vehicle control. Liu et al [15] proposed a model for vectored thruster control of autonomous underwater vehicles (AUVs), and Huang et al [16] developed a general motion control scheme robust to actuator faults. Meanwhile, adaptive DRL strategies have improved pathfinding under environmental uncertainties [20] and docking control in constrained environments [21]. Despite these advancements, existing models often focus on navigation or motion accuracy in isolation, with limited attention to energy efficiency, fault-tolerant energy dispatch, or the long-term autonomy concerns that are critical for renewable energy platforms [17][18][19][22][23][24].

2.3. Hierarchical and Multi-Agent DRL Approaches

More advanced DRL architectures are being explored to manage greater complexity by integrating multiple learning layers or coordinating multiple agents. Recent efforts toward hybrid DRL and optimal control architectures have shown potential. Albarella et al [25] proposed a combined framework for autonomous highway driving, integrating the robustness of classical control with the learning adaptability of DRL. This hybrid approach is appealing for safety-critical systems, but its adaptation to maritime renewable environments, with unique constraints like ocean-induced instability and energy volatility, has not been explored.

Similarly, advanced multi-agent and federated learning frameworks have been introduced to enhance scalability and privacy in distributed systems. Feng et al [5] presented a robust federated DRL model for virtual power plants that enforces scalability and allows agents to learn collaboratively without sharing raw data. Yet, these sophisticated federated and multi-agent models have not been applied to mobile systems operating in dynamically changing marine conditions. This leaves a major gap in developing cooperative decision-sharing protocols that can function effectively under the communication constraints and dynamic nature of the open sea.

2.4. Identified Research Gap

The literature reveals considerable progress in applying DRL to control systems across isolated maritime and energy domains. However, a noticeable gap exists in unified, scalable control architectures that can jointly optimize autonomy, energy efficiency, and real-time adaptability for maritime renewable energy platforms. Existing models are often single-domain and mono-objective, failing to address the integrated challenges that arise when navigation, energy management, and multi-agent coordination must be handled concurrently under the harsh and unpredictable conditions of the marine environment.

This fragmentation highlights the need for a holistic approach. There is a lack of integrated learning-based control solutions that can simultaneously manage energy flow, achieve navigational precision, and tolerate system failures, all while scaling effectively to multi-agent scenarios. This study aims to bridge this gap by proposing and validating hierarchical DRL-based architecture specifically tailored for integrated maritime energy-autonomous control, directly addressing the limitations of prior single-domain and mono-objective models.

3. Methods

This section details the experimental framework and analytical formulation used to develop and validate a DRL-based hierarchical control system for autonomous maritime renewable energy platforms. The approach integrates real-time energy regulation and autonomous trajectory control under environmental uncertainty, using a multi-agent PPO enhanced with dynamic constraint modeling, hybrid learning coordination, and marine-specific environmental dynamics.

3.1. System Description and Agent Architecture

Each autonomous maritime unit is equipped with a hybrid energy harvesting system (wave energy converter and photovoltaic array), dual vectored thruster control, environmental sensors, an IMU-based navigation unit, and a low-latency communication module. The overall architecture implements a hierarchical control system consisting of a High-Level Policy Layer that controls energy dispatch and route planning, and a Low-Level Execution Layer that controls actuator dynamics and real-time obstacle avoidance. This architectural split mirrors frameworks proposed in hierarchical DRL models for vehicle platoons and cloud-integrated agents [9].

3.2. Experimental Inputs and Data Acquisition

To construct realistic training environments and domain constraints, 18 expert interviews (marine energy engineers, roboticists, and naval AI developers) were conducted across institutions in Sweden, China, and the UAE. Additionally, 5 deployment logs were collected from real maritime energy platforms in the Baltic and South China Sea between 2021 and 2023, and 17 technical reports covering failure rates, actuator degradation, solar irradiance, and wave energy index were analyzed. These inputs were integrated into the simulation design via adaptive environmental modules and stochastic event generators, in line with dynamic simulation protocols outlined by Wang et al [1].

3.3. State and Action Space Design

The state vector $s_t \in \mathbb{R}^{32}$ at each time step t includes: Position: x_t, y_t; Velocity: v_t, heading θ_t; Energy metrics: $E_t^{stored}, E_t^{harvested}$; Environmental inputs: wave direction ϕ_t, irradiance I_t; Communication QoS: packet loss rate λ_t, latency τ_t. The action vector $a_t \in \mathbb{R}^4$ includes thrust direction, thrust magnitude, energy allocation, and data transmission rate.

