

# Deep Reinforcement Learning for Robust USVs Navigation in Diverse Environmental Scenarios

Rahim Ullah<sup>1</sup>, Wajid Ali<sup>1</sup>, Usman Ghanni<sup>1</sup>

<sup>1</sup>Department of Computer Science, Air University, Islamabad, Pakistan

## Article Info

### Article history:

Received July 16, 2025

Revised September 24, 2025

Accepted October 01, 2025

### Keywords:

Generative Adversarial  
Proximal Policy Optimization  
Reinforcement Learning  
Unmanned Surface Vehicles  
Environmental Factors

## ABSTRACT

Collision avoidance is essential for the safe operation of unmanned surface vehicles (USVs) in marine environments. While existing studies have addressed USV collaboration and navigation, they often overlook environmental challenges. In this research, we develop a novel deep reinforcement learning model and train USVs to navigate safely under different situations, such as rain, wind, and their combined factors that can disrupt control and increase collision risk. We apply Generative Adversarial Imitation Learning (GAIL) and Proximal Policy Optimization (PPO) to improve the agent's performance using expert demonstrations. The experimental results demonstrate that the proposed framework significantly outperforms the baseline across multiple metrics: the maximum episode length increased from 15 to 350 steps, the cultivated reward improved by 158%, the extrinsic reward increased by 175%, and the number of collisions decreased by 75-78% across all environmental conditions. Moreover, the policy loss stabilized after 1.1 million training steps, confirming efficient convergence. These results show that our model performs better than the baseline model in mean reward, episode length, value estimates, and collision reduction. In the end, some concluding remarks and future directions are added by the authors.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



**Corresponding Author:** Wajid Ali (e-mail: wajidali00258@gmail.com)

## 1. INTRODUCTION

In this era, Machine learning is an emerging research area in computer science and artificial intelligence. Machine learning uses models to enable computers to learn from data and make predictions or decisions based on data [1]. Supervised, unsupervised, and reinforcement learning are three primary categories of machine learning. Supervised learning works on labelled data, unsupervised learning analyses unlabeled data to find patterns, and reinforcement learning, where an agent learns through trial and error using feedback from its actions [2]. Reinforcement learning may apply in various domains, like robotics locomotion [3] and video games [4]. To process high-dimensional input data such as videos and images, the traditional reinforcement learning algorithms may struggle. However, deep reinforcement learning models deal with these types of complex input data [5] effectively. Deep reinforcement learning (DRL) is a subfield of machine learning that combines the use of deep learning algorithms with reinforcement learning techniques [6]. Figure 1 explains how a deep reinforcement learning model is designed and works. In the last few years, many scholars have explored DRL models and applied these approaches for the solution of difficult tasks, including playing games [4], robot locomotion [3], autonomous driving [7], and healthcare [8].

In the application of unmanned surface vehicles (USVs), DRL shows a vital role in making USVs intelligent, enabling them to avoid collisions, navigate through dangerous environments, and complete their missions safely [9]. USVs are small surface ships that autonomously move on the surface of marine systems and complete their tasks without any human intervention. Unmanned underwater vehicles (UUVs), unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), and USVs are significant components of autonomous systems [10], and their collaborative behaviours collectively create a remarkable and efficient marine system. Once equipped with advanced control sensors and communication devices, USVs will become

more flexible and intelligent, capable of completing various missions such as marine detection and water quality measurement.

Previous studies have primarily focused on developing techniques for collision and obstacle avoidance for USVs in maritime systems [11], overlooking environmental factors. Unmanned Surface Vehicles have been widely used in marine systems for various applications such as environmental monitoring and research, marine surveillance and security, search and rescue operations, military and defence, offshore and oil and gas industry, and recreational and commercial use. Due to their broad range of applications, USVs are crucial in current research. In [11], the authors proposed an approach for collision avoidance and navigation control in marine systems cooperatively. In [12], [13], the authors proposed a DRL-based technique for collision and obstacle avoidance. However, they did not consider environmental factors such as rain, wind, and storms, which can significantly affect the control and navigation of USVs in challenging and dynamic environments. Therefore, there is a need to develop an approach that considers these environmental factors to ensure that USVs can effectively control their navigation under such circumstances.

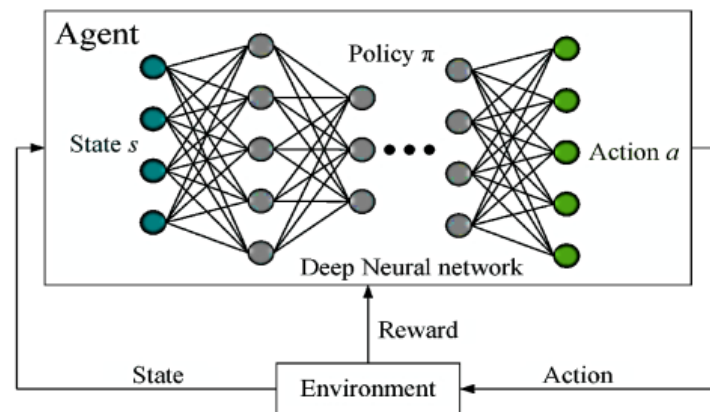


Figure 1. Deep Reinforcement Learning

### 1.1. Contribution of the proposed study

This paper adds several significant insights to the field of Unmanned Surface vehicle navigation and control. It considers the impact of wind velocity on the USV mission completion, highlighting the impact of this environmental factor on USV operations. Further, this study addresses the challenges posed by rain and storms, providing a detailed analysis of USVs' navigation. Notably, this study proposes a novel technique leveraging deep reinforcement learning to tackle the mentioned challenges during the USVs' navigation in the marine system. This model enhances the USV's reliability, robustness, and control during dangerous weather and ensures safe operation in marine systems.

