IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

Ground Water Quality Analysis Using Machine Learning Techniques

Dr Vasavi Ravuri[1], Jakkula Teja Venkata Raju[2], Kommu Sai Rohith[3],

Sheelam Chandrashekhar Reddy[4], Rajavardhan Rao[5]

Assistant Professor, Department of Computer Science and Engineering, VNRVJIET, Hyderabad, Telangana[1]

Computer Science and Business Systems, VNRVJIET, Hyderabad, Telangana[2-5]

Abstract: The study explores the application of IOT sensors, including pH, turbidity, conductivity, temperature, and humidity, for sampling water from diverse sources. By leveraging these sensors, the research aims to predict water portability using the random forest algorithm. This approach involves training the model with existing datasets and subsequently testing it on samples collected via IOT sensors. The abstract suggests that such an approach could provide insights into efficient and accurate methods for assessing water quality in both confined and open water systems. Additionally, comparative analysis with other machine learning algorithms may further elucidate the optimal method for determining water portability.

Keywords: Water Quality, pH, turbidity, conductivity, Random Forest algorithm, PyCharm IDE, sensors.

INTRODUCTION

In recent years, the integration of Internet of Things (IOT) technology has revolutionized various sectors, including environmental monitoring and water quality assessment. With advancements in sensor capabilities, it has become feasible to collect real-time data on physiochemical properties of water from diverse sources efficiently. This project aims to leverage IOT sensors, such as pH, turbidity, conductivity, temperature, and humidity sensors,[3] to collect comprehensive datasets for water quality analysis[9].

The primary objective of this research is to develop a predictive model for assessing water potability using machine learning algorithms, specifically focusing on the random forest algorithm. By training the model with existing datasets and testing it on samples collected through IOT sensors, the study seeks to provide insights into the effectiveness and accuracy of this approach in determining water quality.

The significance of this project lies in its potential to offer efficient and reliable methods for assessing water quality in both confined and open water systems. By combining advanced sensor technology with machine learning algorithms, researchers aim to address the growing concerns surrounding water contamination and its implications for human health and environmental sustainability.

This introduction sets the stage for exploring the methodology, results, and implications of using IOT sensors and machine learning algorithms for water quality assessment, ultimately contributing to the advancement of environmental monitoring practices and public health initiatives.

RELATED WORK

Previous research has investigated the development and implementation of IOT based systems for monitoring water quality parameters such as pH, conductivity, turbidity, temperature, and humidity. Studies by authors such as S.A.Nadhir and S.Y.Nyarko (2020) have demonstrated the effectiveness of IOT sensors in continuously monitoring water quality in various environments [2].

Several studies have utilized machine learning algorithms for predicting water quality parameters and assessing water potability. For instance, the work of Osim Kumar Pal (2022) applied support vector machines (SVM), Random Forest algorithm for water quality prediction based on sensor data. Similarly, the research by Hossam A Zaqoot and Adnan M Aish (2017) employed neural networks for real-time water quality monitoring and prediction. The random forest algorithm has been used in many applications due to its ability to handle large data sets. Previous studies such as the work of M.Anbchezhian, Dr.R.venkataraman, and V.Kumuthavalli. (2018) have applied random forest models for predicting water quality parameters and assessing the health of aquatic ecosystems [8].

Comparative studies that evaluate the performance of different machine learning algorithms for water quality prediction [4] can provide valuable insights. Research by Anthony Ewusi, Issac Ahenkorah, Derrick Aikins (2021) compared the

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

performance of Support Vector machine (SVM), and other algorithms for predicting water quality parameters, highlighting the strengths and weaknesses of each approach.

