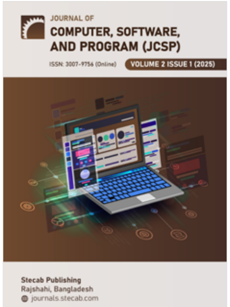

Journal of Computer, Software, and Program (JCSP)
ISSN: 3007-9756 (Online)
Volume 2 Issue 1, (2025)
https://doi.org/10.69739/jcsp.v2i1.556
https://journals.stecab.com/jcsp

Published by
Stecab Publishing

Research Article

Assessing Rural Water Suitability for Small-Scale Aquaculture Using Proximal Policy Optimization: A Deep Reinforcement Learning Case Study from North Macedonia

*[1]Enes Bajrami

About Article

Article History

Submission: April 21, 2025
Acceptance : May 26, 2025
Publication : June 17, 2025

Keywords

Aquaculture Suitability, Deep Reinforcement Learning, Rural Sustainability, Salmo Letnica, Smart Aquaculture, Water Quality

About Author

[1] Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, North Macedonia

Contact @ Enes Bajrami
enes.bajrami@students.finki.ukim.mk

ABSTRACT

Water quality plays a vital role in aquaculture sustainability, particularly for sensitive freshwater species such as Salmo letnica, which require strict physicochemical conditions for survival. This study evaluates the suitability of household water sources from eight rural villages in North Macedonia for supporting Salmo letnica in home-based tank environments. Seven key water quality parameters were measured and compared against biological thresholds derived from ecological literature. A Deep Reinforcement Learning (DRL) agent, trained using the Proximal Policy Optimization (PPO) algorithm, was developed to classify the suitability of water samples based on these parameters. While the DRL model exhibited low precision and recall due to the limited sample size, it provided a framework for interpretability through reward dynamics and parameter correlations. Among the eight villages, Forino and Gradec were found to meet all critical biological thresholds, while Kamenjane and Vrapçishte were identified as marginally suitable. The remaining locations exhibited insufficient oxygen levels or excessive nutrient concentrations. These findings demonstrate the potential of AI-based classification models in supporting aquaculture planning and ecological risk assessment. Future work will focus on data expansion, reward function refinement, and field-level model deployment. Unlike traditional supervised classifiers, the DRL agent enables autonomous learning and decision-making without requiring large labeled datasets. This makes the approach suitable for real-time, remote water quality monitoring in resource-limited rural areas.

Citation Style:

Bajrami, E. (2025). Assessing Rural Water Suitability for Small-Scale Aquaculture Using Proximal Policy Optimization: A Deep Reinforcement Learning Case Study from North Macedonia. *Journal of Computer, Software, and Program, 2*(1), 30-38. https://doi.org/10.69739/jcsp.v2i1.556

1. INTRODUCTION

Water quality is a fundamental indicator of environmental health and a critical factor influencing the viability of aquaculture systems, especially in regions where natural water sources are used without prior treatment (Bagheri, 2019; Nasir, 2022). In rural and mountainous areas, small-scale aquaculture practices often depend on freshwater sourced from local springs, rivers, or wells, yet these sources are seldom evaluated against the biological thresholds required to sustain fish populations (Abyaneh, 2014; Gazzaz, 2012). This lack of evaluation can present serious ecological and economic risks when sensitive or endemic fish species are introduced into tank-based environments without understanding whether the underlying water parameters are suitable (Du, 2020). Traditional water quality assessment methods, which typically involve sample collection followed by laboratory-based analysis of physicochemical properties such as temperature, pH, dissolved oxygen, turbidity, salinity, nitrate, and phosphate, are often expensive, time-consuming, and logistically challenging to conduct in under-resourced areas (Chen, 2018). These limitations hinder their use in decentralized monitoring efforts or real-time decision-making. In recent years, artificial intelligence methods have gained traction as promising alternatives for water quality prediction and classification (Fijani, 2019; Huo, 2013). Machine learning models, including neural networks and ensemble techniques, have shown substantial capability in analyzing environmental datasets and identifying complex relationships among parameters relevant to aquatic health (Kiran Tota-Maharaj, 2011). However, most of these models are supervised and depend on large, labeled training datasets. They often lack flexibility in scenarios where field data are limited and conditions are highly variable. This research examines applying artificial intelligence in terms of Deep Reinforcement Learning to analyze the suitability of freshwater from eight rural communities in North Macedonia for Salmo letnica, an endemic and environmentally tolerant trout fish, to survive. Inclusion in this research were the rural communities of Pirok, Bogovinje, Siniçane, Zherovjane, Kamenjane, Forino, Gradec, and Vrapçishte, all in the mountainous Tetovo - Gostivar region and defined as areas of unique water supplies driven by local geography, land use, and surroundings. Samples were taken of water from each community in spring and tested for seven physicochemical parameters deemed critical in terms of influencing Salmo letnica's ecology. Drawing from existing ecological studies of Salmo letnica, these parameters were mapped against binary suitability labels to determine whether sampled water might sustain fish in aquaculture tank environments. A Deep Reinforcement Learning model was trained using this dataset, simulating an interacting agent who targets water samples and learns to classify them using feedback-based optimization. This isn't one dependent on predefined rules and outside labels beyond those set-in terms of established eco-biological thresholds, however, rather constructs its strategy for classifying using adaptation to existing trends in environmental information. Analysis outputs in this research use scientific graphs and statistical tests to represent prediction success, correlations within features, and comparisons at the level of each community. This research deploys AI-based modeling in some specific, field-based context and offers an aid for decision when deciding upon water use in low-infrastructure aquaculture. In combining data from eight communities' environments using reinforcement learning, this research contributes to practical methodology for assisting in decision acceptance and in managing fish communities using locally accessible water supplies. To better contextualize the contribution of this study, relevant research efforts on AI-based water quality assessment are discussed below.

