

# IoT-Based Smart Aquaculture Water Quality Monitoring and Health Prediction Using Machine Learning



Kamruzzaman Mithu<sup>1\*</sup>, Md. Nesar Uddin<sup>1</sup>, Md. Ataur Rahman<sup>1</sup>, Sayed Rokibul Hossain<sup>1</sup>, A.K.M. Muzahidul Islam<sup>1</sup>

## Abstract

**Background:** Maintaining stable aquatic environmental conditions remains a persistent challenge because fluctuations in water quality parameters can rapidly affect fish health, productivity, and ecosystem stability. Recent advances in Internet of Things (IoT) technologies and machine learning have created new opportunities for developing automated and predictive aquaculture monitoring systems capable of supporting more intelligent water-quality management. **Methods:** In this study, a low-cost IoT-based aquaculture monitoring and prediction framework was developed using an ESP8266 microcontroller integrated with temperature, pH, total dissolved solids (TDS), and electrical conductivity (EC) sensors. Environmental data were continuously collected from aquaculture samples and transmitted through Wi-Fi communication to a cloud-hosted MySQL database. A web-based dashboard was developed for real-time visualization of environmental conditions and predictive outputs. Machine learning algorithms including Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression were implemented using the Scikit-learn framework to classify aquaculture health conditions based on the collected environmental dataset. **Results:** The proposed system successfully performed real-time environmental monitoring, wireless data

transmission, cloud storage, and predictive analysis throughout the experimental period. Comparative evaluation demonstrated that Logistic Regression achieved the highest overall classification accuracy of approximately 91%, while Random Forest and SVM produced comparatively strong F1-score performances of 86.5% and 84.8%, respectively. **Conclusion:** The findings demonstrate that integrating IoT infrastructure with machine learning techniques can provide an affordable and operationally practical framework for intelligent aquaculture monitoring.

## Keywords

Internet of Things (IoT), Smart aquaculture, Water quality monitoring, Machine learning, Environmental prediction

## 1. Introduction

The rapid growth of aquaculture over the past few decades has gradually transformed fish farming from a largely traditional practice into one of the most important food-producing sectors worldwide. Increasing pressure on natural fisheries, combined with the rising global demand for protein-rich food, has forced many countries to rely more heavily on aquaculture systems for sustainable food

**Significance |** Low-cost IoT and machine learning integration enables practical real-time aquaculture health prediction, supporting sustainable fish farming and environmental monitoring.

## \*Correspondence.

Kamruzzaman Mithu, Department of Computer Science & Engineering United International University, Dhaka, Bangladesh. Email: kamruzzamansmithu@gmail.com.

Editor Md Golam Rabiul Alam, Ph.D. And accepted by the Editorial Board Aug 14, 2024 (received for review Jun 10, 2024)

**Author Affiliation:** <sup>1</sup>Department of Computer Science & Engineering United International University, Dhaka, Bangladesh.

**Please cite this article as:** Mithu, K., Uddin, M. N., Rahman, M. A., Hossain, S. R., Islam, A. M. (2024). "IoT-Based Smart Aquaculture Water Quality Monitoring and Health Prediction Using Machine Learning", Data Modeling, 5(1), 1-18, 10759

© 2024 (Data Modeling), a publication of Eman Research, USA.  
This is an open access article under the CC BY-NC-ND license.  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)  
(<https://publishing.emanresearch.org>)

production. According to the Food and Agriculture Organization of the United Nations (UNFAO, 2012), global fishery production contributes substantially to human nutrition, supplying nearly 15% of animal protein intake worldwide, while average annual fish consumption per capita continues to remain high across many populations (UNFAO, 2012). This growing dependence on aquaculture is not merely an economic issue; rather, it is closely tied to food security, nutritional stability, and the broader sustainability of agricultural ecosystems. In many developing regions especially, aquaculture represents both livelihood and survival, making the maintenance of healthy aquatic environments increasingly critical for long-term productivity and public health (Zhu, 2010).

Despite the expansion of aquaculture technologies, maintaining a stable and healthy aquatic environment remains remarkably challenging. Fish are extremely sensitive organisms, and even relatively small fluctuations in water quality can create physiological stress, suppress immune response, reduce growth rates, and, in severe cases, lead to mass mortality events. Unlike terrestrial farming systems, aquatic ecosystems operate through a delicate balance of chemical, biological, and physical interactions. Parameters such as temperature, dissolved oxygen (DO), pH, turbidity, conductivity, salinity, alkalinity, nitrate concentration, and total dissolved solids (TDS) continuously influence one another in complex ways (Bhatnagar & Devi, 2013). A change in one variable often triggers a cascade of secondary effects across the entire aquatic environment. Consequently, aquaculture management increasingly depends on continuous monitoring rather than occasional manual observation.

Traditionally, water quality assessment in fish farming environments has relied heavily on manual inspection and periodic laboratory testing. While such approaches may provide acceptable measurements under controlled conditions, they are often time-consuming, labor-intensive, and economically impractical for many small- and medium-scale farmers. More importantly, intermittent measurements fail to capture real-time environmental fluctuations, which are common in aquaculture ponds. Water temperature, pH, and conductivity may vary substantially throughout the day due to photosynthetic activity, feeding behavior, rainfall, and microbial processes (Boyd, 1982). In practical farming scenarios, these rapid changes can occur long before a farmer becomes aware of the deteriorating conditions. As a result, the lack of real-

time monitoring often delays intervention and increases production losses.

The emergence of the Internet of Things (IoT) has introduced a promising alternative to conventional aquaculture monitoring systems. IoT technologies allow interconnected sensors, microcontrollers, and communication devices to gather environmental data continuously and transmit those data remotely through wireless networks. Recent advances in low-cost microcontrollers and embedded systems have significantly accelerated the integration of IoT into agricultural and aquacultural applications. Devices such as ESP8266-based microcontrollers provide affordable Wi-Fi-enabled platforms capable of collecting and transmitting sensor information in real time, thereby reducing the need for expensive industrial monitoring infrastructure. This evolution has gradually shifted aquaculture management toward more intelligent and automated frameworks.

Several earlier studies have explored the application of IoT technologies in aquaculture monitoring. Raju and Varma (2018), for example, proposed a knowledge-based real-time aquaculture monitoring system utilizing sensors for dissolved oxygen, ammonia, salinity, nitrate, temperature, and pH. Their work demonstrated the potential of IoT-driven monitoring systems for improving aquaculture management efficiency. However, systems incorporating numerous sensors often become economically expensive and technically cumbersome for practical implementation, particularly in low-resource farming environments (Raju & Varma, 2018). Similarly, Kayalvizhi et al. (2015), Simbeye and Yang (2014), and Patil et al. (2015) investigated water monitoring approaches emphasizing parameters such as pH, turbidity, dissolved oxygen, and temperature. These studies contributed important insights into aquatic monitoring systems, yet many of them primarily focused on environmental sensing rather than predictive analysis of aquaculture health conditions (Kayalvizhi et al., 2015; Simbeye & Yang, 2014; Patil et al., 2015).

At the same time, IoT technologies have expanded broadly across agricultural domains beyond aquaculture. Israni et al. (2015) proposed an IoT-based agricultural framework emphasizing cloud-assisted accessibility of farming data, while Gondchawar and Kawitkar (2016) introduced smart agricultural systems incorporating automated irrigation, humidity sensing, and environmental monitoring. Although these studies differ from aquaculture-

focused systems, they collectively highlight the increasing importance of remote sensing, automation, and cloud-based agricultural intelligence in modern food production systems (Israni et al., 2015; Gondchawar & Kawitkar, 2016).