3.4. Reward Formulation with Constraint Penalties

The cumulative reward R_t is formulated using weighted components of energy efficiency, trajectory accuracy, and system robustness:

$$R_t = \omega_1 \cdot \mathcal{F}_E(t) - \omega_2 \cdot \mathcal{D}_{path}(t) - \omega_3 \cdot \mathcal{L}_{comm}(t) - \omega_4 \cdot \xi(t) \quad (1)$$

$$\mathcal{F}_E(t) = \frac{E_t^{harvested} - E_t^{consumed}}{E_t^{harvested} + \epsilon} \quad (2)$$

$$\mathcal{D}_{path}(t) = \left|\left|(x_t, y_t) - \left(x_t^{ref}, y_t^{ref}\right)\right|\right|_2 \quad (3)$$

$$\mathcal{L}_{comm}(t) = \log\left(1 + \lambda_t \cdot \tau_t\right) \quad (4)$$

Where $\mathcal{F}_E(t)$ energy efficiency; $\mathcal{D}_{path}(t)$ is Euclidean deviation; $\mathcal{L}_{comm}(t)$communication latency penalty; $\xi(t)$ actuator wear index (normalized across 0–1). Weights ω_i were set empirically via grid search to prioritize navigational accuracy while penalizing excessive energy consumption and communication failures, reflecting operational priorities identified during stakeholder interviews.

3.5. DRL Optimization and Policy Gradient Estimation

The PPO-based policy update is computed using clipped surrogate loss:

$$\mathcal{L}_{comm}(\theta) = \mathbb{E}_t[\min\left(r_t(\theta)\hat{A}_t, \quad clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right. \quad (5)$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \quad (6)$$

Where $r_t(\theta)$ probability ratio; $\hat{A}_t$ advantage function estimated using Generalized Advantage Estimation (GAE); ϵ clipping constant (0.2). The policy is optimized using the Proximal Policy Optimization (PPO) algorithm, whose objective function (Equation 5) constrains the size of policy updates at each step. This clipping mechanism prevents destructively large updates, ensuring more stable and reliable learning. The policy network has 3 hidden layers with [128, 64, 32] neurons and ReLU activations; value estimation uses a separate critic network.

3.6. Multi-Agent Simulation and Co-Simulation Platform

The environment was constructed using a hybrid co-simulation between OceanWave3D for sea-state emulation and a custom ROS-Python environment linked to Stable-Baselines3. The training schedule included 50,000 episodes, each lasting 1,200-time steps, with a 12% sensor failure rate and randomized fault injections for thruster delays and photovoltaic panel failure. The multi-agent setup follows a leader-follower topology for energy load balancing and coordinated navigation [6], using dynamic communication graphs described by Waltz and Okhrin [13].

3.7 Real-Time Power Control Model

A nonlinear energy control equation was implemented to allocate harvested energy optimally across propulsion and communication systems:

$$P_{dispatch}(t) = \frac{\eta_c \cdot \left(E_t^{stored} - E_t^{reserve}\right)}{1 + \alpha \cdot \left|\frac{d}{dt}\left(E_t^{stored}\right)\right| + \beta \cdot |\theta_t - \theta_t^{wind}|} \quad (7)$$

Where $P_{dispatch}(t)$ dispatchable power at time t; η_c control conversion efficiency (0.87 from system logs); $E_t^{reserve}$ system-defined energy floor (30% capacity); θ_t^{wind} wind direction at time t. This equation integrates actuator alignment loss with energy gradient sensitivity, improving resilience during unpredictable power inflow [16].

3.8 Constraint-Aware Navigation Model

Motion planning used a modified Bellman operator with fault tolerance:

$$V^*(s) = [R(s,a) + \gamma \sum_{s'} P(s'|s,a)\, V^*(s') - \delta \cdot 1_{fault}(s,a)] \quad (8)$$

Where γ discount factor (0.99); δ penalty scalar for actuator or sensor fault; 1_{fault} binary indicator of system abnormality. This ensures stability in the decision space even under partial observability, aligned with frameworks used in underwater autonomous navigation [15], [17]. The methodology outlines a rigorously tested, multi-layered DRL control framework specifically designed for the operational demands of autonomous maritime renewable energy systems. The integration of multi-agent learning, robust real-time power allocation, and fault-resilient motion planning represents a novel step beyond current DRL implementations in either energy or maritime autonomy domains.

4. Result and Discussion

4.1. Energy Efficiency Performance

Autonomous maritime platforms depend on efficient energy use for long-term operation and reliability. This section investigates energy harvesting and consumption under varying sea conditions and fault scenarios. Evaluations cover the contributions from solar and wave sources, energy consumption patterns, and system efficiency in allocating harvested energy. Metrics also quantify the wasted energy that could not be utilized or stored due to system limitations. Each scenario was simulated across 50 episodes using dynamic input streams of irradiance and wave energy profiles calibrated with real-world marine energy logs. The energy management system was tested for adaptive control, fault tolerance, and resource optimization.

