Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 02
2024

Research Article
https://doi.org/10.56979/702/2024

Water Quality Assessment Through Predictive Modeling Employing Machine Learning Methods

Mehak Afzal[1], Shujaat Ali[1], Hafiz Burhan Ul Haq[1*], Rabia Younis[1], Hamid Ali[1], Amna Kosar[1], and Hafiz Muneeb Akhtar[1]

[1]Department of Information and Technology, Faculty of Computer Sciences, Lahore Garrison University, Lahore, Pakistan.
[*]Corresponding Author: Hafiz Burhan Ul Haq. Email: burhanhashmi64@lgu.edu.pk

Received: June 11, 2024 **Accepted:** August 23, 2024 **Published:** September 01, 2024

Abstract: Water quality declines pose serious problems that need creative methods to ensure proper monitoring. In order to gather and evaluate water quality data in real time, this study presents the Water Quality Measurement Application, which integrates cutting-edge sensors with artificial intelligence (AI). By creating a tool that is easily accessed by environmentalists and scientists, the goal is to enhance existing techniques that depend on labor-intensive and less accurate manual sampling and analysis. The accuracy of current sensor-based systems is frequently restricted, and they are unable to accurately forecast problems with water quality. To get around this, machine learning (ML) techniques are used in the application to assess and forecast water quality situations while IoT sensors are integrated for continuous data collecting. Safely moving data to a cloud platform is made possible by the Blynk IoT framework, which guarantees accessibility and security. When it came to identifying water quality characteristics, the Random Forest (RF) approach outperformed other machine learning models, including K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Multinomial Naive Bayes (MNB). When compared to conventional techniques, this breakthrough yields forecasts that are more trustworthy. Subsequent efforts will concentrate on improving the application by extending the scope of observable metrics and adding user input. Accuracy improvement is another goal of ongoing research into ML algorithms. By providing a more sophisticated, automated method of comprehending and regulating water health, this creative solution benefits environmental professionals as well as labs.

Keywords: Internet of Things (IoT); Sensors; User Interface; Integration; Blynk IoT Application.

1. Introduction

AI allows the machine to be programmed with attributes similar to human cognition, including understanding, thinking, and making decisions for itself in executing functions. Machine learning, which enables the creation of those algorithms and models that are used to mine information from big data, supports this potential by providing predictions and recommendations [1].

This term, namely Internet of Things (IoT), is a single system used massively in different fields, starting from smart clothing to smart cities to smart power grids and intelligent supply chains [2]. On the global level, the study unveils an astonishing According to the source, 1 billion people were deprived of access to safe drinking water; 2. 6 billion people do not have access to even the minimum level of hygiene, not mentioning about other necessities. The decade of 2005-2015 has been declared the water for life decade, underscoring the vital imperative for efficient water management. This initiative study illustrated the centrality of real-time water quality and quantity monitoring, enabling new solutions via edge computing and AI. These technologies depict creativity in preventing waterborne diseases resulting in pollutants and enhance public health immensely. These technologies have a strong potential to lower many diseases associated with water pollution and improve the populace's health tremendously [3].

For water quality finding, train a machine learning model informed by a rich historical water quality database to estimate specific unmeasurable parameters such as bacterial count and organic pollutants based on easily available IoT data such as ph conductivity, etc. [4]. Over the course of the last decade, this has evolved into one of the most refreshing and effective means of monitoring water quality [5], which provides special possibilities in such spheres as in agriculture [6]. Also, in waste management practices [7]. The daily prediction of both the pH and the temperature of a water point using AI-based methods is more accurate when done simultaneously. The daily prediction of both the pH and the temperature of a water point using AI-based methods is more accurate when done simultaneously [8]. Internet of things IoT systems that act as the connection between the digital and physical world can help in the efficient control and consumption of physical assets and help in the concept of conscious preservation in several IoT use cases for smart societies. Water is a scarce and depleting resource, and all can be accessed through the Internet of Things reservoir [9].

The models that have been developed to exhibit good performance in predicting the water quality components, the Multi-Layer Perceptron (MLP) model was identified as the most accurate [10]. This aligns with an application and diffusion type of innovation where the researchers sought to enhance the supply of water through enhanced techniques [11]. The study revolves around the process of control as well as the assessment of water quality by the employment of machine learning algorithms that are most of the time paired with sensor devices to give real-time evaluation of water quality measurements. Water is involved in all aspects of life in the contemporary world in relation to people, plants, and animals. Water is a life essential that is under more pressure due to pollution, climate change, and the constantly expanding population. By using this Water Quality Measurement application, one can easily and effectively evaluate and forecast the quality of water. Serious environmental enthusiasts and the public can easily embrace this program's ability to analyze and predict the state of water for application at any given time in the future. It can also be applied to the management of parameters, water resources, and the real-time evaluation and analysis of results. Thus, along with delivering technical knowledge and skills, the user interface app proves to be of great advantage to the students of institutions of learning and in numerous other ways.

The nature of water quality monitoring systems requires high accuracy, or rather precision, as part of the system. While currently IoT-based systems that depend on the raw data collected from the sensors are still prone to anomalies and outliers, which in turn greatly affect the outcome [12]. For accurate water quality analysis and to protect people's health from waterborne diseases, a much stronger and sounder way is needed [13].