By combining low-cost IoT infrastructure with predictive analytics, this work aims to develop a comparatively affordable and practical framework for intelligent aquaculture monitoring. The overall conceptual workflow of the proposed IoT-driven aquaculture monitoring and prediction framework is illustrated in **Figure 1**.

Still, one issue remains somewhat underexplored. Many existing monitoring systems attempt to measure large numbers of environmental parameters simultaneously, often increasing hardware complexity and operational cost without necessarily improving prediction efficiency proportionally. In real-world aquaculture environments, certain parameters exert stronger systemic influence than others. Temperature, for instance, directly affects dissolved oxygen concentration, conductivity, salinity, and metabolic activity within aquatic ecosystems (Boyd, 1982; Bhatnagar et al., 2004). Likewise, pH is strongly associated with carbon dioxide concentration, ammonia toxicity, and broader biochemical reactions within pond environments (Delince, 2013). TDS and electrical conductivity similarly reflect dissolved mineral content and ionic concentration, both of which influence fish growth, larval development, and aquatic productivity (Bhatnagar & Devi, 2013).

These interrelationships suggest that a smaller but carefully selected set of environmental parameters may provide sufficient predictive insight into overall aquaculture health. Such an approach could reduce system cost while maintaining acceptable predictive performance. This idea forms the conceptual basis of the present study. Rather than constructing a highly sensor-intensive platform, the proposed system focuses on four critical parameters—temperature, pH, TDS, and electrical conductivity—to evaluate aquaculture health conditions. These variables were selected based on their biological importance, interdependency, and relative ease of measurement and control.

Beyond real-time sensing alone, the integration of machine learning (ML) further strengthens the predictive capability of modern monitoring systems. Machine learning algorithms can identify patterns within sensor-

generated datasets and classify environmental conditions more efficiently than manual threshold-based observation. Libraries such as Scikit-learn provide accessible frameworks for implementing classification models including Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression approaches (Pedregosa et al., 2011). In aquaculture applications, such predictive models may help identify deteriorating water conditions earlier, thereby enabling timely intervention before significant biological or economic damage occurs.

In this study, we propose an IoT-driven aquaculture monitoring and prediction system using an ESP8266-based microcontroller integrated with temperature, pH, TDS, and conductivity sensors. Sensor data are continuously collected and transmitted to a cloud-based database through Wi-Fi connectivity. A web-based dashboard enables real-time visualization of environmental parameters and prediction outcomes. Furthermore, machine learning models are employed to classify the healthiness of aquaculture conditions based on collected environmental data. By combining low-cost IoT infrastructure with predictive analytics, this work aims to develop a comparatively affordable and practical framework for intelligent aquaculture monitoring. Ultimately, the proposed approach seeks not only to improve operational efficiency but also to contribute toward more sustainable and data-driven fish farming practices.

## 2. Related Works and Research Gap

The integration of Internet of Things (IoT) technologies into aquaculture and precision agriculture has attracted increasing research attention over the last decade, largely because traditional monitoring approaches are often insufficient for maintaining stable environmental conditions in dynamic aquatic ecosystems. Water quality parameters fluctuate continuously under the influence of biological activity, climatic variation, feeding behavior, and chemical interactions within the pond environment. As a result, researchers have gradually shifted from manual inspection methods toward automated sensing and predictive monitoring systems capable of real-time environmental observation. Nevertheless, despite the growing number of IoT-based aquaculture studies, many existing systems still face practical limitations related to hardware complexity, operational cost, scalability, and predictive intelligence.

One of the earlier notable contributions in this area was

presented by Raju and Varma (2018), who developed a knowledge-based real-time aquaculture monitoring system integrating multiple environmental parameters including dissolved oxygen, temperature, ammonia, salinity, pH, nitrate, and carbonate levels. Their work demonstrated the significant potential of IoT-assisted environmental surveillance for fish farming applications. By continuously monitoring several physicochemical variables simultaneously, the system aimed to improve decision-making and reduce the likelihood of sudden environmental deterioration. However, although the approach was technically comprehensive, the inclusion of numerous sensing modules inevitably increases system complexity, maintenance requirements, calibration burden, and overall implementation cost. In practical aquaculture environments—particularly within small-scale or resource-limited farming systems—such highly sensor-intensive frameworks may become difficult to sustain operationally over long periods. This issue, perhaps somewhat understated in many previous studies, remains an important challenge in the deployment of scalable smart aquaculture infrastructures.

Similarly, Kayalvizhi et al. (2015) explored cyber-aquaculture monitoring systems utilizing Arduino and Raspberry Pi platforms for environmental sensing and remote observation. Their work primarily emphasized parameters such as pH, dissolved oxygen, and turbidity, which are indeed among the most biologically significant indicators of aquatic health. Simbeye and Yang (2014) also proposed wireless sensor network-based water quality monitoring systems designed to improve remote environmental supervision in aquaculture settings. Meanwhile, Patil et al. (2015) investigated GSM-based monitoring approaches focusing on turbidity, pH, and temperature assessment. Collectively, these studies contributed meaningfully to the growing field of intelligent aquaculture monitoring by demonstrating the feasibility of integrating low-cost electronics, wireless communication, and environmental sensing technologies into aquatic management systems.

Yet, despite these technological advances, many previous studies remain primarily focused on environmental observation rather than predictive interpretation of environmental conditions. In other words, sensor measurements are often collected and displayed, but relatively limited attention is given to deriving actionable intelligence from those data streams. This distinction is

important. Real-time sensing alone may not necessarily prevent environmental deterioration unless the collected data can also be interpreted meaningfully and translated into predictive assessments regarding the healthiness of the aquaculture system. Consequently, the integration of machine learning techniques into IoT-driven aquaculture monitoring frameworks represents an increasingly relevant research direction.

At the same time, broader agricultural IoT studies have further highlighted the expanding role of cloud computing and smart automation in precision farming ecosystems. Israni et al. (2015), for instance, proposed an IoT-based agricultural framework emphasizing cloud-assisted accessibility of agricultural information. Their study focused largely on improving the availability and remote accessibility of production-related data through connected sensing systems. However, unlike aquaculture-focused environmental prediction systems, the primary emphasis of their work centered more heavily on agricultural data communication rather than biological water-quality interpretation.

Likewise, Gondchawar and Kawitkar (2016) introduced an IoT-based smart agriculture model integrating automated spraying systems, moisture sensing, and humidity monitoring technologies. Their work illustrated how environmental sensing and automation can improve agricultural productivity through intelligent intervention. Nevertheless, while these systems provide valuable insights into precision agriculture and environmental automation, aquatic ecosystems present fundamentally different challenges. Water quality parameters interact dynamically and often nonlinearly, meaning that small disturbances in one parameter can rapidly propagate across the system and affect fish metabolism, oxygen availability, microbial activity, and overall ecosystem stability.

This interconnected nature of aquatic chemistry has been extensively discussed in earlier environmental studies. According to Bhatnagar and Devi (2013), effective pond management requires continuous awareness of multiple environmental indicators including dissolved oxygen, temperature, turbidity, pH, alkalinity, hardness, conductivity, salinity, TDS, nitrate concentration, nitrite concentration, and biological productivity. These parameters collectively influence fish growth, stress tolerance, reproductive performance, and survival. However, maintaining continuous surveillance of every

possible variable may not always be economically realistic or operationally efficient. In practice, some environmental parameters exert stronger systemic influence than others and may therefore serve as representative indicators of broader environmental conditions.

