ACIS
Advances in Computational and Intelligent Systems
Vol. 1, No. 1, Month July 2025, pp. 10~17
ISSN: 3093-7388, DOI: https://doi.org/10.56313/acis.v1i1.2

Predictive Water Quality Monitoring in Aquaculture Using Machine Learning and IoT Automation

Zulfikhry Bin Zulhairi[1], Razali Muda[1], Adyanata Lubis[2]
[1]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan Campus, 26600 Pekan, Pahang, Malaysia.
[2]Department of Computer Science, Universitas Pasir Pengaraian, Jl. Tuanku Tambusai, Jl. Raya Kumu, Rambah, Kec. Rambah Hilir, Kabupaten Rokan Hulu 28558, Riau, Indonesia

Article Info

Article history:

Received March 6, 2025
Revised March 25, 2025
Accepted March 27, 2025

Keywords:

Machine Learning-Based Aquaculture
IoT-Enabled Water Quality Monitoring
Automated Decision-Making in Smart Farming

Corresponding Author:

Razali Muda
Email: razali@umpsa.edu.my

ABSTRACT

This study presents an IoT and machine learning (ML)-driven intelligent aquaculture system for real-time water quality monitoring and automated decision-making. The system integrates IoT-enabled sensors to measure key parameters such as temperature, turbidity, dissolved oxygen, and water levels, with data processed by a Random Forest Classifier (RFC) for predictive analytics. The ML model classifies water quality conditions into optimal, warning, or critical states, triggering automated responses such as aeration control and water exchange. Experimental validation over six months demonstrates high classification accuracy (92.3%), improved automation, and reduced manual intervention, enhancing sustainability and efficiency in aquaculture management.

1. INTRODUCTION

Aquaculture is rapidly becoming the world's fastest-growing food production sector, projected to supply more than 60% of the global fish consumption by 2030 [1]. Tilapia farming, in particular, plays a crucial role in meeting seafood demand due to its adaptability, low production costs, and high nutritional value. However, traditional aquaculture practices face significant challenges, primarily in maintaining optimal water quality, improving feed efficiency, and ensuring operational sustainability. Water quality mismanagement is one of the leading causes of fish mortality, accounting for losses ranging from 30% to 50% annually, translating into billions of dollars in economic damage [2]. Additionally, inefficient feeding strategies contribute to excessive feed waste, increased water pollution, and rising operational costs, making sustainability in aquaculture a critical issue.

Despite the importance of environmental monitoring in fish farming, many aquaculture operations still rely on manual methods for water quality assessment. These conventional approaches are often time-consuming, labor-intensive, and prone to inaccuracies, leading to delayed responses to environmental fluctuations [3]. Without real-time monitoring and predictive analytics, farmers struggle to detect early signs of water contamination, temperature fluctuations, or oxygen depletion, resulting in suboptimal fish health and reduced production efficiency. Addressing these limitations requires the adoption of intelligent automation systems that can continuously monitor water parameters, provide real-time alerts, and autonomously optimize farm conditions.

Recent advancements in artificial intelligence (AI), the Internet of Things (IoT), and sensor networks have enabled the development of intelligent farm automation systems capable of transforming aquaculture operations. Intelligent farm automation refers to the integration of AI-driven predictive models, IoT-based real-time monitoring, and cloud-based data analytics to enhance decision-making in both aquaculture and agriculture management. In aquaculture, these technologies enable automated water quality assessment, feeding optimization, and environmental monitoring, ensuring improved fish health and farm productivity. Similarly, in agriculture, intelligent automation facilitates precision farming, optimizing irrigation, soil health monitoring, crop growth prediction, and pest control through data-driven insights and automated interventions. By leveraging machine learning algorithms, sensor networks, and cloud computing, intelligent farm automation enhances efficiency, sustainability, and resource optimization across both land-based and aquatic farming systems. [4] Unlike conventional automated systems that operate on pre-set schedules, an AI-powered system can dynamically adjust water aeration, feeding schedules, and environmental controls based on real-time sensor data. The proposed system in this study incorporates IoT-enabled multi-sensor deployment to monitor critical water quality parameters such as temperature, turbidity, dissolved oxygen and water level. It also utilizes AI-based predictive analytics to classify water quality conditions as optimal, warning, or critical, allowing for proactive interventions.

