Water Quality Monitoring Model Using Machine Learning

Dr. A.J. Kadam[1], Alex Sunny[2], Mitali Admuthe[3], Atharva Bhosale[4], Aashay Bhujbal[5]
Department of Computer Engineering All India Shri Shivaji Memorial Society's College of Engineering

Abstract:
Water is one of the vital elements for the existence of life. Quality and accessibility of potable water are growing concerns all over the world. Water from natural sources is usually contaminated with wide variety of substances like pathogenic microorganisms, organic waste, fertilizers, sediments and petroleum that pose health concerns. Taking these factors into consideration, we have designed a system that classifies water as clean or turbid, considering its physical properties. The system makes use of IoT and Machine Learning Technology. It consists of physical and chemical sensors that detect pH, turbidity and TDS to investigate the influencing parameters. The data is collected by sensors and converted into a csv file and then submitted to the machine learning model for analysis. When any of the parameters falls below or above the standard values, the water is classified as turbid. There are some pre-existing similar systems that use sensors water quality. However, the novelty in our proposed model lies in the use of IoT devices and a ML model that will predict the best result. Also, we will be using advanced sensors that can give precise inputs to the ML model. This tracks the water quality and notifies the user about the water sample being examined.
Keywords: Machine Learning, Decision tree, Internet of Things.
DOI: 10.24297/j.cims.2023.6.7

1. Introduction

During the recent years, the monitoring and quality check of water has been a growing concern. Only 2% of the world's water resources are fresh water supplies which are getting polluted because of various human activities. Also, these freshwater resources are rapidly getting depleted due to global warming and other climate change induced phenomena, deforestation, fuel spillages, sewage and wastewater, agriculture contamination etc. Contaminated water is linked to transmission of diseases such as cholera, diarrhea, and many other diseases. According to WHO (World Health Organization), consumption of contaminated water has caused 30 percent deaths worldwide. A water quality monitoring system could be an important primary step in preventing such hazards. This study proposes a system that uses IoT and machine learning technology to analyze if the water is fit for drinking or not. The same when done in a laboratory is time- consuming and requires a lot of effort and is sometimes inefficient too. This could be improved by using computer technology developed recently. This system also aims to provide the immediate users (the common people) with a simplified and swift method to determine if the water to be used is fit for consumption.

The technologies that we are using to realize this system are:

IoT-

The Internet of Things (IoT) describes the network of physical objects that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet.[14]

ML-

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn.[14]

2. Objectives

According to the World Health Organization (WHO), at least two billion people use a drinking water source contaminated with faeces. Microbial contamination of drinking-water because of contamination with faeces poses the greatest risk to drinking-water safety. Safe and readily available water is important for public health, whether it is used for drinking, domestic use, or food production or for industrial applications. Inadequate management of urban, industrial, and agricultural wastewater means the drinking-water of people is dangerously contaminated or chemically polluted. Safe and sufficient water facilitates the practice of hygiene, which is a key measure to prevent not only water-borne diseases, but acute respiratory infections and numerous neglected tropical diseases. Our system aims to provide a handy device that can analyse the quality of water based on its physical parameters in the immediate surrounding itself, without the need for sending the water sample to a laboratory. It will utilize the power of ML technologies to classify if the water is fit for consumption. It thus saves the time and effort of the user. Also, the complete process of checking the water quality is simplified and expedited. With this we have designed a cost-effective web application-water monitoring solution.

3. Scope

In this system, we are using physical properties of water such as pH, turbidity, and TDS (Total Dissolved Solids) to determine thequality of water. This can be improved by using more precise chemical parameters like dissolved oxygen, certain ion concentration etc., however this will then require use of more sophisticated sensors and input devices. There is a scope of improvement where the precise amount andtype of impurities in the water can be shown along with the result. Different methods of improving the quality of water based on the type of impurity can be suggested to customer.

Our system can be used in fields such as:

1. Government water purification system.
2. Hospitals and restaurants.
3. Official authorities like ITC can build a simultaneous larger model in order to inspect the quality of mineral water manufacturers.
4. Mineral water manufacturers can use the system for water quality inspection.
5. The common people
6. It can be used as a preliminary water usability test at the primary level before sending the water sample for extensive investigations.

Everyone has the right to sufficient, continuous, safe, acceptable, physically accessible, and affordable water for personal and domestic use.

Our system has the potential to provide the common people access to a method/device to check the quality of water they are using in a hassle-free way. Children at risk from water-related diseases can be given access to improved sources of water. This would result in better health, and therefore better school attendance, with positive longer- term benefits in their lives.