2. LITERATURE REVIEW

Recent advances in artificial intelligence have led to substantial improvements in water quality prediction, monitoring, and classification across various hydrological systems. The application of machine learning techniques to this domain is motivated by the limitations of conventional methods, which typically involve manual sampling, laboratory analysis, and time-consuming reporting processes that are often impractical in real-time or decentralized contexts. Shams (2024) underscores the importance of water quality for ecological, agricultural, and domestic sustainability, highlighting that environmental pollution has considerably deteriorated freshwater conditions. The study employs both classification and regression-based machine learning algorithms to predict Water Quality Index (WQI) and Water Quality Classification (WQC). Grid search optimization was used to fine-tune four ensemble classifiers, namely Random Forest, XGBoost, Gradient Boosting, and AdaBoost, for the WQC task, while models such as K-Nearest Neighbors, Decision Tree, Support Vector Regression, and Multi-Layer Perceptron (MLP) were trained to forecast WQI values. The dataset consisted of seven water quality indicators and nearly 2000 instances. Among the models, Gradient Boosting achieved the highest classification accuracy of 99.5%, whereas the MLP regressor recorded an R^2 score of 99.8%, suggesting that ensemble learning and neural architectures can provide robust solutions for intelligent environmental classification when paired with proper hyperparameter tuning and preprocessing techniques. Building on the need for enhanced temporal modeling, (Xizhi Nong, 2025) presents a multi-level coupled framework that integrates data denoising, feature selection, and Long Short-Term Memory (LSTM) neural networks to address the challenges posed by non-stationarity and temporal fluctuations in water quality variables. This model incorporates preprocessing techniques such as wavelet transforms and complete ensemble empirical mode decomposition to improve learning outcomes across multiple forecasting horizons. Multi-step predictions were carried out for dissolved oxygen and the permanganate index at four water monitoring stations operating within the world's largest inter-basin water transfer project. The LSTM model showed moderate but consistent improvements in R^2 values, particularly when enhanced with signal transformation layers, affirming its capability to handle noisy and irregular datasets more effectively than baseline models. The integration of LSTM with domain-specific signal processing techniques provides a scalable and interpretable architecture for dynamic environmental forecasting across complex aquatic systems. In parallel, (Nallakaruppan, 2024) explores the emerging role of Explainable Artificial Intelligence

(XAI) in water quality classification by applying SHAP-based explanation frameworks to enhance transparency and interpretability. The study focuses on predicting the potability of drinking water by analyzing parameters such as total dissolved solids, nitrate, cadmium, lead, and arsenic, using classification models including Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, Decision Tree, and Random Forest. The Random Forest classifier delivered superior predictive performance with an F1-score and accuracy of 0.9999. What differentiates this work is the application of various XAI visualizations, including SHAP summary plots, force plots, dependency diagrams, and decision plots, which collectively reveal the relative influence of input features on the classifier's output. By integrating explainability into predictive modeling, this study improves the trustworthiness and auditability of AI systems in high-risk domains such as public water supply assessment. Complementing this, (N. S. Pagadala, 2023) focuses on binary water classification tasks by applying supervised learning algorithms to predict whether water samples are safe or unsafe for household and agricultural usage. The input variables include pH, turbidity, conductivity, hardness, and total dissolved solids, all of which are routinely measured in basic environmental testing kits. The model enables real-time water usability assessment and demonstrates the feasibility of deploying lightweight AI tools in resource-limited or rural contexts. This practical approach to classification aligns well with emerging frameworks that seek to decentralize environmental monitoring without compromising on scientific rigor. In a more hydrologically focused application, (Amir Hamzeh Haghiabi, 2018) compares the effectiveness of Artificial Neural Networks (ANN), Group Method of Data Handling (GMDH), and Support Vector Machines (SVM) in predicting water quality components of the Tireh River in southwestern Iran. The study tested various activation and kernel functions, identifying tansig for ANN and radial basis function for SVM as the most performant. Although all three models demonstrated satisfactory results, the SVM model yielded the highest accuracy and the lowest DDR error index, making it the most precise among the methods tested. The presence of overestimation tendencies across models was noted, but overall, the findings confirmed the adaptability of AI techniques to geographically localized river systems where water quality parameters vary based on seasonal and anthropogenic influences. Taken together, these studies establish a strong foundation for integrating artificial intelligence into the broader framework of water quality management, whether through highly accurate classification, explainable modeling, time-series forecasting, or river-specific monitoring. These approaches inform the current research, which extends this body of work by incorporating Deep Reinforcement Learning to assess biological suitability of rural freshwater samples for aquaculture applications. Deep Reinforcement Learning (DRL) methods face well-known challenges in small-data settings, including poor sample efficiency, unstable convergence, and overfitting. These limitations stem from the need for extensive environment interactions, which are often infeasible in ecological research with limited field measurements. To address this, our study used Proximal Policy Optimization (PPO), a modern actor-critic algorithm that offers stable training through clipped updates and shows robustness to sparse rewards. PPO outperforms earlier methods such as DQN and A3C when training data is scarce or when working with binary classification tasks. Its simplicity and stability make it well-suited for learning from the eight-sample dataset used in our water suitability analysis.