Existing studies have demonstrated the potential of IoT and AI applications in aquaculture, yet several gaps remain unaddressed [5]. Previous research on IoT-based water quality monitoring systems has shown promising results in real-time data acquisition but lacks AI-driven decision-making, requiring manual user intervention [6]. Similarly, AI-based approaches to fish health monitoring and feeding optimization have been explored but without seamless integration with IoT-based real-time automation. Additionally, studies on aeration and water exchange mechanisms in aquaculture highlight their importance in maintaining dissolved oxygen levels, but most systems lack intelligent control mechanisms for adaptive regulation [7]. This study builds upon existing research by integrating real-time IoT monitoring with AI-powered predictive analytics to create a fully automated, data-driven decision-making system for aquaculture.

ML has emerged as a core technology in modern aquaculture, enabling data-driven insights, predictive analytics, and automation. ML algorithms process vast amounts of sensor data collected from aquaculture environments, learning from historical patterns to predict optimal conditions and potential risks. Common ML applications in aquaculture include automated fish behavior analysis, intelligent feeding systems, water quality forecasting, and disease detection. Supervised learning models, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Decision Trees, have been widely used for classification tasks, while unsupervised techniques help in clustering and anomaly detection [8]. These models contribute to reducing manual interventions, improving resource efficiency, and minimizing environmental impact. Among various ML approaches, ensemble learning methods such as Random Forest Classifiers have gained prominence due to their ability to handle non-linear relationships, noisy datasets, and real-time decision-making. In aquaculture, where water quality parameters fluctuate dynamically, robust predictive models ensure precise control over environmental conditions, leading to better fish health and optimized resource utilization.

The Random Forest Classifier (RFC) is a robust ensemble learning method designed for classification and regression tasks. It enhances predictive accuracy by constructing multiple decision trees on bootstrapped datasets while randomly selecting subsets of features at each split. This approach mitigates overfitting, improves generalization, and makes RFC particularly effective for handling non-linear relationships, noisy data, and high-dimensional datasets. Mathematically, RFC employs a bagging technique where each tree is trained on a randomly sampled subset of the original dataset. For classification, the final prediction is determined by majority voting, given by,

$$\hat{y} = argmax \sum_{b}^{B} 1(T_b(x) = y) \quad (1)$$

Where $T_b(x) = y$ is the prediction from the b-th decision tree.

2. METHODOLOGY

Machine Learning (ML) has become a transformative technology in various domains, including aquaculture, where data-driven approaches enhance efficiency, sustainability, and predictive decision-making. Traditional rule-based automation systems rely on pre-set thresholds and manual intervention, limiting their adaptability to dynamic environmental changes. In contrast, ML enables systems to learn from historical and real-time data, recognize complex patterns, and make autonomous decisions based on data-driven insights. ML algorithms can predict water quality deterioration, optimize feeding schedules, and automate environmental adjustments, reducing fish mortality and operational costs.

Among the various ML techniques available, supervised learning models are particularly well-suited for aquaculture applications, as they can classify environmental conditions based on labeled training data [9], [10]. In this study, a Random Forest Classifier (RFC) is used to predict and classify water quality conditions based on temperature, DO and turbidity data. The Random Forest algorithm, an ensemble learning method, constructs multiple decision trees and combines their outputs to improve classification accuracy. Unlike single decision trees, which may suffer from overfitting, the ensemble approach enhances robustness, generalization, and predictive reliability. The RFC model learns from past sensor readings, allowing it to classify water conditions as optimal, warning, or critical and triggering automated responses accordingly.

2.1. System Architecture and Hardware Components

To achieve efficient, automated water quality management in aquaculture, the system is designed to monitor and control critical water parameters using IoT-enabled sensors and embedded microcontrollers. These parameters—temperature, turbidity, dissolved oxygen (DO), and water level—directly influence fish health, metabolic rates, and overall farm productivity. Effective monitoring and regulation of these factors minimize fish stress, prevent mass mortality, and improve feed conversion efficiency, ultimately enhancing sustainability in fish farming.