4. Comparitive Analysis

1. Artificial Intelligence Based Mobile Application for Water Quality Detection
- Naga Siva Kumar Gunda and Sentinel (2019)

The paper ensures the development of an AI-based mobile application platform, which captures the sensor image, built using an in-built smart phone camera, identifies the presence of sensing parameter based on colour. It also demonstrates the use of developed application.

2. Deep learning based water quality estimation and anomaly detection using virtual constellation and cloud computing- Kyle T. Peterson (2020)

This paper applies novel deep learning method for estimation of blue green algae, chlorophyll-a, and dissolved organic matter in water. It uses Landsat- 8 and Sentinel-2 datasets.

3. Smart phone Camera based Water Quality Reflectance measurement and Typical Water quality parameter Inversion - Min Gao (2022)
This paper aims to extract more accurate water reflectance data from smart phone photographs with variable exposure parameters and to test usability of these data in deriving water quality parameters.

4.Use of Satellite Imagery for Water Quality Studies- F. L. Hellweger (2003)
This Paper Uses Satellite Imagery for water quality studies in the \New York harbour region. Ground data from a routine sampling program are compared to imagery from the Landsat Thematic Mapper.it correlates the turbidity of water and the red reflectance in regions affected by Hudson River Sediments based on correlation the estuarine turbidity maximum of the Hudson River is mapped. The study also correlates Chlorophyll-a concentration with the ratio of Landsat TM green to red reflectance.

5.Artificial Intelligence based monitoring system of water quality parameters for early detection of non- specific bio-contamination in water distribution systems. - Silvia Tinelli (2019)
This research aims to simulate bio- contamination risk propagation under real life. conditions in the water distribution systems. It proposes the use of pattern recognizers like Support Vector Machines and innovative sensing technology solutions such as Artificial Neural Network. The paper proposes an SVM able to identify two different classes for 'anomaly' and 'not anomaly' classification.

6. In the project ‘A Biological Sensor System Using Computer Vision for Water Quality Monitoring’, following techniques were used: The classification model based on neural network model is used to categorize the parameters of different water quality environments. It has the following issues: Explanation to the progress of applicability, accuracy and reliability of the systems approach is not given.

The results it provided are: Provided more accurate results in in the case of multi-level classification in compared to shallow neural network.

7.In the project ‘Water Quality Monitoring System using IoT and Machine Learning’, following techniques were used: Wireless sensor technology. It has the following issues: Time consuming as multiple sensor nodes is employed. The results it provided are: The type of sensors and number of sensor nodes provide high scalability, low cost, and low power consumption.

8.In the project ‘IoT Based Realtime River Water Quality Monitoring System’, following techniques were used: IoT, Machine Learning and Cloud Computing technique are used. It has the following issues: Additional quality sensors such as temperature sensor, TDS sensor can detect the chemical and physical parameters which can help in improvement of the results and

5. System Architecture

A number of sensors like pH sensor, turbidity sensor and TDS sensor are connected to an Arduino UNO microcontroller board in the system. This gathers the data from the sensors and sends them to the Arduino board It will generate a database in the form of a csv file that has the readings from sensors, and the findings are transmitted to the Machine Learning model for analysis. Based on the analysis by the ML model, the user is alerted of the quality of water being tested based on the different parameters that have been tested and the result predicted by the model.

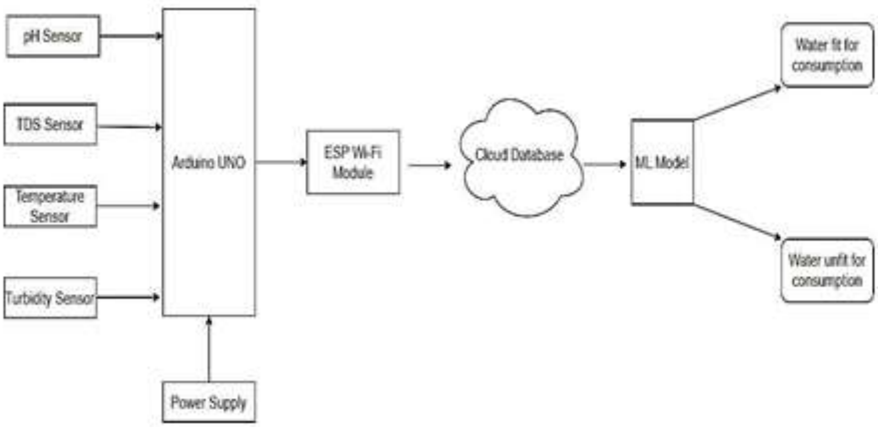


Fig 1. System Architecture

6. Hardware

The hardware includes an Arduino UNO microcontroller. It has a Wi-Fi module that enables easy connection to the user. The pH sensor, TDS sensor and turbidity sensor are all coupled with themicrocontroller in the proposed model.