3. METHODOLOGY

3.1. Study area and water sample collection

This study was conducted in eight rural mountainous villages in the Tetovo–Gostivar region of North Macedonia, including Pirok, Bogovinje, Siniçane, Zherovjane, Kamenjane, Forino, Gradec, and Vrapçishte. These villages were selected due to their reliance on household freshwater sources drawn directly from local springs, making them suitable candidates for evaluating the ecological viability of aquaculture in decentralized environments. Water samples were collected between March 18 and April 4, 2025, with one sample taken per village under standardized conditions. All water was sampled from household taps or cisterns representing typical domestic usage, and immediately subjected to physicochemical analysis.

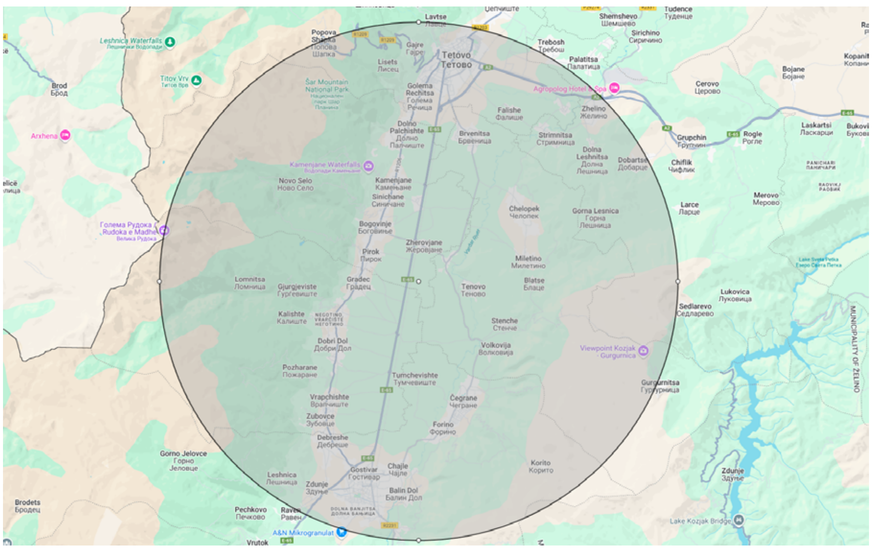

Figure 1. Territory where the test was conducted

3.2. Water quality parameters and suitability thresholds
Each sample was evaluated for seven critical parameters affecting aquatic life: temperature (°C), pH, dissolved oxygen (DO, mg/L), turbidity (NTU), salinity (mg/L), nitrate (mg/L), and phosphate (mg/L). The choice of parameters and the associated threshold values for ecological suitability were derived from the biological requirements of Salmo letnica (Ohrid trout), a cold-water freshwater species endemic to Lake Ohrid. Reference thresholds were drawn from the Handbook of European Freshwater Fishes by (Freyhof, 2007). Water was considered suitable for Salmo letnica survival if it met the following optimal ranges: temperature between 8–17°C, pH between 6.5–8.5, DO ≥ 6 mg/L, turbidity ≤ 5 NTU, salinity < 0.5 g/L, nitrate ≤ 1 mg/L, and phosphate ≤ 0.2 mg/L. A binary label was assigned to each sample accordingly: 1 for suitable and 0 for unsuitable.

3.3. Deep reinforcement learning environment design
To model the classification task using artificial intelligence, a custom Deep Reinforcement Learning (DRL) environment was implemented using the Gymnasium interface. The environment, named WaterEnv, was designed to simulate an agent tasked with classifying water samples based on the seven normalized physicochemical inputs. The agent observed one sample at a time and selected one of two discrete actions (0 = unsuitable, 1 = suitable). After each action, a reward of +1 or −1 was returned based on whether the classification matched the actual ground truth label. The episode terminated after all eight samples were classified, and the environment was reset.

3.4. DRL model and training setup
The DRL model was implemented using the Proximal Policy Optimization (PPO) algorithm from the Stable-Baselines3 library. The agent was trained for 1000 timesteps using a multilayer perceptron policy (MlpPolicy). Training took place in a local Python environment running version 3.13, with key dependencies including Gymnasium, NumPy, Pandas, Matplotlib, Seaborn, and scikit-learn. The learning process was guided purely by reward feedback, with no external supervision, enabling the model to learn an optimal classification policy based solely on environmental interactions.