At the core of the system is the ESP32 TTGO T-Call SIM800L microcontroller, which serves as the central processing unit (CPU). This microcontroller is selected due to its low power consumption, built-in Wi-Fi and GSM/GPRS communication capabilities, and real-time data processing efficiency. By acting as the main control hub, the ESP32 collects data from multiple environmental sensors, processes it locally, and transmits it to a cloud-based monitoring platform for real-time decision-making and predictive analytics. The GSM module within the ESP32 allows for remote access and notifications, ensuring that farmers receive alerts and control system operations from anywhere, even in locations with limited Wi-Fi connectivity. The system integrates IoT-enabled sensors to ensure continuous water quality monitoring and automation.

- Temperature Sensor (DS18B20) - Measures water temperature, essential for fish metabolism and health. Its high precision (±0.5°C) and waterproof design enable real-time tracking to prevent thermal stress and disease outbreaks.
- Turbidity Sensor (DFRobot) - Detects water clarity, classifying it as clear, cloudy, or dirty. High turbidity triggers automated water exchange to maintain optimal conditions.
- Dissolved Oxygen (DO) Sensor - Monitors oxygen levels to prevent fish suffocation. When DO drops below safe thresholds, the system activates aerators to increase oxygen diffusion.
- Water Level Sensor - Detects depth variations to prevent overflow or depletion. If water levels drop critically, the system automatically refills or sends alerts.
- Relay Modules and Aerators - Enable automated aeration and water circulation. When turbidity or DO levels reach unsafe limits, relays activate pumps and aerators, reducing manual intervention.
- LCD Display - Provides on-site real-time visualization of sensor readings, system status, and alerts, ensuring quick access to critical information without cloud dependency.

All hardware components are interconnected through the ESP32 microcontroller, which continuously processes sensor data, executes automated control actions, and transmits real-time information to the cloud-based monitoring system. By leveraging IoT connectivity and AI-driven predictive analytics, the system optimizes fish farming efficiency, reduces operational costs, and enhances environmental sustainability. Figures 1 and 2 show the circuit design and the system workflow.

