Monitoring water quality metrics of ponds with IoT sensors and machine learning to predict fish species survival

Article *in* Microprocessors and Microsystems · September 2023

DOI: 10.1016/j.micpro.2023.104930

Md. Monirul Islam
Hankuk University of Foreign Studies

Salem A. Alyami
Imam Mohammad ibn Saud Islamic University

Microprocessors and Microsystems 102 (2023) 104930

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Monitoring water quality metrics of ponds with IoT sensors and machine learning to predict fish species survival

Md. Monirul Islam [a,*], Mohammod Abul Kashem [b], Salem A. Alyami [c], Mohammad Ali Moni [d]

[a] Department of Software Engineering, Daffodil International University, Daffodil Smart City (DSC), Birulia, Savar, Dhaka 1216, Bangladesh
[b] Department of Computer Science and Engineering, Dhaka University of Engineering and Technology, Gazipur 6700, Bangladesh
[c] Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 13318, Saudi Arabia
[d] AI and Cyber Futures Institute, Charles Stuart University, Bathurst, 2795, NSW, Australia

ARTICLE INFO

Keywords:
Smart fish farming
IoT-based ponds' water monitoring
Machine learning-based fish farming
Fish survival prediction

ABSTRACT

Aquaculture involves cultivating various marine and freshwater aquatic creatures within regulated environments. Monitoring the aquatic environmental conditions in real-time is crucial for successful fish farming. The Internet of Things (IoT) offers significant potential for real-time monitoring, and this paper introduces an IoT framework designed for efficient monitoring and effective control of various water-related aquatic environmental parameters. The proposed system is implemented as an embedded system utilizing sensors and an Arduino microcontroller. In cultivating pond water, diverse sensors such as pH, temperature, and turbidity sensors are deployed, with each sensor connected to an Arduino Uno-based microcontroller board. These sensors collect data from the water, which is then stored as a CSV file in an IoT cloud platform called ThingSpeak through the Arduino microcontroller. To gather data for analysis, we conducted measurements across five ponds, varying in size and environmental conditions. After getting the real-time data, we compared our experimental results with the standard reference values. As a result, we could take the decision of whether a pond is suitable for cultivating fish or not. After that, we labeled the data with 11 fish categories: Katla, sing, prawn, shrimp, rui, tilapia, pangas, karpio, magur, silver carp, and koi. The data was analyzed using 10 machine learning (ML) algorithms, including J48, Random Forest, K Nearest Neighbors (K-NN), K*, Logistic Model Tree (LMT), Reduced Error Pruning Tree (REPTree), Jumping Rule Inference with Pruned Search (JRIP), Partial Decision Trees (PART), Decision Table, and Logit boost. After experimental analyses, it was discovered that only three of the five ponds were ideal for fish farming, and those three ponds only met the required standards for pH, Temperature, Turbidity, and Conductivity. Among the state-of-art machine learning algorithms, Random Forest achieved the highest score of performance metrics as accuracy 94.42%, kappa statistics 93.5%, and Avg. TP Rate 94.4%. In addition, we calculated the Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Dissolved Oxygen (DO) for one scenario. This study includes prototype hardware details of the proposed IoT system.

1. Introduction

Aquaculture means agriculture primarily for the food of aquatic animals or plants or pearls. When it is for only fish farming, then it is called aqua fisheries. There are many types of aqua fisheries, such as extensive fish farming, intensive fish farming, ditch fish farming, cage system, etc. Extensive fish farming occurs in a pond where intensive fish farming is farming in closed circulation water. In a ditch system, water retaining is the main requirement of this system, and it is also a crucial point to keep electrolytes for fish corrected. And cage is a system where it confines the fish in a mesh enclosure. Overall, aquaculture contains cultivating, nourishing, and harvesting fresh and saltwater fish, mollusks, crustaceans, and seedlings. The practice began in China approximately 4000 years ago, and international production remains undermined by China and other Asian states. Aquaculture is used by some of the global poor populations and by major corporations to harvest food. Aquaculture now provides more than half of all human consumption of seafood, a proportion that continues to increase in terms of the global population. Aquaculture produced 3 million tons of food in the seventies, steadily increasing to more than 80 million tons in 2017, per the Food and Agriculture Organization (FAO) [1]. We depend on fish for the various vitamins source. Some of them are

* Corresponding author.
E-mail address: monirul.swe@diu.edu.bd (M.M. Islam).

https://doi.org/10.1016/j.micpro.2023.104930
Received 1 December 2022; Received in revised form 23 August 2023; Accepted 19 September 2023
Available online 23 September 2023

protein, vitamin D and omega-3 fatty acids, which are helpful to fit the body and brain as well as reducing the risk of heart disease and supporting pregnant women's health.

Fish depends on water quality factors. So, maintaining the quality of water plays a vital role in fish farming. There are some metrics of water named PH, temperature, turbidity, BOD, COD, and DO. PH is an important factor in water. It is a scale of 0–14. DO is another important parameter for fish farming. We have to check the level of it for fish survival. The standard recommended value is 5 mg/l for optimum fish health. Most DOs in ponds are generated by aquatic plants and algae during photosynthesis. This is why DO increases before dawn, decreases at night, and is lowest shortly before dawn. DO levels below 5 mg/L can be hazardous to fish, whereas surface gulping air can be seen if DO falls below 2 mg/L. An electronic oxygen meter or a chemical test kit may be used to measure DO. When DO falls below 4 mg/L or ambient circumstances promote an oxygen depletion event, emergency aeration should be provided [2]. Another critical parameter is BOD for fish farming [3]. It is the quantity of oxygen necessary for the biological breakdown of organic matter in bodies of water. Generally, the BOD is a pollution measure used to determine the quality of effluent or wastewater.

If the cultivation system depends on the digital system, it will benefit the people more because it is cost-effective and costs approximately BDT 15 000. IoT can contribute to this field by monitoring quality factors digitally. We used an IoT framework to monitor these quality factors in this paper. Every device is connected to the Internet [4].

We derived the real-time values of water quality by using our IoT framework. After that, we used machine learning to validate the dataset. We can say that machine learning is a sub-field of artificial intelligence. It can learn from data without any explicitly programmed [5]. The work experiments on water's temperature, pH, conductivity, BOD, DO, COD, and turbidity parameters.

Monitoring work holds significant importance and serves as a motivating factor for conducting this work. The health and survival of fish species in ponds rely heavily on the quality of their aquatic environmental factor, including temperature, pH levels, dissolved oxygen, and pollutant concentrations. By deploying IoT sensors, we can continuously and accurately measure these water quality parameters in real-time, enabling timely interventions to maintain optimal conditions for fish survival. Integrating machine learning algorithms with the collected sensor data allows us to analyze complex relationships and patterns that may impact fish species survival. By training predictive models using historical data on water quality and corresponding fish species survival rates, we can develop a reliable tool to forecast potential risks to fish populations. This predictive capability offers crucial insights to pond managers, enabling them to take proactive measures to mitigate adverse conditions and enhance the overall well-being of the aquatic ecosystem.

The contribution of this paper is summarized as follows.

- We created an IoT framework to collect real-time values of five ponds of water using the PH, Temperature, and Turbidity IoT sensors and calculated the conductivity values from the temperature values.
- We studied the life cycle of 11 fish species and compared the real-time values with the traditional values and compared which pond is suitable for fish farming.
- We also predicted the fish survival using ten machine learning algorithms after labeling the data from the cloud. Among these models, Random forest outperforms the performance metrics.
- We briefly presented a background study on hardware and machine learning processes.

The rest of the article is structured as follows. Section 2 presents a literature review. In Section 3, the methodology is presented, divided into subsections like the IoT framework and machine learning parts. In Section 4, the result and discussion are analyzed, and finally, Section 5 concludes the paper.

