Hydro Sense: Empowering Water Quality Monitoring Through IoT And ML

Athira Sankar
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
cs.athira@sbcemail.in

Sajishma S R
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
sajishmasr2001@gmail.com

Alan Raj
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
alanraj416@gmail.com

Vaishnavi A K
CSE Department
Sree Buddha College of Engineering,
Alappuzha, India
vaishnavivanisree@gmail.com

Reshmi S Kaimal
CSE Department
Sree Buddha College of
Engineering,
Alappuzha, India
reshmiskaimal@gmail.com

Abstract— **Clean water is an essential resource in sustaining life, and ensuring the quality of drinking water is crucial for promoting the wellbeing and health of individuals. Water quality monitoring systems are essential for evaluating and guaranteeing the safety of water sources. The current water quality surveillance system lacks real-time information, which is a drawback. Manually checking water quality continuously is impractical. To address this issue, we have developed a cost-effective live-stream water quality monitoring system specifically for consumable water. Key factors such as turbidity, Ph and temperature need to be measured to detect contaminants and prevent water-related illnesses. Our system includes specially designed sensors connected to a microcontroller with an integrated ADC circuit for signal conversion, data processing, and analysis. The hardware component is connected to the main system via a USB cable. The system displays the values of each parameters in the Blynk console and when values are manually given to the trained model it will predict if the water is in consumable form or not. We have trained the model using the Random Forest classification Algorithm to predict if the water is consumable or not**.

Keywords— pH sensor, turbidity sensor, temperature sensor, ESP32, machine learning, Random Forest classification Algorithm

I. INTRODUCTION

The increasing population, industrialization, and environmental challenges underscore the urgent need for comprehensive and sustainable solutions to address the global issue of safe drinking water in the 21st century. The UN and the WHO have realized rightness of humans on potable, affordable, and easily available drinking water for personal and household usage. The Central Pollution Control Board (CPCB) has organized a network of water quality monitoring stations This network enables the Central Pollution Control Board (CPCB) to systematically assess water quality parameters, identify potential sources of pollution, and formulate targeted strategies to safeguard and improve the overall quality of water bodies across the nation. These stations regularly assess the water quality, either on a monthly or yearly basis, to guarantee that it is being preserved or improved to meet the desired standards. Shockingly, recent research by the WHO has revealed that 844 million people lack access to even basic drinking water services, which can lead to life-threatening illnesses like diarrhea, Cholera, Dysentery, Typhoid, and Polio. Hygienic water is a crucial resource for sustaining life, and the quality of drinking water is vital for human health and wellbeing. To combat water pollution, there is the need of water quality surveillance system

A low-cost water quality surveillance system has been developed to ensure the safe supply of consumable water. Manually checking the water quality at all times is an impossible task. Therefore, the need for an automatic real-time monitoring system becomes imperative to

IJERA Volume 04, Issue 01
DOI: 10.5281/zenodo.12519386

safeguard the health of the water stored in our society or apartment's water tank. This system acts as a vigilant guardian, promptly alerting us to any issues that may arise with the reserved water, ensuring our safety and well-being. This system is designed to rigorously monitor water resources' pollution levels, enabling the creation of a conducive environment for ensuring the safety and quality of drinking water.

II. RELATED STUDY

The literature review highlights the increasing significance of employing wireless sensor network (WSN) technology in the development of advanced water quality monitoring systems, showcasing its potential to revolutionize real-time data collection and analysis for improved environmental management. The system outlined in [1] has been developed with the purpose of assessing various water quality factors such as color, pH, temperature, turbidity, dissolved oxygen, conductivity and total natural carbon. This research paper introduces an economical approach to mitigate water pollution in privately owned elevated tanks. Through the integration of Internet of Things (IoT) devices and artificial intelligence (AI) algorithms, the system maintains continuous monitoring of water quality, offering the capability to predict and forecast potential pollution incidents, thereby allowing for proactive and preventive measures to ensure the safety and sustainability of water resources. It utilizes a network of multiple sensors connected to NodeMCU to collect data on water parameters, enabling timely notifications to users before any contamination takes place.

