International Journal of Advanced Trends in Computer Science and Engineering

Cloud-Based Smart Water Quality Monitoring System using IoT Sensors and Machine Learning

Bineet Kumar Jha[1], Sivasankari G.G[2], Venugopal K.R.[3]
[1]CMR Institute of Technology, Bengaluru, India, yoursbineetjha@gmail.com
[2]AMC Engineering College, Bengaluru,India, sivasankarigg@gmail.com
[3]Bangalore University,Bengaluru, India, venugopalkr@gmail.com

ABSTRACT

Low water quality is a major concern in urban as well as rural areas. Consumption of contaminated water leads to several health hazards. Early water quality detection can prevent most of such health-related issues. Parameters such as conductivity, pH, nitrate, biochemical oxygen demand, fecal coliform are significant parameters in deciding the quality of water. These parameters which are collected from groundwater samples at different places are highly correlated to each other. Therefore, machine learning algorithms are used for classification. The data collected from sensors are further analyzed using a cloud-based environment Ubidots to support distributed computing. The cloud environment is connected to display units and mobile devices. To predict the quality of water it is necessary to check the values associated with the quality attributes and for that reason, a decision tree classification model is used. The dataset is broken into subsets that have decision nodes and leaf nodes to decide classifications. The IoT based sensors are deployed in the water tank to measure the quality parameters which are further sent to the cloud. The proposed framework predicts the water quality and assesses the performance of the decision tree classifier. Decision Tree is used to infer decision rules based on various parameters read through sensors.

Key words: Artificial Neural Networks, Cloud, Decision Tree, IoT, Linear Regression, Machine Learning

1. INTRODUCTION

Water is essential for everyone and without it, we cannot imagine life on the earth. The fresh groundwater is used for drinking, production of food and other purposes. As per the World Health Organization (WHO) [10], safe water can keep 1.4 million kids dying from diarrhea, right around 500,000 dying from Malaria, 860,000 youngsters dying from malnutrition. Improving water management, sanitation and hygiene can significantly reduce the health-related issues in human beings and other living bodies. Water contamination and groundwater depletion have been the major problem for urban civilization. This problem is man-made and it is the consequence of overutilization of natural resources and rapid urbanization and industrialization. The use of fertilizers, industry discharges, sewage leakage, non-uniform rainfall and human practices led to water pollution at alarming levels. According to the Central Groundwater Board (CGWB) the depth of soils and subsurface geological formation determines the quality of groundwater. The CGWB checks the chemical content of groundwater every year through the network of wells and underground water in various parts of the country. The water salinity problem in different states of the country is found where water level contains high concentrations of fluoride, iron, arsenic and other heavy metal. This led to the severe health issues in people consuming it such as pain in muscle or joints, mental disorder, vision problems, chronic fatigue and gastrointestinal. Sometimes the water is also mixed with foreign objects such as dead snake, bird or animal which also contaminates water in community tanks in villages.

According to the Water Resource Information System of India-2017, around 5 million people die because of drinking contaminated water worldwide. This problem is more in underdeveloped and developing nations. The quality monitoring system depends on the collection of data collected from different water sources. The monitoring system offers the data analyzing interface for the data collected in real-time and provides solutions or suggestions to resolve the issues.

Our proposed work is to be carried out in two phases wherein the first phase we are going to conduct a survey on the recent water monitoring system and in the second phase for the development of the cloud-based water quality monitoring framework which checks the water nature of the groundwater which is put in overhead or community tanks. The parameters used in this research work to measure water quality- turbidity, TDS, conductivity, B.O.D, nitrate, fecal coliform and pH value. To check the water level and pressure in tanks, a water level sensor is used. The flow sensor is used to detect the leakage in the pipe. Early water quality detection can save people from the adverse effects of drinking contaminated water. People will get an alert message if the water is not safe to drink.