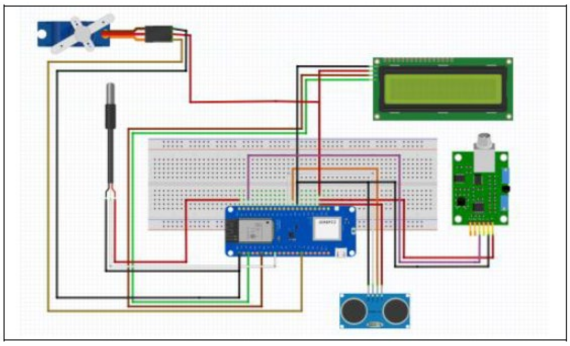

Figure 1. Schematic diagram or the circuit

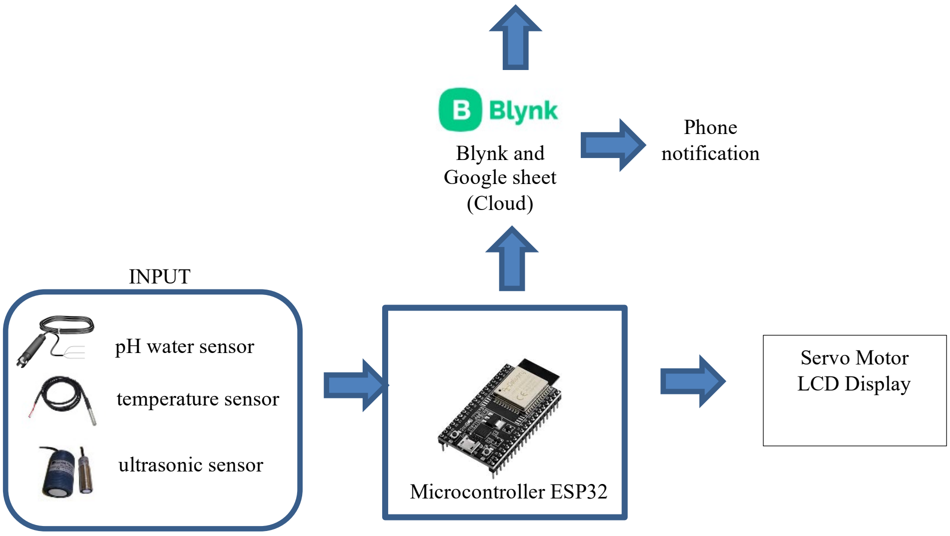

Figure 2. Block diagram of overall system

To enable real-time visualization and remote control, the system utilizes the Blynk IoT platform. Sensor readings are transmitted to the Blynk cloud, allowing users to access real-time dashboards via mobile or web applications. The system workflow includes data acquisition, storage, real-time notifications, and automated control responses based on sensor readings. Additionally, Google Sheets is used for data logging, enabling long-term analysis of water quality trends. The integration of cloud-based storage and IoT connectivity ensures seamless communication between hardware devices, AI algorithms, and the user interface.

2.2. Machine Learning Model Development

Traditional aquaculture monitoring systems typically rely on threshold-based control mechanisms or manual observations, which are reactive in nature and prone to human error. In contrast, ML enables a proactive approach by continuously learning from historical trends and real-time sensor data, allowing for early anomaly detection, dynamic system adjustments, and data-driven decision-making. This study employs a Random Forest Classifier (RFC) as the core ML model for predicting and classifying water quality conditions based on sensor readings. The RFC model is chosen due to its high accuracy, robustness to noisy data, and ability to handle non-linear relationships among environmental parameters [11, 12]. By leveraging an ensemble of decision trees, RFC mitigates overfitting, enhances generalization, and provides reliable multi-class classification, which is essential for real-world aquaculture applications where environmental variability is high. The model development follows a structured machine learning pipeline as shown in the following block diagram of Figure 3.

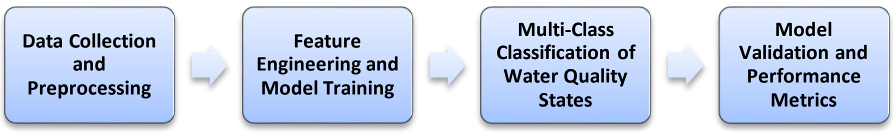

Figure 3. Machine learning model development pipeline for water quality classification, including data acquisition, preprocessing, training, validation, and real-time deployment on ESP32

The machine learning pipeline begins with data collection and preprocessing, where sensor data, including temperature and turbidity, is continuously recorded and stored in a structured dataset. To ensure data reliability, missing values—which may occur due to intermittent sensor errors, communication delays, or power interruptions—are identified and handled using imputation techniques, such as mean substitution or interpolation. Alongside this, outlier detection and normalization are applied to remove noise and standardize input ranges, preventing bias and improving model accuracy. During feature engineering and model training, the dataset is split into 80% training and 20% testing subsets to ensure proper generalization. The Random Forest Classifier (RFC) is trained on historical sensor data to identify key relationships between environmental parameters and water quality states. Feature importance analysis helps determine the most significant factors influencing classification. Once trained, the model performs multi-class classification, categorizing water conditions into three states: Optimal, where no action is required; Warning, where early signs of unfavorable conditions trigger alerts; and Critical, where immediate intervention, such as aeration activation or water exchange, is necessary.

To assess model reliability, validation and performance evaluation are conducted using precision, recall, F1-score, and confusion matrix analysis. Hyperparameter tuning, including grid search and cross-validation, is applied to optimize classification accuracy and improve system responsiveness to real-time water quality fluctuations. An example of the structured dataset used for ML training is shown below:

Table 1. Data collection of all sensors

Timestamp	pH Level	Temperature (°C)	Ultrasonic Sensor (cm)	Water Quality Label
2024-03-05 10:00	7.2	26.5	15	Optimal
2024-03-05 10:30	7.5	27.1	16	Optimal
2024-03-05 11:00	6.8	29.0	14	Warning
2024-03-05 11:30	6.5	30.2	13	Critical
2024-03-05 12:00	7.0	28.5	14	Optimal