### 1.2. Organization of work

The paper is divided into the following sections: Section 2 describes the detailed literature review, explaining the techniques and identifying research gaps. Section 3 explains the proposed model, architecture, and implementation. Section 4 describes the result and evaluation of the developed algorithm and compares it with the base model. Section 5 talks about the conclusion of the paper and future work.

## 2. LITERATURE REVIEW

In dynamic and complex marine systems, USVs must control and navigate themselves to the target without any human intervention. USVs must know about the surrounding obstacles and other moving boats with the help of various kinds of sensors. Given current environmental conditions, USVs ought to be able to effectively make the appropriate choices and complete their predetermined mission.

Path planning and collision avoidance have been popular research topics for a few years. On one hand, the conventional path planning algorithms [14] used environmental structure space for path exploration. The efficiency of path exploration is related to the division of the environmental structure space. Some researchers utilized swarm optimization techniques [15] and genetic algorithms to enhance the effectiveness of path planning. In [16][17], a heuristic path-planning approach was introduced in a continuous state of action space for an autonomous unmanned surface vehicle.

Recent advancements [18] in deep reinforcement learning have introduced novel approaches to addressing the challenge of dynamic obstacles, potentially offering ideal solutions for real-time control in dynamic environments. Q Learning [19], which utilizes a Q-table, and deep Q-network (DQN) [20], which employs a deep neural network, along with other deep reinforcement learning methods [21], show promise in

tackling these challenges. In [22], the authors utilized deep reinforcement learning techniques for multi-agent coordination.

Moreover, DRL made an important development in these years, leading to different achievements in wireless communications, data-driven, and Artificial Intelligence. The relevance of DRL in AI-enabled wireless networks is presented by the authors, with an emphasis on deep Multi-Agent Reinforcement Learning (MARL) [23]. The first portion of the paper [23] goes through the mathematical underpinnings of single-agent RL and MARL in detail. The main purpose of this study is to promote the usage of RL outside of the popular model-free approach of recent years. Consequently, the authors provide a quick review of RL algorithms, such as Model-Based RL (MBRL) and cooperative MARL, and their potential uses in future wireless networks.

The author in [24] offered a path-planning simulator for several automobiles. In a single-agent environment, the author used the proximal policy optimization technique of deep reinforcement learning. Currently, different AI methods have been studied for intelligent transport systems and autonomous vehicles. In [25], a DRL method is applied to attain UAV navigation using a massive multiple-input multiple-output (MIMO) algorithm. By implementing DQN, an optimal location selection policy can be obtained based on the received signal strengths. Unlike prior studies, which primarily show the speed or geographic position for UAV navigation, the proposed techniques converge fast. A decision-making strategy has been presented in [26] to identify decision-making issues in a dynamic robotic soccer game. To collect and distinguish environmental and situational data made up of the specified assessment factors, an Improved Support Vector Machine (ISVM) [27][28] is employed, and adaptive decision-making techniques with reinforcement learning select the approach in an adaptive manner. In this model, there are many agents, and each agent has its own role. The roles of the agent can be modified by the given approach [29][30].

### 3. METHOD

Explaining research in chronological order, including research design, research technique (as algorithms, pseudocode, or otherwise), how to test, and data gathering [4][5]. The summary of the research course should be accompanied by references, so that the explanation can be accepted scientifically [6]. Figure 1 shows an Overview of Prediction modelling.

Now, we present the methodology of the proposed system for training USVs to navigate effectively and complete their mission successfully in a challenging and active sea atmosphere. The designed model was developed by Generative Adversarial Imitation Learning (GAIL) and Proximal Policy Optimization (PPO) techniques to achieve robust control of the USVs. Imitation learning is applied for the collection of behavioural data via an expert agent or human intervention.

In our study, GAIL uses a Generative Adversarial Network (GAN) to train the agent. The agent acts, and a Convolutional Neural Network (CNN) captures observations, and then a GAN compares these observations with the observations of an expert, guiding the agent to mimic the behaviour of the specialist. The GAIL agent's neural network is designed to be relatively shallow to prevent vanishing gradients and improve training quality. ReLU is used as the activation function for the hidden layer, and the output layer has a single output unit, providing binary rewards based on the similarity of the agent's actions to the expert's actions. The details of the hyperparameters for GAIL implementation are shown in Table 1.

Table 1. Training Parameters for PPO Implementation

Parameter	Value
Batch Size	1024
Buffer Size	16384
Learning Rate	0.0003
Max Steps per Episode	5000
Number of Epochs	3
Number of Layers	3
Hidden Units	256
Gamma (Extrinsic Reward Signal)	0.99

Proximal Policy Optimization (PPO) is used as the trainer in this work. The PPO's purpose is to find the best policy to maximise the overall cumulative reward. PPO and Trust Region Policy Optimization (TRPO) algorithms are among the most successful in all deep reinforcement learning algorithms. PPO is chosen in this work due to its simplicity and ease of customisation. PPO employs a CNN with three layers and employs the Adam optimizer for the optimization process. The input layer dimensions correspond to the number of actions

and hidden units, two hidden layers each containing 512 units, and a multi-class output layer, i.e., the output layer caters to four possible actions: left, right, forward, and backward, with ReLU as the activation function. The hyperparameters for PPO implementation are detailed in Table 2.