To the best of our knowledge, there are no comprehensive online datasets that sufficiently include these parameters for model training and evaluation. Therefore, the dataset used for training the AI models is found in the IEEE Data Port. The dataset consists of daily 2792 samples from different user locations with 25 features that cut across location, season, weather, year, depth, ph, etc. [14].

Contamination control using traditional water monitoring is expensive, time-consuming, and has poor coverage; important areas are not monitored while others are partially covered. Thus, sensor-supported mobile applications fill this gap by providing timely information, indicating the impact of the ecosystem, and providing access to water data to anyone. This also helps the organizations and NGOs to know about the water conditions within a short period and make efficient work, offering clean water to the suffering societies.

For improving the performance of monitoring and detail analysis, it is necessary to incorporate artificial intelligence (AI) and Internet of Things (IoT) technology to offer real-time monitoring and measurement facilities to the main water quality parameters in an easy manner. Due to this, the created database must enable users to perform a historical analysis of water quality, which in turn enables long-term evaluations. To offer the benefits of data visualization allowing the management of large and, in this case, water quality data for the specified users to comprehend. Interactive presentation including the group bar charts, proportional colors & symbols, figures, tooltips, etc. The measurement of water quality is useful in lab research as it provides separate advantages besides the usual laboratory analysis. It is used as a tool for strategic decision-making, an early sign of a deviation, and in real-time monitoring of water, which raises the bar in terms of accuracy and success for laboratories. Assessment of water quality is useful in lab research because, although it has different characteristics, it proved to be securing specific benefits in correlation with laboratory analytics. It assists as a tool in the identification of early anomalies and real-time monitoring of water to enhance the effectiveness of laboratories in strategic decision-making. This application has a somewhat real-life utility as a water measurement device, due to which this app will be useful for the students as well as for the environmental institutions that are related to water and aquatic life. This app has the provision of educational tools and approaches to learning.

2. Related work

Artificial intelligence (AI) is most likely to drastically change water management and contribute to the achievement of the water-related Sustainable Development Goals (SDGs). Thus, AI can unlock a sustainable water future that will impact both the economy and ecology by addressing those challenges and scaling up successful applications [15]. The system relies on simple and smooth interplay between hardware and software. Communication and data transfer are important as the app connects with a database called Blynk that is on the cloud, involving an Internet connection to reach the system from an Android device. Blynk is a database to store data. Users get help through the Blynk for secure access to the growing of system called aquatic animal like fish, etc. [16]. Random-forest model shows potential for the water quality forecast system with 85% accuracy. Another drawback of the high precision required and necessity to monitor the resource requirements makes it inapplicable in the real world. In addition, there are the following considerations: The Random Forest model's requirements for precision prevent its real-world implementation. Another related work stays untouched. Thus, future research should consequently focus on higher accuracy, algorithm optimization for edge devices, and real-time processing [17].

This study underscores the attainment of a high prediction accuracy of 94.5% using Cat-Boost, a machine learning method, and achieving 100% accuracy from the ensemble method. Existing AI water quality monitoring systems exhibit a missing metric, undisclosed system latency, blocking real-time decision-making. These systems lack the necessary competitive intelligence and edge friendliness, resulting in a deficiency of user satisfaction. These apps lack visualization for user use; Water Quality Monitoring System (WQMS) deployment becomes imperative [18]. Conventional methods of managing water supply and usage are being revolutionized with the features of machine learning algorithms, like ANNs, DLNNs, and the SVM, which are now fashionable tools for activities such as the prediction of water qualities and wastewater indicators [19]. It also shows how helpful they are in predicting some trends of the future based on data from the real world. This black-box approach aims at providing improvements in water management by discovering heretofore unknown relations and making sophisticated predictions [20].

Contaminants have adversely affected various properties of water, underscoring the need to reduce water pollution and prioritize water quality modeling and prediction. This study explores techniques beyond artificial intelligence to forecast the Water Quality Index (WQI) and categorize water quality (WQC) [21]. A nonlinear operator equation was employed to address parameter inversion in the dissolved biological oxygen demand (BOD) and dissolved oxygen (DO) water quality models. The research illustrates how developing a suitable function transforms the original problem into an optimization task [22]. To achieve the process of building a machine learning model, we used the water_potability.csv file that contains water quality metrics for 3,276 different water bodies. Available, comprehensive, and rigorous water quality datasets are key in conducting machine learning application techniques for the purpose of scientific research. The data are comprehensive and used for research purposes, especially in machine learning applications [23].

There were various technologies used for the water quality detection. One such system is Opti-Event detection system (OptiEDS) by Elad Salomons, which helps to detect anomalous water quality conditions in real time. It is also capable of water quality monitoring and water contamination detection in real time. Bluebox is another system that can identify the behavior of water quality parameters that cause abnormal behavior. It produces a reliable output even if some of the parameters are missing. It initially performs normalization, calculates the points' distance amongst the parameters in each data point and plots the frequency curves of the distances to visualize. However, it is quite expensive, costing up to $92,500, and does not have the capability to directly integrate operational data into event detection [24].