The important hardware components that we propose to use are:

1) Arduino UNO: Arduino Uno is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator (CSTCE16M0V53-R0), a USB connection, a power jack, an ICSP header and a reset button.

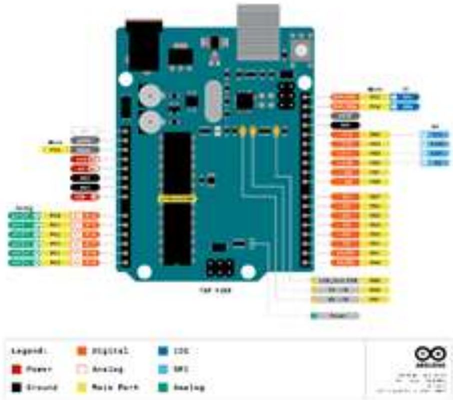


Fig 2. Arduino Board

2) Esp8266 Wi-Fi module: The ESP8266 is a low-cost Wi-Fi microchip, with a full TCP/IP stack and microcontroller capability. This is used for controlling gadgets over the Internet.

3) pH Sensor: The pH sensor measures the activity of the hydrogen ion of the liquid. It helps in determining if the water is acidic or alkaline. The scale of pH is 0-14, where pH 0-6
defined acidic and 8-14 defined alkaline. pH 7 is a neutral solution. pH sensor with Arduino or Raspberry Pi is a very cost-friendly pH sensor and that is very easy to handle. The driver board of the kit support 3.3V and 5V both.

Fig 3. pH graph

4) Turbidity Sensor: The gravity Arduino turbidity sensor is a water quality monitor that measures the level of turbidity of the water. By transmitting and scattering the light it measures

the suspended particle in water. Analog and digital signal output modes can be found by this liquid sensor.

5) TDS sensor: The DS18B20 is a 1- wire programmable TDS sensor for maxim integration. This TDS sensor supports 3.3 ~ 5.5V wide voltage input, and 0 ~ 2.3V analog voltage output, which makes it compatible with a 5V or 3.3V control system or board. Each sensor has a unique address and requires only one pin of the MCU to transfer data. The TDS Measurement Range is 0 ~ 1000ppm.

Fig 4. Temperature graph

7. Software

The data that will be continuously generated by the hardware components will be stored as a csv database. The data gathered by the sensors is given asinput to the Machine learning model (in thiscase a decision tree). It predicts the outcome, i.e., if the water is fit forconsumption or not (predicts the quality of water). Decision trees can capture nonlinear relationships between features and the target variable. They accomplish this by recursively splitting the data based on different feature thresholds, allowing for complex decision boundaries.

Decision trees can provide a measure of feature importance. By observing which features are used near the root of the tree and in higher splits, one can gain insights into the relative importance of different features for making predictions. The model is trained using extensive datasets on Kaggle that alreadyexist. These datasets are routed through the decision tree nodes to produce output.

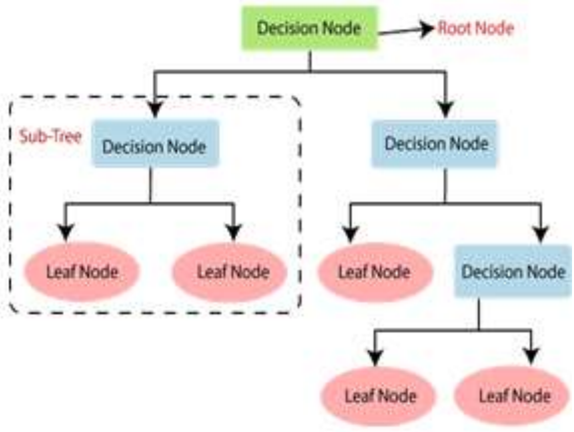


Fig 5. Decision tree architecture

8. Result

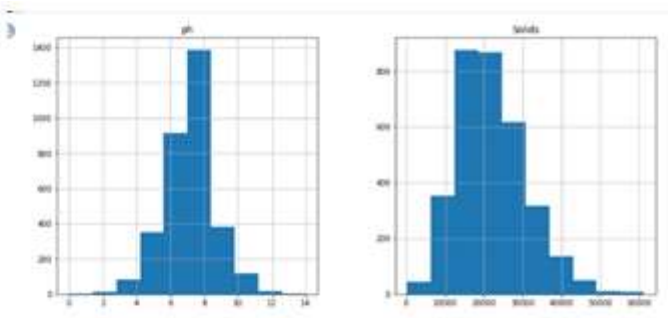
We were able to design a working model that uses the sensors to get inputs from water when it is dipped in a water sample. The readings are converted into a csv database file. This is transferred to the ML model (decision tree) which is coupled with a front end. The model predicts on the basis of the learning it has received. The sensor readings are displayed on the front-end GUI along the result predicted by the ML model- Pure water or Impure water.