Table 2. Training Parameters for GAN Implementation

Parameter	Value
Batch Size	1024
Buffer Size	16384
Learning Rate	0.0003
Max Steps per Episode	5000
Number of Epochs	3
Number of Layers	3
Hidden Units	256
Gamma (Extrinsic Reward Signal)	0.99
GAIL Strength	0.01
Behaviour Cloning Strength	0.5

The training process includes using PPO to handle environmental factors, with GAIL serving additional guidance through expert demonstrations. Expert behaviour is recorded in the simulated environment where a human controls the USVs under environmental factors using a keyboard. These recorded demonstrations are provided as a hyperparameter for GAIL. The training process is episodic, and each episode consists of 5000 time steps. The agent tries new policies and optimizes them based on the feedback. The agent collects new observations and acts accordingly, receiving rewards when the agent's actions move closer to the target. The GAN neural network in GAIL compares the agent's observations with the expert's demonstrations and provides feedback in the form of rewards for each action.

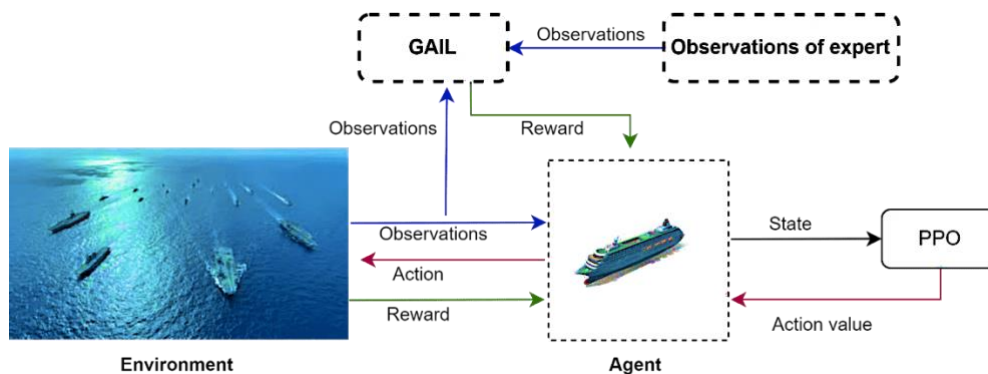


Figure 2. Working of the proposed framework

This proposed model is designed and simulated in a sea environment. The agent navigates under various environmental factors like rain, wind, and storms, utilizing reinforcement learning methods and GAIL rewards for training. The agent learns to control itself by using policies, and it gets positive rewards for successful navigation. If the agent collides with other boats or obstacles, it receives negative rewards. Figure 2 shows the proposed system model.

### 3.1. Learning Environment

The training environment of 200 x 200 meters is created in Unity 3D. The ML-Agents toolkit provides components such as Brain, Academy, and Agent, which are essential for developing the RL model. A wave height of 9-12 meters and a wind speed of 60-75km/h is employed. The atmosphere is made more dynamic by the addition of moving boats and stones as impediments. Vector observations are utilized to monitor the surroundings, and a ray perception sensor is used to sense the surroundings. A Unity component, Rigid Body, is used for the physics properties like weight and drag force. Figure 3 displays the learning environment of the developed system model.

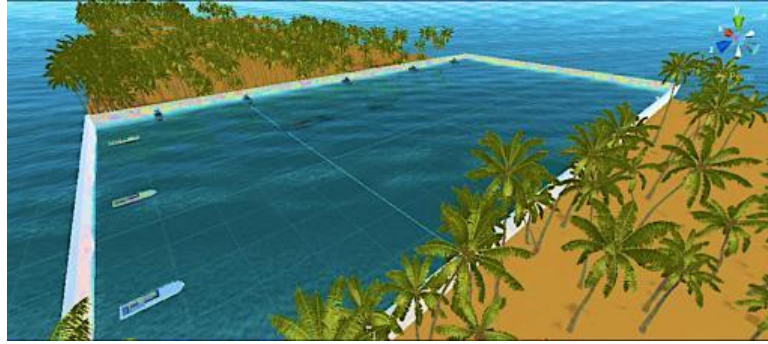


Figure 3. Learning Environment of the developed system

### 3.2. Agent

In deep reinforcement learning, an agent is anything that picks up certain behaviours from actions and observations. In the learning environment, Agents [28] create states, carry out actions, and move between states. Every agent has a brain connected to it that tells it what to do next. The USV agent can move forward, backward, left, and right, and is illustrated in Figure 4. The reward function encourages the agent to achieve the desired behaviour. It provides a positive reward for mission completion and negative rewards for collisions with other boats or obstacles. The total reward is the sum of the successful mission completion reward and the negative rewards for obstacle collisions and boat collisions, as defined in Equation 1. The specific reward values are shown in Table 3.

$$r = r_{destination} - (r_{obstacle} + r_{boat}) \quad (1)$$

Table 3. Agent Rewards

Event	Reward
The agent makes it to the destination	0.7
Getting hit by an obstacle	-0.3
Running into boats	-0.3

The proposed model operates through a series of episodes, each consisting of multiple time steps. The environment is randomized at the start of each episode, and the agent collects observations, selects actions, and receives rewards based on its performance. The process continues until the agent identifies the best policy for controlling the USV.