In IoT, the animals, humans or objects can be connected to other physical devices using sensors and actuators explained

by Pravin et al.[1]. They connect to digital machines and transfer data to each other over the network. The advancement of IoT leads to better interactions between humans and machines described by S. Geetha et al. [2]. The IoT has various application domains and healthcare is one among them. The IoT based health care devices are used by individuals and health-care professionals more often in recent times. A comprehensive survey is done by DaehanKwak et al. [5] on the IoT based technologies and architectures used in health care sectors. They analyze the security and privacy issues for IoT based health care devices. An IoT based weather monitoring system generates alerts based on moisture, carbon monoxide level, wind speed, light intensity, humidity using various sensors. A NodeMCU microcontroller and Ubidots cloud-based system are proposed by YashaswiRahut et al. [8]. The polluted water is harmful for underwater living organisms. To detect the pollutants an IoT based framework is utilized to screen water quality, for example, ammonia, pH, membrane, and oxygen level is explained by A.Ramya et al.[9].

Machine Learning (ML) based techniques are also used to diagnose various ailments in the human body by analyzing medical images. A similar study is conducted by Munya A. Arasi*et al.*[23] in that various ML techniques are studied such as Naïve Bayes (NB), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Decision Tree (DT) for analyzing the medical image. ML algorithms are widely used for training and prediction.

The growth of IoT made it easy to collect real-time data in many applications, Agriculture is one among them. As we know that Indian Agriculture relies upon many factors, for example, the chance of rainfall, soil moisture level, plant disease, etc. Many approaches are developed to assist the farmer with prediction using machine learning algorithms. The various IoT frameworks are studied for agriculture in the research work done by Sujeet More et al. [22].

The follow-on sections contain a literature review on existing approaches used for water quality predictions, proposed system, results and conclusion.

2. RELATED WORKS

Water quality has been a major issue in urban civilization and a lot of initiations have been made by the local and government agencies. Awareness among people has also increased in recent time and they have put a lot of effort into water cleanliness. The advancement of technologies also plays a major role to monitor water quality remotely on a large scale. Nikhil *et al.* [11] implemented the Azure cloud platform based water quality monitoring system which uses the NodeMCU microcontroller to collect data from sensors in Jason format for analysis. The multivariable linear regression technique is used to classify the data.

Satyam *et al.* [12] have proposed the smartphone-based water quality analysis where a water impurity warning sent through SMS, Email and WhatsApp to the consumer. The

handheld embedded system proposed in this research work where the device is connected to different mobile applications and interfaces through a classical Bluetooth model. They have also analyzed the various quality measurement techniques to determine the parameter to ensure water quality. An Artificial Neural Network (ANN) is utilized as a model in the proposed work. The proposed techniques include- Bayesian regulation, Levenberg Marquardt, and conjugate gradient. The power consumption in Bluetooth based localization of device using is one of the challenges in IoT. A Bluetooth low energy technology is proposed by MarcoTerán*et al.*[6] that uses Received Signal Strength (RSS) based technique to determine the proximity and localization of an IoT device.

Mobile devices usually have many in-built sensors that commonly used for navigation and localization purposes. The research work carried by Jaime Alonso et al. [7] in that a cloud-based platform Ubidots is utilized to store the received information from the mobile sensors. Initially, the data are stored in a local database associated with a mobile device which is later sent to the cloud. An android based OS is used for this purpose which has a set of APIs which transmit the data using HTTPS to the cloud. The performance is the IoT based communication gradually decreases with the movement from one layer to another in the IoT network to reach the cloud. Dr. J. SasiBhanu et al. [4] have proposed high-performance computing in that the client requests the services from RESTful servers and being offered fast after meeting Service Level Agreement (SLA). A parallel processing approach was suggested for fast processing. The cluster of processors is used to provide multiple services together in the proposed work 8 processors are used for this purpose.

In another research work done by A.N. Prasad *et al.* [3] where the water is collected from various sources and the tradeoff between water quality parameters is evaluated. Mo'tamad Bata *et al.* [13]have proposed a hybrid model dependent on supervised and unsupervised learning for water demand forecasting. The performance of Regression Tree (RT), Self Organizing Map (SOM), Clustering and organizing unlabelled data is evaluated for water demand forecasting. Machine learning techniques are effective in predicting water quality.

In a research work carried out by Amir *et al.* [14] artificial intelligence, support vector mechanism and a group data handling method are addressed. Several parameters are investigated such as SO_4, pH, Mg, Ca, HCO3, Cl, etc. for the water quality concerning impurity and mineral contents. The SVM and ANN were used to predict the Ca whereas to predict pH and SO_4 the SVM technique is used.