2. Literature review

Currently, some researchers have applied IoT devices for monitoring remotely in agriculture, aquaculture, smart city, smart medical, smart home, and so on [6,7]. Here, we discussed the works of some researchers in water quality monitoring for fish farming. The authors implemented an IoT system using only 2 IoT sensors named PH and temperature for fish farming. Authors also measured ammonia and dissolved oxygen using kits [8]. In [9], researchers presented a web-based IoT system for guppy fish farming using only two sensors called PH and salinity. In paper no [10], using 4 sensors named DO, PH, temperature and salinity, the authors implemented an IoT system for only one fish mentioned, Pangasius, in Mekong Delta. Authors proposed an intelligent fish farming system using IoT devices to monitor the 3 water quality factors: pH, temperature, water level, and oil layer [11]. In [12], the authors implemented an IoT system using carbon monoxide, pH, temperature, water level and turbidity sensors for fish farming and utilized machine learning for the secondary water quality dataset. The researchers made an IoT framework for fish farming using pH, temperature, turbidity, conductivity, and depth sensors to check the water quality [13]. This paper presented an IoT framework for monitoring the water quality of a northern part of Padma River and drinking tube-well water using pH, temperature, and turbidity [14]. The authors designed an IoT-based framework using pH, temperature, and dissolved oxygen for Indian Aquaculture [15]. Juan Huan et al. [16] presented a narrow-band IoT framework-based water quality monitoring using temperature, pH, and dissolved oxygen for pond aquaculture. Chien Lee and Yu-JenWang [17] proposed an IoT-based water monitoring system using temperature, depth, dissolved oxygen, and pH value for aquaponics. Chiu, Min-Chie, et al. [18] presented an IoT framework using turbidity, temperature, underwater oxygen level, and pH sensors for fish farming. They also analyzed their real-time dataset using deep learning techniques. The researchers [19] developed an IoT-based water quality monitoring system using temperature, conductivity, water level, pH, turbidity and dissolved oxygen for fish farming. They also analyzed the fish pond data using machine learning models. In [20], the authors implemented a water quality monitoring system using IoT devices containing temperature, pH, turbidity, and dissolved oxygen for aquaculture. In [21], the authors proposed an IoT-based smart fish farming system using oil layer, water level, and temperature sensors. In [22], the researchers implemented a handheld meter for measuring the turbidity values of water. In [23], the authors presented an IoT-based system for monitoring water quality factors like PH, total dissolved solids, turbidity, and temperature in various locations in St. Petersburg, Russia. In [24], the authors proposed a Machine learning approach to look into the effectiveness of water quality (WQ) on prawn fish only in freshwater ponds in Australia. They also collected harvest data as well as WQ data to check the variations of WQ on the harvest outcomes of prawns. Table 1 shows the overall comparison of all the related works.

In this paper, we implemented an IoT framework using the pH, temperature, turbidity, and conductivity sensors for fish farming. In addition, we tested BOD, DO, and COD using the kits at Dhaka University of Engineering and Technology. We also applied machine learning models to analyze the real-time dataset.

3. Methods

Fig. 1 exhibits the proposed methodology. First of all, we introduce all the aquatic sensors such as pH sensor, temperature sensor, turbidity sensor, etc. Each sensor is linked to an Arduino Uno through an Ethernet shield using a different jumper on the breadboard. A cloud server is linked to this system via the Rest-API.

Then, we use 10 machine learning models including J48, K-NN, random forest, K*, LMT, PART, JRIP, decision table, logit boost, and REP-Tree for the fish classification. There are mainly two segments in our task. They are the hardware portion for implementing the IoT architecture and the machine learning portion to predict the survival of fish species.

Table 1
Comparison of our works with existing works.

Source	Parameters	Limitations
[8]	PH and temperature, ammonia, DO	Only four parameters. No ML techniques are used.
[9]	PH and salinity	Only two parameters. No ML techniques are used.
[10]	DO, PH, temperature, and salinity	Only four parameters and only for Pangasius fish. No ML techniques is used.
[11]	pH, temperature, water level	Only three parameters. No ML techniques is used.
[12]	CO, pH, temperature, water level and turbidity	They used secondary dataset for machine learning.
[13]	pH, Temperature, turbidity, conductivity, depth	No primary dataset and no ML techniques is used.
[14]	pH, temperature, turbidity	No primary dataset and no ML techniques is used.
[15]	pH, temperature, and dissolved oxygen	No primary dataset and no ML techniques is used.
[13]	Temperature, pH, dissolved oxygen	No primary dataset and no ML techniques is used.
[14]	Temperature, depth, dissolved oxygen, and pH	No primary dataset and no ML techniques is used.
[15]	Turbidity, temperature, underwater oxygen level, and pH	No primary dataset and no ML techniques is used.
[24]	No sensors are used	Secondary dataset and ML is used

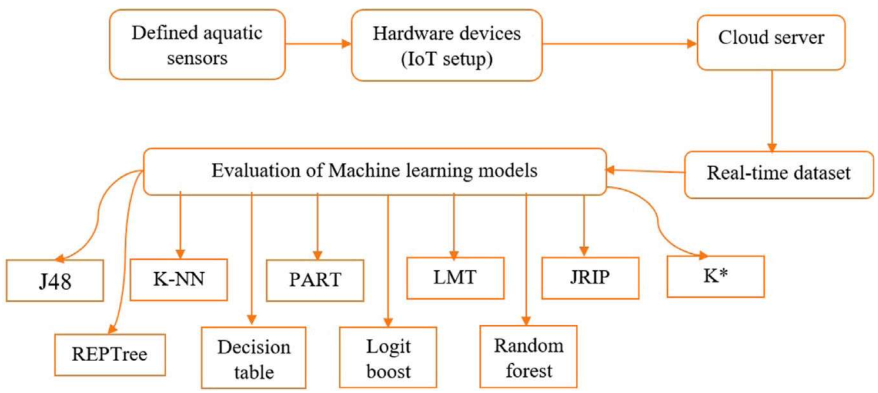

Fig. 1. A detailed block diagram of the methods.

3.1. IoT system implementation

In the IoT system, 4 sensors including a ph sensor, temperature sensor, turbidity sensor, and ultrasonic sensor are used for collecting real-time values of each pond water. Besides, we experimented with the BOD, COD, and DO tests at the Dhaka University of Engineering and Technology (DUET). The visualization of our IoT framework with the proposed procedure is shown in 2. All hardware devices are shown in Fig. 3.

3.1.1. Arduino Uno

In contrast to the ATmega328P, the Arduino Uno refers to a microcontroller board that has 14 modernized data/yield pins, a sixteen mega Hartz quartz crystal, a universal serial board link, a jack power, in-circuit serial programming, and a reset pin. It comes with all where the microcontroller needs to start up by connecting to a PC through USB or an AC-to-DC adapter. It is used in this work for reading data signals from various sensors shown in 3(a).

3.1.2. Ethernet shield

The ethernet shield is shown in Fig. 3(b) allows easily connecting Arduino Uno to the Internet. This shield enables the Arduino to send data to a cloud server and to receive data from sensors with an Internet connection. The Wiznet W5100 ethernet chip is used in this shield. This chip contains the PHY, MAC, IP, and TCP layers. Hardware implementation on the chip is the merits of using the shield through the ENC28J60. Transfer control protocol or internet protocol must be executed on the microcontroller where it is connected to the ENC28J60 chip. It is used in this paper for allowing the Arduino board to interface with the internet.

3.1.3. PH sensor

The potential of hydrogen (pH) meter is a systematic tool that deals with the hydrogen-ion movement in aquatic-based solutions to determine their tartness or alkalinity, which is said as pH. A pH meter is made up of two basic components: a **pointer** that moves against a scale and a **digital meter** that takes value from the resources and displays it numerically via a circuit board. In our work, we have a pH sensor which is a digital meter that we use to test the acidity of the water. We built a circuit board and connect it to an Arduino shown in Fig. 3(c). There is some code in Arduino that works with a pH meter. The meter has a 14 scale from 0 to 14, and it calculates the acidic or alkalinity of a solution. The ideal range of pH for fish farming in a pond is between 6.5 and 8.5. If the pH value is below 4 and above 11, then it is the death point for fish due to the acidity and alkalinity. When the range of pH is 4–5, no reproduction of fish occurs, and when it is in the range of 4–6.5, and 8.5–10, slow growth will be for fish. So it is the major survival element of water.

3.1.4. Temperature sensor

The single wire protocol digital temperature sensor, such as the DS18B20 model can calculate the temperature on the scale of −55 °C to +125 °C with ±5 percent precision. The data established from the 1-wire is in the 9-bit to 12-bit range. This sensor may be controlled by a single pin on a microcontroller since it follows the single-wire protocol. The sensor can be programmed with a 64-bit serial code for the progressive level protocol, allowing multiple sensors to be controlled from a single microcontroller pin. There is 3 colored pin named black color, red color, and yellow color. A black pin is used to connect to the GND. The red pin is known as VCC which varies from 3.3 V or 5 V. Yellow pin supplies the output shown in Fig. 3(d).