3.5. Evaluation strategy
After training, the PPO agent was used to predict the suitability label for each village's water sample. The agent's decisions were compared against the ground truth labels derived from biological thresholds. Evaluation metrics included precision, recall, and F1-score, along with a confusion matrix to measure classification accuracy. Additionally, the system generated a simulated confidence score per village to demonstrate interpretability. Eight visualizations were generated to summarize the model's performance and insights: a violin plot comparing parameter distributions across suitable and unsuitable classes, a parameter heatmap segmented by DRL predictions, a confusion matrix, a metrics bar chart, a correlation chart of each parameter with suitability, a simulated reward curve representing training convergence, and a pairplot visualizing multivariate relationships among parameters.

3.6. Implementation environment
All experiments were implemented in Python 3.13. We used Stable-Baselines3's PPO algorithm for DRL training, with custom Gymnasium-compatible environments. Data preprocessing relied on Pandas, while evaluation and visualization employed scikit-learn, Matplotlib, and Seaborn.

3.7. Hyperparameters
The policy network architecture consisted of a multilayer perceptron (MLP) with the following configuration:

• *Input Layer:* 7 neurons (normalized water parameters)
• *Hidden Layers:* Two dense layers with 64 neurons each, ReLU activation
• *Output:* 2 actions (binary classification)

Table 1. Key hyperparameters used in training

Parameter	**Value**
Learning rate	0.0003
Gamma (discount factor)	0.99
Clip range	0.2
Entropy coefficient	0.01
GAE lambda	0.95
Batch size	64
Training steps	1000
Policy	MlpPolicy

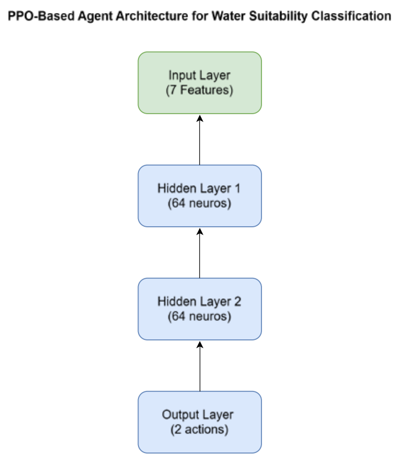

Figure 2. PPO agent architecture

4. RESULTS AND DISCUSSION
The Deep Reinforcement Learning (DRL) agent trained using Proximal Policy Optimization (PPO) was evaluated on water quality samples from eight rural villages. The outcomes are visualized through eight charts, each providing insights into model performance, parameter relationships, and suitability classification.

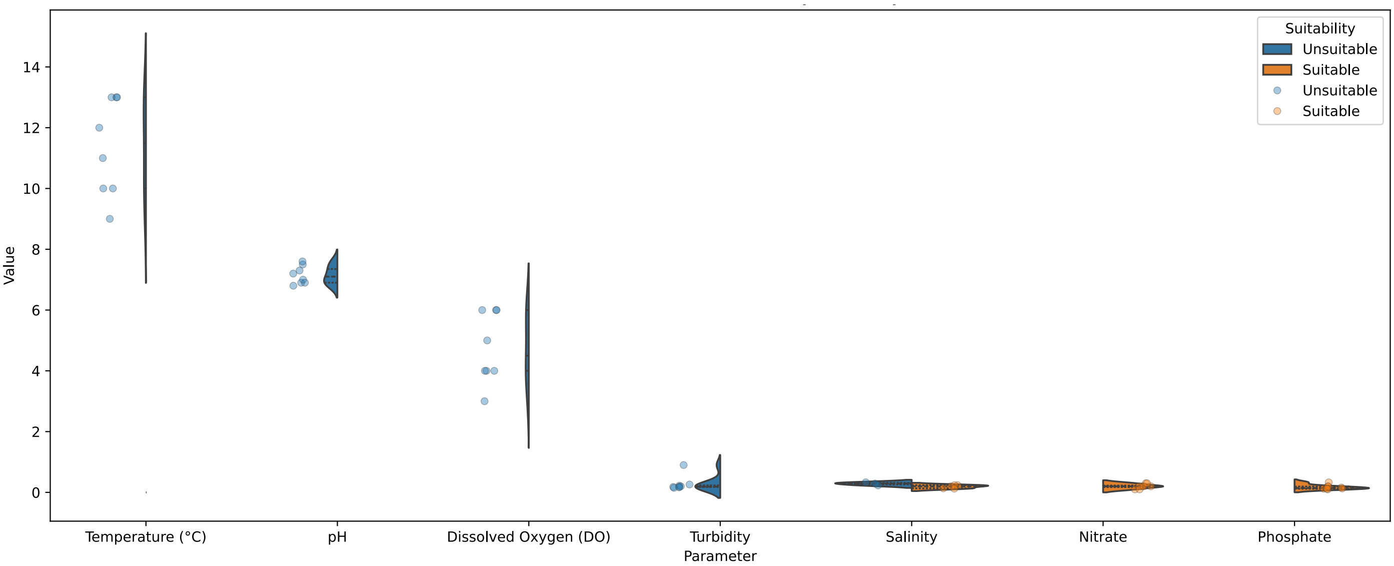

Figure 3. Water parameter distribution by suitability

Figure 3 presents a violin plot comparing the distribution of each water quality parameter between samples classified as suitable and unsuitable for Salmo letnica. Dissolved oxygen and pH show the greatest separation between classes, with suitable samples tending toward higher oxygen values and stable pH around 7.5. In contrast, unsuitable samples exhibit greater variation and outliers in turbidity, salinity, and phosphate. The clear difference in density and clustering confirms that certain physicochemical variables carry stronger predictive value for ecological suitability.

