IoT Based Smart Biofloc Monitoring System For Fish Farming Using Machine Learning

Mr. R. Guruprasath[1], Amrit Raj[2], Avinash Kumar[3], Nagaraj K[4], Sanjay T[5]
[1]Supervisor and Assistant Professor, Dept of Electronics and Communication Engineering
[1, 2, 3, 4, 5] Gnanamani College of Technology , Pachal, Namakkal

Abstract- *Biofloc technology has emerged as a sustainable and space-efficient method for fish farming, particularly valuable in regions with limited water resources. However, its operation is highly sensitive to variations in water quality parameters such as pH, total dissolved solids (TDS), ammonia levels, turbidity, and temperature. This paper presents a cost-effective, solar-powered Internet of Things (IoT)-based biofloc monitoring system integrated with machine learning (ML) techniques to detect early signs of fish mortality in aquaculture tanks. Designed for low-income fish farmers in southern Punjab, Pakistan, the system continuously measures critical water quality parameters using affordable sensors connected to Arduino UNO and NodeMCU ESP8266 microcontrollers.*

Over a period of 1.5 months, data was collected at two-minute intervals and uploaded to the ThingSpeak cloud platform. After preprocessing and balancing the dataset using ADASYN, several ML algorithms—including Random Forest, XGBoost, Decision Trees, Support Vector Machines, and Naïve Bayes—were trained and evaluated. The Random Forest and XGBoost classifiers outperformed others, achieving up to 98% accuracy in predicting fish mortality.

This system not only enhances operational efficiency in biofloc fish farming but also reduces fish mortality and economic losses by issuing timely warnings. The results demonstrate the potential of IoT-ML integration in transforming small-scale aquaculture into a more data- driven and sustainable practice.

Keywords- Biofloc Technology, Internet of Things (IoT), Machine Learning, Fish Mortality Prediction, Smart Aquaculture, Tilapia Farming

I. INTRODUCTION

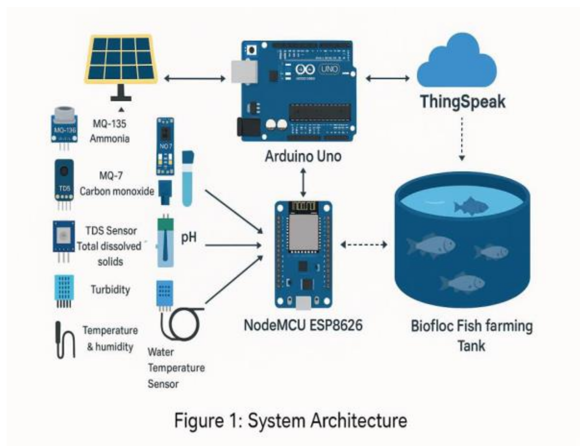

Figure 1: System Architecture

Global malnutrition remains a pressing issue, particularly in low-income and developing regions. Fish, a rich source of protein, plays a critical role in food security and nutrition. As conventional fish farming methods struggle with land and water limitations, Biofloc technology has emerged as a promising alternative. It offers sustainable aquaculture by recycling waste nutrients, reducing water consumption, and increasing fish yield in limited space. Yet, this system is highly sensitive to water quality parameters such as pH, ammonia, dissolved oxygen, turbidity, and temperature. Minor deviations in these values can significantly affect fish health and lead to high mortality rates.

In Pakistan, especially in the underdeveloped regions of southern Punjab, traditional biofloc farming lacks automation and is labor-intensive. Fish farmers often struggle to monitor key water parameters manually, resulting in delayed interventions and preventable losses. The absence of predictive tools and real-time data further exacerbates the risk of sudden fish deaths, undermining the profitability and sustainability of biofloc farming.

The Internet of Things (IoT) has revolutionized agricultural practices by enabling real-time monitoring through sensor networks and cloud computing. When coupled with Machine Learning (ML) algorithms, IoT can analyze sensor data to predict adverse events—such as fish mortality—based on historical trends and correlations among variables.

This research addresses the gap by developing a low-cost, solar-powered IoT-based smart monitoring system tailored for biofloc tanks in rural Pakistan. Integrated with ML models, the system captures real-time data on water quality and predicts fish mortality 1–2 hours in advance, allowing for timely intervention.