In [2] a system focuses on the measurement of water levels in primary tanks, detection of leaks, and monitoring water consumption per unit in a residential complex. Furthermore, the system is capable of detecting the flow rate and turbidity of water. The sensor data is processed by an ESP8266 hall effect sensor and transmitted wirelessly to an Arduino UNO via a Wi-Fi module. Subsequently, the Arduino UNO presents the data on a cloud platform using graphical representations. The system continuously monitors the water levels in the tanks that supply water to the residential units.

In [3] a system introduces an innovative and economical solution to the aquatic monitoring issue in both water supply networks and consumer sites. The core concept of this approach revolves around the development of affordable sensor nodes that can continuously monitor and evaluate water quality within pipes. These sensor nodes consist of multiple electrochemical and optical sensors strategically placed within the pipes. The main objective is to create a lightweight and cost-effective design that can operate reliably for extended periods.

In [4] a system focuses on the development of a sensor that can measure various factors such turbidity, Dissolved Oxygen (DO), ammonia, chloride, nitrate, and hardness too determine the grade of well water. Integrated into an Internet of Things (IoT) framework, the sensor plays a crucial role in a sophisticated water quality monitoring system that utilizes Machine Learning algorithms, enabling precise and automated analysis for enhanced environmental data insights. This system has several advantages over traditional system of water grade checking. This aquatic quality monitoring system is a crucial step towards ensuring access to safe and clean drinking water.

In [5] a system details the development of a water quality monitoring system that utilizes smart sensor networks and Internet of Things (IoT) technology. The system aims to provide real-time notifications to users regarding water quality parameters. The World Health Organization (WHO) reports that approximately 76 million people in our country lack access to safe drinking water, and 21% of diseases are water-related. To address this issue, it is necessary to measure water parameters such as pH, conductivity, turbidity, and temperature, which can help detect water contaminants. The sensors used in this system are connected to a microcontroller with an inbuilt ADC circuit that converts, processes, and analyzes the data.

In [6] the system approach emphasizes harnessing Wireless Sensor Network (WSN) technology for continuous real-time data collection, enabling comprehensive monitoring of vital water quality parameters in freshwater environments such as rivers, lakes, or wetlands, facilitating more accurate and timely environmental assessments. The WSN system serves as an area for monitoring freshwater quality readings, strategically placed at various locations where each node can connect with different water quality sensors. The system is designed to operate on renewable energy sources by harnessing solar power during the day, with an optimized power management system to ensure long-lasting functionality in remote rural areas. Data collected by each node is transmitted to a relay station acting as network manager, and then forwarded to a monitoring central server communicates data seamlessly through the GSM network, ensuring real-time transmission and enabling remote access. The PIC16F886 nano-watt MCU serves as the power source for this system, with RF XBEE 802.15.4, ISM 2.4 GHz modules employed for communication in each node. Additionally, the Coordinator device is equipped with a GSM/GPRS modem and a monitoring LCD display for efficient data transmission and on-site information display.

In [7] a system utilizes the Wireless Sensor Grid , consisting of the Wireless Water Quality Monitoring Network and Offsite Data Center. The hardware part is centered around the wireless microprocessor CC2430. The sensor network is constructed based on the Zigbee wireless transmission protocol. Wireless Sensor Grid collects samples of water qualities and then the data will be transmitted to the Internet using the GPRS DTU which is furnished with TCP/IP protocol. The Offsite Data Center receives real-time water quality data via the Internet, conducts analysis, processing, and records the real time data. By leveraging real-time data provided by the environmental protection department, industries dependent on regional water quality conditions, including the industrial, plant, and aquaculture sectors, can receive timely guidance and make informed decisions to ensure environmental sustainability and compliance with regulatory standards. This approach enhances efficiency and reduces costs significantly.