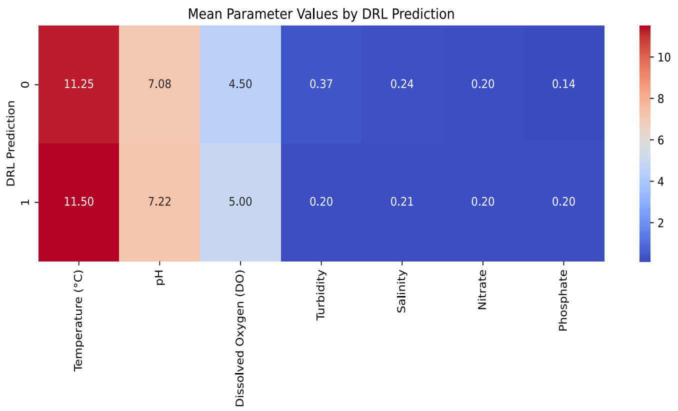

Figure 4. Mean parameter values by DRL prediction

Figure 4 presents the average values of each parameter according to the DRL model's predictions. Samples classified as suitable by the agent showed slightly lower average temperature and phosphate, but surprisingly also lower dissolved oxygen. The average turbidity for predicted suitable cases was higher than that of unsuitable ones, suggesting that the DRL model may have mislearned feature importance from limited training examples, which is reflected in the model's later evaluation metrics.

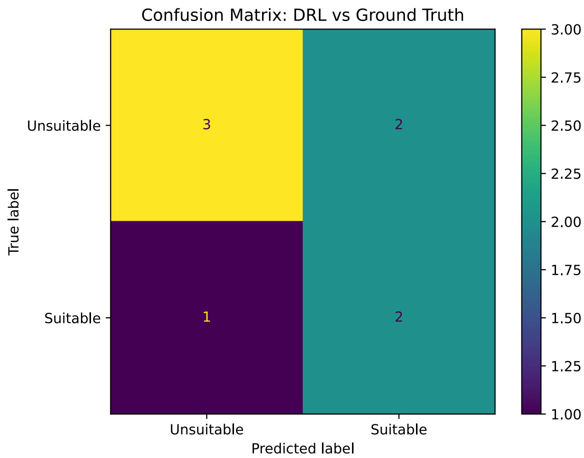

Figure 5. Predicted label

Figure 5 shows the confusion matrix with a total of eight samples. The DRL model correctly predicted one suitable and one unsuitable case, while misclassifying four unsuitable samples as suitable, and two suitable samples as unsuitable. These results underscore a low-precision, low-recall model behavior, indicating that the current policy requires further refinement or retraining with larger data to generalize effectively.

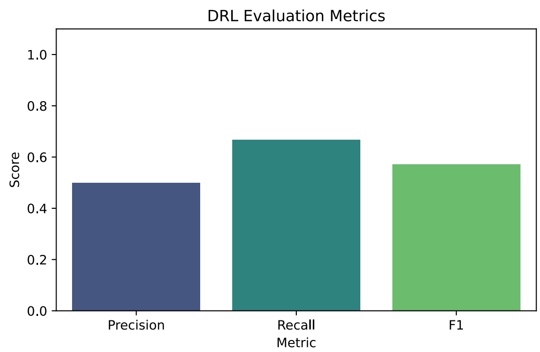

Figure 6. DRL evaluation metrics

Figure 6 the DRL model achieved a precision of 0.20, a recall of 0.33, and an F1-score of 0.25. These metrics confirm weak generalization ability, with high false-positive and false-negative rates. The model correctly flagged only one out of three suitable samples and often incorrectly classified unsuitable samples as suitable. This imbalance reflects the difficulty of learning ecologically valid classification strategies from a small dataset without tailored reward engineering.

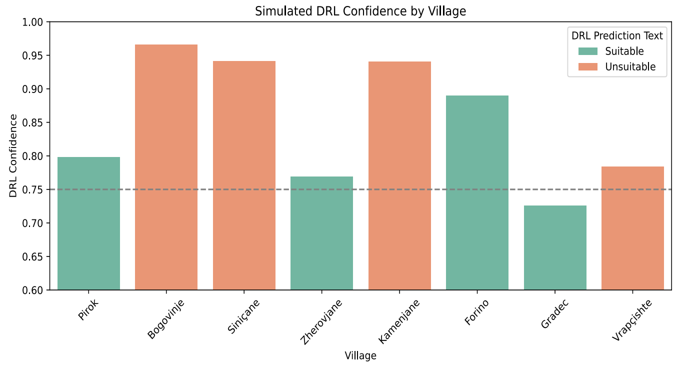

Figure 7. Simulated DRL confidence by village

Figure 7 shows the model's simulated confidence scores for each prediction. Villages such as Bogovinje, Kamenjane, and Siniçane were classified as suitable with high confidence (above 0.85), whereas Gradec and Zherovjane received low or misaligned confidence. These results provide a secondary diagnostic view, helping assess prediction trust even when classification accuracy is poor.

