Journal of Smart Agriculture and Environmental Technology
e-ISSN:3021-8802
Vol. 4, No. 1, April 2026
https://josaet.com | https://doi.org/10.60105/josaet.2026.4.1.1-6

Research Paper

Machine Learning-Driven Smart Aquaculture Technology for Climate-Resilient Water Quality Monitoring

Moses Kazeem Abiodun[1]*, Adesina Babatunde[2], Onashoga S. Adebukola[3], Eniola Kehinde[4], Kehinde Adeniyi[1], Rasheed Babatomiwa Idris[5]

[1] *Department of Computer Science, Bowen University, Iwo, Osun State*
[2] *Department of Agriculture, Landmark University, Omu-Aran, Kwara State, Nigeria*
[3] *Department of Cybersecurity and Data Science, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria*
[4] *I.T Department of Microbiology, Kogi State University, Kabba, Kogi State, Nigeria*
[5] *Department of Computer Science, Landmark University, Omu-Aran, Kwara State, Nigeria*
***Corresponding author**: moses.abiodun@bowen.edu.ng

Article History: Received: November 21, 2025, Accepted: March 11, 2026

Abstract
The aquaculture sector is among the fastest-growing food production industries, playing a critical role in global food security by supplying high-quality protein to millions. However, climate change has introduced severe challenges, disrupting production through altered temperature regimes, unpredictable rainfall, and deteriorating water quality. Key parameters such as pH, dissolved oxygen (DO), and salinity have shown significant fluctuations, which are directly affecting fish health, growth rates, reproduction, and overall pond productivity. To address these challenges, this study proposes an integrated IoT and machine learning (ML) framework designed for real-time water quality monitoring and adaptive management in aquaculture systems. The primary objective is to enhance climate resilience by enabling data-driven decision-making for optimal fish health and production efficiency. A comprehensive dataset was sourced from reputable offline and online repositories, and then partitioned into training (90%) and testing (10%) subsets. Water quality was classified into two categories: "good" and "bad" based on critical thresholds for aquaculture sustainability. Four supervised machine learning algorithms were evaluated for classification performance, including Random Forest (RF) with an accuracy of 100%, demonstrating superior predictive capability, and Logistic Regression (LR) with an accuracy of 57%, indicating moderate performance, Support Vector Machine (SVM) yielded an accuracy of 62%, suitable for certain nonlinear patterns, and Naive Bayes (NB) attained 89% accuracy, offering a balance between speed and reliability. This research paves the way for next-generation smart aquaculture systems, bridging the gap between environmental monitoring and AI-based decision support.

Keywords
Water Quality; Aquaculture; Climate; Machine Learning, Fish Farming

1. INTRODUCTION

The aquaculture industry is one of the fastest-growing food-producing sectors worldwide, promoting household food security by supplying protein to the global population (FAO, 2022, 2024). However, global aquaculture production is experiencing serious challenges due to the increasing impacts of climate change in recent years (Maulu et al., 2021; Ahmed et al., 2019). Changes in temperature and rainfall patterns have affected water quality parameters such as pH, salinity, and oxygen, as well as fish production, which is expected to impact reproduction, growth, survival, and pond productivity in aquaculture production. These environmental changes are a major concern in aquaculture development in recent years; it has disturbed the creation of new job opportunities for farm technicians and skilled labourers (Mehrim and Refaey, 2023).

Recent technological advancements in aquaculture have led to the development of smart monitoring systems that make it easier to tackle climate challenges (Misra, 2022). Over the past few years, breakthroughs such as the Internet of Things (IoT) have been transforming the way we could approach traditional farming (Aquaculture systems that are AIoT-based can simultaneously monitor water quality, microclimate, and provide warning functions while enhancing early warnings and response times (Chiu et al., 2022). These advancements assisted fish farmers to effectively monitor their operations by providing a vast coverage of data from numerous locations, which allow for real-time remedial steps to be implemented (Chiu et al., 2022).