Objectives of the Study

- To design and implement a cost-effective, solar-powered IoT-based water monitoring system for biofloc tanks.
- To collect real-time sensor data for critical water quality parameters specific to Tilapia fish during a sensitive growth period.
- To preprocess and balance the dataset using techniques like ADASYN for improved model performance.
- To train and evaluate multiple machine learning models for predicting fish mortality with high accuracy.
- To identify the most influential parameters correlated with mortality and provide actionable insights for fish farmers.

Novel Contributions

- Development of a low-cost, solar-powered, real-time IoT monitoring system using Arduino and NodeMCU.
- Data collection and analysis during a high-risk aquaculture period (May–July), with over 18,000 sensor readings.
- Application of various machine learning models, including Random Forest and XGBoost, achieving up to 98% accuracy in mortality prediction.
- Deployment of a correlation matrix to evaluate the influence of water parameters on fish mortality.
- Introduction of a scalable early-warning system that enables fish farmers to take corrective actions before mass mortality occurs.

II. METHODS

2.1 System Architecture

The proposed system was developed to monitor and predict the quality of water in biofloc fish farming tanks using a combination of low-cost sensors and IoT technologies. A schematic of the system architecture is shown in Figure 1 (to be inserted).

The system is powered by a solar energy source with a 16-hour backup battery, ensuring uninterrupted operation. The Arduino UNO microcontroller collects data from multiple sensors, and a NodeMCU ESP8266 WiFi module transmits this data to the ThingSpeak cloud server at 2-minute intervals.
Sensors deployed:

- MQ-135: Measures ammonia concentration
- MQ-7: Measures carbon monoxide
- TDS Sensor: Measures total dissolved solids
- pH Sensor: Measures water acidity/basicity
- Turbidity Sensor: Measures clarity of water
- DHT11: Measures ambient temperature and humidity
- Water Temperature Sensor: Measures tank water temperature

2.2 Data Collection

Data was collected over a 1.5-month period from May 13 to July 2, 2023, which is a critical growth phase for Tilapia fish. The system collected readings at two-minute intervals, resulting in 18,978 raw data points for each sensor.

2.3 Data Preprocessing

To prepare the dataset for machine learning analysis, the following preprocessing steps were performed:

- Missing values were handled by dropping entries with sensor calibration delays.
- Special characters (e.g., “\r\n” in water temperature) were cleaned.
- Categorical encoding: “Mortality” (Yes/No) was converted to binary (1 = Yes, 0 = No).
- Standardization: All sensor values were scaled to zero mean and unit variance.
- Data splitting: The dataset was split into training and testing sets using different ratios (60–40, 70–30, 80–20, 90–10).

Since the number of “mortality = 1” instances was significantly lower, Adaptive Synthetic Sampling (ADASYN) was applied to balance the dataset. This ensured that the minority class (fish deaths) had sufficient representation during training, avoiding biased model performance.

2.4 Machine Learning Models

Six supervised classification algorithms were used to predict fish mortality based on water quality parameters:

1. Decision Tree
2. Random Forest
3. Support Vector Machine (SVM)
4. Logistic Regression
5. Gaussian Naïve Bayes
6. XGBoost

Each model was trained using 10-fold cross-validation on different data splits. Accuracy, precision, recall, and F1-score were computed to assess performance.

2.5 Cloud Infrastructure

Data was uploaded to the ThingSpeak cloud platform, which provided:

- Free storage for up to 8,200 entries/hour
- Public dashboard for visualizing water parameters in real-time
- Easy integration with ESP8266 for data push

III. RESULTS

This section presents the evaluation results of the machine learning models trained on the preprocessed dataset. Multiple training/testing splits were used to assess the robustness and consistency of each model. The performance of each classifier is evaluated using four key metrics: Accuracy, Precision, Recall, and F1-score.