	ph	Solids	Turbidity	Potability
0	NaN	20791.31898	2.963135	0
1	3.716080	18630.05786	4.500656	0
2	8.099124	19909.54173	3.055934	0
3	8.316766	22018.41744	4.628771	0
4	9.092223	17978.98634	4.075075	0

Fig 6. Dataset

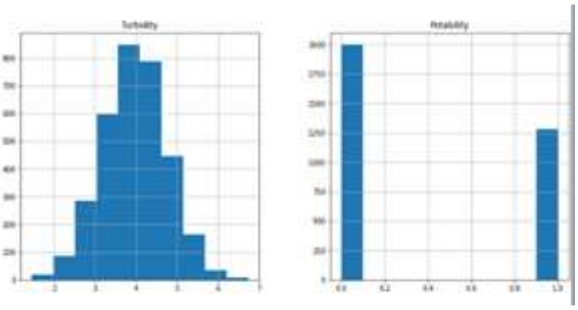
	ph	Solids	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000
mean	7.080795	22014.092526	3.966786	0.390110
std	1.594320	8768.570828	0.780382	0.487849
min	0.000000	320.942611	1.450000	0.000000
25%	6.093092	15666.690300	3.439711	0.000000
50%	7.036752	20927.833605	3.955028	0.000000
75%	8.062066	27332.762125	4.500320	1.000000
max	14.000000	61227.196010	6.739000	1.000000

Fig 7. Dataset



pH vs Count Solids vs Count

Fig 8.1. Comparative graph



Turbidity vs Count Potability vs Count

Fig 8.2. Comparative graphs

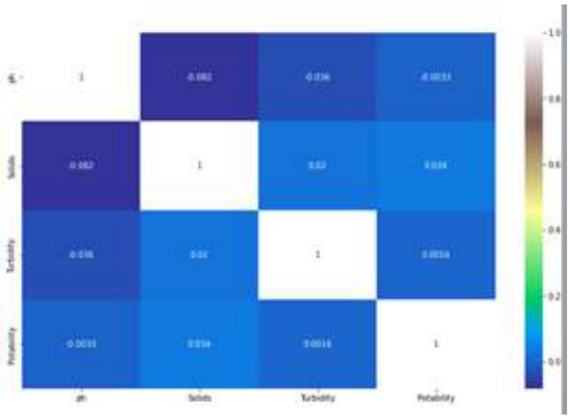


Fig 9. Co-relation Matrix

Fig 10. Model Code

Fig 11.1. Output- User Interface

Fig 11.2. Output- User Interface

9. Conclusion

This study provides an economical technique to assess water quality and eliminate the possible health hazards. IoT devices and machine learning algorithms are used to forecast the potential safety of water. To gather water parameters, the proposed system comprises multiple sensors such as TDS sensors integrated on an Arduino compatible kit.[14] The implemented solution notifies the user about itscontamination. Based on the analysis ofabove cited papers, we believe that our proposed system will overcome their limitations due to the following reasons:

1) We are using a decision tree model to analyze the data generated by the water sensors, which will lead to more accurate predictions based on the analysis of input data.

2) We will be using advanced sensors that detect both physical properties (pH, turbidity) and chemical property (TDS) of water. This will help us to better analyze the quality of water.

3) This proposed model safeguards standard quality detection of water using low price embedded devices like Arduino Uno and ESP Wi-Fimodule