A novel decision tree is proposed by Shiueng-BienYang et al. [15] to create an uncertain decision tree (UDT). In this approach, an Uncertain Genetic Clustering Algorithm (UGCA) is utilized to arrange uncertain information. The algorithm automatically finds the number of branches for every node. The accuracy level of the algorithm is high for

uncertain data. A short term water quality is predicted in the research work proposed by Hongfang et al.[16] where novel extreme gradient boosting and random forest-based denoising procedure- CEEMDAN is proposed for prediction. Water can be also contaminated due to microplastics in the water used for drinking and it also dangerous for marine life. A study on these microplastics is done by Albert A et al.[17] where the quantitative examination of measured quality is done on the structure of microplastics to evaluate the impact on living organisms. Another machine learning approach is described by C. Ashwini et al. [19] where NodeMCU is used to evaluate the water quality parameters- pH, Turbidity, temperature, dissolved oxygen, conductivity, dissolved carbon and watercolor is examined. The neural network-based approach is used for this purpose.

Yafra Khan et al.[20] have used a machine learning-based approach to evaluate the Root Mean Square Error(RMSE), Mean Square Error(MSE) to evaluate the performance of the artificial neural networks in predicting water quality. A wavelet denoising technique is proposed by Ali Najah Ahmed et al. [21] to evaluate the water quality parameters that use the Neuro-Fuzzy Inference System as a model for prediction and evaluation. The various water quality parameters such as turbidity, minerals, temperature, etc. are examined by G. Kanagaraj et al. [24] using a microcontroller connected to the sensors. They have used a webpage to monitor the information received from the sensors.

Many technologies are being used to transmit sensor data for analysis purposes. The Wireless Sensor Network (WSN) is one among them which is used for many applications such as habitat monitoring, sensing gas leakage and localization. It is also used to measure water quality parameters, for example, pH, temperature, pressure and water level using different sensors. A Field Programmable Gate Array (FPGA) based smart water monitoring system proposed by N.Balaji et al. [25]. The design works use Xilinx ISE and Modelsim Simulator. The proposed system uses Xilinx Spartan3 kit for hardware implementation. The Verilog Hardware Descriptive Language is used to interact with the FPGA.

The Water Quality Index (WQI) is derived from water quality parameters described by Andreea-Mihaela et al. [18] which helps to identify pollutants in the river and improve their quality.

3. PROPOSED SYSTEM

Our proposed system works on Cloud-based Ubidots platform for data integration and integration at the dashboard.
Ubidots is a Platform as a Service (PAAS) model in the cloud platform, that provides the web service to the connected sensors and devices. The cloud offers web services to electronic equipment to communicate with each other on the World Wide Web. The communication happens between human to human or machine to machine.

2.1 A Smart Water Tank Monitored by a Controller

In this proposed system a microcontroller is used to maintain using sensors. These sensors are deployed inside the tank to read parameters associated with the quality and the level of water inside the water tank. The pH sensor measures the quality of water in terms of alkalinity or acidity. The pH value ranges from 0-14 depending upon the quality of water. For pure water, the pH water is considered as 7. The pH value is calculated as –

$$pH = -\log 10\ [H^+]$$

The $[H^+]$ represents the concentration of hydrogen ions in mol L -1 (mol/L or M)
The pH value alone cannot guarantee the good quality of water therefore we have considered other parameters such as temperature, turbidity, etc.

The turbidity indicates the concentration of mud inside the water. If the turbidity is more than 5 NTU it is considered to be unsafe, for water which has a lot of mud content this range goes upto 100 NTU whereas the groundwater where it goes through the multiple filtration layers the turbidity normally found as less than 5.The temperature is another factor that decides the quality of water such as odors, viscosity, solubility and palatability. The temperature range from $10\text{-}15^o\,C$ is considered to be safe for drinking.

The Electrical Conductivity (EC) defines the ability of water for electric conductivity. The conductivity of water increases with an increase in H^+ ions. The pure water is considered to have low electrical conductivity. To measure the electrical conductivity two units are considered US Unit is micromahos/cm whereas the SI unit is milliSiemens/m or deciSiemens/m (dS/m). Table 1 shows the recommended EC values for various water sources.

Table 1: Recommended EC values

Water Type	EC Value
Pure Water	0.005–0.05 S/m
Ultra- Pure water	5.5×10^{-6} S/m
Seawater	5 S/m

The display unit such as LCD monitor is attached with the microcontroller to display the received values from sensors. The microcontroller used in the proposed system is T1CC3200.