Although these smart solutions exist, a lot of fish farmers don't have access to such resources, particularly, in developing countries like Nigeria. They are used to more traditional approaches and lack the skill to operate smart solutions. These fish farmers face the challenge of constant monitoring of the water supply and water changing capacity in such a way that water quality is compromised regularly (Karim et al., 2021). Understanding how intelligent aquaculture technology work can reduce the impacts of climate change on water quality for ensuring the future of sustainable fish farming in the developing world (Li and Liu, 2020).

This study explores how smart aquaculture systems, powered by machine learning, can be used to monitor water quality, especially in the face of climate change. The aim is to see how these technologies can help boost fish production in areas affected by shifting environmental conditions. This research matters because it tackles both the environmental and economic challenges fish farmers face, and offers practical ways to build resilience and support sustainable aquaculture in the long term.

1.1 The Role of Water Quality in Aquaculture

Aquaculture has become a necessity worldwide due to the growing demand for fish and seafood. The vast majority of traditional aquaculture systems, though, employ labour-intensive procedures for monitoring water quality, such as measuring temperature, pH, and dissolved oxygen manually. While they have worked in the past, they are time-consuming, error-prone, and often unable to react to trouble in real time. This means problems like water quality go undetected until they cause extensive harm, such as fish health deterioration or production losses. On top of this, climate change is also making it more difficult for farmers to manage these systems effectively, with unpredictable fluctuations in the quality of water happening more and more often. It's clear that traditional methods alone are no longer enough to meet the challenges of modern aquaculture.

Economically, aquaculture contributes significantly to livelihoods, especially in coastal and rural communities. It supports millions of jobs globally, from farming operations to supply chains, and helps ensure food security for a growing population. By offering a sustainable method of seafood production, aquaculture bridges the gap between rising protein demands and environmental conservation.

Water quality is a natural parameter for aquaculture success since it directly influences the health, growth, and productivity of the aquatic animal. The optimal water conditions for the majority of species are the correct levels of dissolved oxygen, pH, temperature, and reduced amounts of toxins like ammonia and nitrites. The parameters must be maintained in species-specific ranges of optimum for proper growth and absence of disease outbreaks (Boyd et al., 2020; Hussenot et al., 2023).

Dissolved oxygen (DO) is the most important parameter as it is used for respiration by fish and other aquatic organisms (Rahman et al., 2023). Deficiency of DO results in hypoxia, causing stress, reduced growth rate, and vulnerability to disease (Boyd et al., 2020). DO needs to be monitored and controlled from time to time so that oxygen remains within the tolerable optimal range for the cultured species (Hussenot et al., 2023).

Ammonia, particularly its non-ionized form, is toxic to aquatic animals even in very small quantities (Rahman et al., 2023). It is primarily due to the excretion of waste products by fish and decomposition of unpicked food (Xie et al., 2020). There should be proper waste management measures, such as appropriate feeding practices and regular removal of organic matter, for the removal of ammonia accumulation and water quality maintenance (Hussenot et al., 2023).

Water pH regulates the toxicity and solubility of most chemicals, including ammonia (Rahman et al., 2023). pH variation to promote higher or lower values can trigger fish physiological stress, reducing their immune status and growth (Boyd et al., 2020). Constant or stable pH in the range optimal for the cultured species is required for health and production (Xie et al., 2020).

Temperature impacts the metabolic rate, growth, and reproduction of aquatic animals (Xie et al., 2020). A deviation from the ideal range of temperatures has been shown to result in stress, inhibited growth, and disease susceptibility (Rahman et al., 2023). Using temperature control measures such as shading and aeration is able to obtain the required water temperatures to support effective aquaculture production (Hussenot et al., 2023).

Good water quality is essential to aquaculture system productivity and health. Monitoring and control of critical water quality parameters like dissolved oxygen, ammonia, pH, and temperature must be carried out continuously to ensure a suitable environment for aquatic animals (Boyd et al., 2020). Management of good water quality is not only focused on enhancing fish production and health but also on creating aquaculture farms as profitable and sustainable business (Rahman et al., 2023).