Table 1. Literature survey of some methodologies and work

Author	Approach	Methodology	Strength	Weakness
Simeon et al. [4]	Sensor /ML/SSL/Decision Tree	A mobile AI-powered interactive app is developed to evaluate the water quality instantaneously based on sensor measurements.	It's reliable with a precision of 99.36%, an accuracy of 99.46%, and 0.0214 prediction error.	This paper does not tell features how the results should be visualized to the user, and it does not tell features of educational resources for people, etc. Lack of app knowledge. And user interface.
Nguyen et al. [25]	XGBoost/ ecurrent neural network (RNN) /(Long short-term memory)LSTM	Getting accuracy for collecting WQI-specific parameters for specific rivers.	accuracy ranging from 84% to 96%	This paper does not talk about real-time data, only specific rivers, and has an error in the prediction of data that was not known.
Priya et aI. [26]	Sensors and Cloud based IOT technology ThinkSpeak.	This can be done by using Internet of Things (IoT).	The data is also stored on the Cloud using IoT based Thing Speak. The resources involved in water quality monitoring are greatly reduced and the parameters are analyzed in real-time.	This paper has a lack of anomaly detection in water quality and a lack of safety measures.

Wang et al. [1]	General regression neural network (GRNN)/ multivariate polynomial regression (MPR)	Identify those parameters that can be measured with an IOT sensor and develop an IOT system. Develop an AI model to estimate parameters that cannot be measured by IOT.	Overall performance against the laboratory is valid and acceptable accuracy, but less than 0.2 mg error.	This paper does not talk about machine learning models. Perfor mance is good only at the river site. For accuracy and timeline balance Machine learning algorithms need to be implanted for further quality enhancement.

3. Materials and Methods

A machine learning working flow diagram with the help of sensors is mentioned in figure 1. Highly advanced artificial intelligence technology will be able to solve problems with real-time data processing and prediction software. As basically we are using AI branch machine learning algorithms through real-time processing of sensor data and instantaneous user ratings of water quality, machine learning is integrated with the backend of the app. All users might monitor and evaluate water quality measures with ease and precision thanks to the app's user-friendly design. This project can combine machine learning models with representation through charts and colors, anomaly detection, and predictive insights.

The starting point of sensors collects data as shown in Figure 1. Different studies show different parameters through which we convert the raw data of sensors into readable form.

Sensors connect with the software Arduino IDE and its board NodeMCU 1.0 (esp-12e module), which is most suitable for hardware IOT devices with a serial port used in the board is COM 4. This port is a communications COM port; it's basically a hardware interface on a computer that allows us to connect external devices for data transfer.

Arduino IDE 2.3.2 is faster and even more powerful. In addition to a more modern editor and a more responsive interface, it features auto completion, code navigation, and even a live debugger for water quality measurement.

The machine learning algorithm used to enhance the efficiency, accuracy, and depth of analysis in water quality measurement of data will convert raw data into understandable form, contributing to effective resource management and environmental conservation. Real-time water quality analysis, analysis of historical data, taking safety measures, and checking predictions.

When machine learning algorithms are utilized by researchers and organizations to analyze sensor data, providing insights into pollutant levels, water contamination, and overall water health, it becomes the most reliable method for water quality monitoring. Many diseases that occur because of poor water quality can be treated early by analyzing the water condition at an early stage. When sensors take data, the Arduino will communicate with the database, and Blynk the Arduino will generate the tokens, such as IDs and passwords, and that helps the Arduino integrate with the app backend where the ML algorithms are using. This innovative solution can help in the environmental profession, laboratories, as well as the general public, who can benefit by learning about the importance of water resources and making strategies for them. [13].