3.2 Proposed Architecture

Figure 1 shows the proposed framework design that has various sensors kept inside the water tank. These sensors are connected to the microcontroller using Zigbee/Wi-Fi technology. The microcontroller is directly connected to the LCD to read the result. IoT using inbuilt Wi-Fi can be used to connect microcontrollers to the cloud or mobile devices. In our proposed technique Ubidots is used as a cloud platform to read sensor data using mobile and remote devices. The following architecture diagram shows the overall working of

the proposed architecture. The data stored in the Cloud is further analyzed using a decision tree. The sensor data is being collected in a CSV file which is being analyzed further.
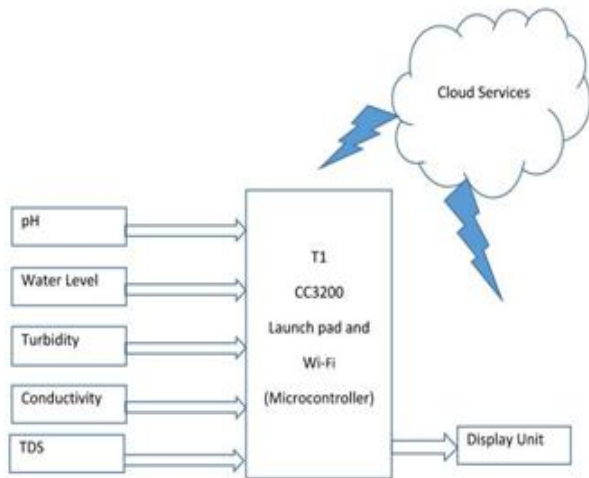

Figure 1: The proposed System Architecture

The CC3200 Launcher is connected to the desired sensors. The data is gathered from sensors and kept in the cloud which is further processed by a Machine Learning algorithm or Ubidots analytics tool. The Ubidots is connected to the Decision Tree-based analytics tool which is running in the background to predict sensed data.
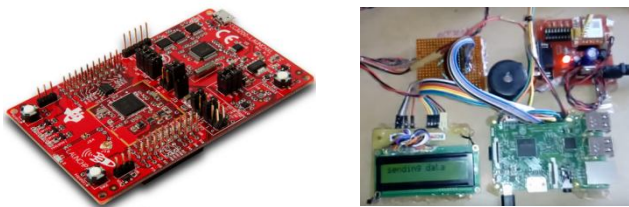

Figure 2: Experimental Arrangement with CC200 Microcontroller

Figure 2 shows the hardware setup is where the terminals are connected to sensors to measure water quality and Ubidots to represent data and analyze them. The measured quality and alert are also sent to mobile using integrated GSM and Bluetooth modules.

3.3 Sensor Data Reading and Updating

The data received from sensors are updated frequently in the Ubidots cloud. The cloud has various interfaces that provide easy access to received data and analyze them. The cloud is also accessible by mobile users. Figure 3 shows the data flow in the system. The data which are collected from sensors are then compared with the standard threshold which is recommended by WHO, if it exceeds the standard value the alert message will be sent to the users.
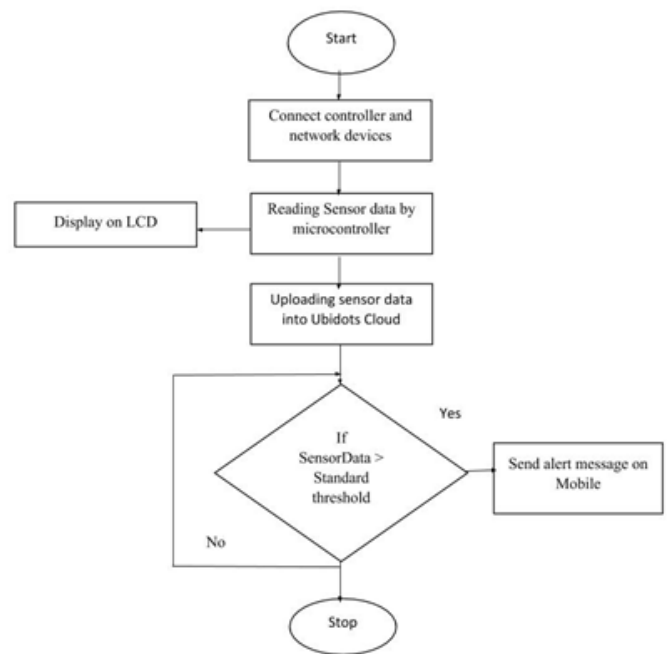

Figure 3: Reading and Storing Sensor Data

4. RESULTS

In this section, we have described the design of the graphical user interface and analyzed the result and performance of the proposed work is actualized in Python 3.7. Information is gathered from sensors with the assistance of a controller. The application program uses collected data to classify them into training and test data. The confusion matrix of the current research work is given in table 2.