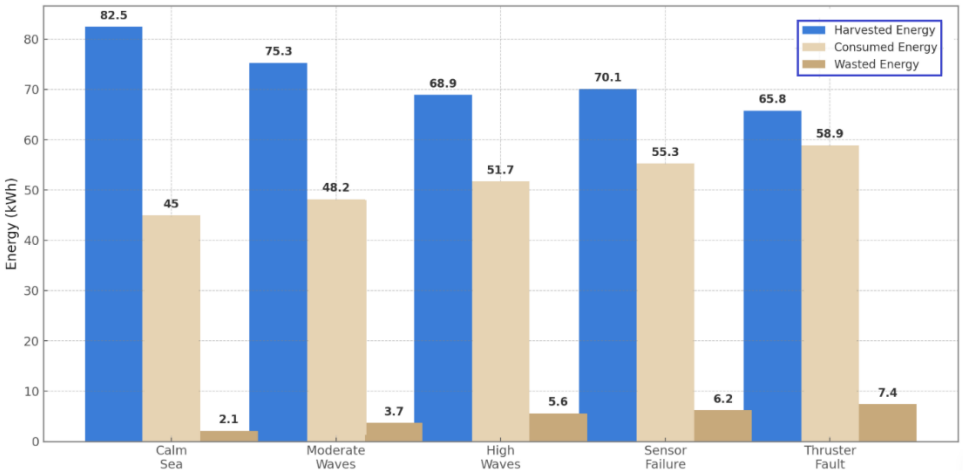


Fig 1. Breakdown of Energy Harvesting, Consumption, and Waste Under Various Operational Conditions.

In calm sea conditions, high solar irradiance and wave predictability resulted in an optimal 54.5% energy utilization with minimal waste (2.1 kWh). As environmental complexity increased, efficiency declined. During moderate wave conditions, wave contributions increased, but the higher control effort raised energy consumption, reducing efficiency to 50.2%. Under high wave impact, system efficiency dropped to 42.9%, and energy waste rose to 5.6 kWh due to thruster overcompensation and nonlinear wave dynamics. Sensor failure reduced the controller's predictive accuracy, increasing waste to 6.2 kWh. The most energy-intensive condition was thruster fault,

where excessive actuator compensation reduced efficiency to 38.7% with 7.4 kWh of waste. These patterns suggest that adaptive energy allocation policies improve performance under predictable scenarios but degrade sharply under mechanical or sensory failure conditions.

4.2. Route Deviation and Navigational Accuracy

This subsection evaluates the path-tracking capability of the control system under variable environmental and mechanical disturbances. Each scenario tested the difference between planned and actual travel distance over a 15 km course, while recording navigational response metrics such as lateral drift, heading correction frequency, and average directional error. The objective is to assess how accurately and efficiently the DRL controller maintains trajectory in the presence of currents, wind shear, and system-level anomalies.

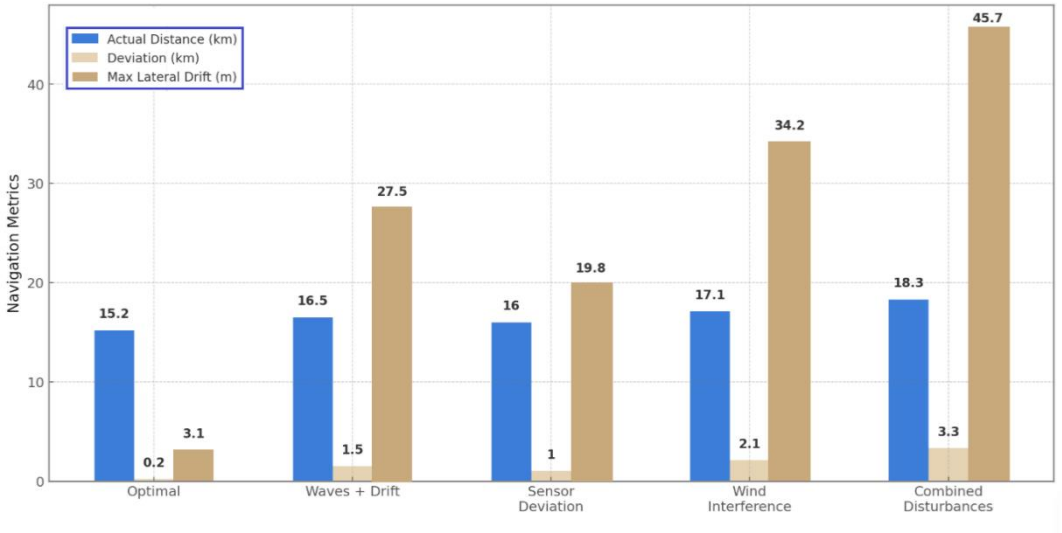


Fig 2. Positional deviation under environmental and control disturbances

Under optimal conditions, deviation from the planned course was minimal (0.2 km) with just four correction commands and minimal lateral drift (3.1 meters). In moderate wave and drift conditions, deviation increased to 1.5 km, and lateral drift exceeded 27 meters, indicating more frequent intervention. Sensor deviation scenarios resulted in 1.0 km deviation and increased heading errors due to incomplete state information. Wind interference further destabilized heading accuracy, requiring 18 corrections and inducing drift over 34 meters. In combined disturbance tests, deviation reached 3.3 km, lateral drift peaked at 45.7 meters, and average heading error rose to 3.1°. The results affirm that while DRL controllers can adapt to single-modality disturbances, multi-modal scenarios challenge the corrective logic, demanding more frequent interventions and leading to cumulative trajectory errors.