Temperature, for example, is widely recognized as one of the most influential environmental variables in aquaculture systems. Boyd (1982) reported that temperature substantially affects both biological and chemical reaction rates within aquatic environments, with reaction intensity often doubling following a 10°C increase. Sudden temperature changes may induce physiological stress or mortality in fish populations. Furthermore, temperature directly influences dissolved oxygen concentration, salinity, conductivity, and pH balance (Bhatnagar & Devi, 2013; Boyd, 1982; Delince, 2013). Because of these cascading interactions, temperature frequently acts as a foundational environmental parameter influencing multiple secondary variables simultaneously.

The role of pH is similarly complex and biologically significant. Pond water pH varies continuously throughout the day due to photosynthetic activity and fluctuations in dissolved carbon dioxide concentration. During daylight hours, phytoplankton consume carbon dioxide during photosynthesis, often increasing pH levels, whereas nighttime respiration causes pH reduction before sunrise (Boyd, 1982). Delince (2013) further emphasized that pH regulation is closely associated with ammonia toxicity and hydrogen sulfide dynamics in aquatic systems. Excessive pH fluctuations may therefore destabilize fish physiology and microbial balance within aquaculture environments. Importantly, pH also maintains direct or indirect relationships with several additional environmental variables, making it a particularly informative parameter for predictive monitoring systems.

Total dissolved solids (TDS) and electrical conductivity (EC) additionally serve as important indicators of aquatic chemical composition. TDS reflects the total concentration of dissolved salts and minerals within water, and abnormal concentrations may negatively affect fish growth, osmoregulation, and ecosystem stability (Bhatnagar & Devi, 2013). Excessively high TDS levels may also contribute to algal bloom formation and deteriorating water quality conditions. Electrical conductivity, meanwhile, reflects the concentration of charged ions and is strongly associated with salinity and dissolved mineral content. Previous studies have shown that conductivity influences

fertilization success, embryonic development, larval growth, and aquatic population density (Boyd, 1982). Moreover, conductivity and TDS are themselves interrelated, suggesting that selected combinations of environmental parameters may provide sufficient predictive representation of broader aquatic health conditions.

These observations collectively suggest that effective aquaculture monitoring may not necessarily require extremely large numbers of sensing modules if carefully selected parameters can adequately capture broader environmental behavior. Nevertheless, many previous IoT-based systems continue to rely primarily on extensive multi-sensor architectures without fully exploring parameter optimization strategies or predictive modeling efficiency. In addition, relatively few studies integrate machine learning techniques directly into low-cost IoT infrastructures for real-time aquaculture health prediction.

Based on previously established aquaculture water-quality guidelines, acceptable and unhealthy threshold ranges for temperature, pH, TDS, and electrical conductivity were identified to support environmental classification and predictive modeling. The healthy, moderate, and unhealthy ranges of the selected parameters are summarized in **Table I**.

To better contextualize these limitations, a comparative analysis of selected related studies is presented in Table II. The comparison highlights differences in parameter selection, sensing architecture, and machine learning integration among previously reported systems and the proposed framework. In contrast to several earlier studies that focused primarily on environmental observation, the present work integrates temperature, pH, TDS, and electrical conductivity monitoring with machine learning-based predictive analysis while maintaining a comparatively lightweight and cost-effective sensing architecture. This balance between affordability, environmental representation, and predictive capability forms the central research motivation underlying the proposed system.

### 3. Methodology

#### 3.1 System Design Overview

The present study was designed to develop a low-cost Internet of Things (IoT)-based aquaculture monitoring and prediction framework capable of continuously observing

critical water-quality parameters and predicting the health condition of an aquaculture environment in near real time. The overall architecture of the proposed system integrates environmental sensing, wireless data transmission, cloud-based storage, machine learning–based prediction, and web-based visualization into a unified monitoring framework (Figure 1). The methodological design was intentionally structured to remain comparatively lightweight and economically feasible while still preserving sufficient predictive capability for practical aquaculture applications.

The proposed system consisted of four major operational components: (i) sensor-based environmental data acquisition, (ii) wireless communication and cloud storage, (iii) machine learning–based predictive analysis, and (iv) real-time dashboard visualization. Sensor nodes continuously collected water-quality measurements from the aquaculture environment and transmitted the data through a Wi-Fi-enabled microcontroller to a remote database server. The collected data were subsequently processed using machine learning algorithms to classify the healthiness of the aquaculture system.

### 3.2 Selection of Water-Quality Parameters

Environmental parameter selection was performed following an extensive review of previously reported aquaculture monitoring studies and established water-quality management guidelines (Bhatnagar & Devi, 2013; Boyd, 1982; Delince, 2013). Although numerous environmental variables may influence aquatic ecosystems, the present study focused specifically on four parameters: temperature, pH, total dissolved solids (TDS), and electrical conductivity (EC). These parameters were selected because of their strong biological significance, relative interdependence, ease of measurement, and practical applicability in low-cost monitoring environments.

Temperature was considered a primary parameter because it directly influences metabolic activity, dissolved oxygen concentration, conductivity, and broader biochemical interactions within aquatic systems (Boyd, 1982; Bhatnagar et al., 2004). Similarly, pH was selected due to its close association with carbon dioxide concentration, ammonia toxicity, and aquatic chemical balance (Delince, 2013). TDS and EC were included because both parameters reflect dissolved ionic composition and mineral concentration, which significantly affect fish growth and environmental stability (Bhatnagar & Devi, 2013).

The healthy, moderate, and unhealthy threshold ranges used for environmental classification are summarized in Table I. These threshold values were adapted from previously published aquaculture water-quality guidelines and environmental management studies (Bhatnagar & Devi, 2013; Boyd, 1982).

### 3.3 Hardware Configuration

#### 3.3.1 Microcontroller Unit

The central processing and communication component of the proposed system was the LOLIN D1 Mini v2.3.0 microcontroller board manufactured by WEMOS, which is based on the ESP8266EX Wi-Fi chipset (Figure 2). The ESP8266 platform was selected primarily because of its comparatively low cost, integrated TCP/IP networking capability, low power consumption, and suitability for embedded IoT applications (Espressif Systems, n.d.).

The microcontroller contains a 32-bit Tensilica L106 RISC processor with integrated wireless networking functionality and supports communication through Wi-Fi protocols. The board additionally provides digital and analog input/output interfaces required for environmental sensor integration. Programming and firmware deployment were conducted using the Arduino Integrated Development Environment (IDE).

#### 3.3.2 Sensor Configuration

Three primary environmental sensing modules were integrated into the system architecture for real-time data acquisition (Figure 3).

##### 3.3.2.1 pH Sensor

Water acidity and alkalinity were monitored using the DFRobot Gravity Analog pH Meter V2 (SEN0161-V2). The sensor was specifically designed for environmental monitoring applications including aquaculture and aquaponics systems. The module supports operation within 3.3–5.5 V input ranges and provides analog output signals with reduced signal fluctuation and improved stability (DFRobot, n.d.-a). Prior to deployment, calibration was performed using standard buffer solutions according to the manufacturer's operational recommendations.