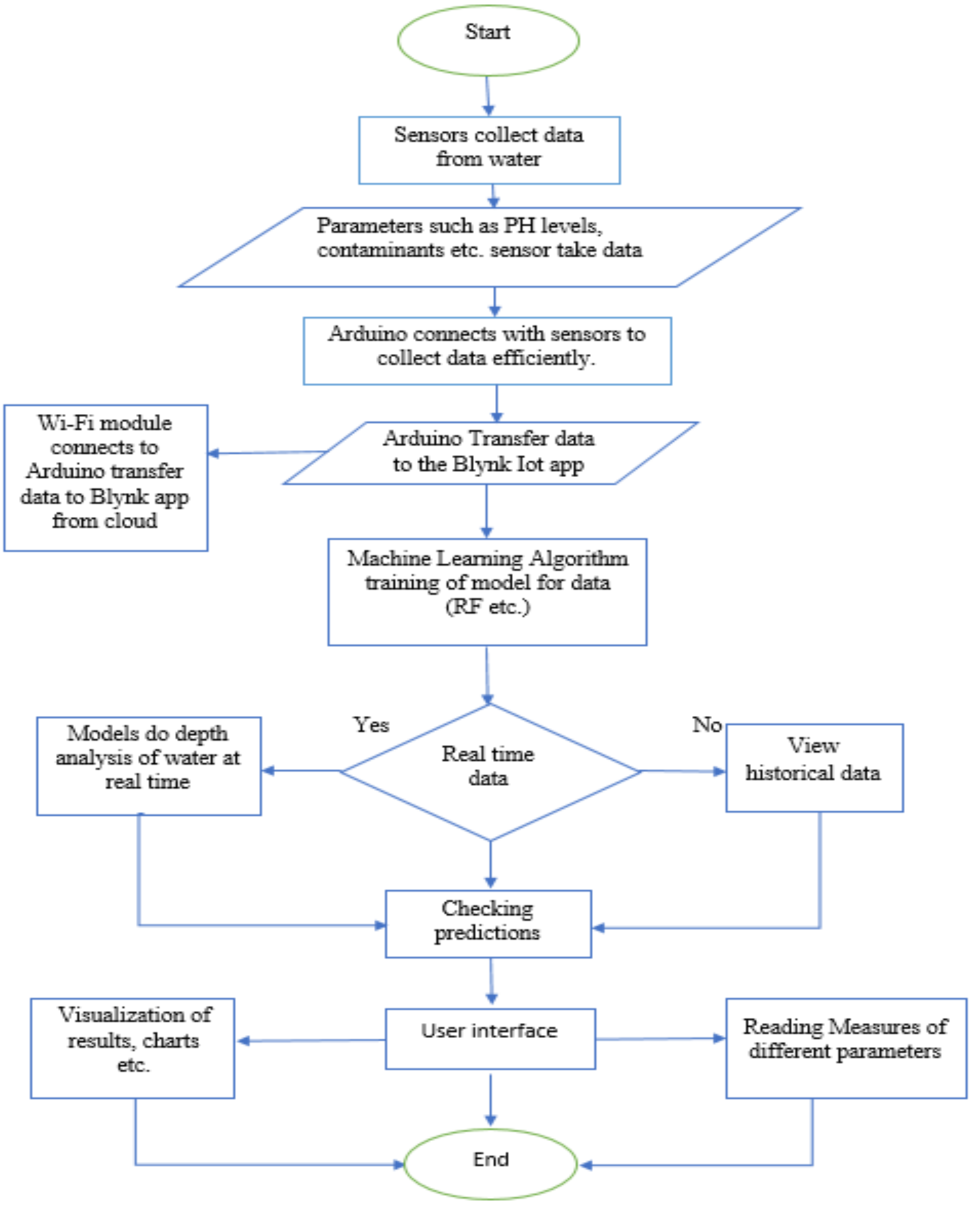

Figure 1. The Flow Diagram of Water Quality Monitoring Technology

3.1. Sensor Define Pins manage through Code of Blynk IOT app

- ONE_WIRE_BUS: Pin for the one-wire temperature sensor
- Sensor Pin: Pin for the pH sensor
- TDS Sensor Pin: Pin for the TDS sensor

In the context of the Blynk IoT platform, "virtual pin" refers to a software-defined pin that is a software control pin that is used to share data from the real hardware, through an ESP8266 microcontroller, to the software Blynk app. So you can send information without connecting the application to the actual hardware GPIO's, virtual pins.

Here's how virtual pins (V pins) are used in code to connect and visualize the water data:

```
Blynk.virtualWrite(V0,Celsius);
Blynk.virtualWrite(V1,TDS);
Blynk.virtualWrite(V2,PHValue);
```

- V0 pin: This pin sends the temperature reading to the Blynk app.
- V1 pin: This pin transfers the TDS (Total Dissolved Solids) data to the Blynk app.
- V2 pin: This pin transfers the pH data to the Blynk app.

These virtual pins help to create a user interface in the Blynk application. Through these pins, I can Display real-time data, manage the devices, and interact with water quality data without being

Limited by the number of physical pins on hardware.

Figure 1. Hardware part of water quality through sensors

In figure 2, we have hardware that includes three sensors that Collect water data seamlessly and an Arduino name as an ESP8266 microcontroller. Breadboard and jumper wires are used to send data electronically with the help of Arduino, and lastly, a water sample collection cup. The water quality measurement system has the IOT Blynk application and AI (artificial intelligence) with sensor technology. Here are the main elements used in this project.

3.1.1. Software

1) Arduino IDE 2.3.2 we write code in the Arduino IDE to read data from these sensors; the code is in C++.
2) Blynk IOT Application

3.1.2. Operating System

While the development environment for this project is Windows-based, the system is designed to be platform-independent, ensuring compatibility with macOS, Linux, and other operating systems, providing flexibility and accessibility for a diverse range of users.

3.1.3. Programming Languages

1) C++: The C++ code will be added to the Arduino to help it to work better.
2) Python: Language is going to be used in this.
3) Machine learning algorithms:
In this study, the algorithms that are used are K-Nearest Neighbors (KNN).
Random Forest
Multinomial Naive Bayes
Gaussian Naive Bayes

3.1.4. IoT Technologies

Such as sensors in IoT technologies we are using
Ph sensor, Total Dissolve Solid sensor, and temperature sensor.
Arduino ESP8266 microcontroller

3.1.5. Parameters

These are the best-quality water ranges taken from WHO as mentioned in table 2.

Table 2. Parameters Table from WHO world health organization for water drinkability

Features	WHO Unit
PH	6.5-8.5
Temperature	25° C
TDS	150-300

Potability	0 or 1

Table 2 has shown the parameters that are considered to evaluate the water quality through WQI, namely, these parameters include:

PH Level: This parameter is an index of the water's pH that determines the solubility of various chemicals and their access to nutrients and metals. It is vital when analyzing the composition of the water and the balance of the chemicals contained in it.

Temperature: It influences the chemistry of water and other reactions as well as the biological activity and solubility of gases. It also affects the rate of metabolism among the aquatic life; hence, it plays a key role in determining the quality of water.