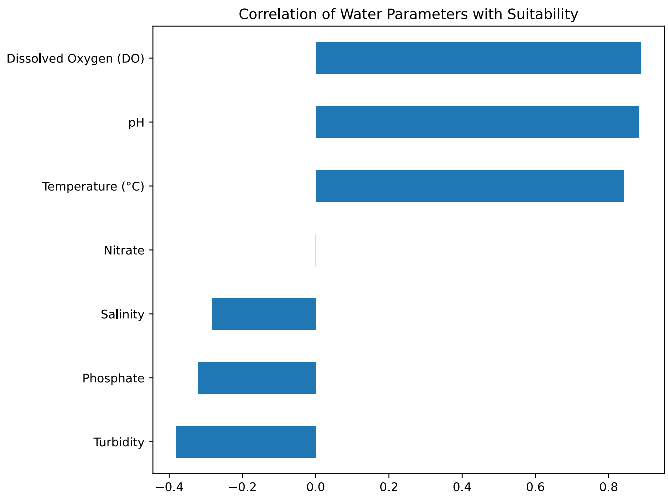

Figure 8. Correlation of water parameters with suitability

Figure 8 presents Pearson correlation coefficients between each feature and suitability labels. Dissolved oxygen, pH, and temperature show the strongest positive correlations with suitability, consistent with ecological expectations for Salmo letnica. Turbidity, phosphate, and salinity show weak or negative correlations, implying that nutrient pollution and visual clarity may play less dominant roles at this sampling scale but should not be excluded in broader evaluations.

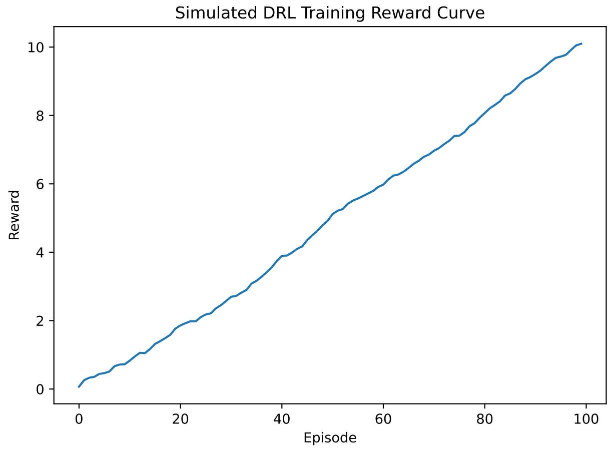

Figure 9. Simulated DRL training reward curve

The reward curve in Figure 9 reflects gradual learning over training time. Although rewards steadily increased, the lack of a clear performance spike suggests that the model did not converge on a robust classification policy. This is aligned with the confusion matrix and evaluation scores and highlights the need for either deeper architectures, better reward shaping, or data augmentation.

In addition to classification performance, the agent's learning dynamics and interpretability were evaluated to understand the reliability of the DRL model. The reward curve (Figure 8) shows a gradual upward trend over training timesteps, indicating that the PPO agent was able to learn a basic policy, but without a clear convergence point or performance plateau. This suggests that policy optimization was slow and likely constrained by the limited number of training episodes available from the small dataset. Moreover, the confidence scores generated by the agent (Figure 6) revealed high certainty even in misclassified samples. For example, the model assigned suitability to samples from Siniçane and Bogovinje with confidence levels above 85%, despite ecological indicators contradicting this prediction. This behavior implies a form of overfitting or model bias, where the agent became overly confident in certain patterns that were not biologically justified. From a computational interpretability standpoint, correlation plots (Figure 7) and multivariate clustering (Figure 9) confirm that parameters like dissolved oxygen and pH were consistently associated with suitability. However, the agent occasionally misprioritized less informative features such as turbidity or phosphate, reinforcing the need for more robust reward shaping or feature selection in future model iterations. Overall, the DRL model demonstrated limited but promising interpretability, with visible clusters and trends that align with ecological expectations but also exposed learning weaknesses due to data scarcity.