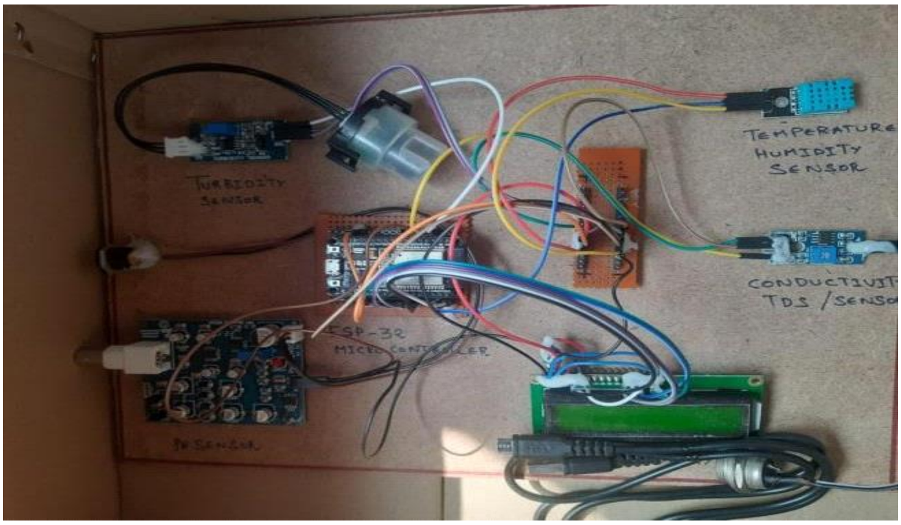
Explore studies focusing on advancements in sensor technologies relevant to water quality monitoring. Research by Monira Mukta, Samia Islam, Ahmed Wasif Reza, Md. Emon Miea, Surajit das Barman, M Saddam Hossain Khan (2019) could be referenced for their investigation into novel sensor designs or materials that enhance the accuracy and reliability of measurements for parameters like pH, turbidity, conductivity, temperature, and humidity[9].

Investigate real-world applications and case studies where IOT sensors and machine learning have been successfully applied in water quality monitoring. Case studies from different regions or industries, such as agriculture, urban water management, or industrial wastewater treatment, can provide practical examples of how these technologies are being implemented and their impact on decision-making processes.

PROPOSED SYSTEM

The proposed system aims to present the design and methodology is to assess the portability of water. The methodology and design involve the following steps:

• *IOT Sensor Network* - The proposed system involves the implementation of IOT (Internet of Things) sensors strategically positioned in various water sources to continuously monitor key water quality parameters such as pH, turbidity, temperature, humidity and conductivity. These sensors are designed to gather real-time data from water samples, this data is passed to a system on same network as the connected with micro controller (ESP-32) which is shown in the diagram.



Dia1. Sensors connected with microcontroller(ESP-32)

• *Data Collection*- The data set of water quality parameters is collected from Kaggle *(Kaggle.com)*. The overall size of data set is about 525KB with 3277 records. The data set consists of 10 parameters.

```
1. ph: pH of 1. water (0 to 14).
2. Hardness: Capacity of water to precipitate soap in mg/L.
3. Solids: Total dissolved solids in ppm.
4. Chloramines: Amount of Chloramines in ppm.
5. Sulfate: Amount of Sulfates dissolved in mg/L.
6. Conductivity: Electrical conductivity of water in μS/cm.
7. Organic_carbon: Amount of organic carbon in ppm.
8. Trihalomethanes: Amount of Trihalomethanes in μg/L.
9. Turbidity: Measure of light emiting property of water in NTU.
10. Potability: Indicates if water is safe for human consumption.
Potable -1 and Not potable -0
```

The data set of water quality parameters is collected from Kaggle *(Kaggle.com)*. The overall size of data set is about 525KB with 3277 records. The data set consists of 10 parameters.

• *Pre-processing*- The data set consists a few null values (NAN), these NAN values can be found using isna(). These values will be replaced by using few techniques such as data parameter. fillna() to fill the NAN values with median values .

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

Random Forest Algorithm:

The Random forest algorithm is utilized to predict the portability of water samples based on their parameters. The Random Forest algorithm is made by constructing multiple decision trees based on our specification , where each tree is built using a random subset of the training data which is divided into multiple parts and features. This process is called bagging, each decision tree is trained on a sample of the original data, ensuring diversity . During prediction, each tree independently makes a prediction, and the final output is determined by a voting mechanism for classification tasks . Random Forest is highly effective . Additionally, it provides insights into feature importance, aiding in feature selection and model interpretation.

The system is implemented using the PyCharm IDE, which provides a convenient development environment. The user interface is designed to allow farmers to input their resource constraints, such as pH, conductivity, turbidity and few other parameters.