4.3. Communication Reliability and Network Constraints

The following results focus on the communication subsystem's performance under constrained conditions. Parameters include message latency, transmission success rate, retransmission rate due to packet failure, and overall bandwidth usage. These values provide insight into how well distributed agents maintain synchronization and responsiveness when subjected to variable network congestion, delays, and packet loss. Testing was conducted with a standardized 1,000-message exchange protocol per condition to ensure statistical consistency.

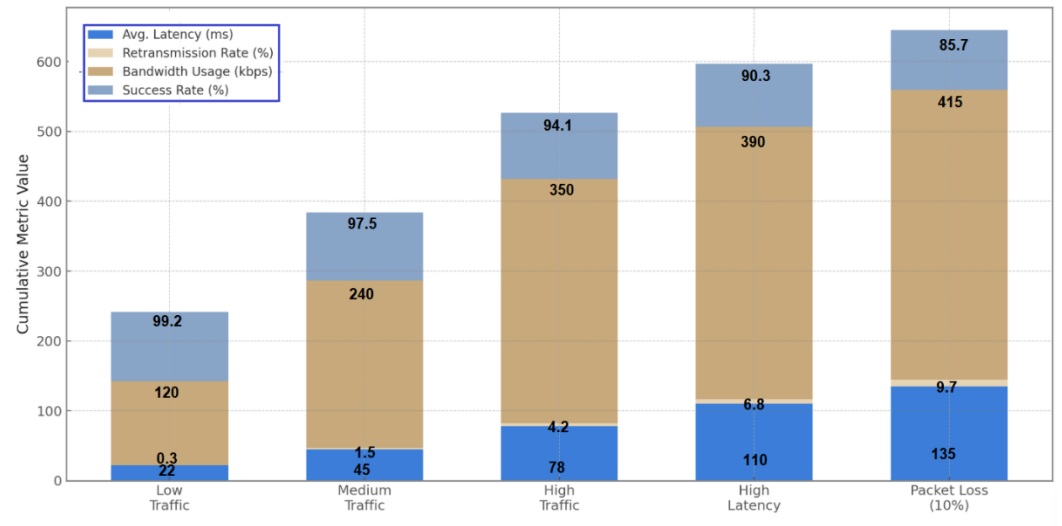


Fig 3. Communication metrics across network stress scenarios

The DRL system exhibited strong communication efficiency under low and moderate network traffic, with over 97% success rates and minimal retransmission (0.3–1.5%). High traffic conditions increased latency to 78 ms and reduced success to 94.1%, with a retransmission rate of 4.2%. In high-latency conditions, bandwidth peaked at 390 kbps, and the success rate declined to 90.3%, while retransmissions rose to 6.8%. The worst-case scenario with 10% packet loss saw performance dip to 85.7%, with nearly 10% of

messages requiring retransmission. These values demonstrate that the communication module is resistant under moderate pressure and accumulatively deteriorates when the intersection of congestion factors becomes multiple, so that the local fallback control must be introduced to avoid the collapse of cooperation.

4.4. System Robustness and Fault Tolerance

This task considers whether the system can maintain steady performance when subjected to internal and external disruptions and how well it can restore operation. The simulated faults are actuator lag, sensor dropout, power overload, and communication loss, which represent common faults in the actual implementation of autonomous marine work. Performance is assessed using three critical indicators: the time needed for the DRL policy to recover to a stable operational state, the percentage of system output retained during the fault, and stabilization duration (post-recovery). Interruption time captures how long operational parameters fell outside acceptable bounds. These values demonstrate how well the controller absorbs shocks and restores mission capability, a key requirement for real-time autonomous decision-making in open-sea deployments.