Once trained and validated, the Random Forest model is deployed on the ESP32 microcontroller to enable real-time decision-making for water quality management. The system continuously receives live sensor inputs, preprocesses the data, and feeds it into the trained model for classification. The classified water quality states (e.g., optimal, warning, or critical) trigger automated responses, ensuring proactive intervention without manual supervision. The decision-making process follows a structured pipeline as shown in the following pseudocode in figure 4.

```
Algorithm: Water Quality Monitoring and Decision-Making
1. Initialize ESP32, sensors, and cloud connectivity.
2. Continuously acquire real-time sensor data:
   - Read pH, temperature, and turbidity values.
3. Perform data preprocessing:
   - Normalize sensor readings.
   - Detect and remove outliers.
4. Classify water quality using ML inference:
   - If (pH, temperature, and turbidity within safe range) → State = "Optimal"
   - Else if (mild deviations detected) → State = "Warning"
   - Else (severe deviations detected) → State = "Critical"
5. Execute automated responses based on classification:
   - If State == "Optimal": No action taken.
   - If State == "Warning": Send alert notification via Telegram.
   - If State == "Critical":
      - Activate aerators to increase oxygen.
      - Initiate water exchange system.
      - Update LCD display for on-site monitoring.
6. Repeat every predefined time interval.
```

Figure 3. Pseudocode representation of the ESP32-based decision-making process for real-time water quality classification and automated system response.

3. EXPERIMENTAL SETUP

A six-month experimental study was conducted in a Recirculating Aquaculture System (RAS) for tilapia farming, integrating IoT sensors and an AI-powered microcontroller (ESP32) for real-time water quality monitoring and automated decision-making. The system collected temperature, turbidity, dissolved oxygen, and water level data, enabling ML-driven predictions and automated aeration and water exchange adjustments.

Sensor data was recorded hourly and transmitted to the Blynk cloud platform for remote monitoring, visualization, and alerts. Historical data was stored in Google Sheets for trend analysis and model training. This setup validated the effectiveness of machine learning in optimizing fish health, improving farm efficiency, and reducing manual intervention, ensuring a scalable and sustainable aquaculture management system. The following Figure 4 illustrates the hardware setup used in this experiment.

Over the six-month experimental period, the system continuously monitored key water quality parameters, ensuring data integrity and reliability for machine learning model training. Sensor readings were recorded at hourly intervals, generating a comprehensive dataset for further analysis. The monitored parameters included:

- Temperature (°C): Maintained within the optimal range to support tilapia metabolism and growth.
- Turbidity (NTU): Measured to assess water clarity and detect contamination levels, ensuring suitable conditions for fish health.
- Dissolved Oxygen (mg/L): Monitored to evaluate oxygen availability, a critical factor for fish respiration and metabolic efficiency.
- Water Level (cm): Tracked to prevent excessive evaporation, maintain a stable aquatic environment, and ensure adequate water circulation.

Figure 4. Complete hardware setup of the intelligent aquaculture monitoring system, integrating IoT sensors, ESP32 microcontroller, LCD display, and automated control components for real-time water quality management.

4. RESULT AND DISCUSSION CONCLUSION

The implementation of machine learning-driven water quality monitoring in tilapia aquaculture was evaluated over a six-month experimental study, during which the system continuously collected, processed, and analyzed sensor-based environmental data. The results demonstrate the effectiveness of machine learning (ML) in predictive water quality classification, anomaly detection, and automated decision-making. Throughout the six-month monitoring period, the system collected real-time data on temperature, turbidity, dissolved oxygen (DO), and water levels at hourly intervals. The data was transmitted to the cloud-based monitoring platform and subsequently analyzed for trend identification and machine learning training.

The system continuously monitored temperature, turbidity, and dissolved oxygen (DO) levels, ensuring optimal conditions for tilapia farming. Variations in these parameters were recorded over time, influenced by daily cycles, seasonal changes, feeding activities, and organic waste accumulation. The integration of IoT sensors and AI-driven decision-making allowed for real-time adjustments, minimizing fish stress and optimizing farm conditions.

Temperature was maintained within 25°C to 30°C, with automated aeration triggered when levels exceeded 30°C, preventing heat stress and metabolic inefficiencies. Turbidity fluctuations were detected and classified into clear (<30% NTU), cloudy (30%-60% NTU), and dirty (>60% NTU) states, with automated water exchange initiated when necessary to maintain clarity. Dissolved oxygen (DO) levels, ranging between 4.5 mg/L and 7.5 mg/L, were regulated through automated aerator activation, ensuring adequate oxygen availability. Below is a summary table presenting the overall results of the water quality monitoring and automation system.