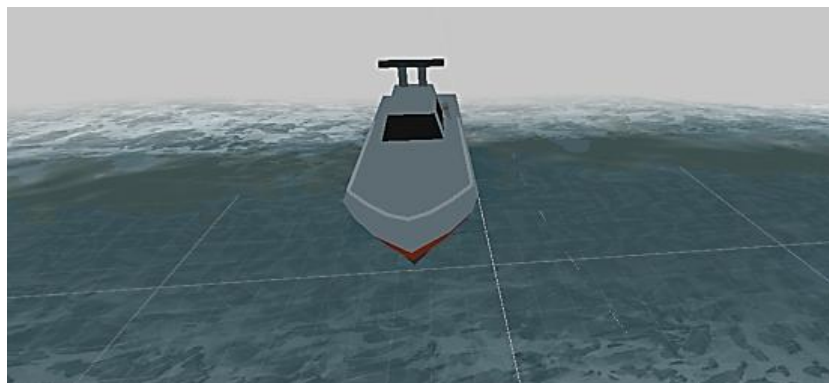


Figure 4. An agent has developed a Deep reinforcement learning approach

Algorithm 1 shows the pseudocode of the proposed system model. The developed model is relevant in various real-world scenarios where USVs are employed for missions such as naval operations, search and rescue, smuggling deterrence, and tugboat functions. These missions often involve challenging environmental conditions, making the proposed model essential for effective USV control.

**Algorithm 1: Pseudocode of the proposed model**

- 1: while episodes do not end, do
- 2:   while Time steps do not end do
- 3:     Collect observations
- 4:     Select action  $a_i$  from actions according to Policy  $\pi_\theta$
- 5:     Take action  $a_i$
- 6:     Compute Local reward
- 7:     Compute observations and state  $s_{t+1}$
- 8:     Compute GAIL reward using the discriminator
- 9:     Update Policy  $\pi_\theta$  based on GAIL reward
- 10:    Update Discriminator
- 11:   end while
- 12: end while

**4. RESULTS AND DISCUSSION**

This part of the paper describes the results of the developed model and the parameters that were assessed throughout the study. The value estimate is the first parameter that was assessed in this study. Value estimate means the prediction of the future state, either a good prediction or a bad prediction. Figure 5 shows that our proposed model better assesses the value of the state compared to the baseline model (ANOVA).