##### 3.3.2.2 Temperature Sensor

Water temperature measurements were acquired using the waterproof DS18B20 digital temperature sensor. This sensor supports 9–12-bit temperature readings and

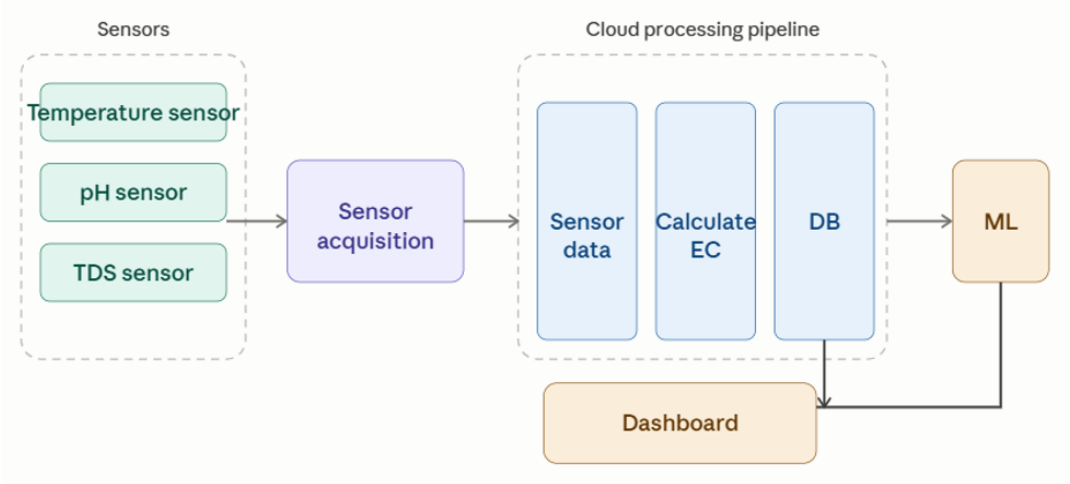


Figure 1. Conceptual architecture of the proposed IoT-based aquaculture health monitoring and machine learning prediction system.

Table 1. Classification thresholds of the selected aquaculture water-quality parameters used for environmental health prediction. The table summarizes the healthy, moderate, and unhealthy ranges of temperature, pH, total dissolved solids (TDS), and electrical conductivity (EC) utilized for environmental monitoring and machine learning-based aquaculture health classification.

| # | Parameter        | Healthy Range | Moderate Range | Unhealthy Range |
|---|------------------|---------------|----------------|-----------------|
| 1 | Temperature (°C) | 20–30         | 15–35          | <15, >35        |
| 2 | pH               | 6.5–9.0       | 7.0–9.5        | <6.5, >9.5      |
| 3 | TDS              | 85–164        | 75–200         | <75, >200       |
| 4 | EC               | 138–274       | 120–320        | <120, >320      |

Table 2. Comparative analysis of previously reported IoT-based aquaculture monitoring systems and the proposed framework. The table highlights differences in monitored environmental parameters and machine learning integration among related studies and the present work.

| Study                  | Temperature | pH | TDS | EC | DO | ML Integration |
|------------------------|-------------|----|-----|----|----|----------------|
| S. Saha et al.         |             |    | ×   |    | ×  | ×              |
| S. U. Kiruthika et al. |             |    | ×   | ×  | ×  | ×              |
| D. S. Simbeye et al.   |             |    | ×   | ×  |    | ×              |
| Present Study          |             |    |     |    | ×  |                |

operates through a one-wire digital communication interface, allowing reliable signal transmission over comparatively long distances (DFRobot, n.d.-b). The sensor was submerged directly within the aquaculture environment to allow continuous real-time temperature monitoring.

### 3.3.2.3 TDS Sensor

Total dissolved solids were measured using the SEN0244 TDS sensor module. The sensor estimates dissolved mineral concentration based on electrical conductivity properties and supports analog voltage output within 0–2.3 V ranges compatible with ESP8266-based systems (DFRobot, n.d.-c). The sensor was immersed within the aquaculture water body during experimental operation to enable continuous ionic concentration monitoring.

Electrical conductivity values were derived through conductivity-related sensor measurements integrated within the monitoring framework. The physical arrangement and integration of the sensors with the microcontroller are illustrated in Figure 3.

### 3.3.3 Sensor Deployment and Data Acquisition

The sensors were connected directly to the ESP8266-based microcontroller through analog and digital input channels according to the manufacturer's interfacing guidelines. Sensor readings were collected periodically through Arduino-based firmware developed specifically for this study.

The Arduino program continuously initialized sensor modules, acquired environmental measurements, processed the raw sensor signals, and transmitted the data to the cloud server through Wi-Fi communication protocols. The program structure followed the conventional Arduino framework utilizing `setup()` and `loop()` functions for initialization and repeated execution processes, respectively. A representative screenshot of the Arduino-based data acquisition environment is shown in Figure 5.

Environmental data collection was conducted continuously for approximately one week across three aquaculture samples used for experimental analysis and predictive testing. Sensor measurements were acquired repeatedly throughout the monitoring period to capture environmental fluctuations under varying aquatic conditions.

## 3.4 Cloud Database Integration and Data Storage

To enable centralized storage and remote accessibility, collected sensor measurements were transmitted to a cloud-hosted MySQL database using a PHP-based insertion script. The insertion script received incoming sensor data through HTTP requests generated by the ESP8266 module and subsequently stored the data within structured database tables.

PHP was used as the server-side scripting language because of its compatibility with lightweight web applications and relational database management systems. JQuery and CSS were additionally utilized for front-end dashboard implementation and visualization purposes. The insert script responsible for database communication is illustrated in Figure 6, while representative stored environmental records are shown in Figure 7.

The web-based dashboard was designed to provide users with real-time access to sensor measurements and prediction outputs through internet connectivity. The dashboard displayed continuously updated environmental conditions and machine learning prediction results. A representative interface of the monitoring dashboard is presented in Figure 8.

## 3.5 Dataset Preparation and Preprocessing

Environmental data stored within the MySQL database were exported into comma-separated value (CSV) format for machine learning analysis. The dataset consisted of four primary predictor variables: temperature, pH, TDS, and electrical conductivity. Each observation was associated with a corresponding health-status classification label based on the predefined threshold ranges described in Table I.

Prior to model development, the dataset was reviewed for incomplete entries, abnormal values, and measurement inconsistencies. Environmental observations were then organized into structured input matrices suitable for machine learning implementation. Data preprocessing and analytical procedures were conducted using Python programming language within the Jupyter Notebook computational environment using the Pandas data-processing library.

## 3.6 Machine Learning Implementation

Machine learning analysis was conducted using the Scikit-learn framework, an open-source machine learning library developed for Python-based analytical environments

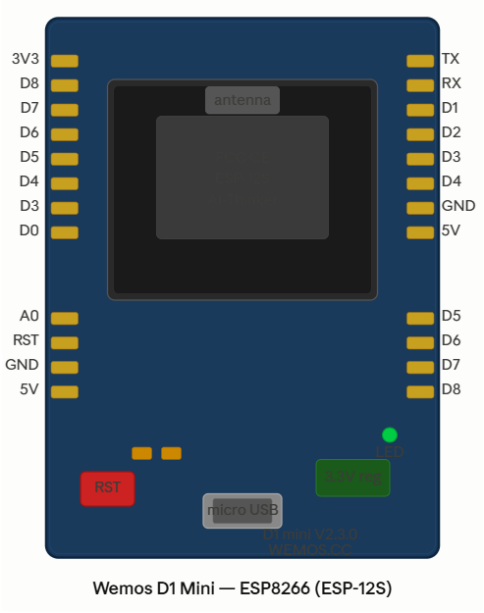


Figure 2. LOLIN D1 Mini v2.3.0 microcontroller based on the ESP8266EX Wi-Fi chipset used for sensor interfacing, wireless data transmission, and real-time environmental monitoring within the proposed IoT-based aquaculture system.

