International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

H_2O Caliber: An IoT Enabled Surface Water Pollutant Assessment System with Deep Learning

T. Tamilselvi[1], R. Saravanakumar[2], R. Arunkumar[3], K. Revathi*[4]

Submitted: 13/11/2022 **Accepted**: 15/02/2023

Abstract: Water is one of the natural sources that indicate the health of all living organisms like plants, animals, human being, etc., of our ecosystem. For humans, it helps to maintain their body temperature. Further it protects our tissues and organs from shock and damage by cushioning our joints. The quality of water is determined by its biological, chemical, microbiological and physical characteristics. India's water bodies are becoming increasingly hazardous as the country develops and urbanizes. Around 70% of India's surface water is unsafe for human consumption, according to estimates given by commission of pollution control board (PCB). The proposed work is aimed to produce a regulatory water monitoring system in place through incorporating the appropriate wireless sensors that evaluates the quality which is achieved possibly using technology named as Internet of Things (IoT). From the IoT hardware unit designed, the quality data of water are gathered, the pollutant level and its readiness for drinking is assessed with the help of deep learning mechanism. In order to improve the prediction accuracy, an enhanced random forest algorithm is implemented and evaluated against conventional machine learning algorithms. The water bodies' quality of present as well futuristic scenario which is to be maintained as per the standards set by PCB and other legislative support schemes is the predominant application of this work.

Keywords: *Deep learning, internet of things, random forest, regulatory monitoring, surface water pollutants*

1. Introduction

Water pollution is mostly caused by human activities such as littering, improper disposal of industrial and production waste, and poor waste management in general. Trashes that we toss into the water or that we throw on the streets and end up in water are examples of common water pollutants. Sewage and industrial waste are two more surprise sources. Untreated sewage is a major problem in many poor countries, according to United Nations Educational, Scientific, and Cultural Organization (UNESCO).The water pollution assessment by Statista against the percentage of death reveals that India is positioned the first place and caused death of 2.33 million persons lives in the last year (2021) because of the diseases raised from the water pollution.

The contamination of water bodies such as ponds, rivers, lakes, seas, and others is known as water pollution. In turn, this degrades quality of the water, harms aquatic life beneath the surface, and occasionally upon reaching our own domestic water supplies, putting our health at high risk. Every year, nearly 1 billion people become ill as a result of contaminated water. Low-income groups are particularly vulnerable because their dwellings are frequently located near polluting enterprises. Disease-causing viruses and bacteria from the waste of animal and human are a primary cause of illness from contaminated drinking water. Cholera, giardia, and typhoid are among the diseases carried by contaminated water. Infectious aquatic illnesses are the leading cause of death among young children worldwide. Every year, poisoned water kills more people than war and other types of violence combined.

1 Associate Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai – 600123, INDIA, tamilselvime@gmail.com, https://orcid.org/0000-0001-6614-9412
2 Dayananda Sagar Academy of Technology and Management, Bangalore – 560082, INDIA, dr.saravanakumar@dsatm.edu.in, https://orcid.org/0000-0002-4622-5745
3 Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai - 600124, INDIA,arunkumar.r@ritchennai.edu.in https://orcid.org/0000-0002-9136-8083
4 Associate Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Chennai – 600123, INDIA, neyadharshini@gmail.com, https://orcid.org/0000-0003-3411-9264
** Corresponding Author Email: neyadharshini@gmail.com*

1.1. Related Work

Kamaruidzaman et al. [1] conducted a deep survey in demonstrating the role of IoT and available tools in monitoring water quality as great interest of their study. Dimas Adiputra et al. [2] developed a real time water monitoring system by ascertaining the predominant parameters in use for evaluating quality with IoT technology. Daigavane and Gaikwad [3] built an IoT prototype for smart and real time water quality monitoring by incorporating prominent sensors. Cho Zin Myint et al. [4] designed a cost effective system to screen water quality. Chen et al. [5] addressed the importance of water quality assessment system and significant usage of IoT technology of high scope. Hrithik Yadav et al. [6] assessed the quality of water through IoT based kit over various water sources.

Li Liang [7] proposed a kind of deep belief network in evaluating the quality of water. It worked on data retrieved from real time IoT kit installed in location of interest. Jianhong Li et al [8] formulated a machine learning framework for predicting water pollutants by extracting the data from exact exploitation of IoT sensors in association with spectral technology. He formulated a novel version of SVM through annotating radial basis and least square equations into it in order to increase the prediction rate. Dilek Düstegör et al. [9] carried out an exhaustive survey on IoT based water quality monitoring. The authors inherently investigated the role of IoT and machine learning technologies in bringing the smart and intelligent system for remote monitoring of water quality. The system also presented the associated challenges in installing the same in domestic applications.