3.1.5. Turbidity sensor

The amount of suspended particles in a stream is measured by its turbidity. Our water may contain soil, or our milkshake may have chocolate chips. Despite our desire for chocolate in our beverages, earth particles are not at all acceptable. Water is utilized in a broad variety of industrial and residential contexts in addition to potable

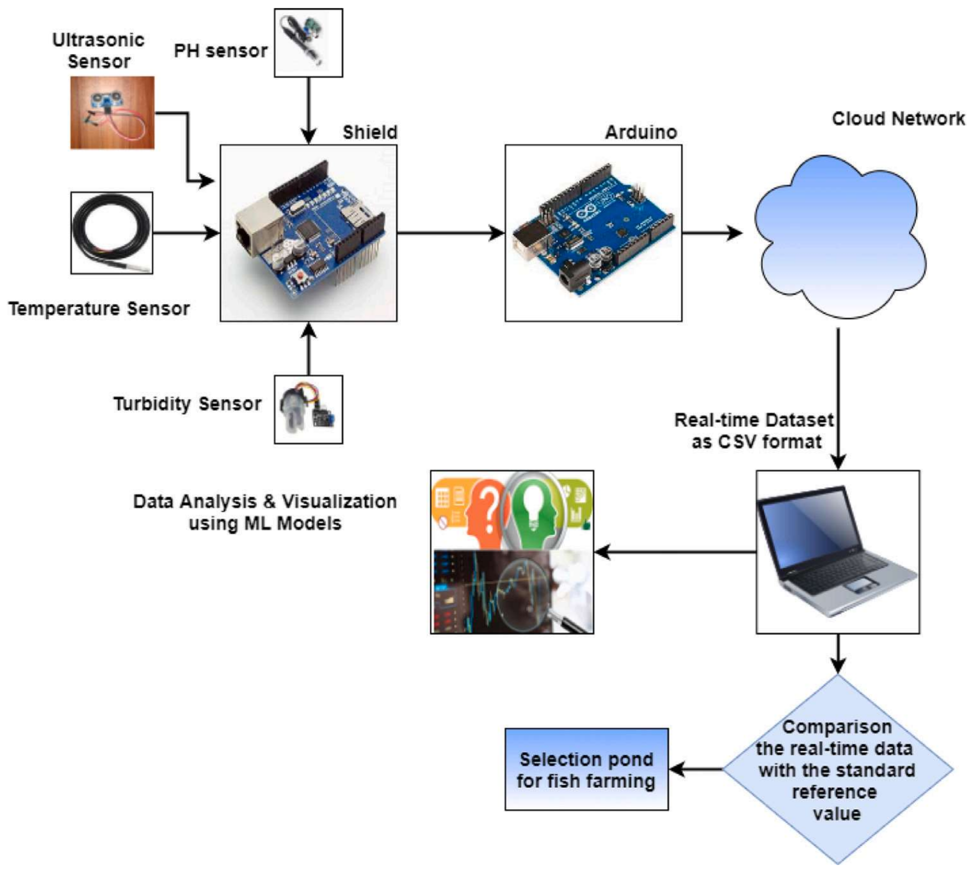

Fig. 2. Visualization of IoT framework with methodology.

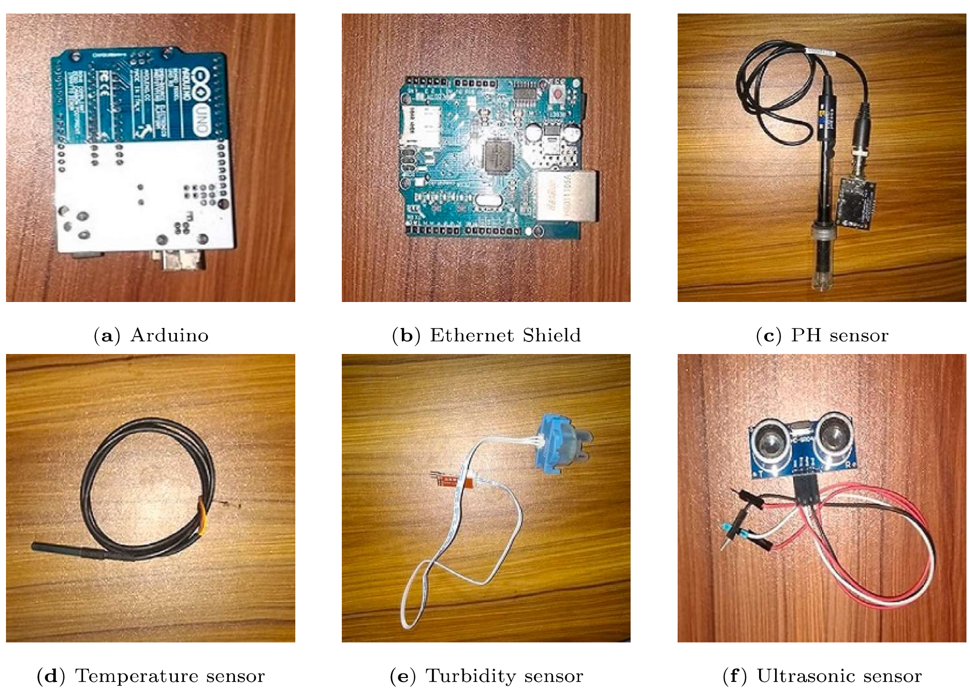

Fig. 3. Hardware devices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

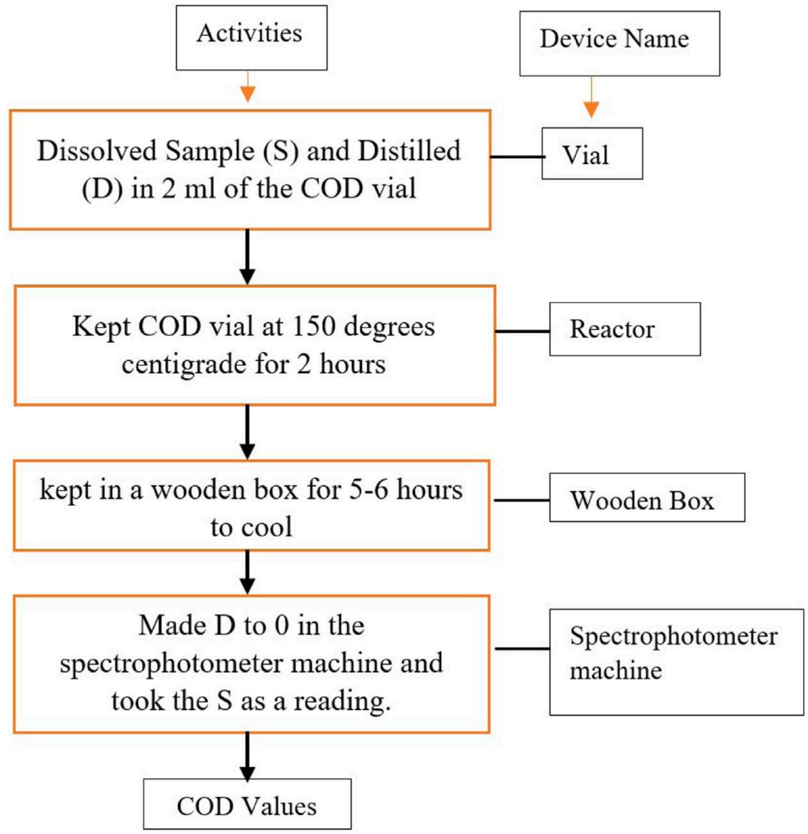

Fig. 4. Step wise COD experiments.

purposes. For instance, water is used to clean a car's windshield, to cool a power plant's reactors, and to the extent that washing machines and dishwashers depend on it, as do fish. It is used in this work to find out the suspended particles in water. This sensor is shown in Fig. 3(e).

3.1.6. Ultrasonic sensor

Ultrasonic sensors are excellent instruments for measuring distance without making physical contact, and they are used in a variety of applications such as water level measurement, distance measurement, and so on. This is a quick and accurate way to calculate small distances. We used this sensor to assess the distance of a hindrance from the sensor in the thesis. The ECHO principle is the foundation of ultrasonic distance measurement. When sound waves are transmitted in the area, they return to their source as ECHO after colliding with an obstacle. So, after striking the barrier, we just need to measure the travel time of both sounds, i.e., the outgoing time and the return time to the origin. Since we know the speed of sound, we can measure the distance with a little math. This type of HC-SR04 is a sensor shown in Fig. 3(f) that is primarily handled to assess the distance between the goal object and the sensor. It uses non-contact technology, which means there is no direct contact between the sensor and the object being measured.

3.1.7. DO, COD and BOD

In addition, we find out the biochemical oxygen demand, chemical oxygen demand, and dissolved oxygen of water in Pond 1. Dissolved Oxygen (DO) is oxygen that is dissolved in water. After the experiment with the DO, we get 6.79 mg/L.