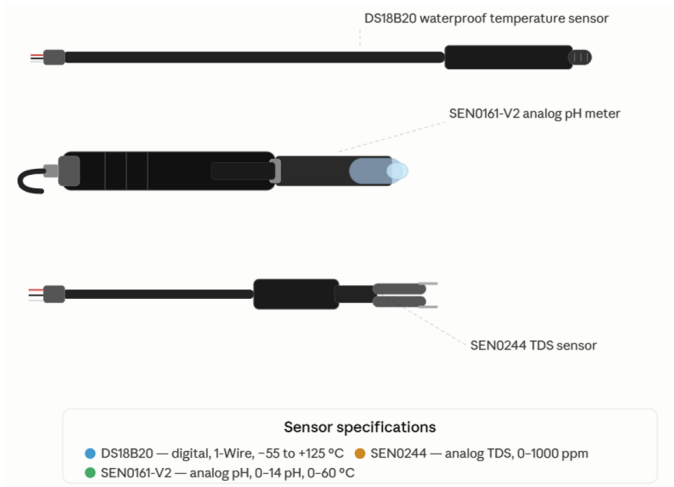


Figure 3. Environmental sensing modules integrated into the proposed monitoring framework, including the pH sensor, DS18B20 waterproof temperature sensor, and TDS sensor used for continuous water-quality acquisition.

(Pedregosa et al., 2011). Multiple classification algorithms were implemented to evaluate predictive performance under identical environmental datasets.

The machine learning models evaluated in this study included:

- Decision Tree
- Support Vector Machine (SVM)
- Random Forest
- Logistic Regression

The dataset was divided into training and testing subsets using a 60:40 partitioning strategy, where approximately 60% of the observations were used for model training and the remaining 40% were reserved for predictive testing and validation. Model training was performed iteratively to evaluate the classification capability of each algorithm for predicting aquaculture health conditions.

A representative machine learning implementation interface developed using Scikit-learn is illustrated in Figure 9.

### 3.7 Performance Evaluation

Model performance was evaluated using confusion matrix–derived classification metrics including accuracy, precision, recall, and F1-score (Figure 10). These metrics were selected because they collectively provide a broader evaluation of classifier behavior beyond simple accuracy measurements alone.

Accuracy was calculated as the proportion of correctly classified observations relative to the total number of observations. Precision represented the proportion of correctly predicted positive classifications among all positive predictions, whereas recall reflected the proportion of actual positive observations correctly identified by the classifier. The F1-score was calculated as the harmonic mean of precision and recall to provide a balanced evaluation metric for classification performance.

The confusion matrix used for performance evaluation is presented in Figure 10.

### 3.8 Comparative Model Analysis

To evaluate the effectiveness of the proposed framework, predictive results obtained from the implemented machine learning models were comparatively analyzed. The

performance of Decision Tree, Random Forest, SVM, and Logistic Regression classifiers was evaluated using identical environmental datasets and classification criteria.

A comparative summary of classification accuracy, precision, recall, and F1-score values is presented in Table III, while a graphical comparison of classifier performance is shown in Figure 11.

The final prediction output generated through the Logistic Regression model, which demonstrated comparatively superior predictive consistency within the experimental dataset, is illustrated in Figure 12.

## 4. Results and Discussion

### 4.1 Performance of the IoT-Based Aquaculture Monitoring System

The proposed IoT-driven aquaculture monitoring framework successfully performed continuous environmental sensing, wireless data transmission, cloud-based storage, and machine learning–based health prediction throughout the experimental observation period. The integrated architecture enabled real-time acquisition of temperature, pH, TDS, and electrical conductivity measurements from the aquaculture environment while simultaneously maintaining remote accessibility through the developed dashboard interface. The overall system workflow, including sensor integration, database communication, and prediction generation, operated with comparatively stable performance during the monitoring period.

The environmental data acquired from the deployed sensors were transmitted through the ESP8266-based communication module and stored within the cloud-hosted MySQL database without substantial interruption. Representative stored sensor observations retrieved from the database are shown in (Figure 7), while the web-based monitoring dashboard displaying real-time environmental conditions and predictive outputs is presented in (Figure 8). The dashboard interface allowed continuous observation of environmental fluctuations, thereby reducing dependence on manual water-quality inspection procedures.

From an operational perspective, the use of the ESP8266 platform appeared particularly advantageous because of its comparatively low hardware cost, integrated Wi-Fi capability, and relatively simple deployment requirements. Earlier aquaculture monitoring studies frequently relied on

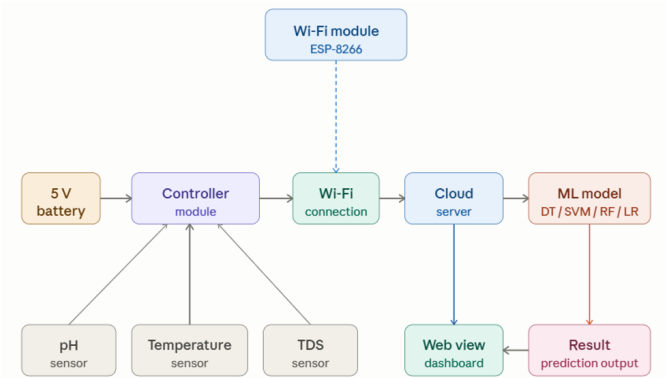


Figure 4. Overall system architecture of the proposed IoT-driven aquaculture monitoring and prediction framework showing sensor integration, ESP8266-based wireless communication, cloud database storage, machine learning analysis, and dashboard visualization.

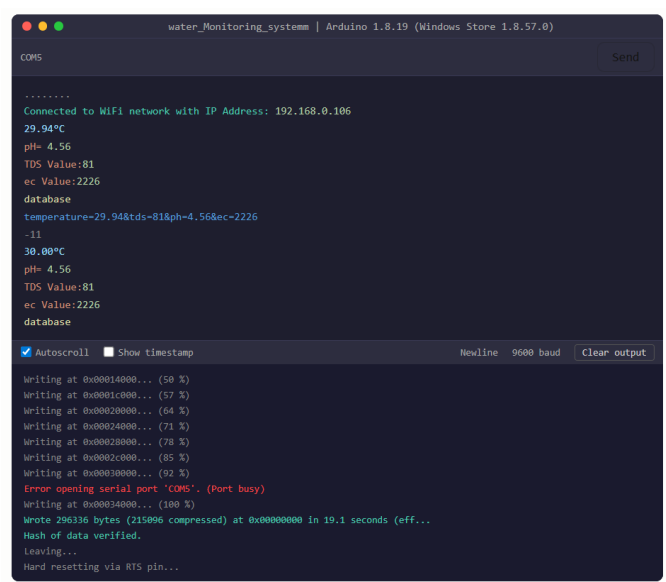
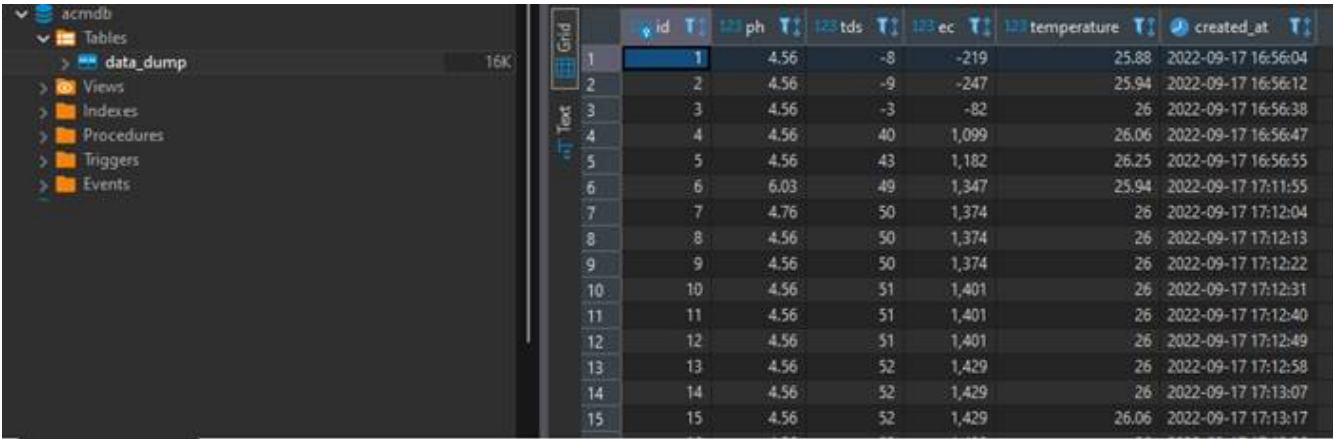