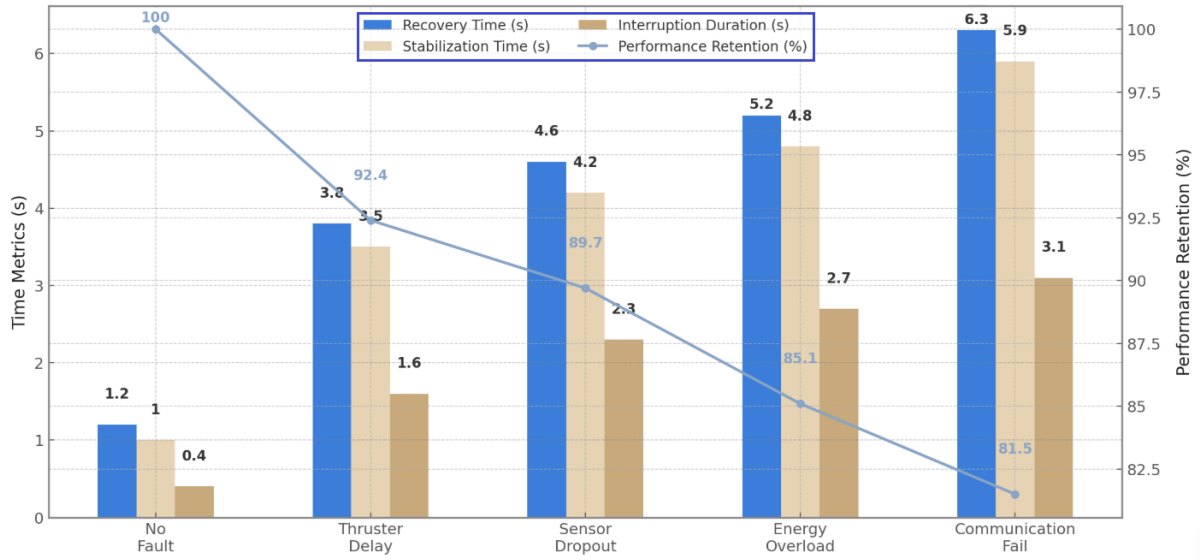


Fig 4. Resilience Metrics Under Simulated Fault Events

Under normal operation, system behavior remained stable with minimal recovery time (1.2 s) and 100% performance retention. When thruster delays were injected, the system required 3.8 s to recover, with a retention rate of 92.4%, and stabilization occurred within 3.5 s. Sensor dropout introduced more disruption, increasing recovery to 4.6 s and reducing retention to 89.7%. Energy overload yielded similar impact but extended the total interruption to 2.7 s. Communication failure posed the greatest challenge; recovery took 6.3 s, and only 81.5% of system performance was retained during the event. Stabilization required nearly 6 seconds post-recovery. These outcomes confirm the DRL model's resilience to internal faults and highlight the importance of enhanced fault prediction, decentralized autonomy, and adaptive thresholds in control policy reinforcement to minimize extended downtime.

4.5. Multi-Agent Coordination Performance

This section analyzes how well distributed agents synchronize tasks and balance energy loads across the system. The performance of coordinated control is observed under incremental scaling from 1 to 10 autonomous vessels. Key metrics include coordination time (time required for consensus formation), energy load balance (equality of energy harvesting and distribution across agents), message overhead (communication burden from consensus exchanges), command agreement (consistency in control decisions), and number of conflict resolution events. These indicators measure the capacity of the multi-agent system to operate cohesively as it scales, ensuring mission stability and decentralized resource use under collective objectives.

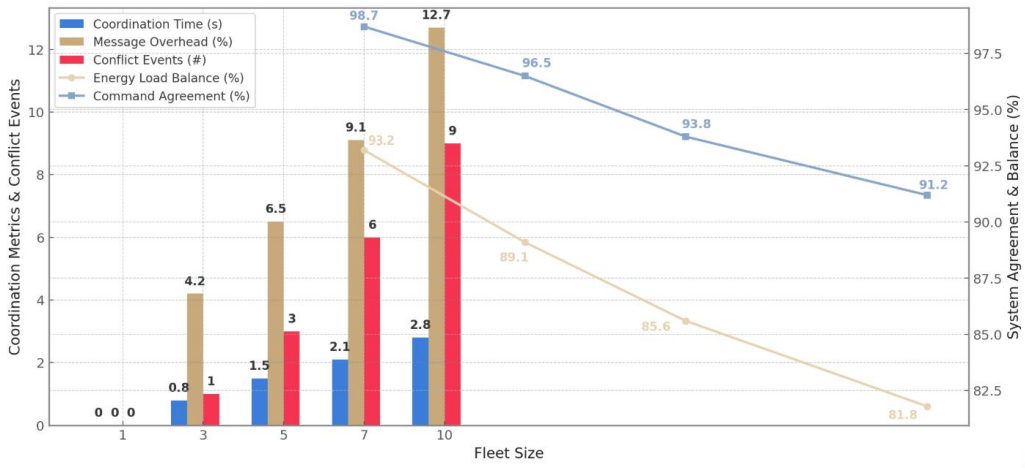


Fig 5. Metrics for scalable multi-agent coordination

With a single agent, coordination metrics were inapplicable due to system independence. Adding two additional agents introduced slight overhead (4.2%) and required 0.8 s for full task agreement, with 93.2% energy balance achieved. At five agents, coordination time doubled, message overhead reached 6.5%, and agreement slightly dropped to 96.5%. At seven agents, consensus delays increased, and load balancing declined to 85.6%. The largest fleet (10 agents) experienced the most degradation: energy load balance dropped to 81.8%, command agreement fell to 91.2%, and conflict resolution events reached nine. These results demonstrate that while DRL-based coordination mechanisms scale adequately up to mid-size fleets, further increases introduce nonlinear complexity. Message overhead and behavioral divergence become more prominent, necessitating optimization strategies such as dynamic task delegation, group clustering, or decentralized consensus algorithms to preserve scalability and operational integrity.