Table 2: Confusion Matrix

	precision	recall	f1-score	support
False	0.85	0.80	0.82	44
True	0.82	0.88	0.85	48
accuracy			0.84	92
macro avg	0.84	0.84	0.84	92
weighted avg	0.84	0.84	0.84	92

The decision tree obtained for the research work is shown in Figure 4. The decision tree algorithm selects the best feature to split the tree in the root node to get the pure subsets of the dataset. In our approach, the Biochemical Oxygen Demand (B.O.D) is selected at the root level. The decision is taken at every intermediate decision node. If the threshold met then it goes to the left side to check other attributes otherwise it goes right to evaluate other sets of attributes. Finally, the leaf nodes decide whether water is drinkable or not.
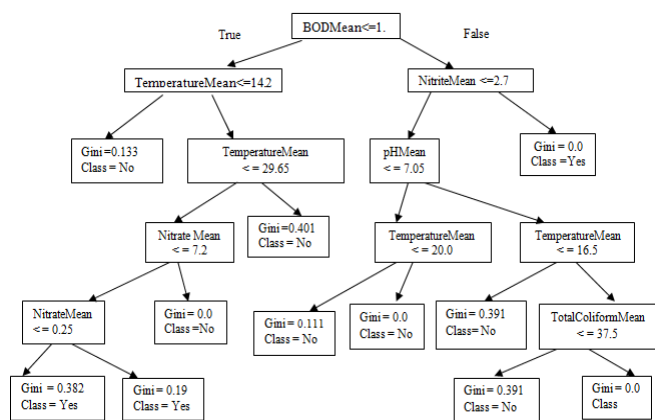
Figure 4: Decision Tree for the Proposed System

4.1 Graphical User Interface

The cloud platform Ubidots has the facilities to customize the display area as per requirements. It uses the decision tree classifier to analyze and predict water quality.

4.2 Performance Evaluation

The decision tree algorithm is utilized to predict the quality of water with an accuracy of 84% for the dataset of 307 records and it is even higher when the training dataset increases. The Accuracy level of ID3 is 83.5 is lowest whereas CART has an accuracy level of 98.67%. The average performance is seen for C4.5 for an accuracy of prediction 95.25%.

The *Decision_Tree_Classifier*gives parameters, for example, *minSampleLeaf*and *maxDepth*to avoid overfitting. The post pruning of the decision tree gives control on the size of a tree. In *Decision_Tree_Classifier*the pruning method is provided with a cost multifaceted nature parameter, *ccpAlpha*. More prominent estimations of *ccpAlpha*increment the quantity of hubs pruned. There are greater impact of *ccpAlpha*on controlling the trees and how to pick a *ccpAlpha*dependent on approval scores.

Node count in the final tree is: 1 with ccpAalpha: 0.0897411699372182

The pruning Path gives the effective value of alphas and the related accumulated impurities at the leaf in every step of the pruning steps. The greater values of alpha indicate greater the pruned tree, and the total impurity increases of its leaves.
The tradeoffs between training and test the dataset are shown in Figure 5. The accuracy is calculated post pruning and it reduces for both training and testing the dataset with increased values of alpha. The Mean errors are computed for Regression model and the obtained values are-
Mean Absolute Error (MAE): 0.26968833373756973
Mean Squared Error (MSR): 0.1481562984055075
Root Mean Squared Error (RMSE): 0.38491076680901964

a. Impurities vs. effective alpha   b. Nodes and depth of tree vs. effective alpha

