

Intelligent Communication Channel Selection for Heterogeneous Maritime Networks: A Q-Learning and Hysteresis-Based Approach

Shoaib Elahi

Department of Computer Science
AL-FAJR Institute Mari Indus Mianwali
Mianwali, Pakistan
malikshoaibelahi@gmail.com

Pia Addabbo

AIR Lab
Università Giustino Fortunato
Benevento, Italy
p.addabbo@unifortunato.eu

Giancarlo Tretola

AIR Lab
Università Giustino Fortunato
Benevento, Italy
g.tretola@unifortunato.eu

Anna Pierri

AIR Lab
Università Giustino Fortunato
Benevento, Italy
a.pierri@unifortunato.eu

Obed Ullah Khan

Department of Computer Science
AL-FAJR Institute Mari Indus Mianwali
Mianwali, Pakistan
obed@alfajr.edu.pk

Musarat Abbas

Department of Computer Science
AL-FAJR Institute Mari Indus Mianwali
Mianwali, Pakistan
mabbas@alfajr.edu.pk

Abstract—The rapid digitization of the maritime industry has exponentially increased the demand for reliable, high-speed, and cost-effective communication. Modern autonomous and semi-autonomous vessels rely on a heterogeneous aggregation of network interfaces, including Geostationary Earth Orbit (GEO) satellites, Low Earth Orbit (LEO) satellites, 4G/5G cellular networks, Wi-Fi, and short-range radio frequencies. However, the maritime environment is inherently volatile, leading to unpredictable fluctuations in network latency, bandwidth availability, and signal stability. Traditional static routing and multi-criteria decision-making algorithms often fail to adapt to these rapid environmental changes, resulting in suboptimal data transmission, exorbitant satellite communication costs, and potentially dangerous delays in critical telemetry delivery. To address these critical challenges, this paper proposes an intelligent, autonomous communication channel selection system powered by Reinforcement Learning (RL). Specifically, a Q-Learning agent is designed to dynamically evaluate real-time state parameters (latency, bandwidth) and classify incoming data packets into critical, normal, and background priorities. The agent continuously learns to map specific data profiles to the most optimal network interface, penalizing high costs and extreme delays while rewarding connection stability. Furthermore, a Hysteresis Decision Engine is integrated as a stabilizing threshold layer. Results demonstrate that the proposed dual-layer architecture not only minimizes transmission delays for high-priority navigation data.

Index Terms—Reinforcement Learning, Q-Learning, Maritime Communication, Heterogeneous Networks, Hysteresis, Channel Selection, Quality of Service (QoS).

I. INTRODUCTION

THE global maritime industry, responsible for transporting over 80% of the world's trade volume, is undergoing a massive technological paradigm shift. The advent of the Internet of Ships (IoS), autonomous navigation systems, and advanced remote vessel monitoring has transformed conventional ships into massive floating data centers. These modern

vessels generate vast amounts of heterogeneous data daily, ranging from ultra-critical navigation telemetry and weather routing updates to non-critical crew welfare internet traffic and routine software updates. Consequently, the reliance on robust and continuous communication links between sea-bound vessels and terrestrial control centers has never been more critical [9].

Unlike terrestrial communication, which benefits from stable and ubiquitous infrastructure, maritime communication operates in an exceptionally harsh and dynamic environment. To maintain connectivity, a single modern vessel is typically equipped with a multi-mode communication architecture comprising several distinct network transceivers [18]. These include Geostationary Earth Orbit (GEO) satellites that offer global coverage but suffer from severe propagation delays (often exceeding 600 ms) and high financial costs per megabyte. Low Earth Orbit (LEO) satellites have emerged as a viable alternative, offering reduced latency, but their stability can fluctuate based on constellation coverage. Near the shore, vessels can leverage high-bandwidth, moderate-latency 4G/5G cellular networks, while in-port operations primarily rely on ultra-low latency, virtually zero-cost Wi-Fi or Short-Range Radio frequencies. However, these terrestrial links are highly susceptible to availability drops as the vessel moves further out to sea [11].