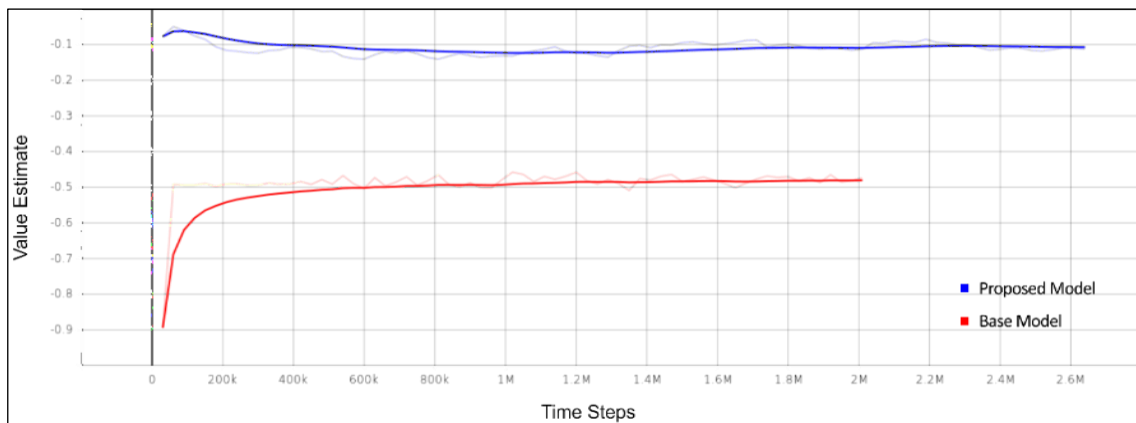


Figure 5. Value estimate

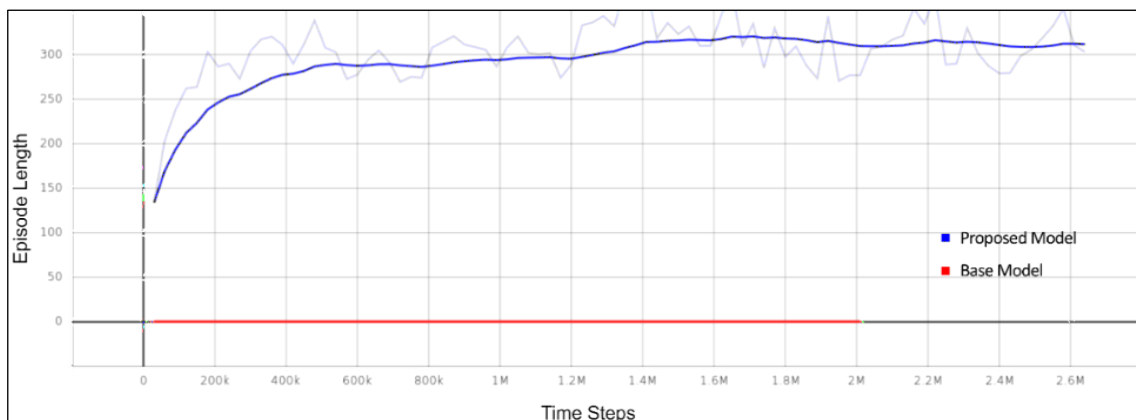


Figure 6. Episode length

The episode length shows how long the agent continues its movement without colliding with any other object. Figure 6 depicts that our proposed model is superior to the base model. It indicates that, compared to the base model, the agent trained using the suggested model experienced longer episodes. The baseline model's maximum episode duration is between 10 and 15, while the suggested model's maximum episode length is 350.

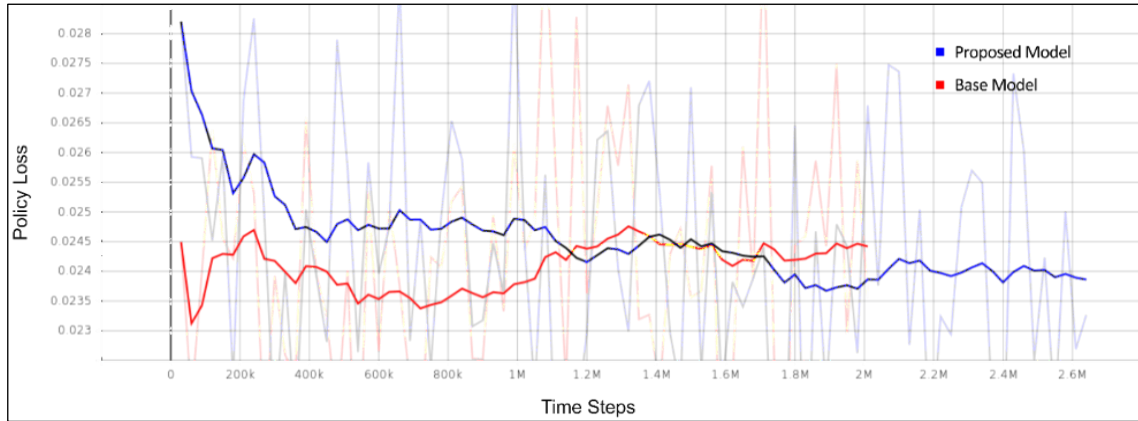


Figure 7. Policy Loss

Policy loss is a measure that represents how much improvement is needed in the training process to achieve better results. According to Figure 7, the suggested model had a large policy loss at first, but after 1.1 million steps, both the proposed and base models performed similarly.