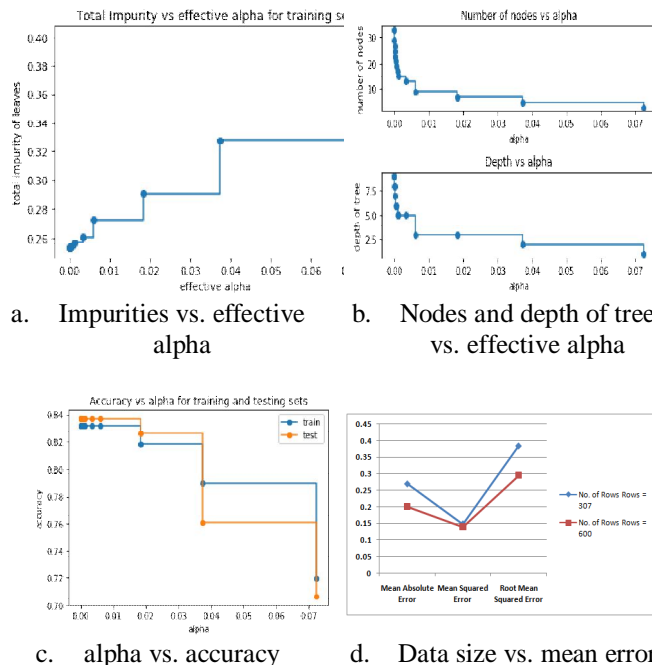c. alpha vs. accuracy   d. Data size vs. mean error

Figure 5: Performance Measurement with Alpha

5. CONCLUSION

In our research work we have inspected the groundwater quality in different spots. A cloud-based setup is created with the help of CC3200 microcontroller. The various sensors are placed in overhead tank to measure quality groundwater stored in the tank. The information gathered by sensors is taken in the cloud with the assistance of a microcontroller. The data which are collected stored in CSV file which is further sent to the machine learning module. In our proposed system a decision Tree-based classifier is used to classify the instance into drinkable i.e. positive or not drinkable i.e. negative instances. The microcontroller is also connected to display system and mobile devices through GSM module and Bluetooth. We have also compared the various algorithm used for the classification of the dataset. The research can be further extended for the distributed environment using big data stream processing in spark framework.

REFERENCES

1. Pravin Pachorkar1,Vidhi Dholu, Pritam Sanghavi, Pranit Mutha, Yatin Suryawanshi, **Real Time Water Monitoring System using IoT**, *International Journal for Modern Trends in Science and*

Technology, Vol. 04, Issue No. 10,PP.30-32, October 2018.

2. S. Geetha and S. Gouthami, **Internet of things enabled real time water quality monitoring system,** *Springer*,PP.1-19, 2017. https://doi.org/10.1186/s40713-017-0005-y

3. A.N. Prasad, K. A. Mamun, F. R. Islam, H. Haqva , **Smart Water Quality Monitoring System**, in *Proc. 2ⁿᵈAsia-Pacific World Congress on Computer Science and Engineering*, USA, 2015.

4. Dr. J.SasiBhanu, Dr. JKR Sastry, P. Venkata Sunil Kumar, B. VenkataSai, K.V.Sowmya, **Enhancing Performance of IoT Networks through High Performance Computing**, *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, Vol. 8, No.3, PP. 432-442, May -June 2019. https://doi.org/10.30534/ijatcse/2019/17832019

5. DaehanKwak, MD. HumaunKabir S. M. Riazul Islam,**The Internet of Things for Health Care:A Comprehensive Survey,***IEEE Access,* Vol. 3, PP. 678-708, June 2015. https://doi.org/10.1109/ACCESS.2015.2437951

6. M. Ter´an, J. Aranda, H. Carrillo, D. Mendez, and C. Parra, **IoT based system for indoor location using bluetooth low energy,** in *Proc. IEEE Colombian Conference on Communications and Computing (COLCOM)*, Colombia, 2017.

7. Jaime Alonso, Carlos Bayona, Omar Rojas, Marco Ter´an, Juan Aranda, Henry Carrillo, Carlos Parra,**IoT Solution for Data Sensing in a Smart Campus using Smartphone Sensors**, in *Proc. Communications Conference (COLCOM), IEEE Colombian*, Colombia, 2018.

8. YashaswiRahut, RimshaAfreen, DivyaKamini, **Smart weather monitoring and real time alert system using IoT**, *International Research Journal of Engineering and Technology (IRJET)*,Vol.5, Issue 10, PP.848-854, Oct 2018.