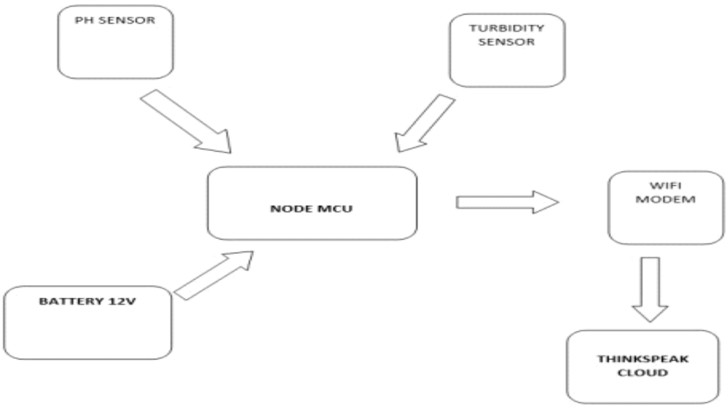
The Random Forest algorithm is executed on the backend to provide crop recommendations based on the input constraints.

Results and Evaluation

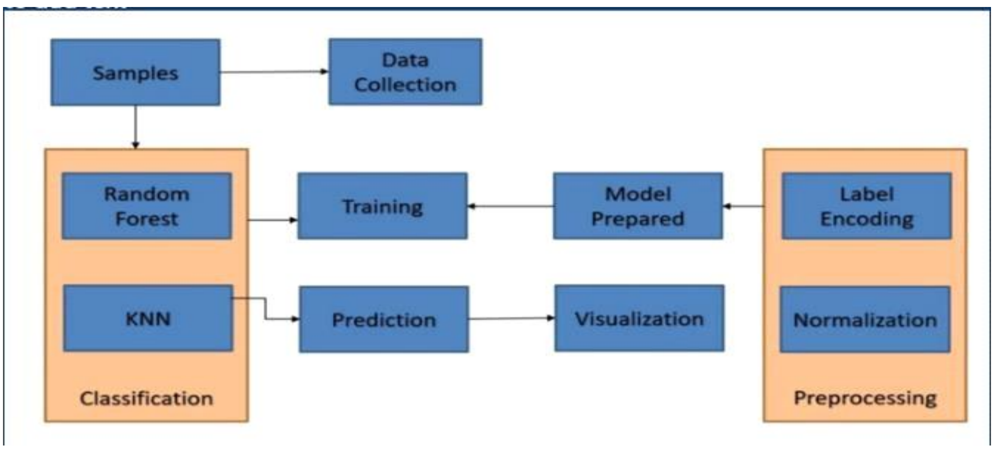
The proposed system is evaluated using a existing data set of water parameters and portability. The evaluation metrics include accuracy, precision, recall, and F1-score, which assess the performance of the water quality detection system.

Comparative analysis with other existing approaches can be conducted to demonstrate the effectiveness of the Random Forest algorithm.

System Architecture



Dia 2. System Architecture



System Architecture (Dia1) is a conceptual model that outlines the organization, perspectives, and actions of a system. An official description and illustration of a system that is structured to make it easier to reason about its structures and behaviors is referred to as a description of the architecture[3]. A system architecture is made up of built sub-systems and system components that work together to implement the whole system[3]

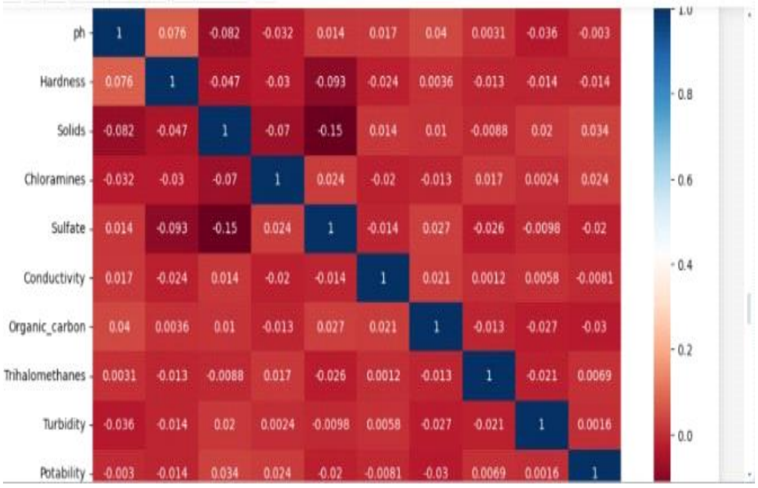
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

RESULTS



Here are some images containing results of our proposed model.As there are no sensors for the other parameters ,the average values from the data set are considered for the other parameters as input. Correlation analysis is made for the different attributes present in the data set , a correlation matrix is generated to find the correlation among the attributes and portability.