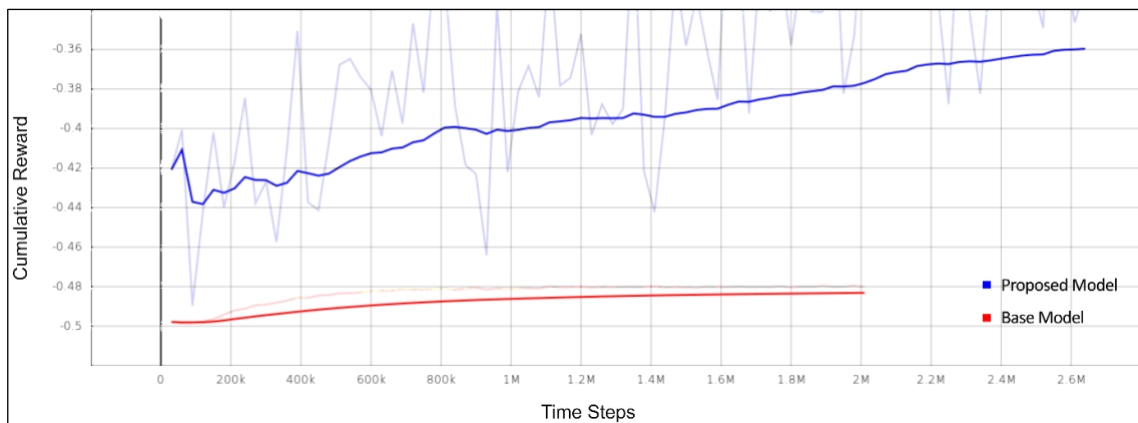


Figure 8. Cumulative Reward

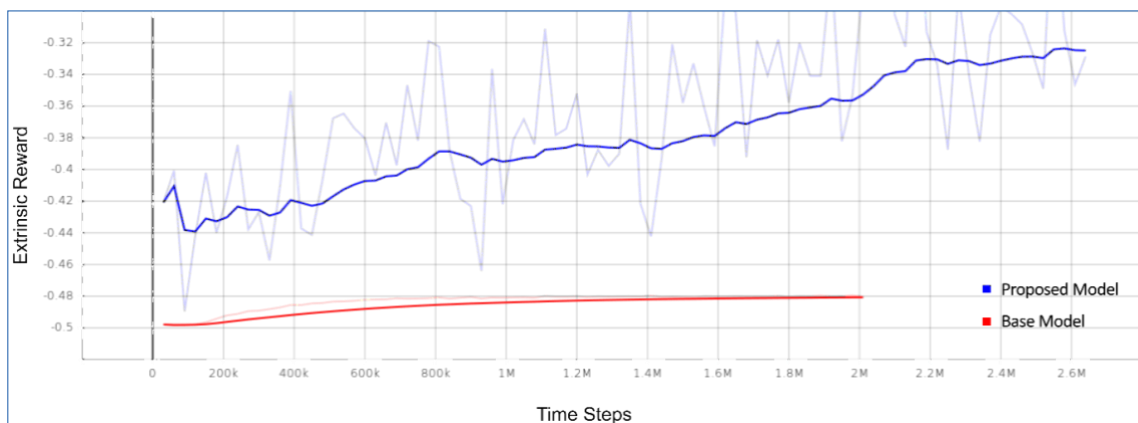


Figure 9. Extrinsic Reward

The cumulative reward is the total score of the agent during their training process. Figure 8 depicts that the suggested model surpasses the base model in terms of the cumulative reward. The agent trained in the suggested model gets a higher cumulative reward than the agent trained in the base model.

An extrinsic reward is like the points or bonuses you earn in a game for completing certain objectives set by the game. These rewards come from the environment and help guide the agent to make better decisions to achieve higher performance. Figure 9 shows that the developed approach outperformed the base model in

terms of the extrinsic reward. It means the agent trained during the proposed model got a higher extrinsic reward from the environment than the base model agent.

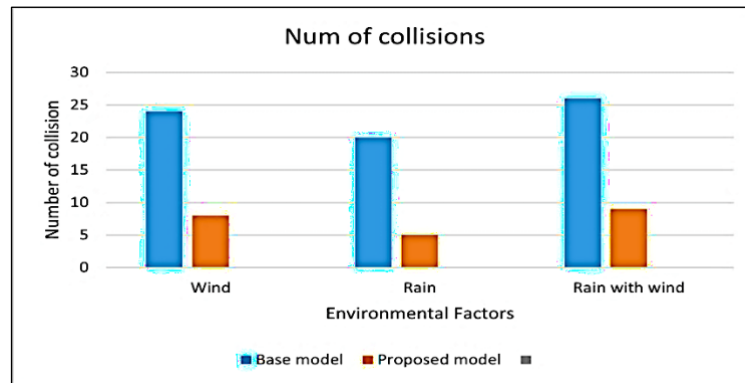


Figure 10. Number of collisions.

The number of collisions is another metric used in this investigation. After training the agent, it is tested in wind, rain, and wind with rain conditions in a sea environment. Figure 10 shows that the agent performed better in the proposed model environment than in the base model. The agent trained during the proposed model has minimal collisions in wind, rain, and wind with rain conditions.

## 5. CONCLUSION

Deep reinforcement learning has various applications in robotics, healthcare, autonomous driving, and real-world scenarios. Unmanned Surface Vehicles (USVs) utilize deep reinforcement learning for different tasks in the sea environment, like navigation and control. However, it is found that the existing studies have not fully considered all environmental factors that can affect USV navigation and control. In this view, all the previous research discussed in the literature review is considered. To overcome this issue, we proposed this model in which the environmental factor is considered, and the environment is created in Unity 3D. The two powerful and strong algorithms, PPO and GAIL, were applied for the training process. Expert demonstrations of states and actions are used to train the model. The paper evaluated performance based on mean reward, extrinsic reward, episode duration, policy loss, value estimation, and the number of collisions with obstacles. The proposed model outperformed the base model in the various matrices. Furthermore, the agent in the proposed model has fewer collisions than the base model agent. Overall, the results indicated that the proposed approach enhances USV control in various environmental situations by effectively learning from expert behaviour. In the future, we will develop some more combined deep reinforcement learning models for multicriteria decision making for fuzzy logic and deep learning techniques and apply them to real-world problems.

## DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest in this work.

## REFERENCES

- [1] Z.-H. Zhou, "Introduction," in *Machine Learning*, Singapore: Springer Singapore, 2021, pp. 1–24. doi: [10.1007/978-981-15-1967-3\\_1](https://doi.org/10.1007/978-981-15-1967-3_1).
- [2] T. Jo, *Machine Learning Foundations*. Cham: Springer International Publishing, 2021. doi: [10.1007/978-3-030-65900-4](https://doi.org/10.1007/978-3-030-65900-4).
- [3] D. Han, B. Mulyana, V. Stankovic, and S. Cheng, "A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation," *Sensors*, vol. 23, no. 7, p. 3762, Apr. 2023, doi: [10.3390/s23073762](https://doi.org/10.3390/s23073762).
- [4] K. Souchleris, G. K. Sidiropoulos, and G. A. Papakostas, "Reinforcement Learning in Game Industry—Review, Prospects and Challenges," *Appl Sci*, vol. 13, no. 4, p. 2443, Feb. 2023, doi: [10.3390/app13042443](https://doi.org/10.3390/app13042443).
- [5] R. Liu, F. Nageotte, P. Zanne, M. de Mathelin, and B. Drespe-Langle, "Deep Reinforcement Learning for the Control of Robotic Manipulation: A Focussed Mini-Review," *Robotics*, vol. 10, no. 1, p. 22, Jan. 2021, doi: [10.3390/robotics10010022](https://doi.org/10.3390/robotics10010022).
- [6] M. Krichen, "Deep Reinforcement Learning," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2023, pp. 1–7. doi: [10.1109/ICCCNT56998.2023.10306453](https://doi.org/10.1109/ICCCNT56998.2023.10306453).