Total Dissolved Solids (TDS): TDS also measures the sum total of soluble matter in the water, which can be organic as well as inorganic. High TDS causes water to taste different, and possibly dangerous chemicals may also be present in the water.

We want to set up with the help of IoT sensors monitoring these parameters, and the machine learning algorithms for data analysis are the real-time analysis and the predictive evaluation of the water quality. Through this method, one can alert any existing water quality issues and help in containing further complications on water quality for diverse uses.

Integrating machine learning in this study greatly enhances its ability to analyze data, enabling efficient processing, classification, and predictive modeling. This empowers the system to deliver insightful and accurate assessments of water quality parameters, ensuring more effective and timely water management solutions.

Blynk: Blynk bridges the physical and digital realms of water quality monitoring, delivering real-time insights straight to your smartphone. By installing the Blynk library, you can connect your Arduino to the Blynk server and create a visual dashboard in the app. This setup allows you to link virtual pins for data flow, ensuring sensor readings are relayed in real-time. Users are empowered with a mobile command center for monitoring water quality and exploring historical data. The cloud-based Blynk Cloud ensures seamless communication and data storage, providing a robust and accessible monitoring experience.

3.2. The Developed Interface

Figure 2 shows us that at the Blynk IoT app, sensor data is managed and visualized through this interface. On the other hand, when codes run and values show on the serial monitor of the Arduino IDE through sensors, those values transfer to the Blynk IOT app, where we have to manage the parameters of water, such as temperature and TDS and ph values. On the Blynk IoT app, those values are managed in the gauge representation figures that we have utilized and managed on the interface.

3.2.1. Temperature Test analysis

Here is the first representation testing of temperature value 30°C that tells us that this temperature of water drinkability is good and absolutely healthy to drink at this degree because it's at room temperature.

3.2.2. TDS Test Analysis

Next representation testing of TDS total dissolved solids (391 parts per million (ppm)) means that this water doesn't have any undissolved particles, and these values of TDS show that water is safe to drink and healthy for humans. What if this value of TDS is so high that it crosses the 500 ppm that we call a addition to these health complications, high TDS can result in an unpleasant taste and odor in water, reducing your water intake, which can potentially lead to dehydration. high TDS of water? It can contribute to kidney stones, heart diseases, diabetes, and gastrointestinal issues such as stomach pain and diarrhea. In addition to these health complications, high TDS can result in an unpleasant taste and odor in water, reducing your water intake, which can potentially lead to dehydration.

3.2.3. PH Test Analysis

Here is the representation of PH values of water. This test value of water is 7.02 ph, which is safe to drink for the human body. If water has a low pH level, it's very acidic. Drinking acidic water regularly can have adverse effects on your health. Regular consumption of acidic water can cause acid reflux and frequent heartburn—it also doesn't taste very good. Water with an extremely low pH level can even corrode certain metals. So the pH level of water always should be between 6.5 ph and 8.5 ph, which is safe to drink.

Figure 3. Water quality Blynk IOT application where sensors collected data shows in gauge forms

These employs using the Blynk app for the display of data in real time. In this aspect, it is capable of reading the temperature, TDS, and pH values using sensor connections and transferring them to the Blynk application for smartphone use to read the parameters. In the code that integrates in the Blynk IoT platform with an ESP8266 microcontroller to monitor temperature, TDS (total dissolved solids), and pH values.

4. The research findings and discussion

Sensors receive information from water about several facets of its quality, like temperature, turbidity, pH, TDS, and the presence of impurities. This study shows the classification and prediction methods that will be analyzed by machine learning algorithms. Arduino interfaces with the sensors to collect and read the data. Arduino transmits the processed sensor data to an app backend from a cloud database. Result of Water Drinkability Based on temperature and TDS:

Observation and analyses of Figure 4:

- Green dots represent water drinkability.
- Red dots represent not water drinkability.

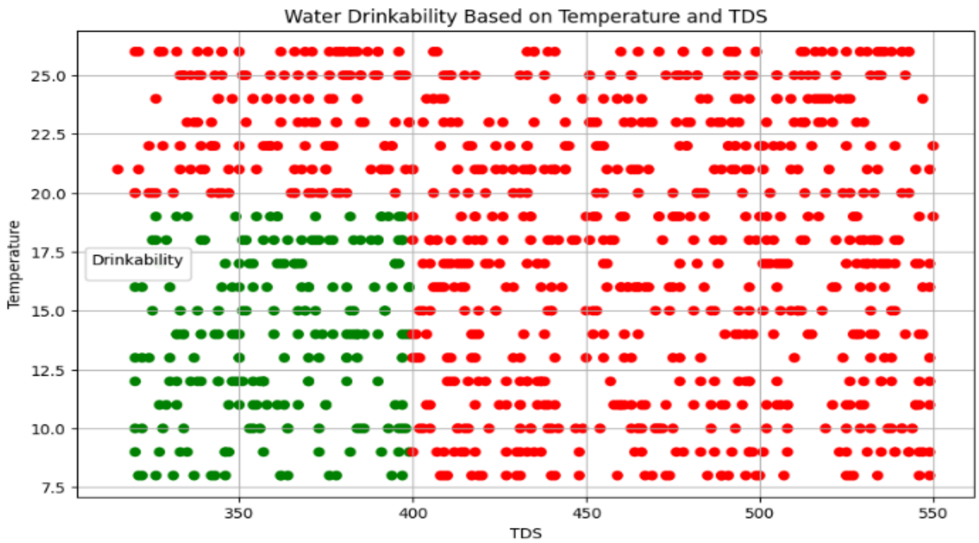

Figure 4. Results of water drinkability based on parameters

4.1. Random Forest

It is used to analyze complex patterns in the data, providing a reliable method for understanding and predicting water quality trends based on these parameters.