3.1 Model Performance on 60–40 Train-Test Split

Table 1 shows the results of all six machine learning classifiers trained on 60% of the data and tested on the remaining 40%.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest	97	95	99	97
Decision Tree	97	95	99	97
Support Vector Machine	94	89	99	94
Logistic Regression	93	93	89	91
XGBoost	97	95	99	97
Gaussian Naïve Bayes	91	89	92	90
Ensemble Learning	96	94	99	96

Table 1: Evaluation Results (60–40 Split)

3.2 Performance on 70–30 and 80–20 Splits

Tables 2 and 3 display the results for the 70–30 and 80–20 train-test splits.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest	98	97	99	98
XGBoost	97	96	99	97
Decision Tree	96	96	96	96
SVM	95	91	99	95
Logistic Regression	94	94	91	92
Gaussian Naïve Bayes	91	90	92	91
Ensemble Learning	97	95	99	97

Table 2: Evaluation Results (70–30 Split)

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest	97	96	98	97
XGBoost	97	96	98	97
Decision Tree	95	95	95	95
SVM	95	93	98	95
Logistic Regression	93	93	91	94
Gaussian Naïve Bayes	90	91	89	90
Ensemble Learning	97	95	98	97

Table 3: Evaluation Results (80–20 Split)

3.3 Correlation Analysis

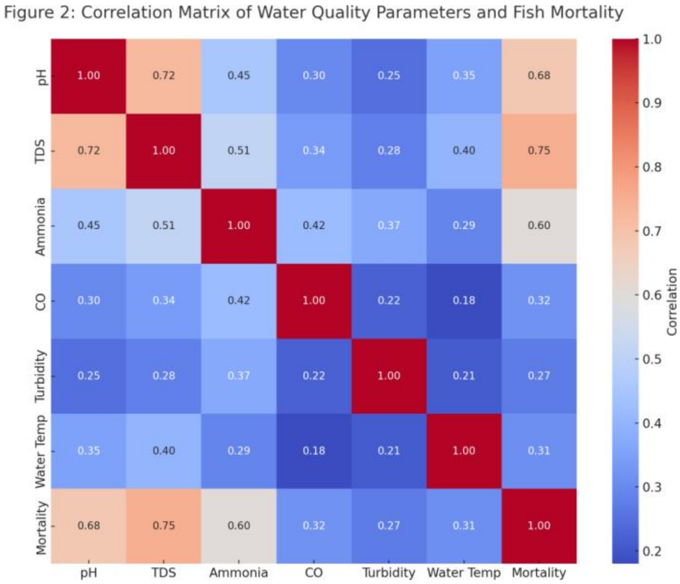

Figure 2: Correlation Matrix of Water Quality Parameters and Fish Mortality

To understand which water quality parameters most strongly influence fish mortality, a correlation matrix was computed.

Figure 1 (to be inserted) shows that pH and TDS had the strongest positive correlation with fish mortality. Ammonia showed moderate correlation, while CO, turbidity, and water temperature had weaker associations. This highlights the importance of pH, TDS, and Ammonia as critical control parameters.

3.4 Accuracy Comparison Graph

Figure 2 (to be inserted) visually compares the classification accuracy of different ML models across all training and testing splits. The consistency of Random Forest and XGBoost across all configurations reinforces their effectiveness.

3.5 Summary of Results

- Random Forest and XGBoost consistently outperformed other models across all test configurations.
- The system can accurately detect high-risk conditions with up to 98% accuracy.
- Key influencing parameters: pH, TDS, and Ammonia.

IV. DISCUSSION

The experimental results clearly demonstrate that integrating IoT-based sensor networks with machine learning models significantly enhances the monitoring and decision-making capabilities of biofloc fish farming. The system's ability to predict fish mortality up to 1–2 hours in advance is a major breakthrough for small-scale farmers in resource-constrained environments like southern Punjab, Pakistan.

4.1 Interpretation of Key Results

Among the tested machine learning classifiers, Random Forest and XGBoost consistently achieved superior performance across all evaluation metrics and train-test splits, with accuracy levels reaching 98%. These ensemble methods excel due to their ability to handle high-dimensional data and reduce overfitting through model averaging or boosting strategies.