Chemical Oxygen Demand (COD) is the amount of oxygen necessary to oxidize all soluble and insoluble organic compounds existing in a volume of water. Its value is usually stated in milligrams per liter of water (mg/L). For calculating the value of COD, in the vial of COD, the sample (S) and distilled (D) are dissolved in 2 ml of the COD vial and kept in the COD reactor at 150 degrees centigrade for 2 h. The vial is kept in a wooden box for 5–6 h to cool. To get the COD value, we make D to 0 in the spectrophotometer machine and take the S as a reading. Thus we get the COD values. we get the 12 milligrams per liter of water (mg/L) as the COD value. The step-wise COD experiment is shown in Fig. 4.

For microbial breakdown (oxidation), Biochemical Oxygen Demand (BOD) quantifies the quantity of oxygen required or consumed in water. For measuring the BOD values, we take the sample according to the COD values. Because BOD depends on COD values. In the OxiTop bottle, we keep the sample of BOD and there is a black dropper in the cap of the bottle. Then, two granules of sodium hydroxide are placed in it. After then, we do the zero level of the bottle scale. After labeling zero the OxiTop, it is kept in the reactor of BOD at the constant 20 degrees centigrade for 5 days. This is the BOD5 test. Thus, we get the BOD values. After experimenting with the BOD, we get 7 mg/L. The BOD experiment is shown in Fig. 5.

3.2. Cloud server

The real-time data which we got from the experimental setup is stored in a cloud server named ThingSpeak IoT server using rest-API. ThingSpeak is an (IoT) stage that allows us to dissect and imagine the information in MATLAB except by purchasing a permit from Mathworks. It permits us to gather and store sensor information in the cloud and create IoT applications. It communicates sensor data to ThingSpeak via Arduino, ESP8266 Wifi Module, Particle Photon and Electron, BeagleBone Black, mobile and web applications, Raspberry Pi, Twilio, Twitter, and Matlab. The ThingSpeak is generally centered around sensor logging, area following, triggers and cautions, and examination. The authors used a cloud server protocol to store the real-time values [25].

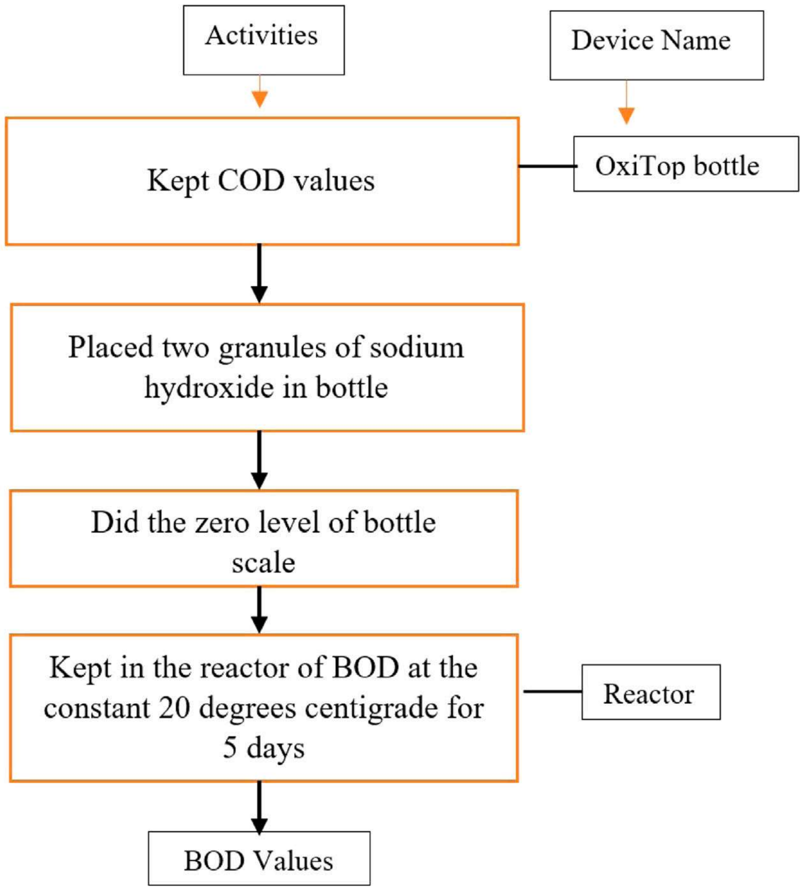

Fig. 5. Step wise BOD experiment.

Table 2
Real-time data stored in the cloud.

Parameters name	Data types
Field 1 data: Turbidity	JSON XML CSV
Field 1 data: Temperature	JSON XML CSV
Field 1 data: PH	JSON XML CSV
Field 1 data: Depth	JSON XML CSV

Table 3
Ideal range of pond water quality parameter.

Parameter	Ideal range
pH value	6.5–8.5
Temperature value	16–24 °C
Turbidity value	Below 10 ntu
Conductivity	970–1825 µS/cm

3.3. Real-time dataset

The real-time data is stored in the ThingSpeak channel. It is stored in CSV, XML, and JSON format, as depicted in Table 2.

3.4. Machine learning part

Fig. 6 describes the detailed block diagram of machine learning classifiers for predicting the fish species. Authors prepare the real data with fish categories according to the standard values of water shown in Table 3. After that, we make preprocessing and test-train splitting. Then, we apply 10 machine learning algorithms to classify fish. Aquatic parameters are the independent variable and fish categories are the dependent variable.

Table 4
Sample water quality dataset for fish survival.

ph	Temperature	Depth-feet	Turbidity-ntu	Fish
6	27	6.5	7	katla
7.5	29	3.5	6	prawn
6.1	31	5	4.9	rui
7.1	23	4.3	5.5	koi
7.5	32	7	7.3	katla
7.7	22	5.1	6	rui
7.9	29	4.9	5.5	rui
5.5	18	4.5	5	koi
6.2	19	5.2	6.1	koi
8.2	27	4	8.5	prawn

3.4.1. Dataset

After getting the real-time dataset from the cloud, we do labeling as per the standard values of fish species. We take 11 kinds of fish species. Table 4 shows some records of our dataset.

3.4.2. Test-train splitting

Test-train splitting is the process of keeping some portion of the dataset for training for the machine and some other portion for testing the machine. In Weka, we used 10-fold cross-validation for test train splitting.

3.4.3. Machine learning models

We prepared our dataset after collecting real-time values using sensors from the cloud server of our system. In addition, we labeled the dataset according to the fish categories including katla, sing, prawn, rui, koi, pangas, tilapia, silvercarp, carpio, magur, etc. The optimum temperature of growing fish is (20–26) °C, (15–25) °C, (18–30) °C,

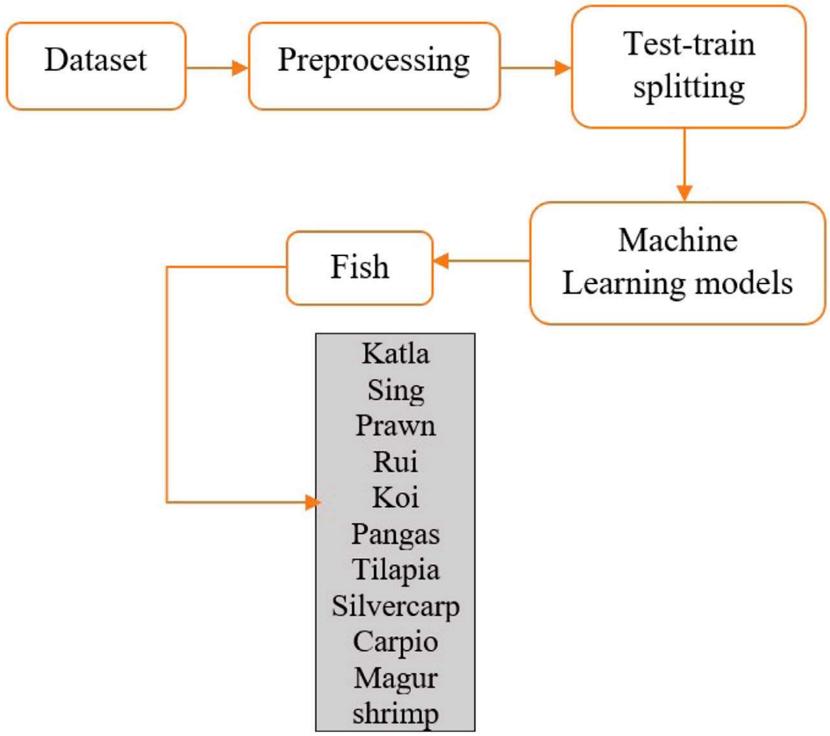

Fig. 6. A block diagram of the proposed methodology.

and (20–25) °C respectively for rui fish, koi fish, silvercarp fish, karpio fish [13]. During the analysis, the authors used fish as a target variable and water parameters as feature values. We applied 10 classification ML models because of available in the WEKA application tool as well as for the collected real-time data for the classification problem.