Figure 5. Arduino IDE environment illustrating real-time sensor data acquisition, preprocessing, and wireless transmission of environmental measurements from the ESP8266 microcontroller to the cloud database server.

```
if ($SERVER["REQUEST_METHOD"] == "POST") {  
    // Create connection  
    $conn = new mysqli($servername, $username, $password, $dbname);  
    // Check connection  
    if ($conn->connect_error) {  
        die("Connection failed: " . $conn->connect_error);  
    }  
    $temperature=$POST['temperature'];  
    $tdsValue=$POST['tds'];  
    $ph=$POST['ph'];  
    $ec=$POST['ec'];  
    $datetime = date('Y-m-d H:i:s');  
  
    $sql = "INSERT INTO data_dump (ph, tds, ec, temperature, created_at)  
VALUES ('$ph', '$tdsValue', '$ec', '$temperature', '$datetime')";  
    if ($conn->query($sql) == TRUE) {  
        echo "New record created successfully";  
    } else {  
        echo "Error: " . $sql . "  
";  
    }  
}
```

Figure 6. PHP-based server-side insertion script responsible for receiving sensor measurements through HTTP requests and storing environmental data within the MySQL cloud database.



|    | id | ph   | tds | ec    | temperature | created_at          |
|----|----|------|-----|-------|-------------|---------------------|
| 1  | 1  | 4.56 | -8  | -219  | 25.88       | 2022-09-17 16:56:04 |
| 2  | 2  | 4.56 | -9  | -247  | 25.94       | 2022-09-17 16:56:12 |
| 3  | 3  | 4.56 | -3  | -82   | 26          | 2022-09-17 16:56:38 |
| 4  | 4  | 4.56 | 40  | 1,099 | 26.06       | 2022-09-17 16:56:47 |
| 5  | 5  | 4.56 | 43  | 1,182 | 26.25       | 2022-09-17 16:56:55 |
| 6  | 6  | 6.03 | 49  | 1,347 | 25.94       | 2022-09-17 17:11:55 |
| 7  | 7  | 4.76 | 50  | 1,374 | 26          | 2022-09-17 17:12:04 |
| 8  | 8  | 4.56 | 50  | 1,374 | 26          | 2022-09-17 17:12:13 |
| 9  | 9  | 4.56 | 50  | 1,374 | 26          | 2022-09-17 17:12:22 |
| 10 | 10 | 4.56 | 51  | 1,401 | 26          | 2022-09-17 17:12:31 |
| 11 | 11 | 4.56 | 51  | 1,401 | 26          | 2022-09-17 17:12:40 |
| 12 | 12 | 4.56 | 51  | 1,401 | 26          | 2022-09-17 17:12:49 |
| 13 | 13 | 4.56 | 52  | 1,429 | 26          | 2022-09-17 17:12:58 |
| 14 | 14 | 4.56 | 52  | 1,429 | 26          | 2022-09-17 17:13:07 |
| 15 | 15 | 4.56 | 52  | 1,429 | 26.06       | 2022-09-17 17:13:17 |

Figure 7. Representative environmental observations stored in the MySQL database following real-time acquisition and wireless transmission from the IoT-based aquaculture monitoring system.

larger multi-sensor infrastructures or more hardware-intensive systems (Raju & Varma, 2018; Simbeye & Yang, 2014). In contrast, the present framework attempted to maintain a balance between affordability and predictive functionality by selecting a smaller number of environmentally representative parameters.

Interestingly, despite the comparatively simplified sensing architecture, the system still demonstrated substantial predictive capability. This observation perhaps reinforces the idea that carefully selected environmental parameters may provide sufficient ecological representation for preliminary aquaculture health prediction without necessarily requiring excessively large sensor networks.

#### **4.2 Environmental Parameter Behavior and Aquaculture Health Interpretation**

Throughout the monitoring period, temperature, pH, TDS, and electrical conductivity values exhibited measurable fluctuations across the aquaculture samples. These fluctuations were expected because aquatic ecosystems are naturally dynamic environments influenced by biological metabolism, photosynthetic activity, dissolved mineral interactions, and surrounding environmental conditions.

Among the monitored variables, temperature remained one of the most influential parameters affecting the broader environmental stability of the aquaculture system. Previous studies have already established that temperature directly affects dissolved oxygen concentration, chemical reaction rates, metabolic activity, and aquatic stress responses (Boyd, 1982; Bhatnagar et al., 2004). During the present study, temperature variations appeared to correspond with changes in other measured parameters, particularly conductivity and pH. Although the current experimental framework did not directly quantify dissolved oxygen concentration, the observed environmental relationships still aligned with previously reported aquatic ecological behavior.

The pH measurements also demonstrated moderate temporal variability. This pattern was not entirely unexpected, since pond water pH is known to fluctuate according to phytoplankton photosynthesis and carbon dioxide dynamics within aquatic ecosystems (Delince, 2013). In some instances, elevated pH conditions appeared to coincide with higher conductivity and TDS values, suggesting broader chemical shifts within the monitored environment. Such interactions further support

earlier findings that pH is closely associated with multiple environmental variables and may serve as a meaningful predictive indicator of aquatic health conditions (Bhatnagar & Devi, 2013).

TDS and electrical conductivity measurements additionally reflected changes in dissolved ionic concentration within the aquaculture samples. Higher TDS levels generally indicate increased dissolved mineral or salt content, which may influence fish metabolism and aquatic stability if concentrations exceed acceptable thresholds (Bhatnagar & Devi, 2013). Conductivity measurements demonstrated similar environmental sensitivity, likely because EC and TDS maintain strong physicochemical relationships through ionic concentration dynamics. These observations collectively suggest that the selected parameters were sufficiently responsive to environmental variation and therefore appropriate for machine learning–based predictive analysis.

The healthy, moderate, and unhealthy classification thresholds used during environmental interpretation are summarized in (Table I). These predefined ranges allowed the environmental observations to be categorized systematically prior to machine learning implementation.

#### **4.3 Machine Learning Classification Performance**

The machine learning component of the proposed framework was developed to determine whether environmental sensor data could be used not merely for observation, but also for predictive classification of aquaculture health conditions. Four classification algorithms—Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression—were implemented and comparatively evaluated using identical environmental datasets.

The machine learning implementation environment developed using Scikit-learn is illustrated in (Figure 9), while the confusion matrix used for classification analysis is shown in (Figure 10).