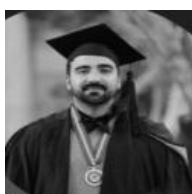
- [7] D. Wang, W. Li, L. Zhu, and J. Pan, "Learning to control and coordinate mixed traffic through robot vehicles at complex and unsignalized intersections," *Int J Rob Res*, vol. 44, no. 5, pp. 805–825, Apr. 2025, doi: [10.1177/02783649241284069](https://doi.org/10.1177/02783649241284069).
- [8] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement Learning in Healthcare: A Survey," *ACM Comput Surv*, vol. 55, no. 1, pp. 1–36, Jan. 2023, doi: [10.1145/3477600](https://doi.org/10.1145/3477600).
- [9] X. Xu, Y. Lu, X. Liu, and W. Zhang, "Intelligent collision avoidance algorithms for USVs via deep reinforcement learning under COLREGs," *Ocean Eng*, vol. 217, p. 107704, Dec. 2020, doi: [10.1016/j.oceaneng.2020.107704](https://doi.org/10.1016/j.oceaneng.2020.107704).
- [10] R. Bloss, "Autonomous unmanned vehicles take over on land, sea and in the air," *Ind Robot An Int J*, vol. 40, no. 2, pp. 100–105, Mar. 2013, doi: [10.1108/01439911311297676](https://doi.org/10.1108/01439911311297676).
- [11] C. Chen, F. Ma, X. Xu, Y. Chen, and J. Wang, "A Novel Ship Collision Avoidance Awareness Approach for Cooperating Ships Using Multi-Agent Deep Reinforcement Learning," *J Mar Sci Eng*, vol. 9, no. 10, p. 1056, Sep. 2021, doi: [10.3390/jmse9101056](https://doi.org/10.3390/jmse9101056).
- [12] X. Wu *et al.*, "The autonomous navigation and obstacle avoidance for USVs with ANOA deep reinforcement learning method," *Knowledge-Based Syst*, vol. 196, p. 105201, May 2020, doi: [10.1016/j.knosys.2019.105201](https://doi.org/10.1016/j.knosys.2019.105201).
- [13] N. Yan, S. Huang, and C. Kong, "Reinforcement Learning-Based Autonomous Navigation and Obstacle Avoidance for USVs under Partially Observable Conditions," *Math Probl Eng*, vol. 2021, pp. 1–13, May 2021, doi: [10.1155/2021/5519033](https://doi.org/10.1155/2021/5519033).
- [14] C. Lamini, S. Benhlila, and A. Elbekri, "Genetic Algorithm Based Approach for Autonomous Mobile Robot Path Planning," *Procedia Comput Sci*, vol. 127, pp. 180–189, 2018, doi: [10.1016/j.procs.2018.01.113](https://doi.org/10.1016/j.procs.2018.01.113).
- [15] W. Li *et al.*, "Crowd intelligence in AI 2.0 era," *Front Inf Technol Electron Eng*, vol. 18, no. 1, pp. 15–43, Jan. 2017, doi: [10.1631/FITEE.1601859](https://doi.org/10.1631/FITEE.1601859).
- [16] Rahim Ullah, Muhammad Saeed, W. Ali, Junaid Nazar, and Fakhra Nazar, "A Cooperative Heterogeneous Multi-Agent System Leveraging Deep Reinforcement Learning," *Knowl Decis Syst with Appl*, vol. 1, pp. 112–124, Mar. 2025, doi: [10.59543/kadsa.v1i.13931](https://doi.org/10.59543/kadsa.v1i.13931).
- [17] E. Raboin, P. Svec, D. Nau, and S. K. Gupta, "Model-predictive target defense by team of unmanned surface vehicles operating in uncertain environments," in *2013 IEEE International Conference on Robotics and Automation*, IEEE, May 2013, pp. 3517–3522. doi: [10.1109/ICRA.2013.6631069](https://doi.org/10.1109/ICRA.2013.6631069).
- [18] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cogn Robot*, vol. 3, pp. 54–70, 2023, doi: [10.1016/j.cogr.2023.04.001](https://doi.org/10.1016/j.cogr.2023.04.001).
- [19] M. Ghasemi and D. Ebrahimi, "Introduction to Reinforcement Learning," Dec. 2024, [Online]. Available: <http://arxiv.org/abs/2408.07712>
- [20] D. Ruhela and A. Ruhela, "Tuning Apex DQN: A Reinforcement Learning based Deep Q-Network Algorithm," in *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, New York, NY, USA: ACM, Jul. 2024, pp. 1–5. doi: [10.1145/3626203.3670581](https://doi.org/10.1145/3626203.3670581).
- [21] H. Van Hasselt, A. Guez, and D. Silver, "Deep Reinforcement Learning with Double Q-Learning," *Proc AAAI Conf Artif Intell*, vol. 30, no. 1, Mar. 2016, doi: [10.1609/aaai.v30i1.10295](https://doi.org/10.1609/aaai.v30i1.10295).
- [22] Q. Wu, W. Wang, P. Fan, Q. Fan, H. Zhu, and K. B. Letaief, "Cooperative Edge Caching Based on Elastic Federated and Multi-Agent Deep Reinforcement Learning in Next-Generation Networks," *IEEE Trans Netw Serv Manag*, vol. 21, no. 4, pp. 4179–4196, Aug. 2024, doi: [10.1109/TNSM.2024.3403842](https://doi.org/10.1109/TNSM.2024.3403842).
- [23] A. Feriani and E. Hossain, "Single and Multi-Agent Deep Reinforcement Learning for AI-Enabled Wireless Networks: A Tutorial," *IEEE Commun Surv Tutor*, vol. 23, no. 2, pp. 1226–1252, 2021, doi: [10.1109/COMST.2021.3063822](https://doi.org/10.1109/COMST.2021.3063822).
- [24] K. Ahmic, J. Ultsch, J. Brembeck, and C. Winter, "Reinforcement Learning-Based Path Following Control with Dynamics Randomization for Parametric Uncertainties in Autonomous Driving," *Appl Sci*, vol. 13, no. 6, p. 3456, Mar. 2023, doi: [10.3390/app13063456](https://doi.org/10.3390/app13063456).
- [25] H. Huang, Y. Yang, H. Wang, Z. Ding, H. Sari, and F. Adachi, "Deep Reinforcement Learning for UAV Navigation Through Massive MIMO Technique," *IEEE Trans Veh Technol*, vol. 69, no. 1, pp. 1117–1121, Jan. 2020, doi: [10.1109/TVT.2019.2952549](https://doi.org/10.1109/TVT.2019.2952549).
- [26] H. Shi, Z. Lin, K.-S. Hwang, S. Yang, and J. Chen, "An Adaptive Strategy Selection Method With Reinforcement Learning for Robotic Soccer Games," *IEEE Access*, vol. 6, pp. 8376–8386, 2018, doi: [10.1109/ACCESS.2018.2808266](https://doi.org/10.1109/ACCESS.2018.2808266).
- [27] D. Sisodia, S. K. Shrivastava, and R. C. Jain, "ISVM for Face Recognition," in *2010 International Conference on Computational Intelligence and Communication Networks*, IEEE, Nov. 2010, pp. 554–559. doi: [10.1109/CICN.2010.109](https://doi.org/10.1109/CICN.2010.109).
- [28] F. L. Da Silva, G. Warnell, A. H. R. Costa, and P. Stone, "Agents teaching agents: a survey on inter-agent transfer learning," *Auton Agent Multi Agent Syst*, vol. 34, no. 1, p. 9, Apr. 2020, doi: [10.1007/s10458-019-09430-0](https://doi.org/10.1007/s10458-019-09430-0).
- [29] Y. Mushtaq, W. Ali, U. Ghani, R. U. Khan, and A. Kumar Adak, "Advancing Aviation Safety and Sustainable Infrastructure: High-Accuracy Detection and Classification of Foreign Object Debris Using Deep Learning Models," *Int J Sustain Dev Goals*, vol. 1, pp. 82–98, May 2025, doi: [10.59543/ijisdg.v1i.14279](https://doi.org/10.59543/ijisdg.v1i.14279).
- [30] A. Saha, B. K. Debnath, P. Chatterjee, A. K. Panaiyappan, S. Das, and G. Anusha, "Generalized Dombi-based probabilistic hesitant fuzzy consensus reaching model for supplier selection under healthcare supply chain framework," *Eng Appl Artif Intell*, vol. 133, p. 107966, Jul. 2024, doi: [10.1016/j.engappai.2024.107966](https://doi.org/10.1016/j.engappai.2024.107966).