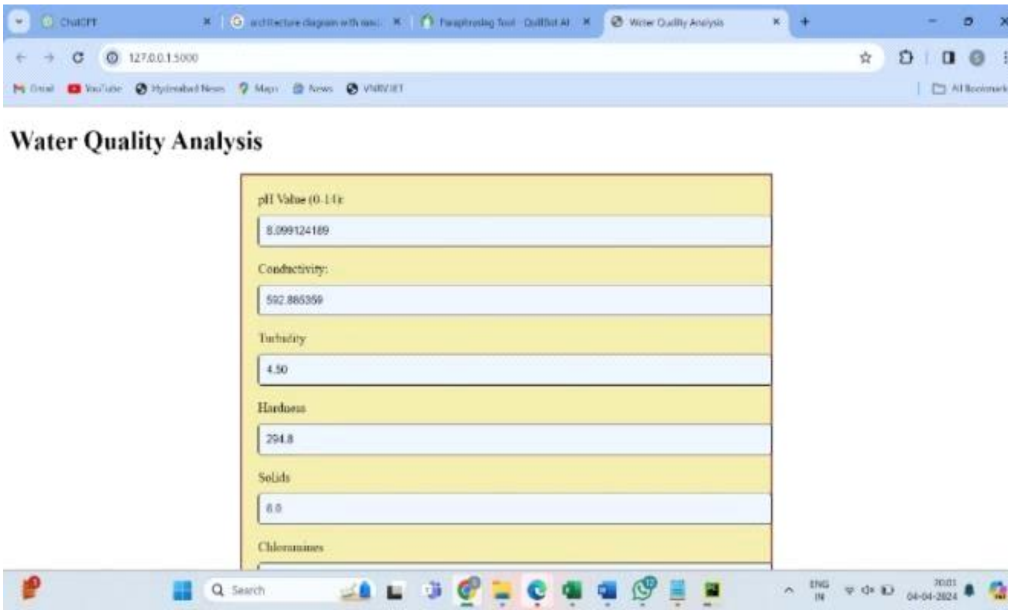


fig 7.1 display to enter Readings from the Sensors

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 9, September 2024
DOI: 10.17148/IJARCCE.2024.13917