References

1. Ping Liu Jin Wang Arun Kumar Sangaiah Yang Xie Xinchun Yin, 2019. " Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment," Sustainability, MDPI, Open Access Journal, vol. 11(7), pages 1- 14, April.
2. K. Gopavanitha and S. Nagaraju —A low cost system for real time water quality monitoring and controlling using IoT||, IEEE Conference paper, August 2017.
3. Manish Kumar Jha ; Rajni Kumari Sah; M. S. Rashmitha ; Rupam Sinha ; B. Sujatha ; K. V. Suma — Smart Water Monitoring System for Real-Time Water Quality and Usage Monitoring||,IEEE Conference paper,july 2018
4. Asma Al Khaili, Aisha Al Mamari,Hoda Amer and Walid Ibrahim—An Affordable system for water quality detection
5. Vaishnavi V. Daigavane, Dr. M.A Gaikwad. ||Water Quality Monitoring System Based on IOT|| 2017 Advances in Wireless and Mobile Com- munications, Nov 2017 ISSN 0973-6972
6. Neural Network Basics using Matlab,http://staff.ttu.ee/ jmajak/Neuralnetworksbasics.pdf
7. Thair S.K. , Abdul Hameed M. J. , and Ayad S. M., Prediction of water quality of Euphrates river by using artificial neural network model ,International Research Journal of Natural Sciences Vol.2,No.3,pp.25-38, September 2014
8. Nikhil Kumar Koditala ; Purnendu Shekar Pandey,|| Water Quality Monitoring System Using IoT and Machine Learning||, IEEE paper, August 2018
9. Najah,A. El-Shafie,O. A. Karim,Amr H. El-Shafie —Application of artificial neural networks for water quality prediction||, May 2013, Volume 22, pp 187– 201.
10. Resources||, Resources — Internet of Things World Forum, www.iotwf.com/resources/72
11. S.Geetha ,S.Gouthami, "Internet of things enabled real time water quality monitoring system." , DOI 10.1186/s40713-017-0005-y (2017).
12. Jianhua D, Guoyin W, Huyong Y, Ji X, Xuerui Z "A survey of smart water quality monitoring system." , EnvironSciPollut Res 22(7):4894906 (2015).
13. JAYTI BHATT, JIGNESH PATOLIYA "IOT BASED WATER QUALITY MONITORING SYSTEM." ,ISSN:2347-6982(2016).
14. Manya Kakkar∗, Vansh Gupta† , Jai Garg and Dr. Surender Dhiman § Electronics and Communication Engineering Department, Dr. Akhilesh Das Gupta Institute of Technology & Management∗ , New Delhi, India ISSN: 2278-0181
15. S. Angel Vergina, Easwari Engineering College, Chennai, Tamil Nadu Dr.S.Kayalvizhi, Easwari Engineering College, Chennai, Tamil Nadu Dr. R.M. Bhavadharini, Easwari

Engineering College, Chennai, Tamil Nadu Kalpana Devi. S, Easwari Engineering College, Chennai, Tamil Nadu ISSN 2515-8260

16. Fei yuan , yifan huang , xin chen , and en cheng (2018), “A Biological Sensor System Using Computer Vision for Water Quality Monitoring” ,IEEE Access,vol.37,pp.151- 161.
17. Nikhil Kumar Koditala, Dr.Purnendu Shekar Pandey(2018), “Water Quality Monitoring System using IoT and Machine Learning” ,IEEE Transactions,vol.64,pp.1172-1173.
18. Segun O. Olatinwo and Trudi-H. Joubert (2019), “Energy Ef¿cient Solutions in Wireless Sensor Systems for Water Quality Monitoring: A Review” , IEEE sensors journal, vol.19, pp.259-271.
19. Segun O. Olatinwo and Trudi-H. Joubert (2019), “Enabling Communication Networks for Water Quality Monitoring Applications: A Survey” , IEEE Access, vol.34, pp.2562-2571.
20. Soundarya Pappu, Prathyusha Vudatha and Niharika.A.V, Soundarya Pappu, Prathyusha Vudatha and Niharika.A.V (2018) “Intelligent IoT Based Water Quality Monitoring System” , IEEE Journal, vol.12, pp.5447-5454.
21. Swapan Shakhari, Indrajit Banerjee (2019), “A multi-class classification system for continuous water quality monitoring” , Elsevier,vol.36,pp.1172-1181.
22. Yafra Khan, Chai Soo See (2017), “Predicting and Analyzing Water Quality using Machine Learning: A Comprehensive Model” Conference, vol.20, pp. 944 - 952. 10. Youchao Wang, S. M. Shariar Morshed Rajib, Chris Collins, and Bruce Grieve (2018), “Low-Cost Turbidity Sensor for Low-Power Wireless Monitoring of FreshWater Courses” , IEEESensor Journal, vol. 37, pp. 151-161. 11. S. K. S. Raja, R. Rishi, E. Sundaresan and V. Srijit, "Demand based crop recommender system for farmers," 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, 2017, pp. 194-199, doi: 10.1109/TIAR.2017.8273714.
23. Gunda N. S. K. Gautam S. Mitra S. K., in Meeting Abstracts, 56, Abstract MA2018-02 1997. The Electrochemical Society (2018).
24. Kyle T. Peterson, Vasit Sagan & John J. Sloan (2020) Deep learning based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing, GIScience & Remote Sensing, 57:4, 510-525, DOI:10.1080/15481603.2020.1738061