Table 2. Overall results of the water quality monitoring and automation system

Parameter	Recorded Range	Optimal Range	Automated Response	Impact on Aquaculture
Temperature (°C)	25°C - 30°C	25°C - 30°C	Aeration activated when >30°C	Prevented heat stress, maintained optimal growth
Turbidity (NTU)	10 - 70 NTU	<30 NTU	Water exchange triggered >60 NTU	Improved water clarity, reduced fish stress
Dissolved Oxygen (mg/L)	4.5 - 7.5 mg/L	>5 mg/L	Aerator activated <5 mg/L	Ensured proper oxygen levels, reduced mortality
AI Decision-Making	Classification Accuracy: 92.3%	–	Automated alerts and interventions	Reduced manual intervention, improved response time
System Efficiency	87% success in water exchange, 93.5% in aeration activation	–	Automated real-time adjustments	Enhanced sustainability and operational efficiency

4.1 Machine Learning Model Performance Evaluation

The Random Forest Classifier (RFC) was trained using historical water quality data, enabling the model to classify water conditions into optimal, warning, or critical states. The model's classification accuracy, precision, recall, and F1-score were assessed to ensure its reliability for real-time aquaculture management. The Random Forest Classifier achieved an overall accuracy of 92.3%, indicating high predictive reliability in classifying water quality conditions. The model's classification breakdown is as follows:

Table 3. ML performance evaluation

Water Quality Condition	Precision (%)	Recall (%)	F1-Score (%)
Optimal	94.5	92.0	93.2
Warning	89.7	90.5	90.1
Critical	91.3	93.1	92.2

5. CONCLUSION

The six-month experimental study validates the effectiveness of ML-driven aquaculture automation. The Random Forest Classifier achieved high classification accuracy (92.3%), demonstrating reliable water quality prediction. The system successfully automated aeration, water exchange, and anomaly detection, significantly reducing manual workload and operational inefficiencies. These findings highlight the transformative potential of AI and IoT in precision aquaculture, paving the way for intelligent, sustainable fish farming systems. However, real-world deployment may face practical challenges, including sensor drift, biofouling, or hardware failure over extended use, which can impact data accuracy. In remote locations, network connectivity issues may also hinder real-time monitoring and control. To mitigate these, future work will focus on local data buffering, edge processing, and robust sensor maintenance protocols. Additionally, the system will be enhanced by expanding sensor integration (e.g., pH, ammonia) and refining AI-driven forecasting models for even greater precision in automated aquaculture management.

REFERENCES

[1] D. Nallaperuma *et al.*, "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.

[2] S. Schulz, M. Becker, M. R. Groseclose, S. Schadt, and C. Hopf, "Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development," *Curr Opin Biotechnol*, vol. 55, pp. 51–59, 2019, doi: 10.1016/j.copbio.2018.08.003.

[3] C. Shang and F. You, "Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era," *Engineering*, vol. 5, no. 6, pp. 1010–1016, 2019, doi: 10.1016/j.eng.2019.01.019.

[4] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 288–305, 2019, doi: 10.26599/BDMA.2019.9020007.

[5] M. Huang, W. Liu, T. Wang, H. Song, X. Li, and A. Liu, "A queuing delay utilization scheme for on-path service aggregation in services-oriented computing networks," *IEEE Access*, vol. 7, pp. 23816–23833, 2019, doi: 10.1109/ACCESS.2019.2899402.

[6] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of things in marine environment monitoring: A review," *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–21, 2019, doi: 10.3390/s19071711.

[7] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, *Smarter traffic prediction using big data, in-memory computing, deep learning and gpus*, vol. 19, no. 9. 2019. doi: 10.3390/s19092206.

[8] S. Leonelli and N. Tempini, *Data Journeys in the Sciences*. 2020.

[9] N. Stylos and J. Zwiegelaar, *Big Data as a Game Changer: How Does It Shape Business Intelligence Within a Tourism and Hospitality Industry Context?* 2019.

[10] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *ArXiv*, vol. 13, no. 1, 2017.

[11] Rubén Baena-Navarro, Yulieth Carriazo-Regino, Francisco Torres-Hoyos, "Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture: Integration of Machine Learning and Internet of Things for Sustainable," *ArXiv*, vol. 1, no. 1, 2023. doi.org/10.3390/w17010082