**BIOGRAPHIES OF AUTHORS**

**Wajid Ali** is a Lecturer (VFM) at Air University Islamabad and completed his Ph.D. in Mathematics in March 2024. His doctoral research focused on generalizations of fuzzy sets, rough sets, and three-way decision models, with strong applications in medical diagnosis, investment decision-making, sustainable systems, and data classification. He has authored 25+ publications in reputable SCI/Scopus-indexed journals, reflecting deep expertise in fuzzy algebra, rough set theory, and advanced decision-making algorithms. He also contributed to an HEC-funded project on developing Urdu Sign Language gloves for speech conversion to support mute individuals. His research interests include Fuzzy Sets and Their Extensions, Rough Sets and Decision-Theoretic Rough Sets, Three-Way and Multi-Granulation Decision Models, Artificial Intelligence and Machine Learning, Deep Learning and Computer Vision, Reinforcement Learning & Autonomous Navigation, and Graph Theory and Intelligent Systems. He has strong practical experience with Python, MATLAB, Neo4j, and Excel, and is passionate about building intelligent, application-oriented solutions for complex, uncertain environments. He can be contacted at email: wajidali00258@gmail.com



**Raheem Ullah** received a master's degree in computer science. From a prestigious COMSATS University Islamabad, with a major in Artificial Intelligence. My research interests lie in AI and Data Science. Currently, I work as a Flutter Developer at Isloo Tech, Islamabad. I have also previously worked as a WordPress website developer and graphics designer at ASK Development, Islamabad. He can be contacted at email: rahimullah.cs@gmail.com



**Usman Ghani** did a Master of Science in Mathematics and Cryptography from Air University. He specialized in Cryptography and focused on modern encryption methods and mathematical security systems. Currently, He is working as a Lecturer at Air University, where he teaches and guides students in Mathematics, Cryptography, and Cybersecurity. His goal is to help students build a strong foundation in mathematical thinking and secure digital communication. He can be contacted at email: usman.ghani@au.edu.pk