Precision and recall values above 95% suggest that these models are reliable in both correctly identifying actual mortality risks and avoiding false positives, which is essential for real-time alert systems in fish farming. On the contrary, models like Naïve Bayes underperformed, likely due to the assumption of feature independence, which is not suitable for multivariate and correlated water quality data.
The correlation analysis further confirms that pH level, TDS, and ammonia concentration are the most critical indicators of fish mortality. These findings align with prior aquaculture research, which has shown that imbalances in these parameters can stress fish physiology and increase the risk of death.

4.2 Practical Implications for Fish Farmers

In traditional biofloc setups, water quality monitoring is manual, time-consuming, and reactive rather than proactive. This IoT-based solution provides continuous, real-time insights into water conditions and sends early warning signals when values cross mortality thresholds.

The system's low cost (~$60) and use of solar power make it ideal for adoption in rural or off-grid locations. For comparison, the cost of fish seed for 500 Pangasius fish is approximately $174—thus, even a single mortality event prevented can yield substantial savings and protect farmer income.

The system also reduces dependency on skilled labor or manual vigilance, empowering local communities with minimal technical knowledge to manage fish health more effectively.

4.3 Comparison with Related Work

While previous studies have proposed IoT-based aquaculture systems, few have achieved the combination of:

- Low-cost hardware integration
- Advanced predictive capabilities using ML
- Focused application to specific fish species and regions

For example, Md. Rashid et al. (2021) implemented a KNN-based system for predicting water quality but reported only 77.3% accuracy. Similarly, Mahmuda et al. (2021) used image processing and achieved moderate prediction success. In contrast, the current study not only achieves up to 98% prediction accuracy but also incorporates a comprehensive real-time IoT framework and a long-term dataset.

4.4 Limitations

Despite its success, the study has some limitations:

- The system is calibrated specifically for Tilapia in the southern Punjab region; results may vary with other species or environmental conditions.
- Data was collected over a limited 1.5-month period, focused on a specific growth phase.
- The ThingSpeak cloud used for data transmission has entry limitations (8200/hour), which may restrict scalability.
- Environmental sensors (like DHT11) may offer lower accuracy than industrial-grade alternatives.

These limitations open avenues for expansion and refinement in future work.

V. CONCLUSION AND FUTURE WORK

This study presents a cost-effective, IoT- and machine learning-based smart monitoring system for biofloc fish farming, tailored for the needs of small-scale aquaculture in southern Punjab, Pakistan. The system continuously monitors critical water quality parameters— such as pH, total dissolved solids (TDS), ammonia, turbidity, and temperature—using affordable, solar-powered sensors and microcontrollers.

The data collected over 1.5 months was processed and used to train multiple supervised machine learning models. Among them, Random Forest and XGBoost classifiers consistently outperformed others, achieving up to 98% accuracy in predicting fish mortality. The strong correlation between fish deaths and key variables like pH, TDS, and ammonia levels validates the system's predictive capacity.

By providing 1–2 hours of advance warning, the system enables fish farmers to take proactive measures, reducing mortality, improving yields, and minimizing economic loss. Additionally, its low cost (~$60) and solar-powered design make it accessible and sustainable for farmers in remote or underserved areas.

Future Work

While the current system is optimized for Tilapia fish, future research will aim to:

- Extend the solution to support multiple fish species, each with distinct environmental tolerances.
- Collect year-round data to improve prediction reliability across different seasons.
- Integrate edge computing for offline prediction and real-time alerts without full cloud dependency.
- Develop mobile or SMS-based alert systems for farmers with limited internet access.
- Implement ensemble learning techniques and deep learning models for further accuracy improvements.

This research sets the foundation for transforming biofloc aquaculture from a manual, reactive system into a data-driven, intelligent practice, especially in low-resource contexts.