J48 classification algorithm The algorithm starts by building a decision tree out of the available training data's attribute values. It recognizes the attribute that specifically distinguishes the different instances when it happenstances a collection of items-training set. This function provides information about data instances so that we can identify them appropriately. It is claimed to provide the most information benefit. If there is a number for this function for which no uncertainty is available, i.e. all data instances in the category have the same value for the target variable, this value is ended and the target value acquired is assigned to it [26].

K-Nearest Neighbors (K-NN) K-NN is a non-parametric distance-based algorithm. The algorithm computes the K number of distances between neighboring data points and discovers the best K for the dataset. It is used for regression as well as classification, but it is mostly used for classification issues. KNN is regarded as an uninterested learner calculation because it does not gain from the preparation set immediately, but rather supplies the dataset and performs an activity on it at the time of order [27].

K algorithm* K* is a classifier based on an instance, expressed as the resemblance of the class of training instances [28].

Logistic Model Tree (LMT) algorithm LMT is the classification tree that includes logistic regression functions on the leaves, the classifier is used. The method supports target variables in binary and multi-classes, number, name, and absent value [29].

Reduced-Error Pruning Tree (REPTree) REPTree is yet another Weka-specific algorithm. It is a quick decision tree learner which has been optimized for simplicity and speed. Reduced-error pruning with back fitting is used in the algorithms to find the smallest representation of the most accurate subtree for the pruning set [30].

JRIP JRip (RIPPER) is a basic and widely used algorithm. Classes are examined as to their size increases, and An initial set of class rules is constructed using incremental error reduction. JRip (RIPPER) begins by giving all examples of a specific decision in the training data as a class and determining protocols that covers all members' class. It then moves on to the next class and repeats the process until all classes have been enclosed [31].

PART PART is a law learner who divides and conquers. It produces "decision lists", which are defined rules. The new data is compared to each rule in the list in turn and a class of the first matching rule is applied to the object. In each iteration, PART constructs a partial C4.5 decision tree and converts the "best" leaf into a rule [32].

Decision table Each class should have its own set of decision rules. A decision table is commonly used to describe the rules. Rough sets can be used to pick attribute subsets as well. Weka classifiers are used to find the algorithm decision table in Rules. The best way to describe machine learning output is to view it in the same format as the input [33].

Logitboost algorithm This class is intended for the regression of additives. This class is classified by regression as a basic learner and can tackle difficulties in several classes [34].

Random forest Random forest is a kind of democratic algorithm. In this algorithm, the decision is made by voting. Such an algorithm is called ensemble learning. Random forests are made up of many trees or trees. Just as there are many trees in the forest, there are many decision trees in the random forest. The decision that most trees make is considered the final decision [35]. Table 5 shows the result of the Random Forest model according to the fish classes. It shows the following statistics: average TP rate of 0.944, FP rate of 0.006, precision of 0.949, recall of 0.944, F-measure of 0.945.

4. Result analysis and discussion

We used the WEKA tool to analyze the data set in this section. The Waikato Environment for Knowledge Analysis (Weka) is an ML algorithms software suite established by New Zealand's University of

Table 5
Class wise result performance of random forest model.

	TP rate	FP rate	Precision	Recall	F-Measure	Class
	0.966	0.006	0.949	0.966	0.957	katla
	0.918	0.015	0.849	0.918	0.882	sing
	0.929	0.012	0.650	0.929	0.765	prawn
	0.980	0.002	0.990	0.980	0.985	rui
	0.733	0.002	0.917	0.733	0.815	koi
	0.923	0.008	0.947	0.923	0.935	pangas
	0.953	0.006	0.976	0.953	0.965	tilapia
	0.945	0.002	0.981	0.945	0.963	silverCup
	0.909	0.009	0.857	0.909	0.882	karpio
	0.909	0.000	1.000	0.909	0.952	magur
	0.980	0.000	1.000	0.980	0.990	shrimp
Weighted Avg.	0.944	0.006	0.949	0.944	0.945	

Waikato. It is engraved in the Java programming language. As a performance metric to analyze the water quality, we have used accuracy, kappa statistics and average TP rate. From the Confusion Matrix, we get the accuracy, TP rate, and Kappa statistic. A confusion matrix is characterized by four terms named true positive (TP), false negative (FN), false positive (FP) and true negative (TN).

Accuracy, on the other hand, can be defined as the ratio of corrected predictions to the total input samples.

$$Accuracy = (No.\ of\ correct\ predictions)/(Total\ no.\ of\ predictions) \quad (1)$$

TP rate, also called Recall is the number of correctly identified cases from all the positive representations.

$$Recall, R = TP/(TP + FN) \quad (2)$$

Cohen's **Kappa Statistic** (CK) is applied to assess the degree of agreement between two raters who categorize objects into mutually exclusive groups which are shown mathematically in Eq. (3).

$$CK = \frac{(p_o - p_e)}{(1 - p_e)} \quad (3)$$

Here, p_o is the relative agreement of raters' observation. p_e denotes the theoretical probability of random agreement. We can calculate p_o and p_e between the raters by using the Eqs. (4)–(7).

$$p_o = \frac{TP + TN}{Total\ number\ of\ predictions} \quad (4)$$

$$p_e = probability\ of\ Positive + probability\ of\ Negative \quad (5)$$

Here,

$$Probability\ of\ Positive = \frac{TP + FP}{TP + TN + FP + FN} \times \frac{TP + FN}{TP + TN + FP + FN} \quad (6)$$

and

$$Probability\ of\ Negative = \frac{FP + TN}{Total\ number\ of\ predictions} \times \frac{FN + TN}{Total\ number\ of\ predictions} \quad (7)$$

Cohen's Kappa coefficient ranges from 0 to 1, where a value of 0 indicates no agreement and a value of 1 represents the complete agreement between the two raters. In the case of all models, Cohen's Kappa coefficient indicates nearly perfect agreement between the actual observations and the predicted outcomes.

4.1. Pond suggestion

5 ponds are chosen for experimenting with the water quality. The ponds are located in Jamalpur district under Mymensingh division, Bangladesh. After collecting water from the ponds, we experimented to get real-time values. Fig. 7 shows the IoT framework for the experiments.

Fig. 7. Experimenting water of all ponds using sensors.

There are approximately 5 h for operating the experiments in all ponds water. We suggest a 9600 baud rate for the serial starting function in the Arduino suite.

4.1.1. Result analysis of real-time values of PH

Fig. 8 displays the pH values for each pond for 25 different occurrences taken from the actual environment. For ponds 1, 2, 3, 4, and 5, the pH ranges are 6.02 to 8.39, 8.5 to 8.87, 6.01 to 8.30, and 3.84 to 3.95, respectively. Considering the optimal standard pH range for fish production, pond 1's pH range is suitable (6.5–8.5). The pH value that was obtained from pond 2 (8.57–8.87) is higher than the desired value. As a result, fish farming is not a good fit for this pond. The pH range for Pond 3 was between 6.00 and 7.83. 6.5–8.5 is almost the perfect range. Additionally, Pond-4 offers adequate pH readings. For pond-5, the pH range is not ideal for fish cultivation. Fish species are killed as a result of this.

4.1.2. Result analysis of real-time values of temperature

A graph of the temperature readings in real-time from all of the trials is shown in Fig. 9. For ponds 1, 2, 3, 4, and 5, respectively, the temperature ranges are 17.50–17.75 °C, 17.75–18.00 °C, 20.87–21.06 °C, 21.06–21.44 °C, and 21.06–21.25 °C. Different fish species require different temperatures. The (16–24) °C temperature range is considered to be the acceptable range for ponds. The real-time sensor that was received from all experiments is ideal for fish farming.

4.1.3. Result analysis of real-time values of conductivity

A graph of the conductivity values in real-time from all of the trials is shown in Fig. 10. The range of conductivity values for ponds 1, 2, 3, 4, and 5 is 989 to 1003, 1003 to 1017, 1179 to 1190, 1193 to 1215, and 1190 to 1201 accordingly. The real-time sensor that was received from all trials is ideal for fish farming.