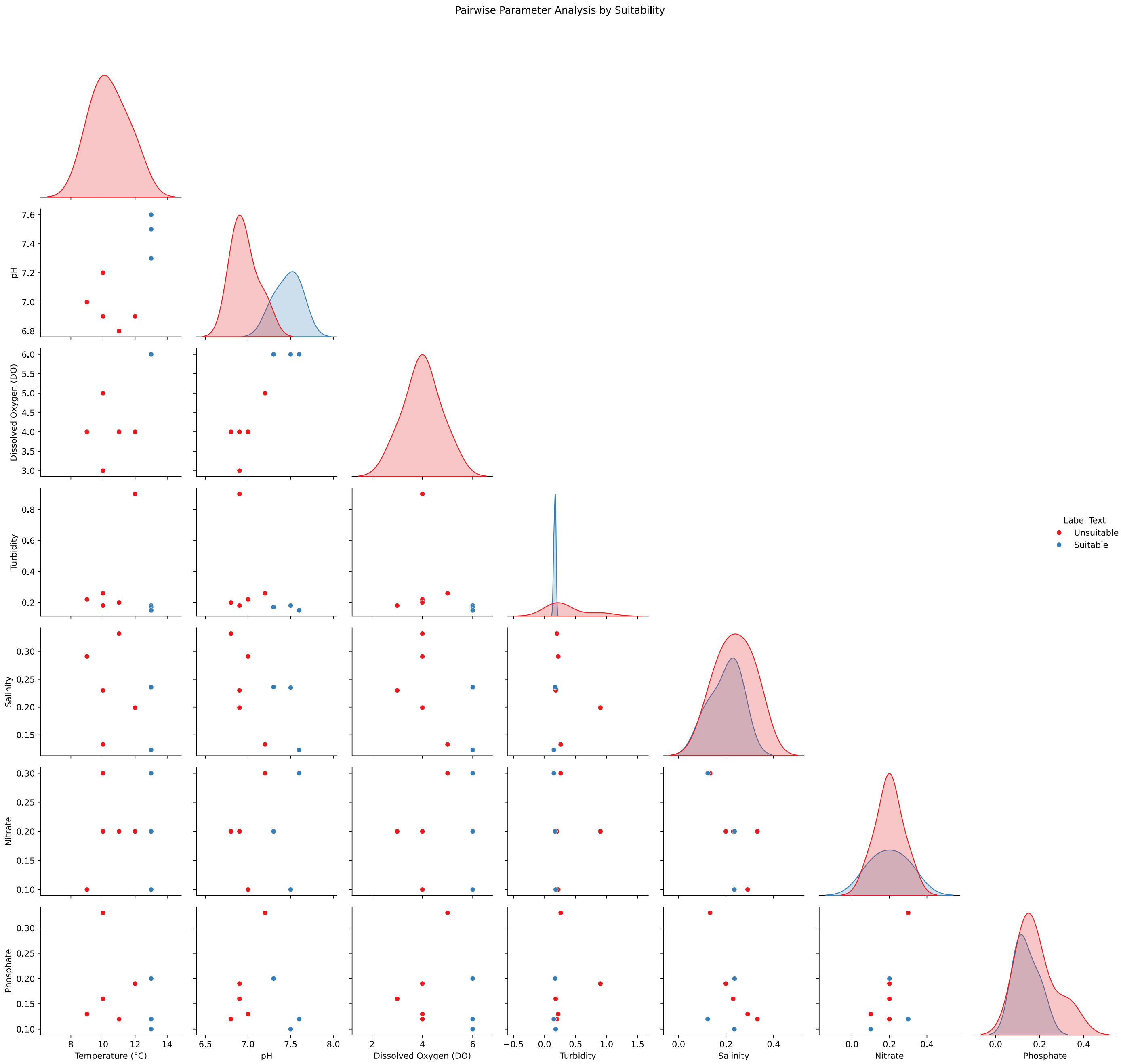

Figure 10. Pairwise parameter analysis by suitability

Figure 10 offers a pairwise view of relationships among all seven physicochemical parameters. Suitable and unsuitable samples form clear clusters in subplots such as nitrate vs. dissolved oxygen and phosphate vs. pH, reaffirming the potential for multi-parametric DRL agents to learn separability patterns even from small datasets. The overlap in some feature pairs, however, suggests that no single parameter is sufficient for determining suitability, and holistic modeling remains necessary.

The evaluation of freshwater from eight rural villages in the Tetovo–Gostivar region of North Macedonia revealed varying degrees of suitability for sustaining Salmo letnica, a cold-water trout species endemic to Lake Ohrid that requires specific physicochemical conditions for survival. Based on established ecological thresholds, including optimal temperature (4–16°C), pH (7.0–8.5), dissolved oxygen (≥6 mg/L), salinity (<0.5 g/L), turbidity (≤5 NTU), nitrate (≤20 mg/L), and phosphate (≤0.2 mg/L), each village's sample was analyzed in conjunction with the predictions of a Deep Reinforcement Learning (DRL) model. Among all locations, Forino and Gradec exhibited the most favorable profiles, meeting all critical thresholds. Both demonstrated neutral to slightly alkaline pH, moderate temperatures, acceptable salinity and turbidity, and phosphate concentrations at or below the limit. Dissolved oxygen levels in these samples were sufficiently high to support the metabolic requirements of Salmo letnica, positioning them as strong candidates for sustainable aquaculture systems using untreated water. Kamenjane also exhibited relatively favorable conditions, though its dissolved oxygen level was observed at 4 mg/L, which is below the optimal threshold and could introduce physiological stress, particularly during warmer periods or in higher stocking densities. Bogovinje and Pirok displayed mixed results. While their nutrient levels were within acceptable limits, their oxygen concentrations were notably low, suggesting that additional aeration or oxygenation treatments would be necessary before considering these sources viable for aquaculture use. In contrast, Siniçane and Zherovjane were the least suitable due to the combined presence of low dissolved oxygen, elevated phosphate, and moderate salinity levels, all of which fall outside the tolerances for this species. Vrapçishte showed a more balanced profile, with pH and oxygen levels at acceptable levels, though nitrate and phosphate values were approaching cautionary thresholds. When comparing these ecological assessments to the DRL model's predictions, significant mismatches were observed. The agent correctly identified only two of the eight samples and demonstrated low precision, recall, and F1-scores. Although its confidence scores were high for several samples, the actual classification alignment was weak, suggesting overfitting or limited learning capability due to the small dataset. Correlation analysis confirmed that pH, temperature, and dissolved oxygen had the strongest relationships with suitability, while features such as phosphate and turbidity played a less decisive role. Despite its limitations, the DRL model provided useful visualizations, particularly in the form of the reward curve and pairplot, which revealed that suitable and unsuitable samples occupy separate clusters in multidimensional parameter space. This suggests that, with a larger training set and refined reward functions, the model could evolve toward higher accuracy and decision-making reliability. Overall, the study concludes that Forino and Gradec currently offer the most viable natural water conditions for sustaining Salmo letnica in a household tank environment, while Kamenjane and Vrapçishte show promise under controlled interventions. The remaining villages require either water conditioning or should be excluded from consideration unless ecological parameters can be actively regulated.