2. EXPERIMENTAL SECTION

2.1 System Analysis and Design Strategy

The project adopts a modular and layered approach to system design, leveraging object-oriented programming, simulation-based input processing, and visualization tools. The architecture consists of four primary layers:

i. Sensor Layer (simulated inputs),
ii. Processing Layer (data cleaning and prediction),
iii. Decision Layer (recommendations),
iv. Interface Layer (Dashboard).

The system design strategy was driven by principles of modularity, scalability, and real-time responsiveness. Each module is developed independently to ensure easy extension in the future for instance, replacing the simulated inputs with live IoT sensor feeds. The methodology follows the Waterfall Model, progressing sequentially from

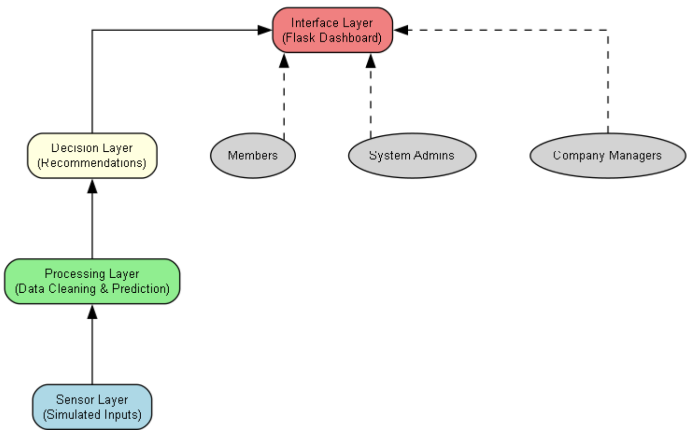

Figure 1. System Architecture Diagram

	Timestamp	Temp	pH	DO	Turbidity	Ammonia
0	2025-01-01 00:00:00	26.857729	6.721152	5.614082	3.592670	0.309487
1	2025-01-01 01:00:00	25.662372	6.833792	6.447713	4.065397	0.160202
2	2025-01-01 02:00:00	27.175901	6.587400	6.741560	3.313212	0.300093
3	2025-01-01 03:00:00	29.068143	6.843936	7.053561	4.039257	0.413428
4	2025-01-01 04:00:00	25.771129	6.053988	5.381207	4.781631	0.593049

	Quality	Aerator	Heater	pH_Regulator
0	1	OFF	OFF	OFF
1	1	OFF	OFF	OFF
2	1	OFF	OFF	OFF
3	0	OFF	OFF	OFF
4	0	OFF	OFF	ADD_BASE

Figure 2. Data Collection and Unpacking

requirements gathering to implementation and evaluation as shown in Figure 1. This was selected for clarity and structure, appropriate for a simulation-driven academic prototype. Additionally, Python was chosen as the primary development language due to its support for machine learning, simulation, and dashboarding through libraries like scikit-learn, matplotlib, and Flask.

2.2 Implementation and Results

Data Collection and Unpacking

To create the illusion of real-time aquaculture monitoring without live sensor feeds, synthetic data was generated to emulate readings fairly closely to those typically collected by IoT sensors in an aquatic environment. The dataset included the following water quality parameters relevant for aquaculture management: Temperature (Temp), pH, Dissolved Oxygen (DO), Turbidity, Ammonia, and Timestamp - provided in hourly increments from a designated start date. Each parameter was sampled from a normal distribution, which ensured that the synthetic data reflected a reasonable range and variation found in real aquaculture conditions. In order to achieve a more realistic dataset, random noise was also added to represent sensor error and natural variation to simulate a real-time monitoring system. One important attribute added to the dataset was the "Quality" label for use as a target variable in machine learning classification.

```
df.head()
✓ 0.0s
```