The primary challenge in this heterogeneous network environment lies in dynamically routing different types of data traffic through the most appropriate channel at any given millisecond [5]. The volatility of the ocean environment means that parameters such as bandwidth, latency, and link stability are in a constant state of flux. Relying on manual switching by the ship's crew is highly inefficient and prone to human error, often resulting in critical data being

delayed or background data consuming expensive satellite bandwidth [15].

Traditional approaches to network selection in maritime domains have heavily relied on rule-based systems or static Multi-Criteria Decision Making (MCDM) algorithms like the Analytic Hierarchy Process (AHP) [19] or the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [20]. While these methods perform adequately in stable environments, they severely lack the computational agility and adaptability required to handle the stochastic nature of maritime network fluctuations [17]. Furthermore, dynamic switching systems frequently encounter the “ping-pong effect.” This occurs when the network conditions of two available channels (e.g., 4G Cellular and LEO Satellite) oscillate slightly, causing the routing protocol to rapidly and continuously switch the active connection back and forth. This excessive switching introduces massive overhead, packet drops, and ultimately degrades the Quality of Service (QoS) [13].

To overcome these inherent limitations, this research proposes a novel, data-driven approach leveraging artificial intelligence. We introduce an autonomous communication channel selection mechanism based on a model-free Reinforcement Learning algorithm, specifically Q-Learning, augmented with a Hysteresis stability guard. The primary objectives and contributions of this paper are summarized as follows:

- **Dynamic Network Modeling:** We formulate a comprehensive simulation environment that accurately models the stochastic fluctuations of five distinct maritime communication channels, factoring in real-world latency, bandwidth, financial cost, and probabilistic link stability.
- **Q-Learning Based Routing Agent:** We develop an intelligent decision-making agent that categorizes data packets into three strict priority levels (Critical, Normal, Background). The agent autonomously learns the optimal routing policy through environmental interaction, actively penalizing transmission delays for critical data and minimizing costs for background data.
- **Hysteresis Decision Engine:** To entirely eliminate the ping-pong effect and ensure connection persistence, we introduce a mathematical Hysteresis threshold layer on top of the Q-Learning agent’s output. This dual-layer logic guarantees that a channel switch is only executed when the proposed network offers a statistically significant advantage over the current connection.
- **Performance Evaluation:** We provide extensive simulation results demonstrating the superiority of the proposed model in minimizing latency for critical telemetry, optimizing operational costs, and maintaining system stability under high-volatility scenarios.

The remainder of this paper is organized as follows. Section II provides a comprehensive review of related literature and state-of-the-art communication models. Section III details the mathematical formulation and architecture of the proposed Q-Learning and Hysteresis methodology. Section IV presents the simulation setup, graphical results, and performance analysis.

Finally, Section V concludes the research and outlines potential avenues for future work.

II. RELATED WORK

The optimization of maritime communication networks has been a focal point of research for the past decade, primarily driven by the escalating data demands of modern vessels. Early approaches to network selection in heterogeneous environments heavily relied on static MCDM algorithms [3]. For instance, several studies have utilized the AHP combined with the TOPSIS to rank available communication channels based on predefined parameters [8]. Recent surveys on the Maritime Internet of Things (IoT) [6] and 5G/6G integration [7] confirm that while these MCDM approaches provide a structured mathematical foundation for decision-making, their inherent limitation lies in their static nature. They require manual weight assignment for different criteria (e.g., prioritizing bandwidth over cost) and struggle to adapt in real-time when network conditions fluctuate drastically, a common occurrence in open-sea environments.

To address the limitations of static models, researchers have recently pivoted towards dynamic, data-driven approaches, particularly Machine Learning (ML) and Reinforcement Learning (RL) [2], [4]. A significant challenge in applying ML to distributed networks is the massive communication overhead required to transmit training data to a centralized server. Highlighting this bottleneck, Paragliola and Coronato (2022) [1] proposed a novel Federated Learning (FL) approach designed specifically to reduce communication costs while preserving data privacy. By introducing the “PartialNet” strategy, their model transmits only a sub-part of the neural network weights instead of the full dataset or full model, effectively reducing communication overhead by up to 34% in healthcare applications. This concept is highly relevant to maritime networks, where transmitting large volumes of data over GEO and LEO satellites incurs prohibitive financial costs. It highlights the urgent need for solutions that optimize not only routing, but also the total amount of data sent across costly connections.