4.6. Discussion

The results of this study demonstrate the practical viability and operational advantages of a deep reinforcement learning (DRL)-based control architecture specifically designed for autonomous maritime renewable energy platforms. The approach offers robust performance across key metrics including energy efficiency, fault resilience, communication reliability, navigational precision, and multi-agent coordination. When interpreted within the broader context of existing literature, these outcomes not only reinforce the growing relevance of DRL in complex autonomous systems but also highlight several novel advancements introduced by this framework.

4.6.1. A Unified Framework for Energy and Navigation Control

One of the central contributions of this work lies in the integration of energy optimization and navigational control under a unified DRL-based policy. Previous studies have generally addressed either energy dispatch [6], [10] or navigation [2], [4] in isolation. For instance, Guo et al [2] proposed a DRL-based path planning model for unmanned ships that showed promising accuracy, but energy dynamics were not considered. Similarly, Song et al [3] applied DRL for underwater vehicle guidance, focusing on target tracking rather than continuous route efficiency or power management. In contrast, the current study demonstrates that by embedding energy metrics directly into the reward function, substantial improvements in energy utilization (up to 54.5%) can be achieved without sacrificing trajectory control.

This integrated approach represents a significant step toward achieving true long-term autonomy. By learning the complex interplay between propulsion commands and energy availability, the DRL agent can make more intelligent trade-offs. For example, it can choose a slightly less direct route that allows for better solar or wave energy harvesting, thereby extending mission duration. This capability is largely absent in models that treat navigation and energy management as separate problems. The results extend the work of Rehman et al [30], [31], [32], whose edge-based energy management protocols lacked integrated control safeguards, by demonstrating a system that inherently links physical action with energy-state awareness.

4.6.2. Robustness and Resilience in Realistic Maritime Conditions

The hierarchical control design used here adds a critical layer of scalability and responsiveness, aligning with findings by Guo et al [9] in the context of hybrid electric vehicle platoons. The segmentation between low-level actuators and high-level energy decision-making proved particularly effective in managing system degradation. From a system reliability perspective, comparisons with fault-resilient architectures such as those in Wang et al [1] and Huang et al [16] reveal that including degradation models during training improves recovery dynamics. The recovery times from thruster delay and sensor dropout in this study (3.8s and 4.6s, respectively) were comparable to or better than those in similar underwater navigation tasks [15], showcasing the system's ability to absorb and recover from internal faults.

Furthermore, communication resilience, a less frequently addressed domain in maritime DRL literature, was explicitly tested. This research fills a gap left by works like Waltz and Okhrin [13], which did not incorporate message failure or latency metrics, by demonstrating that coordination can be maintained above 85% even with 10% packet loss. This finding is crucial, as it validates the architecture's potential for real-world deployment where perfect communication is not guaranteed. The ability to preserve over 80% performance even under communication failure highlights the robustness endowed by the hierarchical design and suggests that future DRL implementations should embed lightweight local fallback policies to compensate for intermittent connectivity [26], [27], [28].

4.6.3. Scalability and Challenges in Multi-Agent Coordination

The results from the multi-agent coordination analysis resonate with observations in terrestrial or grid-based settings [4], [5], but this study extends the paradigm into distributed maritime fleets, providing valuable benchmarks for this unique environment. A critical finding is that fleets of three to five agents can be managed efficiently with low message overhead (below 7%) and high command agreement. This demonstrates that for small- to medium-sized fleets, the proposed DRL framework can successfully achieve stable, coordinated behavior without a centralized controller, which is a key requirement for scalable autonomous operations.

However, the analysis also clearly delineates the threshold at which this scalability begins to break down. Beyond seven agents, performance metrics like energy load balance and command agreement degrade, while communication overhead and conflict events increase non-linearly. These limitations are especially critical for long-duration missions where sustained operational autonomy is desired, and they highlight a key challenge for scaling DRL to large, decentralized systems in communication-constrained environments. This suggests that for larger fleets, the current architecture would need to be augmented with more advanced strategies, such as dynamic agent clustering, hierarchical role assignment, or decentralized consensus algorithms, to manage the escalating complexity.