	Timestamp	Temp	pH	DO	Turbidity	Ammonia	Quality	Aerator	Heater	pH_Regulator
0	2025-01-01 00:00:00	26.857729	6.721152	5.614082	3.592670	0.309487	1	OFF	OFF	OFF
1	2025-01-01 01:00:00	25.662372	6.833792	6.447713	4.065397	0.160202	1	OFF	OFF	OFF
2	2025-01-01 02:00:00	27.175901	6.587400	6.741560	3.313212	0.300093	1	OFF	OFF	OFF
3	2025-01-01 03:00:00	29.068143	6.843936	7.053561	4.039257	0.413428	0	OFF	OFF	OFF
4	2025-01-01 04:00:00	25.771129	6.053988	5.381207	4.781631	0.593049	0	OFF	OFF	ADD_BASE

Figure 3. Data Sample after Cleaning

The target label was binary, signified with either Good (1) or Poor (0) water quality. A label of 1 indicates that all the parameters were within acceptable limits, whereas a label of 0 indicates that at least one parameter was outside the acceptable operational limit. This labeled dataset provided a controlled but real environment to be used for training and evaluation of model performance. See Figure 2.

Data Cleaning

As part of the preprocessing of the study's smart aquaculture monitoring system, we wanted to demonstrate a pre-processing approach to express cleaning and transforming the data into a suitable dataset for machine learning modelling and simulation. Although the synthetic data generation was a controlled process and any anomalies were limited during the data construction process and concrete steps taken to mimic a realistic pre-processing pipeline, there were still progressive steps taken to add realism when relevant. In this case, outliers - which exist in all sensor data in the real world - were accounted for. In this case, we were introduced to controlled noise on the extreme values; this allowed us to maintain some level of founded variability as well as minimize the chance of distortion in the model. Moreover, as the dataset was synthetically constructed, our instance was not burdened by missing data. However, in a true-world situation, missing data can be accomplished with interpolation where values are estimated between two known values, or through mean imputation where missing entries would be taken from the averaged observed values. Also at the feature engineering stage, we also simulated real-time actuator responses based on the sensor data. The sample data after cleaning is displayed in Figure 3. Three key actuator features were introduced: the Aerator, triggered by Dissolved Oxygen (DO) levels; the Heater, activated in response to Temperature readings; and the pH_Regulator, engaged based on pH values. These engineered features enriched the dataset by mimicking how a real aquaculture system would autonomously respond to changing environmental conditions, supporting both monitoring and decision-making functionalities.

Inputs

In this study, we carefully selected the input features to train and test machine learning models that encompassed the most important input parameters influencing water quality in aquaculture systems. The input features included Tem-

Temp	pH	DO	Turbidity	Ammonia
26.857729	6.721152	5.614082	3.592670	0.309487
25.662372	6.833792	6.447713	4.065397	0.160202
27.175901	6.587400	6.741560	3.313212	0.300093
29.068143	6.843936	7.053561	4.039257	0.413428
25.771129	6.053988	5.381207	4.781631	0.593049

Figure 4. Input Features

perature (Temp), pH, Dissolved Oxygen (DO), Turbidity, and Ammonia as shown in Figure 4. Each of these variables is vital to determining the health and stability of the aquatic environment, and thus important for real-time monitoring and prediction. The values of each of these features from the synthetic dataset were fed into various machine learning classification models as input features. By evaluating the patterns and relationships in these inputs, the machine learning models classified overall water quality into two pre-determined classifications (Good or Poor-quality water). The selection of these input features represented not only practical significance in aquaculture systems but was also meaningful with regard to actual scientific data, which the machine learning algorithms could classify. The input parameters either represented ordinal or categorical values so that various classifiers could be used to compare and evaluate the model's performance in accurately classifying the quality of water samples, and five supervised learning algorithms were tested for high performance for classifying the quality of water samples, such as logistic regression, decision trees, random forests, and support vector machines. In this way we created an input-output format to support the development of an intelligent system to help guide decision-making for farmers and aquaculture managers for implementing the best data-focused decisions to preserve optimal water quality.