Water Quality= $\frac{1}{n}\sum_{i=1}^{n} T_i(\text{TDS, Temperature, pH})$ Eq (1)

Where:

- n: The number of decision trees in the forest.
- T_i: The prediction from i – th tree.

The final water quality prediction is the average (or majority vote) of the predictions from all the trees.

Figure 5 describes how the Random Forest classifier predicted all instances correctly, with 100% accuracy. This is indicated by the diagonal line from top-left to bottom-right (1s) and zeros elsewhere.

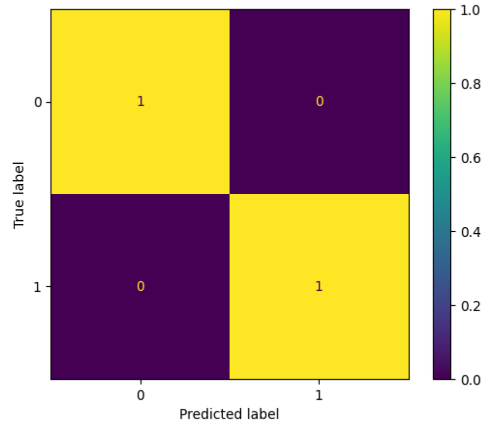

Figure 5. Confusion Matrix of random forest

4.2. Gaussian Naive Bayes

The Naive Bayes classifier works on the idea that the features we use to make predictions are independent and don't affect each other. In reality, features often interact with one another in shaping the outcome, but the Naive Bayes classifier overlooks these relationships. This simplification and ease of handling, however, allows the model to train more quickly and in time.

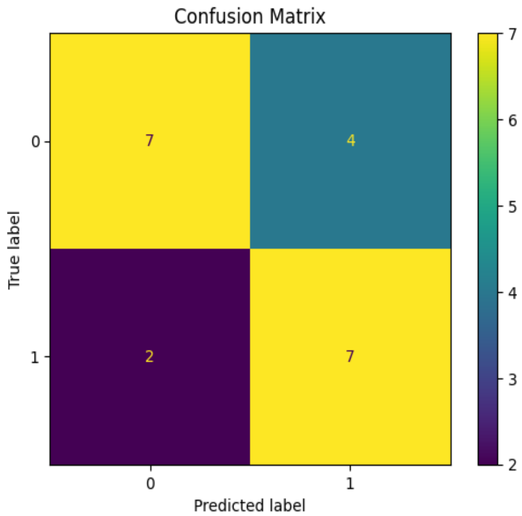

Figure 6. Confusion Matrix of Gaussian Naive Bayes

$$P\ 1\text{Water quality} = \mid Features) = \frac{P(Features|Water\ Quality=1)\cdot P(Water\ Quality=1)}{P(Features)} \quad \text{Eq (2)}$$

Where:

- P (Water quality = 1|Features): The probability of having good water quality given the features.
- P (Features Water Quality=1): The likelihood of the features given good water quality.
- P (Water Quality=1): The prior probability of having good water quality.
- P (Features): The probability of the features, used to normalize the result.

Figure 6 describes how the Gaussian Naive Bayes model has more errors compared to the Random Forest. It misclassified 4 negative instances as positive (FP) and 2 positive instances as negative (FN).

4.3. K-Nearest Neighbors (KNN)

$$\text{Water quality} = \frac{1}{k}\sum_{i=1}^{k} Q_i \quad \text{Eq (3)}$$

Explanation:

k The number of nearest neighbors.

Q_i The quality label (e.g., 0 or 1) of the i nearest neighbor.

The algorithm predicts water quality based on the average of the k nearest neighbors in the feature space (TDS, Temperature, and pH).

Figure 7 shows about The KNN model has a fair number of misclassifications. It incorrectly predicted 6 negative instances as positive (FP) and 1 positive instance as negative (FN).

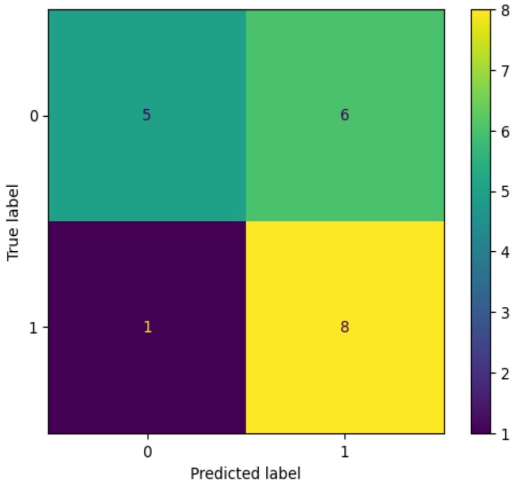

Figure 7. Confusion Matrix of K-Nearest Neighbors (KNN)