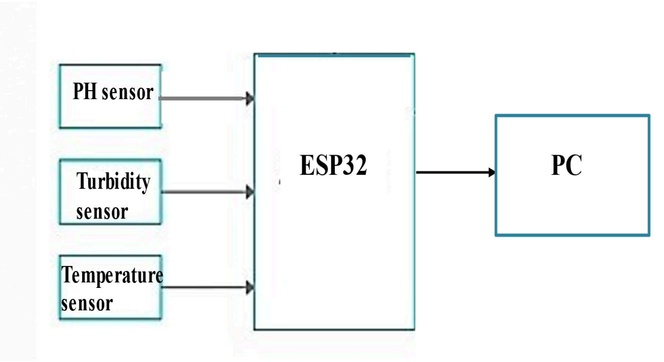

Fig 1: Block diagram of water quality monitoring system

The presented paper underscores the significance of real-time water quality monitoring in an Internet of Things (IoT) environment, detailing the comprehensive block diagram of the proposed method where various sensors (temperature, pH, turbidity) which are connected to a core controller. The core controller accesses sensor values, processes the data, and transfers it through the internet, allowing users to view the water quality parameters on the Blynk console and can be also viewed in trained model. The system aims to ensure that the measured water parameters conform to the specified World Health Organization (WHO) standard range, as illustrated in the table below.

SL.No	Parameter	Range	Unit
1	Temperature	15-40	°C
2	pH	6.5-8.5	pH
3	Turbidity	0-3	NTU

Table1: water quality physical parameter standard value table

B. Components

a. Ph Sensor : The pH value of a solution determines its acidity or alkalinity, with lower pH indicating higher hydrogen ion concentration and greater acidity, while higher pH indicates lower hydrogen ion concentration and more alkalinity. The solution in question falls within the mildly acidic to slightly alkaline range, with a pH range of 6 to 8.5. This versatile solution can be used for various purposes, including food preparation, soil analysis, and water testing. By incorporating an Arduino microcontroller, real-time monitoring and data reporting of pH levels are made possible, which is particularly beneficial for applications requiring precise pH control, such as hydroponics or aquaponics systems.

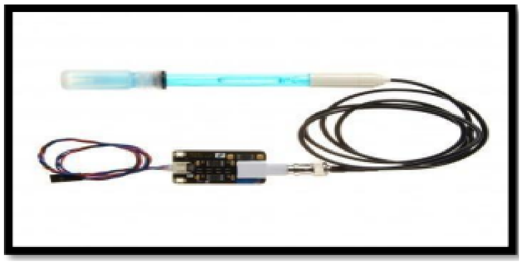

Fig 2: pH sensor

b. Turbidity sensor: Turbidity, the measurement of how cloudy or hazy water is, plays an important role in identifying the clearness and grade of water bodies. It is an important indicator for assessing the health and condition of aquatic ecosystems. One of the main reasons why turbidity is a reliable measure for water quality is because it can show the presence of suspended particles in the water. These particles can include sediment, organic matter, algae, and other pollutants. When turbidity levels are high, it means there is a higher concentration of these particles, which can be harmful to aquatic life and the overall health of the ecosystem. One significant consequence of turbidity is that it interferes with the growth and survival of submerged aquatic vegetation. These plants, like seagrasses and macroalgae, rely on sunlight for photosynthesis. However, when the water becomes turbid, the suspended particles scatter and absorb light, reducing the amount of sunlight that reaches the vegetation. As a result, the growth and productivity of submerged aquatic vegetation are hindered, leading to a decrease in biodiversity and ecological balance.

Fig 3: turbidity sensor

c. Temperature sensor: Temperature sensors like the DS18B20 are extensively employed across diverse fields including environmental monitoring, industrial processes, and HVAC systems for accurate temperature measurement. Renowned for its precision and consistency, it is favored by engineers and researchers alike for its reliability, establishing it as the preferred option in water quality assessments.

This sensor is designed to be waterproof and can easily be integrated into different systems through its digital interface. It can accurately measure water temperature in extremely cold or hot conditions due to its wide temperature range. Its compact dimensions and minimal energy requirements make it an ideal choice for use in remote locations or applications powered by batteries. In summary, the DS18B20 temperature sensor is a versatile tool for monitoring water temperature in different environments. Utilizing a one-wire communication protocol, this digital sensor requires only a single data line (along with GND) to interface with an Arduino. Additionally, each DS18B20 sensor possesses a distinct 64-bit serial code, facilitating the connection of one or more sensors to a single data wire.