The comparative classification results demonstrated that predictive performance varied noticeably among the implemented algorithms (Table III). Among the evaluated classifiers, Logistic Regression achieved the highest overall classification accuracy at approximately 91%, whereas Random Forest and SVM demonstrated comparatively strong F1-score performance values of approximately 86.5% and 84.8%, respectively. Decision

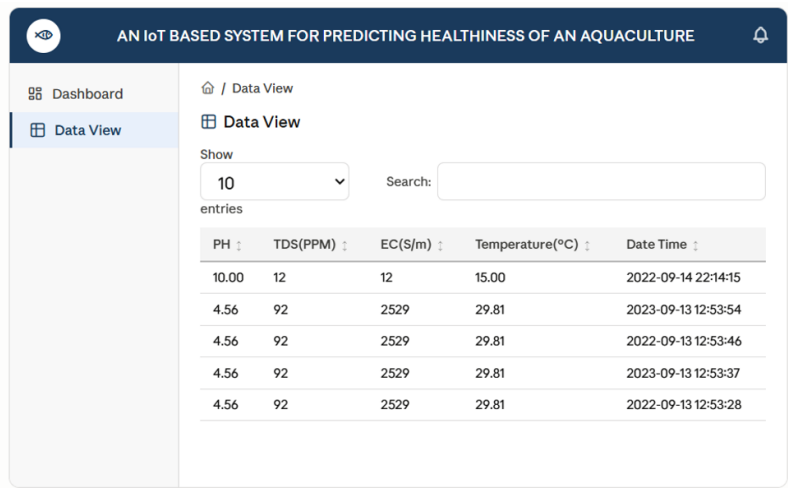


Figure 8. Web-based monitoring dashboard displaying real-time visualization of environmental parameters and machine learning–based aquaculture health predictions generated by the proposed framework.

```
1 import pandas as pd
2 from sklearn import linear_model
3
4 df = pd.read_csv("water_monitoring_dataset.csv")
5
6 X = df[['ph', 'tds', 'ec', 'temperature']]
7 y = df['result']
8
9
10 reg = linear_model.LinearRegression()
11 reg.fit(X, y)
12
13 result = reg.predict([[ph, tds, ec, temperature]])
```

Fig. 9. Machine learning implementation environment developed using the Scikit-learn framework within Jupyter Notebook for aquaculture health classification and predictive analysis.

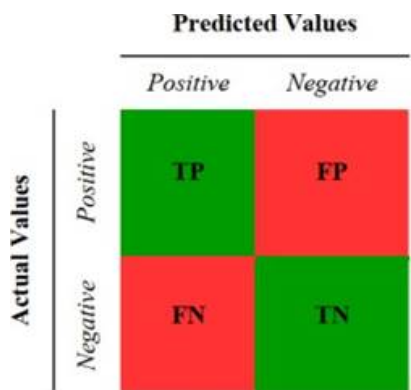


Figure 10. Confusion matrix used for evaluating classification performance of the implemented machine learning algorithms based on accuracy, precision, recall, and F1-score metrics.

Table 3. Comparative classification performance of the implemented machine learning models for aquaculture health prediction. The table summarizes the accuracy, precision, recall, and F1-score values obtained using Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression classifiers.

| #Classifiers        | Accuracy | Precision | Recall | F1 Score |
|---------------------|----------|-----------|--------|----------|
| Decision Tree       | 70.34 %  | 63 %      | 64.6 % | 64.5 %   |
| SVM                 | 72 %     | 81 %      | 84.4 % | 84.8 %   |
| Random Forest       | 71.22 %  | 85.6 %    | 86.5 % | 86.5 %   |
| Logistic Regression | 91 %     | 81 %      | 81.6 % | 81.9 %   |

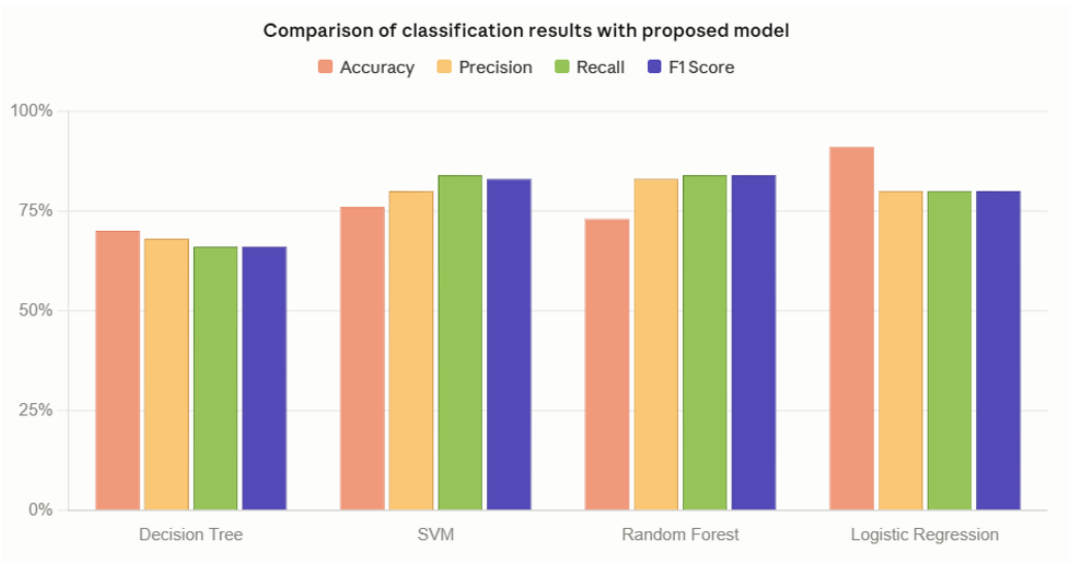


Figure 11. Graphical comparison of classification performance among the implemented machine learning algorithms based on predictive accuracy and evaluation metrics for aquaculture health assessment.

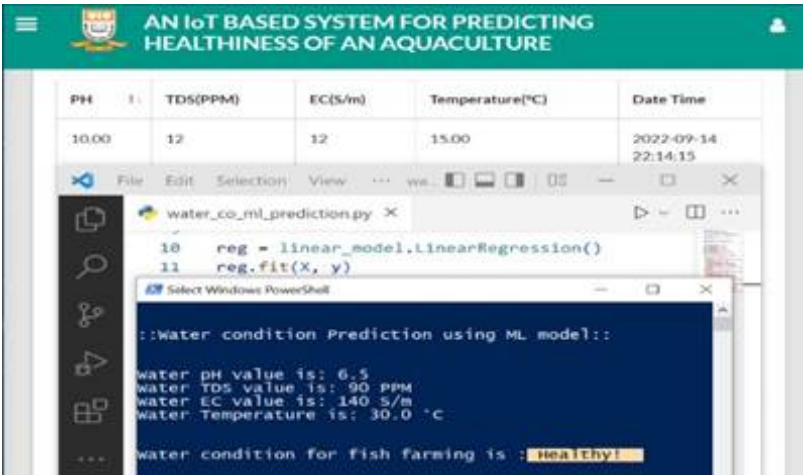


Figure 12. Representative machine learning prediction output demonstrating real-time classification of aquaculture health conditions generated using the proposed IoT-based monitoring and predictive framework.

Tree classification produced lower overall predictive performance relative to the other models.

Interestingly, although Random Forest and SVM demonstrated relatively balanced precision and recall values, Logistic Regression still achieved superior overall classification accuracy within the present dataset. This result may initially appear somewhat unexpected because ensemble-based methods such as Random Forest are often considered highly effective for environmental classification tasks. However, the comparatively limited dataset size and the relatively structured distribution of the selected environmental variables may have favored the simpler decision boundary characteristics of Logistic Regression.