4.1.4. Result analysis of real-time values of turbidity

Fig. 11 shows a graph of the turbidity readings in real-time from each experiment. The range of turbidity readings for ponds 1, 2, 3, 4, and 5 is 3.55–3.57 NTU, 3.41–3.50 NTU, 3.31–3.49 NTU, 3.60–3.62 NTU, and 3.56–3.58 NTU, respectively. The optimal value states that the turbidity range for fish farming is less than 10 NTU.

Here, we have shown all the real-time values of aquatic environmental quality parameters from all ponds in Table 6. Table 6 illustrates the summary of all received real-time values for each pond. The ideal range of all parameters including pH: 6.5–8.5, turbidity: below 10 ntu, temperature: (16–24) °C and conductivity: 970–1825 μS/cm.

This record shows that ponds 1, 3, and 4 are suitable for fish farming. The second pond is not ideal for growing fish. 8.57–8.87 is higher than the optimal levels due to the pH value. It is the reason why

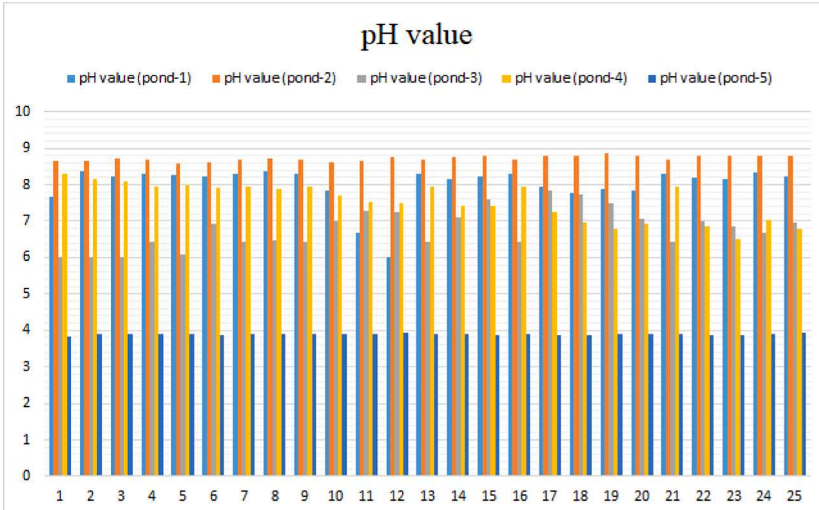

Fig. 8. pH values for each pond for 25 different occurrences.

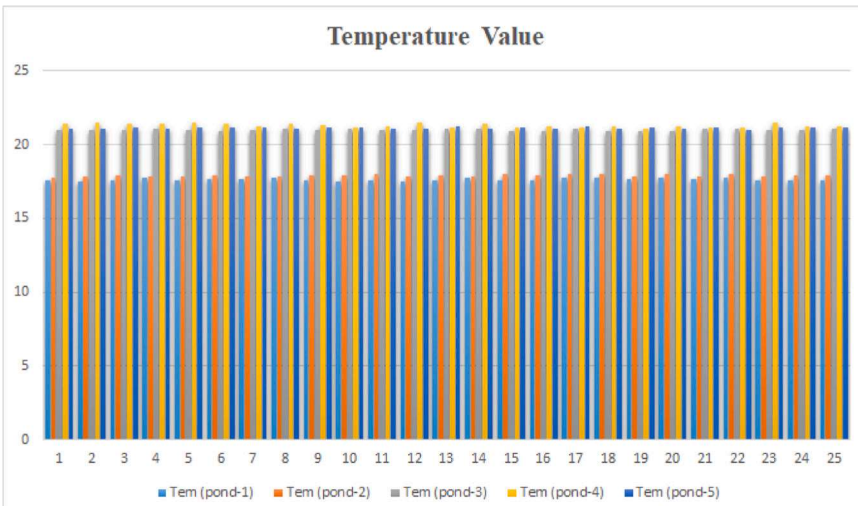

Fig. 9. Temperature values for each pond for 25 different occurrences.

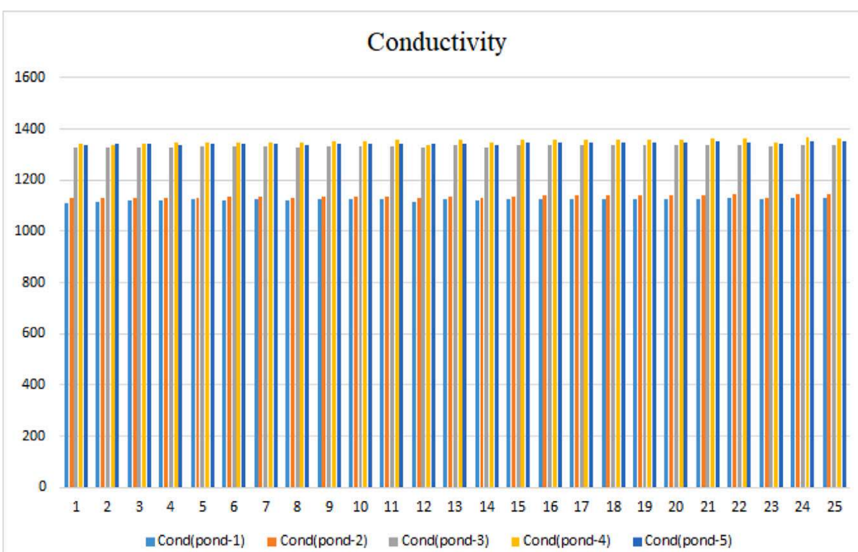

Fig. 10. Conductivity values for each pond for 25 different occurrences.

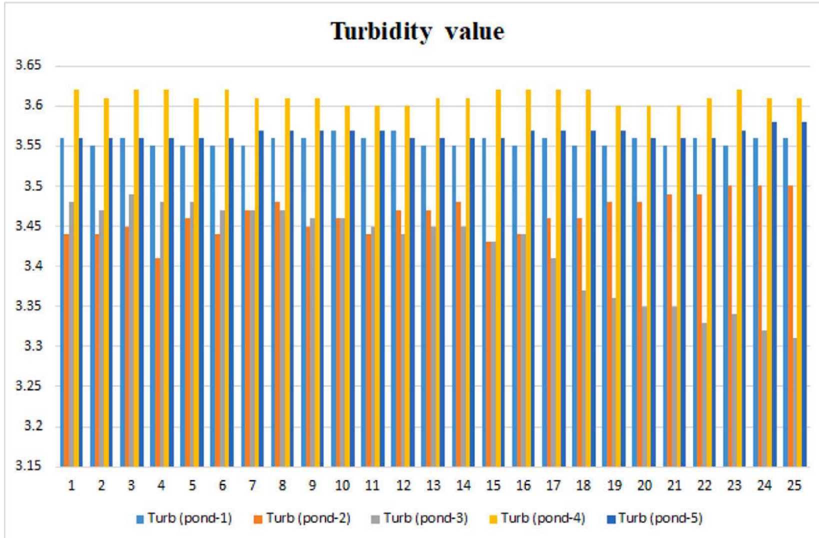

Fig. 11. Turbidity values for each pond for 25 different occurrences.

Table 6
Pond suggestion using the real-time dataset.

Pond number	pH	Temperature (°C)	Turbidity (ntu)	Conductivity (μS/cm)	Remarks
1	6.02–8.39	17.50–17.75	3.55–3.57	989–1003	Recommended
2	8.57–8.87	17.75–18.00	3.41–3.50	1003–1017	Not recommended
3	6.00–7.83	20.87–21.06	3.31–3.49	1179–1190	Recommended
4	6.51–8.30	21.06–21.44	3.60–3.62	1193–1215	Recommended
5	3.84–3.95	21.06–21.25	3.56–3.58	1190–1201	Not recommended

Table 7
Comparison of performance metrics among machine learning models.

SL. no.	ML model	Accuracy (%)	CK (%)	Avg. TP rate (%)	Position
1	J48	90.19	88.7	90.2	5th rank
2	Random forest	94.42	93.5	94.4	1st rank
3	K-NN	93.4	92.4	93.4	2nd rank
4	K* Algorithm	89.85	88.37	89.8	6th rank
5	LMT	92.22	91.08	92.2	3rd rank
6	REPTree	83.93	81.5	83.9	9th rank
7	JRIP	87.14	85.17	87.1	7th rank
8	PART	90.35	88.92	90.4	4th rank
9	Decision table	80.54	77.5	80.5	10th rank
10	Logit boost	84.60	82.37	84.6	8th rank

fish grow slowly. The lower pH range of 3.84 to 3.95 is not suitable for pond-5. Fish perish because of this. Because of this, we were unable to provide fish farmers' advice for ponds 1 and 2.