Fig 4: Temperature sensor

d. ESP32: The ESP32 is a versatile and efficient microcontroller chip that offers advanced capabilities for a wide range of applications. Manufactured by multiple companies using a low-power 40nm process, it integrates Wi-Fi, Bluetooth, and a dual-core processor on a single chip. Operating at 3.3V, the ESP32 combines low power consumption with impressive performance thanks to its dual-core architecture, making it suitable for battery-powered devices. It provides extensive peripheral support through GPIO, ADC, SPI, I2C, and UART interfaces. The chip also incorporates security features to protect data. With its excellent RF performance, multitasking abilities, and adaptability for diverse applications ranging from wireless sensors to wearable devices, the ESP32 delivers reliable, high-performance solutions in a compact and cost-effective package.

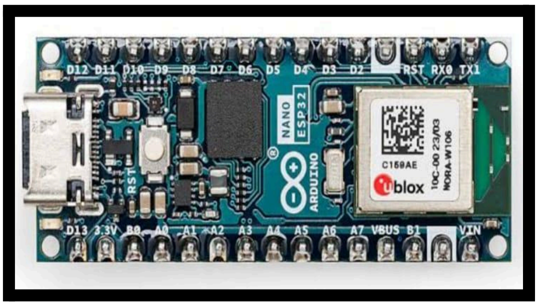

Fig 5: ESP32

C. Algorithm

The Random Forest Model is utilized in a aquatic surveillance system through machine learning to observe and forecast water quality parameters. By training the model on data that includes features like pH, temperature, and turbidity, the Random Forest algorithm can identify patterns and correlations within the dataset. This trained model can then be utilized to predict water quality parameters in real-time based on current input values.

The Random Forest model is a robust group learning technique that boosts prediction accuracy by merging numerous decision trees. During training, it generates numerous decision trees and yields the mode of the classes, thereby improving the model's robustness and ability to generalize. In the realm of water quality monitoring, the Random Forest algorithm can be applied with machine learning to assess and forecast the suitability of water for drinking. Through model training, it gains the capability to offer real-time predictions of water quality parameters based on current input values. This versatile algorithm is commonly used for classification tasks, as it adeptly merges multiple decision trees for precise predictions. Random Forest randomly chooses subsets of training data and features to construct the decision trees, then combines the predictions from all trees to make the final prediction. Notably, this algorithm is well-known for its capacity to handle large datasets and its resistance to overfitting.

Data preprocessing involves managing missing values, encoding categorical variables, and partitioning it into training and testing sets. Subsequently, an instantiation of the RandomForestClassifier class from the scikit-learn library is created based on the task, which may involve classification. The model is then trained on the data using the fit() method and applied to the testing set for predictions using the predict() method. Performance evaluation includes metrics like accuracy, precision, recall, or mean squared error. Necessary adjustments, such as hyperparameter tuning or feature selection, may be implemented to optimize the model. The system undergoes continuous refinement through the integration of enhancements to the feature extraction process, model training, or result interpretation.

IV IMPLEMENTATION

The system is comprised of two primary components: hardware and software. The hardware component encompasses sensors, which facilitate the measurement of real-time values. Moreover, an ESP32 is employed to convert analog values into digital values. The USB facilitates the linkage between the hardware and software components. Programming the ESP32 involves utilizing the Embedded C language. Upon successfully interfacing all the hardware components with the core, programming for the sensor and communication devices becomes essential. To obtain water parameters from the sensor, the module needs to be powered, and each sensor value is measured at various intervals. The sensor values will then be transmitted to the core controller, which is the ESP32. The ESP32 is equipped with an inbuilt Analog to Digital Converter (ADC) that aids in converting the sensor value into a digital value. Subsequently, the digital values will be transmitted to the system using a USB cable. Once the digital value is transmitted to the system, the values of each parameter will be displayed in the Blynk sever and by analyzing the value of each parameters the blynk server shows if the water is consumable or not.

In the Machine Learning part of the system, the trained model using the random forest classification Algorithm predict if the water is consumable or not by analyzing the values of each parameters. Ensemble learning method is used to predict water quality.