The graphical comparison among classifiers presented in (Figure 11) further illustrates the relative performance differences across the evaluated machine learning models.

From a broader analytical perspective, the results suggest that even relatively lightweight machine learning models may provide meaningful predictive capability when trained using carefully selected environmental variables. This is particularly relevant for low-resource aquaculture environments where computational simplicity and operational efficiency remain important practical considerations.

#### 4.4 Prediction of Aquaculture Healthiness

The final prediction output generated by the proposed framework is illustrated in (Figure 12). The predictive system successfully classified aquaculture conditions into predefined health categories based on incoming environmental sensor measurements. These predictions were displayed in near real time through the developed dashboard interface, enabling continuous environmental assessment without requiring manual interpretation of individual sensor values.

One notable aspect of the proposed framework is that predictive analysis was achieved using only four environmental parameters. Earlier monitoring systems often incorporated substantially larger sensor combinations involving dissolved oxygen, nitrate concentration, turbidity, ammonia, salinity, and additional environmental indicators (Raju & Varma, 2018; Simbeye & Yang, 2014). While such comprehensive monitoring architectures undoubtedly provide broader ecological

information, they may also increase operational cost and system complexity.

The present findings suggest that simplified parameter selection strategies may still produce comparatively strong predictive outcomes when the selected variables maintain significant ecological interrelationships. Temperature, pH, TDS, and conductivity are all strongly interconnected within aquatic ecosystems, and changes in one parameter frequently influence broader environmental behavior (Boyd, 1982; Delince, 2013). Consequently, the predictive capability observed in this study may partly arise from these underlying environmental dependencies.

Nevertheless, some limitations should also be acknowledged. The dataset used for machine learning analysis was generated over a relatively short monitoring duration and involved only three aquaculture samples. Although the obtained predictive accuracy was encouraging, larger datasets collected over longer temporal periods and across more diverse aquaculture environments would likely improve model generalizability and predictive robustness. Environmental conditions in real-world aquaculture systems may vary substantially according to climate, species composition, feeding practices, seasonal changes, and microbial dynamics. Therefore, future investigations should incorporate broader datasets and additional environmental parameters to improve predictive stability and ecological representation.

Another important consideration involves the absence of dissolved oxygen sensing within the current framework. DO is widely recognized as one of the most critical indicators of aquatic health and fish survivability (Bhatnagar & Devi, 2013). Although the selected parameters indirectly reflect broader environmental conditions, incorporation of DO sensing in future system versions may improve ecological sensitivity and predictive precision further.

Despite these limitations, the proposed framework demonstrated that low-cost IoT infrastructure integrated with machine learning algorithms can provide practical and comparatively accurate aquaculture health prediction capability. The system successfully combined environmental sensing, wireless communication, cloud-based monitoring, and predictive analytics into a unified operational platform capable of supporting intelligent aquaculture management strategies.

## 5. Conclusion

The present study demonstrated the development and implementation of a low-cost IoT-driven aquaculture monitoring and prediction framework integrating environmental sensing, wireless communication, cloud-based storage, and machine learning-based analysis. By continuously monitoring temperature, pH, total dissolved solids, and electrical conductivity, the proposed system was able to provide real-time environmental assessment and predictive classification of aquaculture health conditions. The experimental findings suggest that carefully selected environmental parameters may provide meaningful ecological representation without requiring excessively complex sensing infrastructures. Among the evaluated machine learning models, Logistic Regression demonstrated the highest overall classification accuracy, achieving approximately 91% predictive performance under the experimental dataset. The integration of IoT technologies with predictive analytics therefore appears promising for improving operational efficiency and supporting more intelligent aquaculture management strategies. Nevertheless, broader datasets, longer monitoring durations, and incorporation of additional environmental variables such as dissolved oxygen may further strengthen predictive robustness and ecological sensitivity in future implementations.

## Acknowledgements

The authors sincerely thank the Department of Computer Science & Engineering, United International University, Dhaka, Bangladesh, for providing the necessary laboratory facilities and technical support throughout this study. The authors also extend their gratitude to all individuals who contributed to data collection and system testing during the experimental phase of this research.

## Conflict of Interest

The authors declare no conflict of interest.

## Financial Disclosure

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors confirm that no financial support or sponsorship influenced the design, conduct, reporting, or publication of this study.

## References

Bhatnagar, A., & Devi, P. (2013). Water quality guidelines for the management of pond fish

culture. *International Journal of Environmental Sciences*, 3(6), 1980–2009.

Bhatnagar, A., Jana, S. N., Garg, S. K., Patra, B. C., Singh, G., & Barman, U. K. (2004). Water quality management in aquaculture. In *Course manual of summer school on development of sustainable aquaculture technology in fresh and saline waters* (pp. 203–210). CCS Haryana Agricultural University.

Boyd, C. E. (1982). *Water quality management for pond fish culture*. Elsevier Scientific Publishing Company.

Delince, G. (2013). *The ecology of the fish pond ecosystem: With special reference to Africa* (Vol. 72). Springer Science & Business Media.

DFRobot. (n.d.-a). SEN0161-V2 analog pH meter. <https://www.dfrobot.com/product-1782.html>

DFRobot. (n.d.-b). DS18B20 waterproof temperature sensor. <https://www.dfrobot.com/product-689.html>

DFRobot. (n.d.-c). SEN0244 TDS sensor. <https://www.dfrobot.com/product-1662.html>

Espressif Systems. (n.d.). ESP8266 Wi-Fi MCU. <https://www.espressif.com/en/products/socs/esp8266>

Gondchawar, N., & Kawitkar, R. S. (2016). IoT based smart agriculture. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(6), 838–842.

Israni, S., Meharkure, H., & Yelore, P. (2015). Application of IoT based system for advance agriculture in India. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(11), 11586–11592.

Kayalvizhi, S., Reddy, K., Kumar, V., & Prasanth, V. (2015). Cyber aqua culture monitoring system using Arduino and Raspberry Pi. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 4, 2320–3765.

Patil, M. K., Patil, M. S., Patil, M. S., & Patil, M. V. (2015). Monitoring of turbidity, pH and temperature of water based on GSM. *International Journal for Research in Emerging Science and Technology*, 2(3), 16–21.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Raju, K. R. S. R., & Varma, G. H. K. (2018). Knowledge based real time monitoring system for aquaculture using IoT. In *Proceedings of the International Conference on Innovations in Science, Engineering and Technology (ICISSET)* (pp. 318–321).

Saha, S., Rajib, R. H., & Kabir, S. (2017). IoT based automated fish farm aquaculture monitoring system. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 201–206). IEEE.

Simbeye, D. S., & Yang, S. F. (2014). Water quality monitoring and control for aquaculture based on wireless sensor networks. *Journal of Networks*, 9(4), 840–849.

United Nations Food and Agriculture Organization. (2012). *The state of world fisheries and aquaculture 2012*. FAO. <https://www.fao.org/3/i2727e/i2727e00.htm>

Varma, G. H. K., & Raju, K. R. S. R. (2018). Knowledge based real time monitoring system for aquaculture using IoT. *International Conference on Innovations in Science, Engineering and Technology*, 318–321.

WEMOS. (2004). LOLIN D1 mini. [https://www.wemos.cc/en/latest/d1/d1\\_mini.html](https://www.wemos.cc/en/latest/d1/d1_mini.html)

Zhu, H.-J. (2010). Global fisheries development status and future trend analysis. *Taiwan Economic Research Monthly*, 33, 1–8.