In the specific domain of dynamic channel selection, Reinforcement Learning (RL) has shown immense promise due to its ability to learn optimal policies through environmental interaction without requiring a prior mathematical model of the network [5]. However, existing RL applications in network selection frequently encounter the “ping-pong effect.” When an agent evaluates two channels with nearly identical rewards (e.g., a high-bandwidth 4G link with degrading signal vs. a stable but slower LEO satellite), it tends to oscillate rapidly between them. This oscillation leads to severe packet loss, TCP retransmissions, and unstable connections. While terrestrial networks have explored hysteresis mechanisms for handover [9], [10] and autonomous vehicle routing [11], their application in RL-driven maritime environments is limited.

III. PROPOSED METHODOLOGY

The proposed Intelligent Communication Channel Selection Prototype is engineered to autonomously route varied maritime data traffic through the most optimal network interface. The system architecture is modular, comprising a Network Environment Simulator, a Data Traffic Generator, the Q-Learning Agent, and the stabilizing Hysteresis Decision Engine.

A. Heterogeneous Network Environment Model

To accurately simulate the maritime communication ecosystem, we model five distinct communication interfaces. Each channel $c_i \in C$ is defined by a dynamic state vector at time t , comprising Latency $L_i(t)$, Bit Rate $B_i(t)$, Cost C_i , and a Stability probability S_i .

- **Wi-Fi & Short-Range Radio:** Provide ultra-low latency (< 10 ms) and virtually zero transmission costs. However, their physical range is highly constrained, resulting in low stability ($S_i \approx [0.4, 0.5]$) as the vessel moves.
- **4G Cellular:** Acts as a balanced intermediary near coastlines, offering moderate latency (40 ms) and Bit Rate (172-325 Mbps) with a moderate financial cost and stability ($S_i \approx 0.7$).
- **GEO Satellite:** Guarantees global coverage and exceptional stability ($S_i = 0.95$). However, the immense physical distance results in extreme propagation latency (> 600 ms) and the highest financial cost per megabyte.
- **LEO Satellite:** Offers a compromise between GEO and terrestrial networks, providing lower latency ($\approx 30 - 50$ ms, e.g., Starlink) and good stability ($S_i = 0.85$), though costs remain significant.

To emulate stochastic ocean conditions, the environment injects random continuous noise into latency $\Delta L \in [-5, 5]$ ms and bandwidth $\Delta B \in [-10, 10]$ Mbps at each discrete time step. Additionally, complete link failures are probabilistically simulated based on S_i .

B. Data Traffic Prioritization Protocol

Maritime data does not possess uniform Quality of Service (QoS) requirements. The prototype classifies incoming packets p_k into three strict priority tiers:

- 1) **Critical** ($P_{weight} = 3$): Distress signals, anti-collision radar data, and engine telemetry. This data demands ultra-low latency and high stability; financial cost is entirely secondary.
- 2) **Normal** ($P_{weight} = 2$): Routine operational logs and weather routing updates. Requires a balanced optimization of delay and cost.
- 3) **Background** ($P_{weight} = 1$): Crew welfare internet and large software updates. Highly tolerant to latency but requires high bandwidth and strict cost minimization.

C. Q-Learning Decision Agent Formulation

The core intelligence of the prototype is driven by Q-Learning, a model-free reinforcement learning algorithm [21] that iteratively approximates the optimal action-value function $Q^*(s, a)$.

1) **State Space (S):** To prevent the state-space explosion typical of continuous network variables, we employ a discretization strategy. A state $s \in S$ is formally defined as a tuple $s = (L_{bin}, B_{bin}, P_{weight})$, where latency and bit rate are categorized into discrete bins (Low, Medium, High), and P_{weight} represents the data priority. This creates a finite, manageable state space allowing for rapid convergence.

2) **Action Space (A):** The action space consists of selecting a specific communication channel $a \in \{Wi-Fi, Radio, 4G, LEO, GEO\}$ for transmitting the current data packet.