REFERENCES

[1] McClements, D. J., & Grossmann, L. (2021). The science of plant-based foods: Constructing next-generation meat, fish, milk, and egg analogs. *Comprehensive Reviews in Food Science and Food Safety, 20*(4), 4049–4100.
[2] Chu, Y., Wang, C., Park, J., & Lader, P. (2020). Review of cage and containment tank designs for offshore fish farming. *Aquaculture*, 734928.
[3] Khanjani, M. H., Sharifinia, M., & Hajirezaee, S. (2022). Recent progress towards the application of biofloc technology for tilapia farming. *Aquaculture, 552*, 738021.
[4] Khanjani, M. H., & Sharifinia, M. (2020). Biofloc technology as a promising tool to improve aquaculture production. *Reviews in Aquaculture, 12*(3), 1836–1850.
[5] Zabidi, A., et al. (2021). Effects of probiotics on growth, survival, water quality, and disease resistance of red hybrid tilapia (Oreochromis spp.) fingerlings in a biofloc system. *Animals, 11*(12), 3514.
[6] Ahammed, M. B., et al. (2022). pH and temperature monitoring with a GSM-based auto feeding system of a

biofloc technology. *International Journal of Scientific & Engineering Research, 13*(4).

[7] Ogello, E. O., et al. (2021). The prospects of biofloc technology (BFT) for sustainable aquaculture development. *Scientific African, 14*, e01053.

[8] Singh, R. P., et al. (2020). Internet of Things (IoT) applications to fight against COVID-19 pandemic. *Diabetes & Metabolic Syndrome, 14*(4), 521–524.

[9] Khanna, A., & Kaur, S. (2020). Internet of Things (IoT), applications and challenges: A comprehensive review. *Wireless Personal Communications, 114*, 1687–1762.

[10] Tun, S. Y. Y., et al. (2021). Internet of Things (IoT) applications for elderly care: A reflective review. *Aging Clinical and Experimental Research, 33*, 855–867.

[11] Landaluce, H., et al. (2020). A review of IoT sensing applications and challenges using RFID and wireless sensor networks. *Sensors, 20*(9), 2495.

[12] Badshah, A., et al. (2023). Towards smart education through Internet of Things: A survey. *ACM Computing Surveys, 56*(2), 1–33.

[13] Goswami, N., et al. (2022). Design and development of smart system for biofloc fish farming in Bangladesh. In *Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1424–1432). Coimbatore, India.

[14] Rosaline, N., & Sathyalakshimi, S. (2019). IoT based aquaculture monitoring and control system. *Journal of Physics: Conference Series, 1362*(1).

[15] Saha, K. K., et al. (2021). Bio-floc monitoring and automatic controlling system using IoT. In *Proceedings of the IEEE International Conference on Internet of Things and Intelligent Systems (IoTaIS)* (pp. 15–21).

[16] Islam, M. M., et al. (2021). Design and implementation of an IoT system for predicting aqua fisheries using Arduino and KNN. In *Lecture Notes in Computer Science, 12616*, 108–118.

[17] Mahmuda, et al. (2021). Image processing-based water quality monitoring system for biofloc fish farming. In *Emerging Technologies in Computing, Communications and Electronics (ETCCE)*.

[18] Mozumder, S. A., & Sagar, S. (2021). Smart IoT-biofloc water management system using decision regression tree.

[19] Rashid, M. M., et al. (2021). IoT based smart water quality prediction for biofloc aquaculture. *International Journal of Advanced Computer Science and Applications, 12*(5).

[20] Blancaflor, E. B., & Baccay, M. (2021). Design of a solar powered IoT (Internet of Things) remote water quality management system for a biofloc aquaculture technology. In *Proceedings of the ACM International Conference Series* (pp. 24–31).

[21] Blancaflor, E. B., & Baccay, M. (2022). Assessment of an automated IoT-biofloc water quality management system in the Litopenaeus vannamei’s mortality and growth rate. *Automatika, 63*(2), 259–274.

[22] Ahamed, I., & Ahmed, A. (2021). Design of smart biofloc for real-time water quality management system. In *International Conference on Robotics, Electrical and Signal Processing Techniques* (pp. 298–302).

[23] Prakosa, J. A., et al. (2022). Development of monitoring techniques and validation of the acidity level of biofloc pond water for optimizing tilapia aquaculture. In *IOP Conference Series: Earth and Environmental Science, 1017*.

[24] Blancaflor, E. B., & Baccay, M. (2021). Economic & operational impact analysis of a solar powered remote water quality management system designed for an indoor biofloc aquaculture setup. In *Proceedings of the ACM International Conference* (pp. 81–86).