Processing / Simulations

To evaluate the effectiveness of predictive modeling in the smart aquaculture monitoring system, four machine learning algorithms were selected and rigorously trained and tested on the prepared dataset. The classification algorithms utilized were Random Forest, Logistic Regression, Support Vector Machine (SVM), and Naive Bayes each for their capacity in classifying tasks. The algorithms were optimized to recognize patterns in the environmental data and classify water quality as either Good or Poor. The data set was split into two partitions: 90% was used for training the models and 10% for testing the generalization performance. The learning process involved using fundamental functions such as *train_test_split()* to separate the data into training and testing sets, then feeding each classifier its respective training set with *model.fit()*. predictions were done with *model.predict()*, and confidence values derived using either *predict_proba()* or *decision_function()* depending on the algorithm. Performance of the model was tested by plotting Receiver Operating Characteristic (ROC) curves with *roc_curve()*, and calculating the Area Under the Curve (AUC) with *auc()*. These measures provided some insight into the performance of each model in discriminating between the two classes. The simulation phase thus enabled a valid comparison of several machine learning approaches within an actual real-world aquaculture context. See Figure 5.

Table 1. Performance of all models

Model	Accuracy	Precission	Recall	F1 Score	ROC AUC
Random Forest	1.000	1.000	1.000	1.000	1.000
Logistic Regression	0.565	0.292	0.091	0.139	0.609
SVM	0.615	0.000	0.000	0.000	0.838
Naive Bayes	0.890	0.951	0.753	0.841	0.965

3. RESULTS AND DISCUSSION

3.1 Results

All four models that we used, Random Forest, Logistic Regression, Support Vector Machine (SVM), and Naive Bayes, output the water quality class either as 0 (Poor) or 1 (Good). For comparing the prediction accuracy of these machine Learning algorithms, the majority of classification metrics were employed: accuracy, precision, recall, F1 score, confusion matrix, Receiver Operating Characteristic (ROC) curve and Area Under the ROC Curve (AUC) as shown in Table 1. Table 1 show how well each of the models could distinguish good water quality from bad water quality conditions. The Random Forest classifier had optimal performance on all the measures in this test, with all the measures registering perfect scores (1.000), meaning it was capable of classifying the synthetic data perfectly. Naive Bayes performed equally well, with good precision (0.951) and recall (0.753), resulting in a very robust F1 measure and good ROC AUC of 0.965.

The SVM did not classify at all, with zero precision, recall, and F1, but did have a respectable ROC AUC measurement of 0.838, indicating the model was highly sensitive to score thresholds but poor at end classification. The Logistic Regression didn't perform well in general, with a poor recall rate (0.091), and both the F1 score and general validity in this test were significantly compromised.

Figure 6 shows the confusion matrix for each of the models used. Random Forest was able to classify the quality of water as either good or bad. While Naïve Bayes and

```python
# Step 6: Machine Learning preparation
X = df[['Temp', 'pH', 'DO', 'Turbidity', 'Ammonia']]
y = df['Quality']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42)

# Step 7: Train models
models = {
    "Random Forest": RandomForestClassifier(n_estimators=100, random_state=42),
    "Logistic Regression": LogisticRegression(max_iter=1000, random_state=42),
    "SVM": SVC(kernel='rbf', probability=True, random_state=42),
    "Naive Bayes": GaussianNB()
}

results = {}
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    y_prob = model.predict_proba(X_test)[:, 1] if hasattr(model, "predict_proba") else model.decision_function(X_test)
    fpr, tpr, _ = roc_curve(y_test, y_prob)
    roc_auc = auc(fpr, tpr)
    
    results[name] = {
        "Accuracy": accuracy_score(y_test, y_pred),
        "Precision": precision_score(y_test, y_pred),
        "Recall": recall_score(y_test, y_pred),
        "F1 Score": f1_score(y_test, y_pred),
        "Classification Report": classification_report(y_test, y_pred),
        "Confusion Matrix": confusion_matrix(y_test, y_pred),
        "FPR": fpr,
        "TPR": tpr,
```