4.2. Comparison of performance metrics among machine learning models

The received real-time values are analyzed using 10 machine learning algorithms. Table 7 shows the comparison of performance metrics including accuracy, kappa statistics and avg. TP rate among the classifiers.

Table 7 demonstrates that Random Forest provides the maximum score possible for each parameter, with an accuracy score of 94.42%, a kappa statistic of 93.11%, and an average True Positive (TP) rate of 94.4%. The KNN model, which reports accuracy as 93.4%, kappa statistic as 92.4%, and TP rate as 93.4%, received the second-highest score. LMT achieves an accuracy of 92.22%, a kappa statistic of 91.08%, and a TP rate of 92.2% to take third place. In terms of performance measures, PART came in fourth with an accuracy score of 90.35 percent, a kappa statistic of 88.92 percent, and a TP rate of 90.4%. J48 ranks fifth with an average TP rate of 90.2%, accuracy of 90.19%, and kappa statistics of 88.7%. K* reports accuracy 89.85%, kappa statistic 88.37%, and average True Positive (TP) rate 89.8% for the sixth score of each parameter. The JRIP model, with accuracy ratings of 87.14%, kappa statistics of 85.17%, and an average TP rate of 87.1%, received the seventh-highest score. By obtaining an accuracy of 84.60%, a kappa statistic of 82.37%, and a TP rate of 84.6%, Logit Boost moves up to the eighth highest rank. With an accuracy score of 83.93%, a kappa statistic of 81.5%, and a TP rate of 83.9%, REPTree ranks ninth in terms of scoring performance parameters.

Random forest gives the highest score for all performance metrics. Because we know that it is a kind of democratic algorithm. In this algorithm, the decision is made by voting. Such an algorithm is called ensemble learning. There are many decision trees in the random forest, the decision that most trees make is considered the final decision.

The graphical representation of the table is shown in Fig. 12. Accuracy is indicated by a blue-colored line, kappa statistics are marked by a dark red-colored line. And the average true positive (TP) rate is demonstrated by the olive-colored line.

If we expanded this work to include more than five ponds with more fish species and enhanced the measures for water quality, it would be more effective. Deep learning or machine learning can be used to analyze real-time values more effectively in Python than in application technologies like WEKA. The records should be kept for a month rather than just a few days in order to get better results.

4.3. Discussion and comparison

There has been significant research on using IoT devices for fish farming to check water quality. In this section, we discuss and compare our proposed work with existing similar work. Table 8 shows the comparison between some existing works with the proposed work.

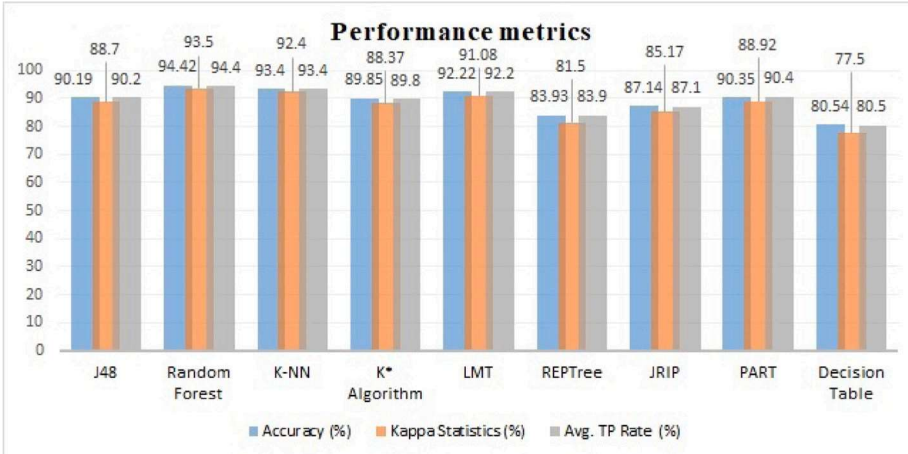

Fig. 12. Graphical presentation of the comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 8
Comparison among some existing works with the proposed work.

SL. No.	Source	Water quality factors	Water type	Fish	Machine learning
01	[8]	pH, temperature, DO, and ammonia	Aquaculture	–	–
02	[9]	PH and salinity	Aquaculture	Guppy	–
03	[10]	DO, PH, temperature and salinity	Aquaculture	Pangasius	–
04	[11]	pH, temperature, water level and oil layer	Aquaculture	–	–
05	[12]	CO, pH, temperature, water level and turbidity	Aquaculture	–	Yes
06	[13]	pH, temperature, turbidity, conductivity, and depth	Aquaculture	–	–
07	[14]	pH, temperature and turbidity	Aquaculture and drinking water	–	–
08	[15]	pH, temperature and DO	Aquaculture	–	–
09	[16]	pH, temperature and DO	Aquaculture	–	–
10	[17]	Temperature, depth, DO, and pH	Aquaculture	–	–
11	[18]	Turbidity, temperature, oxygen level, and pH	Aquaculture	–	Yes
12	[19]	temperature, conductivity, water level, pH, turbidity and DO	Aquaculture	–	Yes
13	[20]	Temperature, pH, turbidity, and DO	Aquaculture	–	–
14	[21]	Oil layer, water level, temperature	Aquaculture	–	–
15	Proposed	pH, temperature, turbidity, conductivity, DO, COD, BOD	Aquaculture	11 types of fishes	Yes

5. Conclusion

In this paper, we implemented an IoT framework for fish survival in pond water. Several water quality metrics are measured and the real-time values are stored in a cloud named ThingSpeak. For our experiments, we used five different pond settings to measure pH, temperature, conductivity, and turbidity in real-time. After examining the real-time results, we discovered that ponds 1, 3, and 4 are ideal for fish farming since they have the ideal values of pH, temperature, turbidity, and conductivity levels. Additionally, ponds 2 and 5 are not ideal for fish farming. Because of this, the end user is able to grow fish in ponds 1, 3, and 4. A farmer can take any action for ponds 2 and 5 to subsequently use them for fish farming. Integrating machine learning algorithms with the collected sensor data, we applied 10 ML algorithms. Among the executed ML algorithms, Random Forest took first place with accuracy 94.42%, kappa statistics 93.5%, and Avg. TP rate 94.4%. We also investigated the BOD, COD, and DO for one scenario. In the future, we can expand it using more than 5 ponds along with more water quality metrics and record the real-time data for a month. For analyzing complex relationships and patterns among the data, we can apply machine learning, and deep learning in Python language.

Declaration of competing interest

No conflict of interest exists in this manuscript, and all authors approve the manuscript for publication. I would like to declare that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

Data availability

Data will be made available on request.

References

[1] FAO, The state of world fisheries and acquaculture, 2020, online at https://www.fao.org/3/ca9231en/CA9231EN.pdf.

[2] R. Francis-Floyd, Dissolved oxygen for fish production, online at http://fisheries.tamu.edu/files/2013/09/Dissolved-Oxygen-for-Fish-Production1.pdf.

[3] Merusonline, What is BOD – Biological Oxygen Demand, online at https://www.merusonline.com/bod-biological-oxygen-demand/.

[4] A. Bouguettaya, Q.Z. Sheng, B. Benatallah, A.G. Neiat, S. Mistry, A. Ghose, S. Nepal, L. Yao, An internet of things service roadmap, Commun. ACM 64 (9) (2021) 86–95.

[5] M.W. Wagner, K. Namdar, A. Biswas, S. Monah, F. Khalvati, B.B. Ertl-Wagner, Radiomics, machine learning, and artificial intelligence—what the neuroradiologist needs to know, Neuroradiology 63 (12) (2021) 1957–1967.

[6] C. Wang, Z. Li, T. Wang, X. Xu, X. Zhang, D. Li, Intelligent fish farm—the future of aquaculture, Aquac. Int. (2021) 1–31.

[7] A. Tchagna Kouanou, C. Tchito Tchapga, M. Sone Ekonde, V. Monthe, B.A. Mezatio, J. Manga, G.R. Simo, Y. Muhozam, Securing data in an internet of things network using blockchain technology: smart home case, SN Comput. Sci. 3 (2) (2022) 167.

[8] A.T. Tamim, H. Begum, S.A. Shachcho, M.M. Khan, B. Yeboah-Akowuah, M. Masud, J.F. Al-Amri, Development of IoT based fish monitoring system for aquaculture, Intell. Autom. Soft Comput. 32 (1) (2022) 55–71.