4.4. Multinomial Naive Bayes

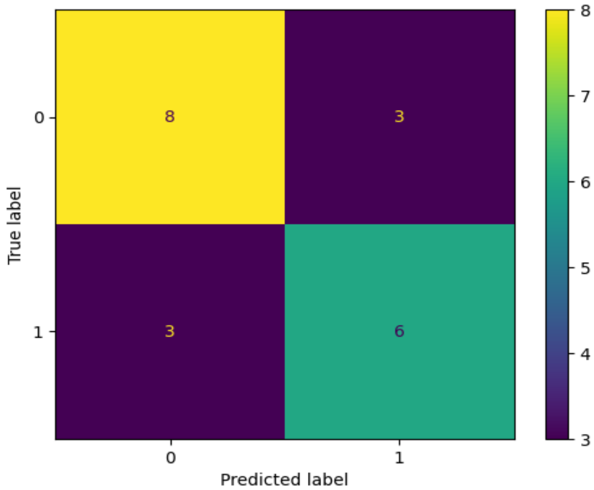

Figure 8. Confusion Matrix of Multinomial Naive Bayes

$$P\text{ (Water quality } = 1 \mid Features) = \frac{P(Water\ Quality=1)\ \prod_{j=1}^{m} \text{P(Featurej|Water Quality=1)}}{P(Features)} \quad \text{Eq (4)}$$

- P(Featurej|Water Quality=1): the probability of each feature (e.g., TDS, Temperature, pH) given good water quality.
- The product of these probabilities is used because Multinomial Naive Bayes assumes that features are multinomially distributed.

Figure 8 shows The Multinomial Naive Bayes model also has some misclassifications. It misclassified 3 negative instances as positive (FP) and 3 positive instances as negative (FN).

Table 3. Evaluation of ML models:

Models	Precision	Recall	Accuracy
Random forest	100.0%	(100%)	100.0%
Gaussian Naive Bayes	63.64%	(77.78%)	70.00%
Multinomial Naive Bayes	66.67%	(66.67%)	70.00%

Evaluation of ML models is considered the most impressive analysis through which we get the most accurate results, and there is no doubt that this method amazingly measures the water health condition.

When the performance of algorithms is studied, it is seen that the Random Forest gives the most successful results than other powerful algorithms such as KNN, Gaussian, and multinomial. Each model balances the recall, precision, and accuracy in this way. As The number of data sets is around 1000 or more, but it is observed that the process is fast.

5. Conclusion

This study presents the development of an IoT and machine learning-based system for real-time water quality monitoring and prediction. By deploying sensors to continuously gather data parameters on pH, temperature, and Total Dissolved Solids (TDS), we have established a comprehensive water quality assessment framework. Machine learning algorithms analyze this data, providing predictive insights and early warnings for potential issues. A critical component is the Blynk IoT app, which offers a user-friendly interface for real-time data visualization, enabling stakeholders to monitor water quality metrics seamlessly. This system enhances the efficiency of water quality management and supports environmentalists and volunteer organizations. Empowers communities to take informed actions to ensure safe water resources by facilitating data-driven decision-making processes. The integration of the Blynk app with machine learning algorithms is crucial for promoting sustainable water management and protecting environmental health through continuous improvement and collaboration in future work related to water. Furthermore, for evaluation purposes, a real-time dataset has been collected with the help of a sensor to get an appropriate result. While in terms of accuracy, Random Forest performs effectively as compared to other ML algorithms such as KNN, GNB, and MNB. In the future, this work will be extended to more parameters to enhance water quality.