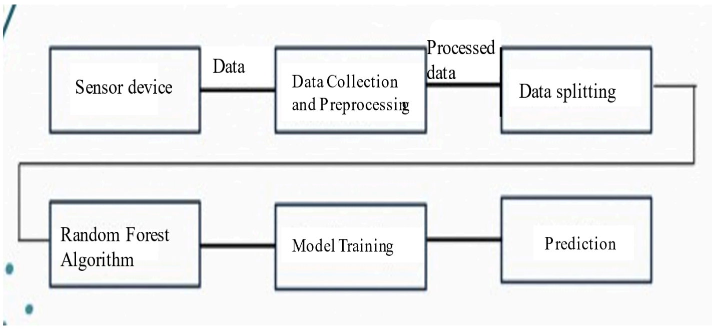

Fig 6: System Architecture

By integrating a Random Forest algorithm into a Machine Learning-based water quality monitoring system, the algorithm can analyze sensor data and make decisions based on predefined rules.

The Random Forest classifier can classify water quality based on different parameters, such as determining if the water is safe or if certain pollutants exceed acceptable levels. To establish an efficient water quality monitoring system, essential steps like gathering and preprocessing data are crucial. These procedures guarantee the accuracy, reliability, and readiness of the collected data for analysis.

The sensor device acquires a large volume of data, which must be refined and optimized for better usability. Throughout the data collection process, the sensor device records data on different factors like pH levels, temperature, and turbidity at regular intervals to track changes over time and ensure precise monitoring. After data collection, it goes through a preprocessing stage to enhance its usability for analysis. This step involves refining the data by eliminating inconsistencies, errors, or outliers that could impact the analysis's accuracy. The data is also structured and organized in a manner that simplifies its use during the analysis phase.

Data refinement includes identifying and eliminating outliers or incorrect data points that may arise from sensor malfunctions or environmental disruptions. It is also crucial to fill in missing data using interpolation techniques, particularly if data transmission is sporadic. During this stage, categorical values are converted into numerical values. Data integration encompasses merging data from different sensors and sources to offer a comprehensive overview of water quality. Normalizing or standardizing data is also vital to align all parameters to a common scale, especially when working with sensors that measure varying ranges. The collected data from the microcontroller is transmit

ed to the system platform for processing using USB cable. To ensure timely transmission of the collected water quality data from sensors to the platform, USB cable. Data processing is crucial in extracting valuable information from raw sensor data within the system, which undergoes regular intervals of processing. Leveraging the scalability, computational power, and storage capabilities of the system, the aquatic quality observing system utilizes predictive analysis with Random Forest algorithms for live monitoring to predict water consumability. The data is formatted appropriately to serve as input for the Random Forest algorithm. When developing predictive models or algorithms for water quality monitoring, it is vital to divide the data for two purposes, one is for data training and another set is for data testing. This is done to evaluate the performance of the model.

Training the Random Forest model with the training dataset constructs multiple decision trees through ensemble learning, enhancing accuracy and mitigating overfitting. Creating a Random Forest classification model is an essential component of any water quality monitoring system as it helps predict water quality.

This algorithm examines the training data to comprehend the correlation between various features and the target variable. Random Forest offers insights into feature importance by evaluating the reduction in impurity achieved by each feature. This information is invaluable in comprehending which features hold the most influence in the classification process. Once the system has learned from the training, it can now predict or estimate values for new, unknown data. Once model is performing well then we can use it to predict results on new data.

The model can take in new information and make predictions or classifications using the knowledge it acquired during training. By utilizing the loaded dataset, the model is trained to identify whether the water is safe for consuming or not. If the condition is true (i.e., data equals 0), it signifies that the water is unsuitable for consumption. Conversely, if the condition is false, it indicates that the water is safe to drink.

V. RESULT AND DISCUSSION

Water samples from different water sources were tested to build a reference on the parameters for each type of sample. The chosen sample of water belongs to tap water, surface water, pool water, etc. The water samples were tested all together at indoor ambient temperature. Readings were taken simultaneously. For security reasons the systems were not installed in the specific areas of interest, instead water samples were collected and tested in a safe controlled environment. As we tested different water Samples, we got some results.