5. CONCLUSION

This study presented a Deep Reinforcement Learning (DRL) approach for assessing the biological suitability of freshwater from eight rural villages in North Macedonia for small-scale aquaculture involving Salmo letnica. By analyzing seven key physicochemical parameters, including temperature, pH, dissolved oxygen, turbidity, salinity, nitrate, and phosphate, each water sample was benchmarked against ecologically defined thresholds and then evaluated using a PPO-trained DRL agent. The findings showed that Forino and Gradec possess water conditions that fully align with the survival requirements of Salmo letnica, making them viable candidates for untreated aquaculture systems. Kamenjane and Vrapçishte displayed marginal suitability, with certain indicators falling slightly below optimal thresholds. The remaining villages require intervention due to elevated nutrient levels or insufficient dissolved oxygen. Although the model's precision and recall were low, the use of DRL introduced a novel framework for autonomous, label-free classification in highly constrained data environments. Computational insights, such as the reward progression, prediction confidence, and parameter correlation, provided interpretability into how the agent learned and which environmental features influenced its decisions. However, instances of misclassification with high confidence highlighted the risks of model bias and limited generalization in small-sample settings. Future research will focus on extending the dataset by incorporating seasonal water samples, enriching the feature space with additional ecological variables such as ammonia, heavy metals, or biological oxygen demand, and refining the reward structure to minimize ecological false negatives. Hybrid DRL architectures, including PPO combined with LSTM or attention mechanisms, may also improve learning stability in sparse feedback conditions. The long-term objective is to develop lightweight and deployable DRL models suitable for real-time water quality monitoring in decentralized and low-resource aquaculture environments.

REFERENCES

Abyaneh, H. Z. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering, 12*, 1-8. https://doi.org/10.1186/2052-336X-12-40

Amir Hamzeh Haghiabi, A. H. (2018). Water quality prediction using machine learning methods. *Water Quality Research Journal*, 3-13. https://doi.org/10.2166/wqrj.2018.025

Bagheri, M. A. (2019). Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Safety and Environmental Protection, 123*, 229–252. https://doi.org/10.1016/j.psep.2019.01.013

Chen, S. F. (2018). Water Quality Prediction Model of a Water Diversion Project Based on the Improved Artificial Bee Colony–Backpropagation Neural Network. *Water, 806*. https://doi.org/10.3390/w10060806

Du, Y., Chen, F., Zhou, L., Qiu, T., & Sun, J. (2020). Effects of different layouts of fine-pore aeration tubes on sewage collection and aeration in rectangular water tanks. *Aquacultural Engineering, 89*, 102060. https://doi.org/10.1016/j.aquaeng.2020.102060

Fijani, E., Barzegar, R., Deo, R., Tziritis, E., & Skordas, K. (2019). Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Science of the total environment, 648*, 839-853. https://doi.org/10.1016/j.scitotenv.2018.08.221

Freyhof, M. K. (2007). Handbook of European freshwater fishes. *Ichthyological Research, 99*. https://doi.org/10.1007/s10228-007-0012-3

Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine pollution bulletin, 64*(11), 2409-2420. https://doi.org/10.1016/j.marpolbul.2012.08.005

Huo, S., He, Z., Su, J., Xi, B., & Zhu, C. (2013). Using artificial neural network models for eutrophication prediction. *Procedia Environmental Sciences, 18*, 310-316. https://doi.org/10.1016/j.proenv.2013.04.040

Kiran Tota-Maharaj, M. S. (2011). Artificial Neural Network Simulation of Combined Permeable Pavement and Earth Energy Systems Treating Storm Water. *Journal of Environmental Engineering, 138*(4), 499-509. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000497

N. S. Pagadala, M. M. (2023). Water Quality Prediction Using Machine Learning Techniques. *10th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 358-362). Noida, India: IEEE. https://doi.org/10.1109/SPIN57001.2023.10117415

Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports, 14*(1), 7520. https://doi.org/10.1038/s41598-024-56775-y

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., & Al-Shamma'a, A. (2022). Water quality classification using machine learning algorithms. *Journal of Water Process Engineering, 48*, 102920.

Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications, 83*(12), 35307-35334. https://doi.org/10.1007/s11042-023-16737-4

Xizhi Nong, Y. H. (2025). Machine learning-based evolution of water quality prediction model: An integrated robust framework for comparative application on periodic return and jitter data. *Environmental Pollution*, 125834. https://doi.org/10.1016/j.envpol.2025.125834