Figure 5. Processing Steps

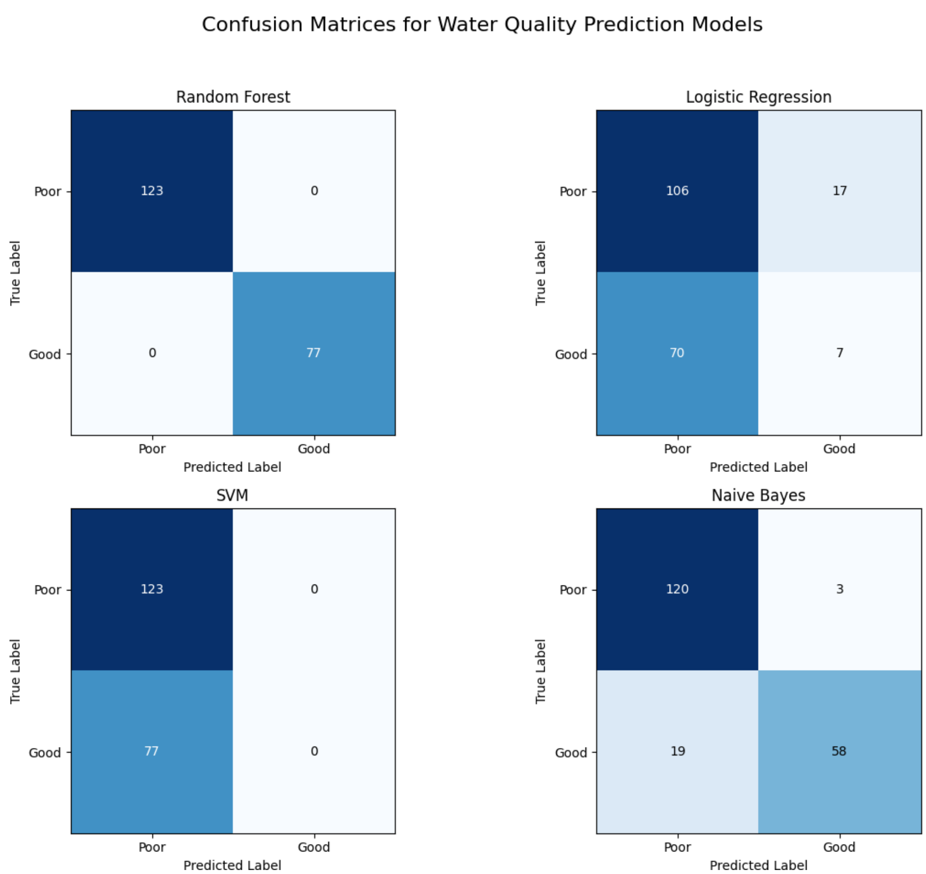

Figure 6. Confusion Matrix of all models

Logistic regression miss the classification of some of the data. SVM perform badly in the classification process.

3.2 Discussion of Results

The performance of the evaluation emphasizes the effectiveness of machine learning in particular, ensemble models like Random Forest in handling high accuracy and credible synthetic aquaculture data. Random Forest outperformed all the other models and delivered flawless classification results. This reflects that the model is highly competent in detecting complex patterns and connections between environmental variables and has high generalization ability even in new data. The implications of such findings are significant for the design of intelligent aquaculture systems. A Random Forest-based model-driven approach enables accurate real-time forecasting of water quality, which can in turn inform autonomous control decisions such as initiating an aerator when dissolved oxygen dips or pH balancing through chemical management. This capability is critical, especially in negating the effect of climate-induced variability in water conditions. The findings explicitly address the shortcomings highlighted in the original problem statement: the ineffectiveness and inefficiency of manual monitoring procedures. By using automation and intelligent forecasting, the study presents evidence for an extensible, data-driven solution that enhances aquaculture operations' sustainability and resilience. Looking ahead, transitioning from simulated environments to real-world deployment involves integrating live IoT sensor data, expanding classification to include more granular quality levels (e.g., Poor, Moderate, Good), and deploying models on low-cost edge devices for accessibility in rural or remote areas.