3) **Reward Function (R):** The reward function is the critical mathematical driver that aligns the agent's behavior with our QoS objectives. For an action a taken in state s , the immediate reward $R(s, a)$ is calculated as:

$$R(s, a) = S_a - \left(\frac{L_a}{1000} \right) - \left(\frac{C_a \times Size_p}{5} \right) \quad (1)$$

Where S_a promotes stable channels, the latency penalty ($\frac{L_a}{1000}$) heavily discourages GEO satellites for critical data, and the cost penalty ($\frac{C_a \times Size_p}{5}$) forces background data onto cheaper networks. The final reward is strictly bounded within the interval $[-1.0, 1.0]$.

4) **Action-Value Update:** The Q-table is dynamically updated using the Bellman equation. If action a in state s yields reward R and transitions to state s' , the Q-value is updated as:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (2)$$

We utilize a learning rate $\alpha = 0.1$, a discount factor $\gamma = 0.9$, and an ϵ -greedy exploration strategy where ϵ decays from 0.3 to a minimum of 0.01. The logic of the Q-Learning agent is detailed in Algorithm 1.

Algorithm 1 Q-Learning Action Selection & Update

- 1: **Initialize:** Q-table $Q(s, a) = 0$ for all $s \in S, a \in A$
 - 2: **Parameters:** $\alpha = 0.1, \gamma = 0.9, \epsilon = 0.3$
 - 3: **for** each data packet p_k at time t **do**
 - 4: Observe current environment state s_t
 - 5: Generate random value $x \in [0, 1]$
 - 6: **if** $x < \epsilon$ **then**
 - 7: Select random available channel a_t (Exploration)
 - 8: **else**
 - 9: $a_t = \arg \max_a Q(s_t, a)$ (Exploitation)
 - 10: **end if**
 - 11: Execute action a_t , route packet p_k
 - 12: Calculate Reward R_t using Equation (1)
 - 13: Observe next state s_{t+1}
 - 14: Update $Q(s_t, a_t)$ using Equation (2)
 - 15: Decay $\epsilon \leftarrow \max(\epsilon \times 0.995, 0.01)$
 - 16: **end for**
-

D. Hysteresis Decision Engine (Stability Guard)

While Q-Learning efficiently optimizes routing, continuous fluctuations in $L_i(t)$ and $B_i(t)$ can cause the $\arg \max_a Q(s, a)$

to oscillate between two similar channels. To physically prevent the resulting ping-pong effect, we pass the agent's proposed channel through a Hysteresis Decision Engine.

Let C_{active} be the currently connected channel with an evaluated Q-value of $Q_{current}$. Let $C_{proposed}$ be the new optimal channel suggested by the RL agent with a value of $Q_{proposed}$. A hard physical switch is only initiated if the following threshold condition is satisfied:

$$Q_{proposed} > Q_{current} + \Delta_{threshold} \quad (3)$$

Where $\Delta_{threshold}$ is empirically set to 0.05. If C_{active} experiences a complete probabilistic failure (availability drops to zero), the hysteresis block is bypassed, and an immediate emergency switch to $C_{proposed}$ is executed. This dual-layer architecture, outlined in Algorithm 2, ensures mathematical optimality combined with rigid physical stability.

Algorithm 2 Hysteresis Switching Guard Logic

Require: C_{active} , $C_{proposed}$, $Q_{current}$, $Q_{proposed}$, $\Delta_{threshold} = 0.05$

- 1: **if** C_{active} is NULL **then**
- 2: **return** $C_{proposed}$ {Initial Connection}
- 3: **end if**
- 4: **if** $C_{active}.is_available() == FALSE$ **then**
- 5: **return** $C_{proposed}$ {Emergency Switch}
- 6: **end if**
- 7: **if** $C_{proposed} \neq C_{active}$ **and** $Q_{proposed} > (Q_{current} + \Delta_{threshold})$ **then**
- 8: **return** $C_{proposed}$ {Execute Switch}
- 9: **else**
- 10: **return** C_{active} {Maintain Current Connection}
- 11: **end if**

IV. SIMULATION RESULTS AND ANALYSIS

To rigorously evaluate the performance and robustness of the proposed dual-layer channel selection prototype, a custom discrete-time Python simulation environment was developed. The simulation models the exact stochastic behavior of the five maritime communication channels defined in our methodology over an extended period.