[9] P. Periyadi, G.I. Hapsari, Z. Wakid, S. Mudopar, IoT-based guppy fish farming monitoring and controlling system, TELKOMNIKA (Telecommun. Comput. Electron. Control) 18 (3) (2020) 1538–1545.

[10] L.V.Q. Danh, D.V.M. Dung, T.H. Danh, N.C. Ngon, Design and deployment of an IoT-based water quality monitoring system for aquaculture in Mekong Delta, Int. J. Mech. Eng. Robot. Res. 9 (8) (2020) 1170–1175.

[11] K.A. Chy, A.K.M. Masum, M.E. Hossain, G.R. Alam, S.I. Khan, M.S. Alam, et al., A low-cost ideal fish farm using IoT: In the context of Bangladesh aquaculture system, in: Inventive Communication and Computational Technologies, Springer, 2020, pp. 1273–1283.

[12] M. Islam, J. Uddin, M.A. Kashem, F. Rabbi, M. Hasnat, et al., Design and implementation of an IoT system for predicting aqua fisheries using arduino and KNN, in: International Conference on Intelligent Human Computer Interaction, Springer, 2021, pp. 108–118.

[13] M.M. Islam, M.A. Kashem, J. Uddin, An internet of things framework for real-time aquatic environment monitoring using an Arduino and sensors, Int. J. Electr. Comput. Eng. 12 (1) (2022) 826.

[14] M. Islam, J.H. Rony, M. Akhtar, S.U.P. Shakil, J. Uddin, et al., Water monitoring using internet of things, in: Internet of Things for Smart Environments, Springer, 2023, pp. 59–69.

[15] Z. Shareef, S. Reddy, Design and development of IoT-based framework for indian aquaculture, in: Intelligent Communication, Control and Devices, Springer, 2020, pp. 195–201.

[16] J. Huan, H. Li, F. Wu, W. Cao, Design of water quality monitoring system for aquaculture ponds based on NB-IoT, Aquac. Eng. 90 (2020) 102088.

[17] C. Lee, Y.-J. Wang, Development of a cloud-based IoT monitoring system for Fish metabolism and activity in aquaponics, Aquac. Eng. 90 (2020) 102067.

[18] M.-C. Chiu, W.-M. Yan, S.A. Bhat, N.-F. Huang, Development of smart aquaculture farm management system using IoT and AI-based surrogate models, J. Agric. Food Res. 9 (2022) 100357.

[19] G. Gao, K. Xiao, M. Chen, An intelligent IoT-based control and traceability system to forecast and maintain water quality in freshwater fish farms, Comput. Electron. Agric. 166 (2019) 105013.

[20] J. Duangwongsa, T. Ungsethaphand, P. Akaboot, S. Khamjai, S. Unankard, Real-time water quality monitoring and notification system for aquaculture, in: 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, IEEE, 2021, pp. 9–13.

[21] A.K.M. Masum, M. Shahin, M.K.A. Chy, S.I. Khan, A. Shan-A-Alahi, M.G.R. Alam, Design and implementation of iot based ideal fish farm in the context of bangladesh aquaculture system, in: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 2019, pp. 1–6.

[22] B.T.W. Putra, L.A. Rocelline, W.N.H. Syahputra, Embedded system in handheld water turbidity meter for smallholders, Microprocess. Microsyst. 93 (2022) 104603.

[23] L.S. Bratchenko, E.A. Rodionova, T.G. Shchetinina, G.Y. Kolev, V.G. Rybin, Embedded water analyzer for unmanned surface vehicles, in: 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), IEEE, 2022, pp. 119–122.

[24] M. Rana, A. Rahman, J. Dabrowski, S. Arnold, J. McCulloch, B. Pais, Machine learning approach to investigate the influence of water quality on aquatic livestock in freshwater ponds, Biosyst. Eng. 208 (2021) 164–175.

[25] M.R. Akbari, H. Barati, A. Barati, An overlapping routing approach for sending data from things to the cloud inspired by fog technology in the large-scale IoT ecosystem, Wirel. Netw. 28 (2) (2022) 521–538.

[26] C. Luu, D.-D. Nguyen, T.V. Phong, I. Prakash, B.T. Pham, Using decision tree J48 based machine learning algorithm for flood susceptibility mapping: A case study in Quang Binh Province, Vietnam, in: CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure, Springer, 2022, pp. 1927–1935.

[27] S. Mustary, M.A. Kashem, M.N.I. Khan, F.A. Jewel, M.M. Islam, S. Islam, LEACH based WSN classification using supervised machine learning algorithm, in: 2021 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2021, pp. 1–5.

[28] R. Duriqi, V. Raca, B. Cico, Comparative analysis of classification algorithms on three different datasets using WEKA, in: 2016 5th Mediterranean Conference on Embedded Computing (MECO), IEEE, 2016, pp. 335–338.

[29] A.D. Amirruddin, F.M. Muharam, M.H. Ismail, N.P. Tan, M.F. Ismail, Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and unmanned aerial vehicles, Comput. Electron. Agric. 193 (2022) 106646.

[30] A. Arabameri, M. Santosh, S. Saha, O. Ghorbanzadeh, J. Roy, J.P. Tiefenbacher, H. Moayedi, R. Costache, Spatial prediction of shallow landslide: application of novel rotational forest-based reduced error pruning tree, Geomat. Nat. Hazards Risk 12 (1) (2021) 1343–1370.

[31] A.A.R. Melvin, G.J.W. Kathrine, S.S. Ilango, S. Vimal, S. Rho, N.N. Xiong, Y. Nam, Dynamic malware attack dataset leveraging virtual machine monitor audit data for the detection of intrusions in cloud, Trans. Emerg. Telecommun. Technol. 33 (4) (2022) e4287.

[32] A. Triayudi, W.O. Widyarto, V. Rosalina, Analysis of educational data mining using WEKA for the performance students achievements, in: Proceedings of the 2nd International Conference on Electronics, Biomedical Engineering, and Health Informatics, Springer, 2022, pp. 1–10.

[33] S. Gautam, C. Sharma, V. Kukreja, et al., Handwritten mathematical symbols classification using WEKA, in: Applications of Artificial Intelligence and Machine Learning, Springer, 2021, pp. 33–41.

[34] M. Hussain, L. Gogoi, et al., Performance analyses of five neural network classifiers on nodule classification in lung CT images using WEKA: a comparative study, Phys. Eng. Sci. Med. (2022) 1–12.

[35] M.M. Islam, M.A. Kashem, J. Uddin, Fish survival prediction in an aquatic environment using random forest model, Int. J. Artif. Intell. ISSN 2252 (8938) (2021) 8938.

Md. Monirul Islam is an Assistant Professor in the Department of Software Engineering at the Daffodil International University in Bangladesh since 4 July 2023. He formerly worked as an Assistant Professor in the Department of CSE and as Director of the ICT cell at the University of Information Technology and Sciences (UITS), a Senior Lecturer and Lecturer at Uttara University and Atish Dipankar University of Science and Technology (ADUST) in Bangladesh. He earned his M.Sc. in engineering in CSE in 2021 from Dhaka University of Engineering and Technology (DUET), Gazipur-1700, Bangladesh, and his B.Sc. in engineering in CSE in 2017 from Pabna University of Science and Technology (PUST), Pabna-6600, Bangladesh. He has published more than 16 articles in reputed Scopus indexed journals and conferences. Recently, he has been awarded in "Research and Publication Award 2022" from Uttara University. Field of interest: Machine learning/Deep learning, Data science, IoT, Computer Network, and Medical Image Analysis. He can be contacted at email: monirul.swe@diu.edu.bd or monir.duet.cse@gmail.com

Mohammod Abul Kashem is a Professor at the Department of Computer Science and Engineering (CSE), Dhaka University of Engineering and Technology (DUET) in the year 2003. He completed his B.Sc. and M.Sc.Engg. degrees from State University Lvivska Polytechnica, Ukraine in 1996 and 1997 respectively. In 2001, he earned Ph.D. in Control Systems and Processes from National University Lviv Politechnic Ukraine. Subsequently, Dr. Kashem completed his Post Doctorate fellowship from University Lumiera Lyon2, France. (Erusmus Mundas Scholarship, European Commission),2016.

Salem A. Alyami is an Assistant Professor of Department of Mathematics and Statistics at Al-Imam Muhammed Ibn Saud Islamic University.

Mohammad Ali Moni is a Senior Research Fellow, Digital Health Lead at Charles Sturt University. Mohammad holds a PhD in Artificial Intelligence and Digital Health Data Science in 2015 from the University of Cambridge.