4. CONCLUSION AND SUGGESTION

This research successfully designed and evaluated a machine learning-enhanced simulation framework for smart aquaculture water quality monitoring. The system used synthetic sensor data and machine learning algorithms to classify water quality and generate actuator recommendations. This work contributes meaningfully toward offering a climate-resilient digital alternative to traditional fish farming practices. This research paves the way for next-generation smart aquaculture systems, bridging the gap between environmental monitoring and AI-based decision support. The future direction of this study will introduce a real-life sensor monitoring system that will enhance water quality in any environment.

ACKNOWLEDGEMENT

We would like to thank our various Universities for giving us an enabling environment to conduct this research work. A colloboration of different researcher from different fields of studies.

Conflict of Interest and Ethics Statements

There is no conflict of interest, and no ethics were violated.

REFERENCES

Ahmed, N., S. Thompson, and M. Glaser (2019). Global aquaculture productivity, environmental sustainabisslity, and climate change adaptability. *Environmental management*, **63**(2); 159–172

Boyd, C. E., C. S. Tucker, A. A. McNevin, K. Bostick, and J. Clay (2020). Indicators of resource use efficiency and environmental performance in fish and crustacean aquaculture. *Reviews in Aquaculture*, **12**(3); 1250–1263. https://doi.org/10.1111/raq.12985

Chiu, M., W. Yan, S. A. Bhat, and N. Huang (2022). Development of smart aquaculture farm management system using IoT and AI-based surrogate models. *Journal of Agriculture and Food Research*, **9**; 100357. https://doi.org/10.1016/j.jafr.2022.100357

FAO (2022). *The State of World Fisheries and Aquaculture 2022*. Food and Agriculture Organization of the United Nations, Rome

FAO (2024). *The State of Food and Aquaculture 2024*. Food and Agriculture Organization of the United Nations, Rome

Hussenot, J. M., G. Benoit, and G. Barroin (2023). Advances in water quality monitoring for aquaculture: Innovative tools and techniques. *Heliyon*, **9**(3); e16753. https://doi.org/10.1016/j.heliyon.2023.e16753

Karim, S., I. Hussain, A. Hussain, K. Hassan, and S. Iqbal (2021). IoT Based Smart Fish Farming Aquaculture Monitoring System. *International Journal on Emerging Technologies*, **12**(2); 45–53

Li, D. and C. Liu (2020). Recent advances and future outlook for artificial intelligence in aquaculture. *Smart Agriculture*, **2**(3); 1–20. https://doi.org/10.12133/j.smartag.2020.2.3.202004-SA007

Maulu, S., O. J. Hasimuna, L. H. Haambiya, C. Monde, C. G. Musuka, T. H. Makorwa, and J. D. Nsekanabo (2021). Climate change effects on aquaculture production: sustainability implications, mitigation, and adaptations. *Frontiers in Sustainable Food Systems*, **5**; 609097

Mehrim, A. I. and M. M. Refaey (2023). An Overview of the Implication of Climate Change on Fish Farming in Egypt. *Sustainability*, **15**(2); 1679

Misra (2022)

Rahman, M. A., A. Shahzad, and S. Ahmed (2023). Challenges and opportunities in aquaculture water quality management. *International Journal of Environmental Research and Public Health*, **20**(1); 1446. https://doi.org/10.3390/ijerph17041446

Xie, Y., X. Zhang, and H. Li (2020). The importance of water quality parameters in aquaculture systems. *Aquaculture Journal*, **2**(1); 3–15. https://doi.org/10.3390/aquacj2010003