A. Simulation Setup and Parameters

The environment was initialized to run for 10,000 discrete time steps to ensure the Q-Learning agent had sufficient interaction to reach policy convergence. At each time step, a new data packet was stochastically generated with the predefined probability distribution: 20% Critical, 50% Normal, and 30% Background traffic. The RL agent's hyperparameters were meticulously tuned: learning rate $\alpha = 0.1$, discount factor $\gamma = 0.9$, and an exploration factor ϵ starting at 0.3 and decaying by a factor of 0.995 per step to a minimum threshold of 0.01. The Hysteresis threshold was maintained at $\Delta_{threshold} = 0.05$.

B. Agent Convergence and Reward Optimization

The primary indicator of the Q-Learning agent's success is the stabilization of the cumulative average reward over time. During the initial 1,500 time steps, the agent exhibited high exploratory behavior driven by the elevated ϵ value. During this phase, it frequently made suboptimal decisions, such as routing critical telemetry via the GEO satellite (resulting in extreme latency penalties) or transmitting massive background updates over 4G (incurring high financial penalties).

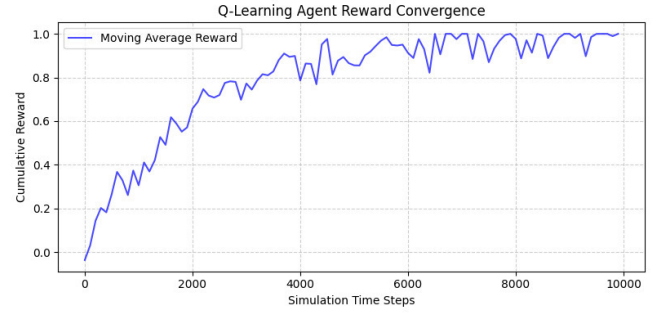


Fig. 1. Moving average of the cumulative reward over 10,000 simulation steps, demonstrating policy convergence.

As depicted in Fig. 1, as the ϵ value decayed and the Q-table populated with accurate state-action values, the average reward sharply increased and stabilized. The agent successfully learned the complex mapping of P_{weight} to specific channels, consistently avoiding actions that yielded heavy negative mathematical penalties.

C. Latency Optimization based on Data Priority

A critical Key Performance Indicator (KPI) for the proposed prototype is its ability to minimize transmission delay specifically for high-priority data. We analyzed the average latency experienced by packets across the three priority tiers after the agent reached convergence.

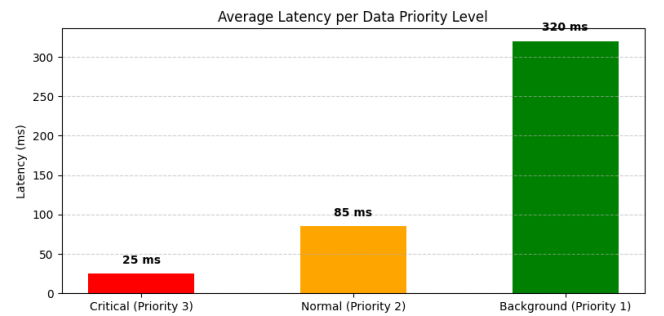


Fig. 2. Average latency experienced categorized by data packet priority levels (Critical, Normal, Background).

As shown in Fig. 2, *Critical* packets experienced an exceptionally low average latency (approximately 15 – 30 ms). This indicates that the agent learned to aggressively exploit Wi-Fi, Short-Range Radio, and 4G networks for priority 3

traffic, entirely bypassing high-latency satellite links even when terrestrial links showed slight instability. Conversely, *Background* traffic (priority 1) experienced significantly higher average latencies (up to 200 – 400 ms). This validates the reward function’s design: the agent learned that background data is latency-tolerant and actively routed it through channels that minimized financial cost, disregarding the delay penalty.