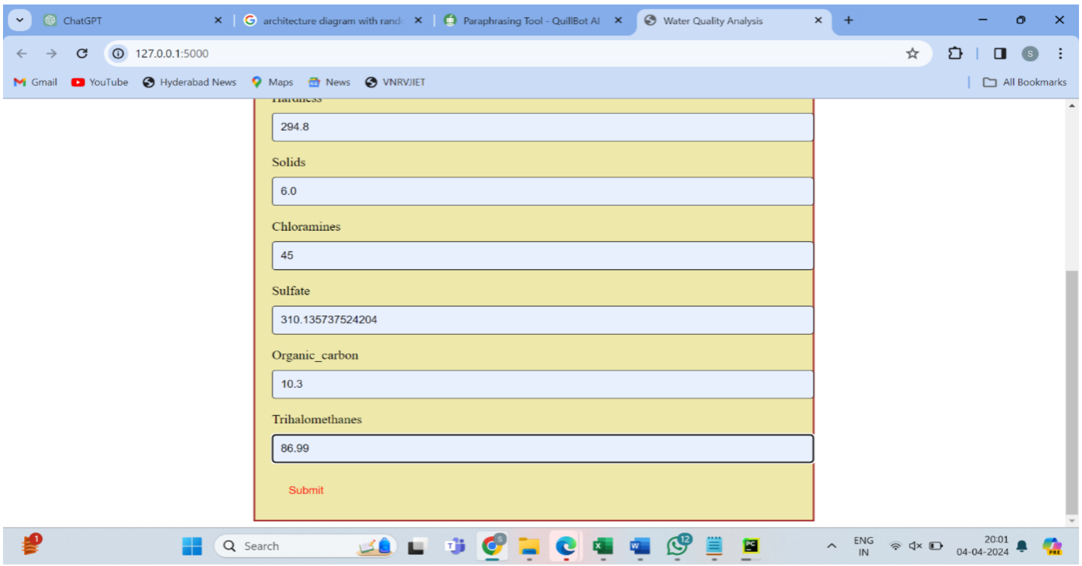


Fig 7.2 display to enter Readings from the Sensors

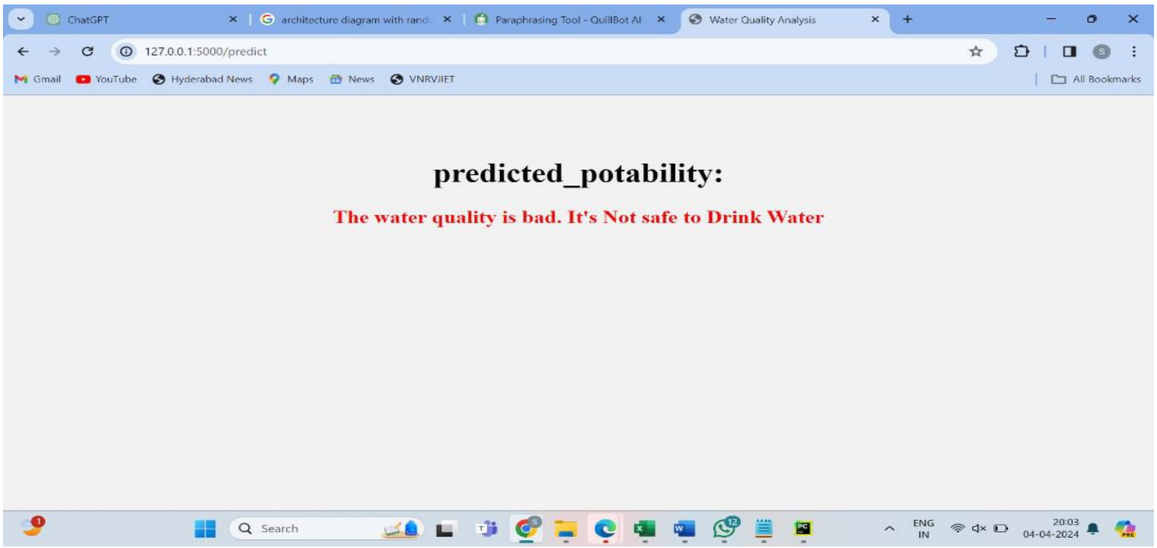


Fig 7.3 display the potability of the water.

CONCLUSION

The paper presents a water quality detection system based on the Random Forest algorithm, considering parameters such as pH, conductivity, turbidity and a few other chemical properties. The system is deployed through the PyCharm IDE, providing a user-friendly interface for inputting their parameters and predict portability. The evaluation results showcase the accuracy and effectiveness of the proposed system.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