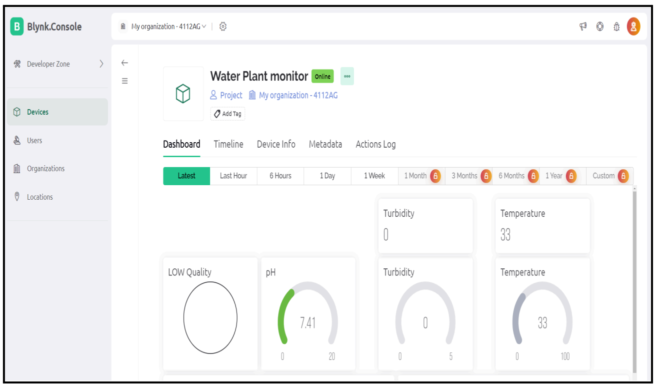

Fig 7: The pH count, water temperature and Turbidity level of a pure water

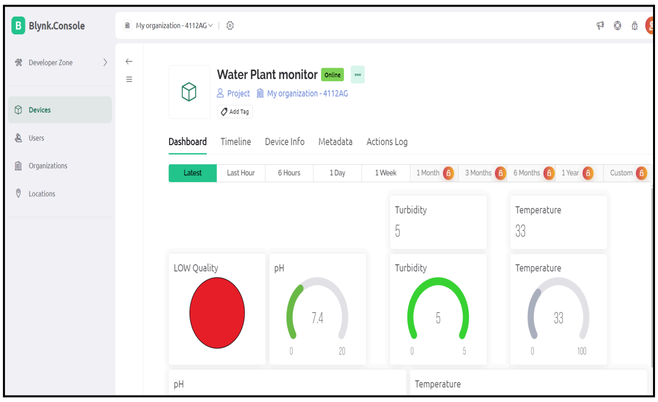

Fig 8: Levels of current water pH count, water temperature and Turbidity of impure water.

In water quality monitoring system, after getting each values of the water parameters it will be entered to the prediction window. If the values of the parameters satisfy the pure water parameter value specified by the WHO then the system will predict "The water is good for Drinking".

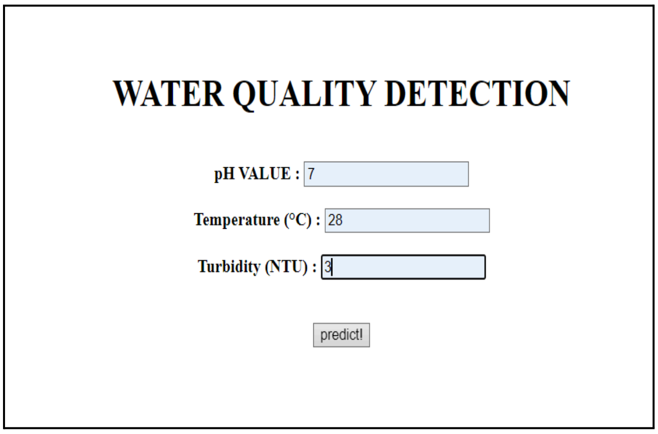

Fig 9: The values of pure water

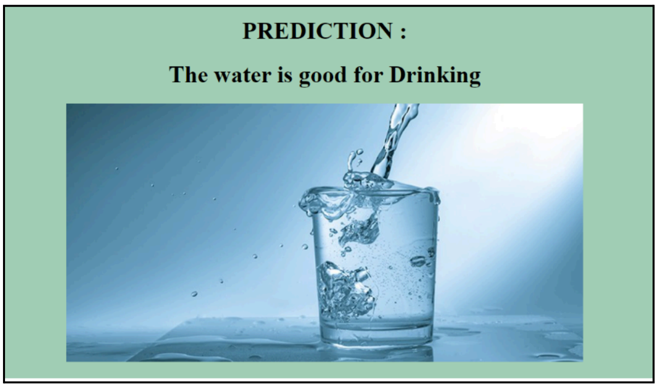

Fig 10: Result showing the water is good for drinking

If the values of the parameters does not satisfy the pure water parameter value specified by the WHO then the system will predict "The water is not good for Drinking".