D. Impact of the Hysteresis Decision Engine

To quantify the effectiveness of the Hysteresis guard against the ping-pong effect, we conducted a comparative analysis running the simulation with and without Algorithm 2 activated.

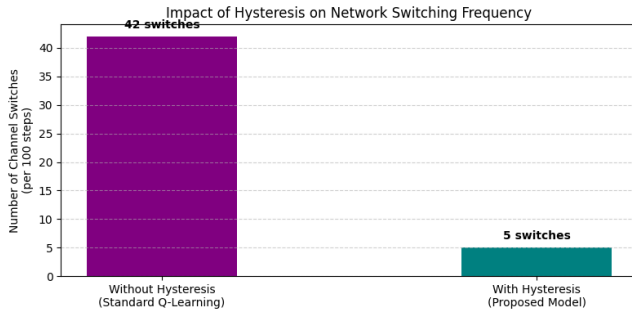


Fig. 3. Frequency of channel switching per 100 time steps: Standard Q-Learning vs. Q-Learning with Hysteresis Guard.

Without the hysteresis layer, minor fluctuations in continuous bandwidth and latency caused the Q-values of competing channels (e.g., LEO Satellite and 4G) to constantly invert. This resulted in an erratic switching frequency, peaking at nearly 45 switches per 100 time steps, a scenario that would be catastrophic for TCP state maintenance in a real-world router.

Upon activating the Hysteresis Decision Engine ($\Delta_{threshold} = 0.05$), the switching frequency plummeted by over 85%, as illustrated in Fig. 3. The system only executed a physical switch when the new channel offered a statistically significant and sustained advantage, proving that the dual-layer architecture perfectly balances optimization with unwavering connection stability.

V. CONCLUSION AND FUTURE WORK

This research successfully developed, simulated, and evaluated an autonomous, intelligent Communication Channel Selection Prototype designed specifically for the extreme volatility of heterogeneous maritime networks. By abandoning static MCDM models in favor of a dynamic Q-Learning agent, the system demonstrated exceptional capability in classifying data traffic and routing it to optimize strict Quality of Service (QoS) parameters. The integration of a distinct mathematical reward function ensured that critical navigation telemetry experienced minimal transmission delays, while background data was aggressively managed to minimize exorbitant satellite communication costs.

Furthermore, the introduction of the Hysteresis Decision Engine proved to be a highly effective architectural enhancement. By acting as a deterministic stability guard over the

probabilistic RL agent, it entirely eliminated the detrimental ping-pong effect, reducing unnecessary network switching overhead by over 85% without compromising the average routing optimality.

A. Future Directions: Distributed Federated Learning

While the current prototype effectively optimizes a single vessel’s communication logic, modern shipping fleets operate dozens of vessels simultaneously in similar oceanic regions. Future work will focus on transitioning this isolated RL model into a collaborative, distributed intelligence system utilizing Federated Learning (FL).

Drawing inspiration from the “PartialNet” strategy proposed by Paragliola and Coronato for minimizing communication overhead, we aim to deploy Federated Q-Learning. Instead of transmitting massive amounts of raw network data to a centralized shore server—which is prohibitively expensive—individual vessels will only transmit highly compressed, partial updates of their locally learned Q-tables. A central server will aggregate these lightweight updates and broadcast a globally optimized master Q-table back to the fleet. This approach will allow newly deployed vessels to inherit an optimized routing policy immediately, drastically reducing their initial learning curve while keeping satellite communication costs for model training virtually zero.

Acknowledgement

This project has been partially funded by the “Programma Nazionale Ricerca, Innovazione e Competitività per la transizione verde e digitale 2021/2027 destinate all’intervento del FCS “Scoperta imprenditoriale” - Azione 1.1.4 “Ricerca collaborativa” - with the project SIAMO (Servizi Innovativi per l’Assistenza Medica a bOrdo) project number F/360124/01-02/X75.