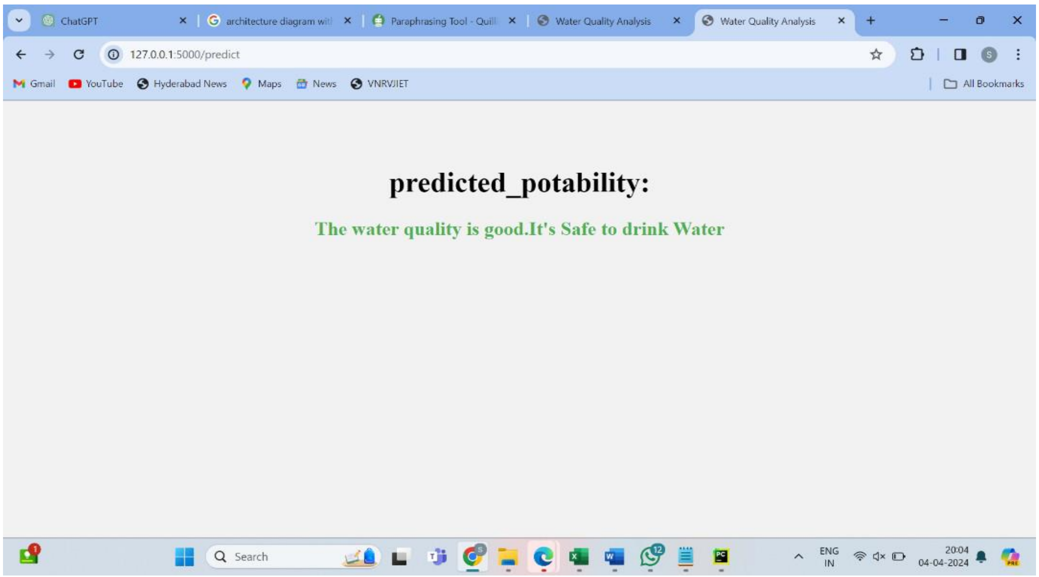
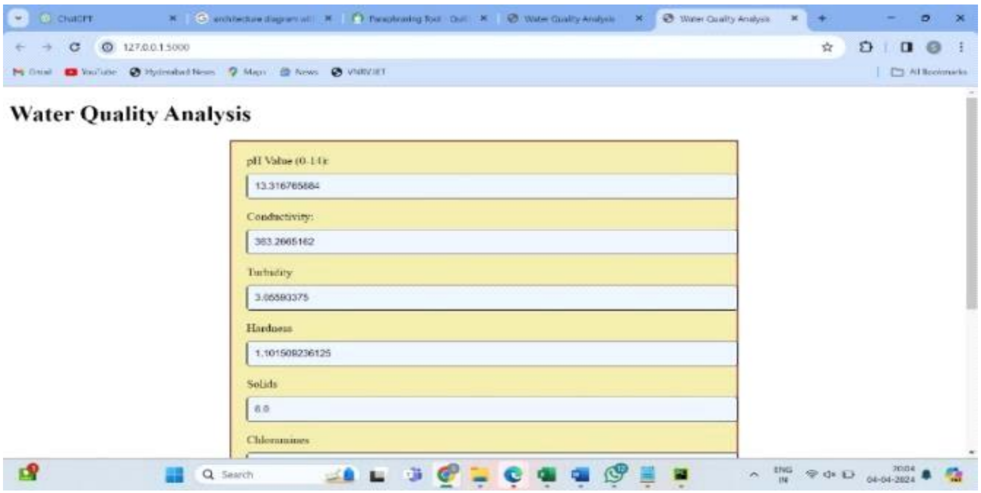


Fig 7.4 page showing the potability of the water.

FUTURE SCOPE



Future research can focus on continuous water monitoring system where sensors can be deployed in a water source itself and continuously collect data and check its portability continuously.

REFERENCES

[1]Osim Kumar Pal: The Quality of Drinkable Water using Machine Learning Techniques(ISSN: 2349-6495) (2022)

[2] S.ANadhir, S.Y.Nyarko, I.Quaicoe, : Heavy Metal Loading in Sutface Sediments along the kawere stream , tarwaka, ghana

[3]] R. Ranjitha, P. Ahuja, R. Shreeshayana and D. Anil, "Edge Intelligence for Traffic Flow Detection: A Deep Learning Approach," *2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, KOTTAYAM, India, 2023, pp. 1-6, doi: 10.1109/iQ-CCHESS56596.2023.10391493

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13917

[4] Essamlali I, Nhaila H, El Khaili M. Advances in machine learning and IoT for water quality monitoring: A comprehensive review. Heliyon. 2024 Mar 13;10(6):e27920. doi: 10.1016/j.heliyon.2024.e27920.

[5] VeeraSekharReddy, B., Sarath, S., Philip, J., Reddy, U.S., Naresh, L., & Tejaswini, K.N. (2023). Water Quality Monitoring System using IoT and Cloud. *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 1036-1042.

[6] Hossam A Zaqoot, Adnan M Aish: Application of artificial neural networks for predicting water quality: Gaza (February 2017)

[7] Sathish Pasika, sai teja gandla: Smart water quality monitoring system with cost effective using IOT(CBIT)

[8]M.Anbchezhian, Dr.R.venkataraman, V.Kumuthavalli: Water quality analysis and prediction using machine learning Algorithm (2018)

[9] Monira Mukta, Samia Islam, Ahmed Wasif Reza, Md. Emon Miea, Surajit das Barman, M Saddam Hossain Khan: IOT based Smart Water Quality Monitoring System(research gate) (2019).

[10] Wong Jun Hong Norazanita , emeroylariffion abas, Rosyzie Anna Apong, Zarifi Masri, Hazwani Suhaimi, Stefan Herwig Goeke and Muhammad Nafi aqmal Noh.: water quality monitoring with auduino based sensors(2021)

[11] Santosh Konde , Dr.S.B.deosarkar.: IOT based water quality monitoring systerJ. (ICCIP-2020)