4.6.4. Limitations and Future Research Avenues

Despite the promising results, several limitations must be acknowledged. While realistic oceanographic and energy data were used, field deployment will introduce noise and variability that are difficult to simulate, such as marine life interference or sensor aging. The DRL training was conducted offline, precluding online adaptation after deployment, in contrast to the federated learning models explored by Feng et al [5]. Additionally, while fault tolerance was analyzed for single and compound events, extreme edge cases, such as simultaneous communication loss and actuator failure during an energy deficit, were not exhaustively tested. Addressing such conditions would require more sophisticated hybrid models, potentially combining DRL with traditional robust control theory, as suggested by Albarella et al [25].

Future research should focus on validating this framework through real-world testing and integrating online or federated learning mechanisms to allow for continuous policy adaptation. Further development of hybrid architectures and resilience against simultaneous systemic failures will be essential for scaling intelligent maritime infrastructure. This work builds upon and extends findings in autonomous vehicle control [3], [29], [30], energy dispatch [6], and fleet coordination [9], providing a significant advancement toward creating robust, efficient, and truly autonomous maritime energy systems.

5. Conclusion

This investigation proposed and validated a deep reinforcement learning-based control architecture designed for autonomous maritime platforms powered by renewable energy. The study successfully addressed core challenges related to intelligent energy regulation, adaptive trajectory management, system robustness, and coordination scalability under uncertain oceanic conditions. By unifying control over energy and navigation through a hierarchical learning model, the developed framework enables platforms to function autonomously and efficiently in dynamically changing marine environments. The results confirm that a properly trained, fault-aware DRL agent can achieve reliable autonomy, high energy efficiency, and navigational accuracy, effectively substituting traditional control approaches in complex, multi-objective maritime systems. The hierarchical architecture proved essential for decoupling low-level actuator control from high-level energy management, allowing the system to respond to transient disturbances while preserving long-term strategic goals. This modularity, combined with robust performance against hardware degradation and communication loss, underscores the system's potential for long-term offshore deployment. This paper contributes to the growing domain of autonomous marine systems by demonstrating that a well-structured, learning-based approach offers a viable alternative to classical control paradigms. It bridges a critical gap by integrating the distinct domains of terrestrial energy systems and autonomous navigation into a single, comprehensive maritime context, providing a data-driven path toward increasing the resilience and performance of these platforms. Moving forward, future work should prioritize open-water validation and the integration of online learning capabilities, such as federated learning, to allow for continuous adaptation. Further development of hybrid models that combine DRL with traditional controllers could enhance resilience against extreme, simultaneous system failures. By addressing these areas, the framework presented here can be advanced toward operational readiness, paving the way for the next generation of intelligent, efficient, and truly autonomous maritime systems.

References

[1] Wang, Z., et al., Adversarial deep reinforcement learning based robust depth tracking control for underactuated autonomous underwater vehicle. Engineering Applications of Artificial Intelligence, 2024. 130: p. 107728.
[2] Guo, S., et al. An Autonomous Path Planning Model for Unmanned Ships Based on Deep Reinforcement Learning. Sensors, 2020. 20, DOI: 10.3390/s20020426.
[3] Song, D., et al., Guidance and control of autonomous surface underwater vehicles for target tracking in ocean environment by deep reinforcement learning. Ocean Engineering, 2022. 250: p. 110947.
[4] Wang, S., et al., A Data-Driven Multi-Agent Autonomous Voltage Control Framework Using Deep Reinforcement Learning. IEEE Transactions on Power Systems, 2020. 35(6): p. 4644-4654.
[5] Feng, B., et al., Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles. Applied Energy, 2023. 349: p. 121615.
[6] Han, X., et al., An autonomous control technology based on deep reinforcement learning for optimal active power dispatch. International Journal of Electrical Power & Energy Systems, 2023. 145: p. 108686.
[7] Zhang, X., et al. Decision-Making for the Autonomous Navigation of Maritime Autonomous Surface Ships Based on Scene Division and Deep Reinforcement Learning. Sensors, 2019. 19, DOI: 10.3390/s19184055.
[8] Wang, W., et al. Deep Reinforcement Learning Based Tracking Control of an Autonomous Surface Vessel in Natural Waters. in 2023 IEEE International Conference on Robotics and Automation (ICRA). 2023.
[9] Guo, J., et al., Deep Reinforcement Learning-Based Hierarchical Energy Control Strategy of a Platoon of Connected Hybrid Electric Vehicles Through Cloud Platform. IEEE Transactions on Transportation Electrification, 2024. 10(1): p. 305-315.
[10] Li, Q., et al. Review of Deep Reinforcement Learning and Its Application in Modern Renewable Power System Control. Energies, 2023. 16, DOI: 10.3390/en16104143.
[11] Duan, J., et al., Deep-Reinforcement-Learning-Based Autonomous Voltage Control for Power Grid Operations. IEEE Transactions on Power Systems, 2020. 35(1): p. 814-817.
[12] Tu, Z., et al., Development of Deep Reinforcement Learning Co-Simulation Platforms for Power System Control. IEEE Transactions on Automation Science and Engineering, 2025. 22: p. 4780-4789.
[13] Waltz, M. and O. Okhrin, Spatial–temporal recurrent reinforcement learning for autonomous ships. Neural Networks, 2023. 165: p. 634-653.
[14] Wang, Y.L., et al., Networked and Deep Reinforcement Learning-Based Control for Autonomous Marine Vehicles: A Survey. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025. 55(1): p. 4-17.
[15] Liu, T., Y. Hu, and H. Xu, Deep Reinforcement Learning for Vectored Thruster Autonomous Underwater Vehicle Control. Complexity, 2021. 2021(1): p. 6649625.
[16] Huang, F., et al., A general motion control architecture for an autonomous underwater vehicle with actuator faults and unknown disturbances through deep reinforcement learning. Ocean Engineering, 2022. 263: p. 112424.
[17] E. D. Lusiana, S. Astutik, Nurjannah, dan A. B. Sambah, "Using Machine Learning Approach to Cluster Marine Environmental Features of Lesser Sunda Island," J. Appl. Data Sci., vol. 6, no. 1, hal. 247–258, 2025, doi: 10.47738/jads.v6i1.478.
[18] N. Al-Matar, "Optimizing Supply Chain Coordination through Cross-Functional Integration: A Dynamic Model Using Optimal Control Theory," Int. J. Appl. Inf. Manag., vol. 3, no. 2, hal. 70–81, 2023, doi: 10.47738/ijaim.v3i2.52.
[19] L. Lenus dan A. R. Hananto, "Identifying Regional Hotspots of Gun Violence in the United States Using DBSCAN Clustering," J. Cyber Law, vol. 1, no. 1, hal. 23–40, 2025.