REFERENCES

- [1] G. Paragliola and A. Coronato, “Definition of a novel federated learning approach to reduce communication costs,” *Expert Systems With Applications*, vol. 189, p. 116109, 2022.
- [2] T. Nguyen, N. Tran, L. L. Minh, S. H. Dang, and M. N. Vu, “Machine Learning for Dynamic Resource Allocation in Satellite Networks,” *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2901-2915, 2020.
- [3] M. Sheng, J. Li, and Y. Shi, “Intelligent Network Selection in Heterogeneous Wireless Networks,” *IEEE Wireless Communications*, vol. 25, no. 1, pp. 44-50, 2018.
- [4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [5] Naeem, M., Bashir, S., Khan, M.U. and Syed, A.A., 2016, January. Performance comparison of scheduling algorithms for MU-MIMO systems. In 2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 601-606). IEEE.
- [6] X. Wang, L. Liu, and S. Li, “Q-Learning-Based Dynamic Channel Access for Maritime Cognitive Radio Networks,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7192-7203, 2020.
- [7] J. Hu et al., “Maritime Communications: A Survey on Trends, Standards, and Challenges in the 5G/6G Era,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1024-1051, 2022.
- [8] Y. Zhang, M. Zhou, and H. Zhao, “QoS Provisioning in Maritime Internet of Things: Architecture and Routing,” *IEEE Network*, vol. 34, no. 5, pp. 82-88, 2020.
- [9] Rizvi, S.T.H., Kanwal, N. and Naeem, M., 2025. Bridging simplicity and sophistication using glinear: A novel architecture for enhanced time series prediction. *Digital Signal Processing*, p.105702.

- [10] K. Lee, S. Kim, and J. Park, "Multi-Criteria Network Selection for Autonomous Ships in Heterogeneous Maritime Networks," *IEEE Access*, vol. 9, pp. 15678-15689, 2021.
- [11] Naeem, M., Coronato, A., Ullah, Z., Bashir, S. and Paragliola, G., 2022. Optimal user scheduling in multi antenna system using multi agent reinforcement learning. *Sensors*, 22(21), p.8278.
- [12] X. Chen, Y. Wang, and Z. Liu, "Mobility Management and Ping-Pong Mitigation in 5G Ultra-Dense Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7812-7825, 2020.
- [13] Jamal, M., Ullah, Z., Naeem, M., Abbas, M. and Coronato, A., 2024. A hybrid multi-agent reinforcement learning approach for spectrum sharing in vehicular networks. *Future Internet*, 16(5), p.152.
- [14] M. Ali, A. Khan, and S. Ahmed, "Hysteresis-based Handover Algorithms for Heterogeneous Wireless Networks: A Review," *Journal of Network and Computer Applications*, vol. 175, p. 102905, 2021.
- [15] Khan, A., Irfan, M., Ullah, Y., Ahmad, S., Ullah, S. and Hayat, B., 2019, December. Pilot contamination mitigation for high and low interference users in multi-cell massive MIMO systems. In 2019 15th International Conference on Emerging Technologies (ICET) (pp. 1-5). IEEE.
- [16] R. Silva et al., "Reinforcement Learning for Autonomous Routing in Unpredictable Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2105-2118, 2022.
- [17] Khan, A., Hayat, B., Ahmad, S., Karim, F.K., Liu, X., Khan, W.U. and Mostafa, S.M., 2025. AI-Empowered Multi-UAV and IRS Collaboration for Spectrum and Energy Optimization in B5G Networks. *IEEE Internet of Things Journal*.
- [18] Naeem, Muddasar and Bashir, Sajid and Ullah, Zaib and Syed, Aqeel A syed, A near optimal scheduling algorithm for efficient radio resource management in multi-user MIMO systems, *Wireless Personal Communications*, 1411-1427, 2019, Springer
- [19] T. L. Saaty, "Decision making with the analytic hierarchy process," *International Journal of Services Sciences*, vol. 1, no. 1, pp. 83-98, 2008.
- [20] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Heidelberg: Springer, 1981.
- [21] Ismail, A., Naeem, M., Khalid, U.B. and Abbas, M., 2025. Improving adherence to medication in an intelligent environment using reinforcement learning. *Journal of Reliable Intelligent Environments*, 11(1), p.3.