References

1. Wang, Y.; Ho, I.W.H.; Chen, Y.; Wang, Y.; Lin, Y. Real-Time Water Quality Monitoring and Estimation in AIoT for Freshwater Biodiversity Conservation. IEEE Internet Things J. 2022, 9, 14366–14374.
2. Okada, Y.; Mertens, M.; Liu, N.; Lam, S.S.W.; Ong, M.E.H. AI and Machine Learning in Resuscitation: Ongoing Research, New Concepts, and Key Challenges. Resusc. Plus 2023, 15, 100348.
3. Thombre, S.; Islam, R.U.; Andersson, K.; Hossain, M.S. Performance Analysis of an IP-Based Protocol Stack for WSNs. In Proceedings of the IEEE Conference on Computer Communications Workshops, 2016; pp. 360–365.
4. Ajakwe, S.O.; Ihekoronye, V.; Udodiri, A.I.; Kim, D.S.; Lee, J.M. Connected Intelligence for Smart Water Quality Monitoring System in IIoT. In Proceedings of the 13th International Conference on ICT Convergence (ICTC), 2022; pp. 2386–2391.
5. Pranata, A.A.; Lee, J.M.; Kim, D.S. Towards an IoT-Based Water Quality Monitoring System with Brokerless Pub/Sub Architecture. In Proceedings of the International Symposium on Local and Metropolitan Area Networks, 2017; pp. 1–6.
6. Chowdury, M.S.U.; Emran, T.B.; Ghosh, S.; Pathak, A.; Alam, M.M.; Absar, N.; Andersson, K.; Hossain, M.S. IoT-Based Real-Time River Water Quality Monitoring System. Procedia Comput. Sci. 2019, 155, 161–168.
7. Ramesh, M.V.; Nibi, K.B.; Kurup, A.; Mohan, R.; Aiswarya, A.; Arsha, A.; Sarang, P.R. Water Quality Monitoring and Waste Management Using IoT. In Proceedings of the Global Humanitarian Technology Conference, 2017; pp. 1–7.
8. Shokoohi, M.; Tabesh, M.; Nazif, S.; Dini, M. Water Quality-Based Multi-Objective Optimal Design of Water Distribution Systems. Water Resour. Manag. 2017, 31, 93–108.
9. Zhang, W.; Wang, Y.; Peng, H.; Li, T.; Tang, J.; Wu, K.B. A Coupled Water Quantity–Quality Model for Water Allocation Analysis. Water Resour. Manag. 2010, 24, 459–484.
10. Yang, L.; Driscol, J.; Sarigai, S.; Wu, Q.; Lippitt, C.D.; Morgan, M. Towards Synoptic Water Monitoring Systems: A Review of AI Methods for Automating Water Body Detection and Water Quality Monitoring Using Remote Sensing. Sensors 2020, 22, 1832.
11. Ajakwe, S.O.; Ihekoronye, V.U.; Kim, D.; Lee, J.M. Pervasive Intrusion Detection Scheme to Mitigate Sensor Attacks on UAV Networks. In Proceedings of the Korean Institute of Communication and Sciences Summer Conference, 2022; pp. 1267–1268.
12. Jia, W.; Cheng, J.; Hu, H. A Cluster-Stacking-Based Approach to Forecasting Seasonal Chlorophyll-a Concentration in Coastal Waters. IEEE Access 2020, 8, 99934–99947.
13. Mehmood, H.; Liao, D.; Mahadeo, K. A Review of Artificial Intelligence Applications to Achieve Water-Related Sustainable Development Goals. In Proceedings of the IEEE/ITU International Conference on Artificial Intelligence for Good, 2020; pp. 135–141.
14. Susanti, N.D.; Sagita, D.; Apriyanto, I.F.; Anggara, C.E.W. Design and Implementation of Water Quality Monitoring System (Temperature, pH, TDS) in Aquaculture Using IoT at Low Cost. In Proceedings of the 6th International Conference of Food, Agriculture, and Natural Resource, 2022; pp. 16.
15. Bhardwaj, A.; Dagar, V.; Khan, M.O.; Aggarwal, A.; Alvarado, R.; Kumar, M.; Irfan, M.; Proshad, R. Smart IoT and Machine Learning-Based Framework for Water Quality Assessment and Device Component Monitoring. Environ. Sci. Pollut. Res. 2022, 29, 46018–46036.
16. Nasir, N.; Kansal, A.; Alshaltone, O.; Barneih, F.; Sameer, M.; Shanableh, A.; Al-Shamma'a, A. Water Quality Classification Using Machine Learning Algorithms. J. Water Process Eng. 2022, 48, 102907.
17. Othman, F.; Alaaeldin, M.E.; Seyam, M.; Ahmed, A.N.; Teo, F.Y.; Fai, C.M.; Afan, H.A.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Efficient River Water Quality Index Prediction Considering Minimal Number of Input Variables. Eng. Appl. Comput. Fluid Mech. 2020, 14, 751–763.
18. Mustafa, H.M.; Mustapha, A.; Hayder, G.; Salisu, A. Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. In Proceedings of the 6th International Conference on Inventive Computation Technologies, 2021; pp. 968–975.
19. Amudha, V.; Lakshmi, S.V. Deep Ensemble-Based Water Quality Index Prediction and Classification Using Randomized Low-Rank Approximation. In Proceedings of the Second International Conference on Augmented Intelligence and Sustainable Systems, 2023; pp. 434–439.
20. Liu, X.G.; Ai, K.F. Inverse Problem of Multiple Parameters Identification for BOD-DO Water Quality Model Using Evolutionary Algorithm. In Proceedings of the International Conference on Computational and Information Sciences, 2013; pp. 1076–1079.

21. Zougmore, T.; Gueye, B.; Malo, S. An AI-Based Approach to the Prediction of Water Points Quality Indicators for Schistosomiasis Prevention. In Proceedings of the IEEE Multi-Conference on Natural and Engineering Sciences for Sahel's Sustainable Development, 2022; pp. 1–6.
22. Mujahid, M. AquaSmart: An IoT-Enabled Smart Water Management Solution. Res. 2020, 1–12.
23. Ahmed, U.; Mumtaz, R.; Anwar, H.; Mumtaz, S.; Qamar, A.M. Water Quality Monitoring: From Conventional to Emerging Technologies. Water Sci. Technol. Water Supply 2019, 20, 28–45.
24. Abuzir, S.Y.; Abuzir, Y.S. Machine Learning for Water Quality Classification. Water Qual. Res. J. 2022, 57, 152–164.
25. Nguyen, D.P.; Ha, D.; Trinh, N.T.; Nguyen, M.T. Application of Artificial Intelligence for Forecasting Surface Quality Index of Irrigation Systems in the Red River Delta, Vietnam. Environ. Syst. Res. 2023, 12, 1–14.
26. Priya, M.D. Smart Water Quality Monitoring System Using IoT. CVR J. Sci. Technol. 2021, 21, 31–38.