9. A.Ramya, R.Rohini, S.Ravi, **IOT Based Smart Monitoring System for Fish Farming,***International Journal of Engineering and Advanced Technology (IJEAT)*, Vol.-8 Issue-6S, PP. 420-424, Augt 2019 https://doi.org/10.35940/ijeat.F1089.0886S19

10. www.who.int

11. Nikhil Kumar Koditala, Dr.PurnenduShekar Pandey, **Water Quality Monitoring System using IoT and Machine Learning**, in Proc.*International Conference on Research in Intelligent and Computing in Engineering (RICE),* Hanoi, 2018.

12. Satyam Srivastava, SaikrishnaVaddadi, Shashikant Sadistap, **Smartphone-based System for water quality analysis,***Applied Water Science*, Springer open Access, PP.1-13, Aug. 2018.

13. Mo'tamad Bata, Rupp Carriveau& David S.-K. Ting, **Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model**, Springer open access, PP. 1-18, 2020.

14. Amir HamzehHaghiabi, Ali HeidarNasrolahi, Abbas Parsaie, **Water quality prediction using machine learning methods**, *Water Quality Research Journal*, Vol. 53, Issue 1, PP. 3-13, February 2018. https://doi.org/10.2166/wqrj.2018.025

15. Shiueng-BienYang, Tai-LiangChen,**Uncertain decision tree for bank marketing classification,***Journal of Computational and Applied Mathematics, Elsevier*, Vol. 371,PP. 1-23, Jan. 2020.

16. Hongfang Lu, Xin Ma, **Hybrid decision tree-based machine learning models for short-term water quality prediction**, *ChemosphereElsevier*,Vol. 249,PP. 1-27, Feb. 2020.

17. Albert A, Koelmansa, NurHazimah Mohamed Nora, EnyaHermsena, MerelKooia, Svenja M. Mintenig, Jennifer De France, **Microplastics in freshwaters and drinking water: Critical review and assessment of data quality**, *Water Research Elsevier*, vol. 155, PP. 410-422, May 2019.

18. Andreea-MihaelaDunca, **Water Pollution and Water Quality Assessment of Major Transboundary Rivers from Banat (Romania)**, *Hindawi Journal of Chemistry*, Vol.2018, PP. 1-8March 2018. https://doi.org/10.1155/2018/9073763

19. C.Ashwini, UdayPratap Singh, EktaPawar, Shristi, **"Water Quality Monitoring Using Machine Learning And Iot"**, International Journal Of Scientific & Technology Research, ISSN 2277-8616, vol. 8, Issue 10, October2019.

20. Yafra Khan, Chai Soo See, **Predicting and analyzing water quality using Machine Learning: A comprehensive model**, in *Proc.Systems, Applications and Technology Conference (LISAT)*, Long Island, April 2016.

21. Ali Najah Ahmed, FaridahBinti Othman, Haitham Abdulmohsin Afan, RusulKhaleel Ibrahim, Chow Ming Fai, Md.Shabbir Hossain , Mohammad Ehterame, Ahmed Elshafie,**"Machine learning methods for better water quality prediction"**, *Journal of Hydrology*, Vol. 578, Aug 2019.

22. Sujeet More, Jimmy Singla, **Machine Learning Techniques with IoT in Agriculture**, *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, Vol. 8, No.3, May -June 2019. https://doi.org/10.30534/ijatcse/2019/63832019

23. Munya A. Arasi, Sangitababu,**Survey of Machine Learning Techniques in Medical Imaging**, *International Journal of Advanced Trends in Computer Science and Engineering(IJATCSE)*, Vol. 8, No.5, September -October 2019. https://doi.org/10.30534/ijatcse/2019/39852019

24. G. Kanagaraj, T. Primya, K. SashiRekha, C. Vinothini, P. Anitha, **IoT-Enabled Water Quality Monitoring System"**, *Inventive Communication and Computational Technologies, Springer, LNNS*, Vol. 89, PP. 275-291, Jan 2020. https://doi.org/10.1007/978-981-15-0146-3_26

25. N.Balaji, S.Vijayalakshmi, K.Durgadevi, K.Mohanraj, T.Mangayarkarasi, **FPGA Implementation of Smart Water Quality Monitoring System**, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 8, Issue-11, September 2019. https://doi.org/10.35940/ijitee.K2500.0981119