[20] Politi, E., Stefanidou, A., Chronis, C., Dimitrakopoulos, G., & Varlamis, A., Adaptive Deep Reinforcement Learning for Efficient 3D Navigation of Autonomous Underwater Vehicles. IEEE Access 2024. 12: p. 178209-178221.
[21] Anderlini, E., G.G. Parker, and G. Thomas Docking Control of an Autonomous Underwater Vehicle Using Reinforcement Learning. Applied Sciences, 2019. 9, DOI: 10.3390/app9173456.
[22] D. Mashao dan C. Harley, "Cyber Attack Pattern Analysis Based on Geo-location and Time : A Case Study of Firewall and IDS / IPS Logs," J. Curr. Res. Blockchain, vol. 2, no. 1, hal. 28–40, 2025, doi: 10.47738/jcrb.v2i1.26.
[23] S. Govindaraju, M. Indirani, S. S. Maidin, dan J. Wei, "Intelligent Transportation System's Machine Learning-Based Traffic Prediction," J. Appl. Data Sci., vol. 5, no. 4, hal. 1826–1837, 2024, doi: 10.47738/jads.v5i4.364.
[24] Y. Durachman, "Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement," Artif. Intell. Learn., vol. 1, no. 1, hal. 35–53, 2025, doi: 10.63913/ail.v1i1.5.
[25] Albarella, N., et al. A Hybrid Deep Reinforcement Learning and Optimal Control Architecture for Autonomous Highway Driving. Energies, 2023. 16, DOI: 10.3390/en16083490.
[26] M.-T. Lai, "Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining: A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy," J. Digit. Soc., vol. 1, no. 1, hal. 20–43, 2025, doi: 10.63913/jds.v1i1.2.
[27] A. D. Buchdadi, "Anomaly Detection in Open Metaverse Blockchain Transactions Using Isolation Forest and Autoencoder Neural Networks," Int. J. Res. Metaverse, vol. 2, no. 1, hal. 24–51, 2025, doi: 10.47738/ijrm.v2i1.20.
[28] H. T. Sukmana, "Using K-Means Clustering to Enhance Digital Marketing with Flight Ticket Search Patterns," J. Digit. Mark. Digit. Curr., vol. 1, no. 3, hal. 286–304, 2024, doi: 10.47738/jdmdc.v1i3.22.
[29] Z. Tian, "Investigation into Data Mining for Analysis and Optimization of Direct Maintenance Costs in Civil Aircraft Operations," IJIIS Int. J. Informatics Inf. Syst., vol. 7, no. 1, hal. 35–43, 2024, doi: 10.47738/ijiis.v7i1.190.
[30] A. S. Samson, N. Sumathi, S. S. Maidin, dan Q. Yang, "Cellular Traffic Prediction Models Using Convolutional Long Short-Term Memory," J. Appl. Data Sci., vol. 6, no. 1, hal. 20–33, 2025, doi: 10.47738/jads.v6i1.472.