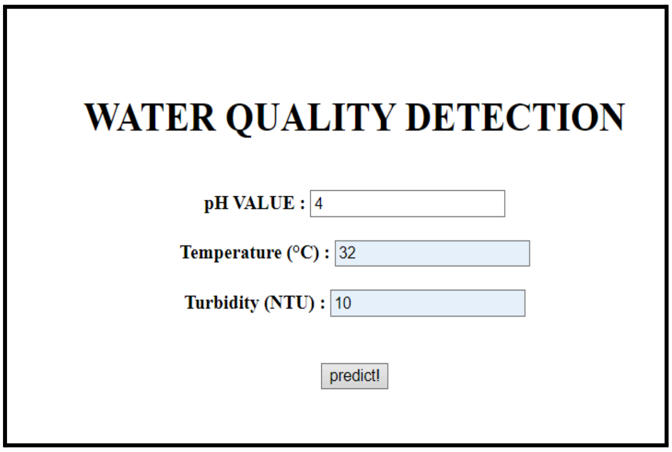

Fig 11: The values of Impure water.

Fig 12: Result showing the water is not good for drinking

VI. CONCLUSION AND FUTURE SCOPE

The proposed intelligent water quality monitoring system improves upon the current manual process by offering automated real-time analysis. This is achieved through the use of a self-contained sensor node, which eliminates the need for human involvement and reduces measurement errors. The collected water quality data serves as the basis for timely notifications, regulatory compliance, and the protection of water resources. Overall, this system achieves automated, precise continuous aquatic quality monitoring to enhance decision-making. Its primary objective is to ensure the safety of water sources for both human consumption and environmental preservation. By incorporating advanced technology and eliminating the manual data collection process, this system offers a cost-effective and efficient method for monitoring water quality. It can promptly alert and inform users of any deviations from established standards, enabling swift response and intervention to prevent water contamination or pollution. This system can be used in many fields like water distribution systems, industries, nuclear power plants and can also be used to measure the water quality parameters of lakes & rivers. This monitoring and controlling process can be performed anytime and anywhere in the world. In future, we can include biological sensors for better detection of contaminants in water and can install the system in several locations for high spatiotemporal coverage.

References

[1] C Ashwini, Uday Pratap Singh, Ekta Pawar, Shristi (2019), 'Water Quality Monitoring Using Machine Learning And IOT' - International journal of scientific & technology research volume 8, issue 10

[2] Anuradha T, Bhakti, Chaitra R, Pooja D "IoT Based Low Cost System for Monitoring of Water Quality in Real Time" IRJET, Vol5, May 2018. www.irjet.net.

[3] T. Lambrou, C. Anastasiou, and C. Panayiotou, "A nephlometric turbidity system for monitoring residential drinking water quality," Sensor Applications, Experimentation, and Logistics, vol. 29, 2010.

[4] Karthick.T, Gayatri Dutt,Tarunjot Singh Kohli and Snigdha Pandey.Prediction of Water quality and smart water quality monitoring system in IOT environment. International Journal of Pure and Applied Mathematics.,Volume 118 No.20 2018,3969-3977.

[5] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 09 | Sep 2020 p-ISSN: 2395-0072 IoT based Real-Time Water Quality Monitoring System using smart Sensors AnanthaNaik G. D1, Dr. Geetha V2

[6] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 09 | Sep 2020 p-ISSN: 2395-0072 IoT based Real-Time Water Quality Monitoring System using smart Sensors AnanthaNaik G. D1, Dr. Geetha V2

[7] M. Nasirudin, U. Za'bah, and O. Sidek, "Fresh water real time monitoring system based on wireless sensor network and Gsm," in IEEE Conference on Open Systems (ICOS), 25- 28 MConference on Open Systems (ICOS), 25- 28 May 2011, Langkawi, Malaysia, 2011, pp. 354–357.

[8] Dong He and Li-Xin Zhang "The Water Quality Monitoring System Based on WSN" "On-line water parameters as indicators of distribution system contamination," Journal of the American Water Works Association, vol. 99, no. 1, pp. 66–77, 2007.

.