Angel Vergina et al. [10] brought the smart design of water quality monitoring model which is experimented in real time. Further the quality attribute data extracted from sensors are processed with the help of machine learning algorithm like SVM and Fuzzy k-means to decide upon the quality of water as drinkable or not. Sathish Pasika et al. [11] developed a cost effective smart water quality system in order to ensure the quality of water which is to be delivered across cities and villages in real through building an IoT hardware placing appropriate sensors in development board. The water quality is assessed in cloud using ThinkSpeak application. Manya Kakkar et al. [12] derived the IoT enabled solution to monitor the pollutants in water using neural networks. It is originally designed to verify the quality of water in home tanks. Further it can apply in any such scenario with the interest in predicting pollutants and facilitating adequate water treatments on time.

Varsha Lakshmikantha et al. [13] built an uninterrupted service to measure the water quality in order to facilitate promising purity water employing IoT and machine learning as technologies of interest. The model used neural network algorithms to decide upon the pollutant levels in water. Manoj et al. [14] inherently conducted a deep study to analyze the features and components of smart water quality monitoring systems realized as past efforts by researchers. Further it also reveals impending features to be considered in the futuristic design. Duarte Maher [15] conducted an exhaustive research in revealing the need for water quality monitoring system in place. He built a sensory circuit and installed the same in Malaren lake. He used the real time extracted from the installed sensors in the area for his study in selecting effective machine learning algorithms for prediction. Through investigating various algorithms from logistic regression, artificial neural network, SVM, decision tree and k-nearest neighbor (k-NN), he chosen hybridization k-NN with SVM for implementation.

Uferah Shafi et al. [16] built an IoT based water monitoring system through incorporating necessary sensors like conductivity, nitrate, pH and turbidity. The prototype built was tested around the data collected over experimentation results arrived out of surface water bodies around Pakistan. The author employed deep neural network algorithm to improve that accuracy prediction. Further a mobile application is developed to remotely monitor and control the system. Similarly Srivastava et al. [17] came up with a mobile application which monitors the surface water pollutants through IoT device. Sorayya Rezayi et al. [18] demonstrated the performance of random forest (RF) over deep neural networks in the process of diagnosing kidney diseases. Likewise Anuradha Vashishtha et al. [19] highlighted the effectiveness of RF approach in their work. Sathya and Balakumar [20] implemented an integrated classifier adopting RF and SVM together to predict fraudulent activities in insurance.

1.2. Research Gap and Objectives

The existing design and solutions to water quality monitoring was found to be ineffective in terms of its efficiency due to lack of sensors, controllers, poor design elements and classification algorithm which predicts the quality of the water.

The primary motive is to bring an smart water quality monitoring system provided all significant functionalities in one application that too in low cost. The parameters that helps to assess the quality of water, falls into three categories namely physical, chemical and biological. The color, taste, odour, temperature, turbidity, solids and electrical conductivity are among the physical properties. The pH, acidity, alkalinity, chlorine, hardness, dissolved oxygen and biological oxygen requirement are examples of chemical parameters. Biological parameters, such as bacteria, algae, and viruses, are the third category of parameter. Dissolved oxygen, for example, is one of the most significant parameters to consider when assessing the water quality of a river. The amount of dissolved oxygen in a water sample determines how filthy it is. Low dissolved oxygen levels suggest that the water is heavily polluted and that organic contaminants are depleting the dissolved oxygen supply. In the proposed water quality monitoring system, the predominant attributes of water to be monitored in implementation are as follows:

- Physical Parameters: Temperature, Turbidity and Conductivity
- Chemical Parameters: pH and Dissolved Oxygen

The novelty of the system lies in achieving cost effective design of IoT Hardware which gathers the data about the water quality remotely and application of optimized fast tree in forecasting the quality in advance. The primary objectives of the system that differentiates the proposed model with existing systems in place are as follows:

- To design a cost effective IoT prototype for smart water quality monitoring system by integrating appropriate sensors depicting significant attributes of water
- To facilitate a handy device design in order to ensure the variety of use cases from domestic stored water diagnosis to running surface water includes lakes, rivers and etc.,
- To improve the performance index with machine learning technologies

2. Materials and Methods

The proposed system is to develop an effective model for regular monitoring of the surface water pollutants and predicting the quality in advance with the help of a handy design of sensory circuit which is cost effective and low energy model.

2.1. An IoT Prototype for Smart Water Monitoring

To bring the working prototype for regulatory monitoring water pollutants, the development board which is the heart of the system is to be decided. The proposed model is planned to use Arduino Uno as its choice of microprocessor. The sensors are to be integrated respectively and other components like charging unit, communication unit are brought into the right place to complete the prototype design as captured in the Fig. 1.

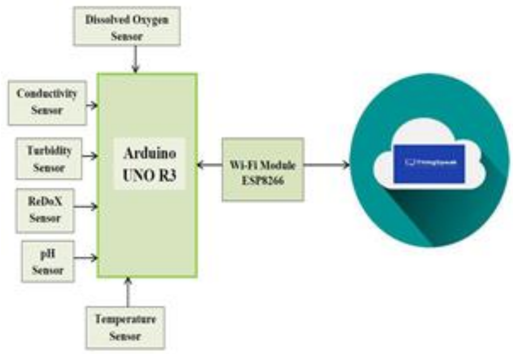

Fig. 1 An IoT design of proposed water monitor

The attributes which determines the quality of water are handpicked as mentioned in Table 1 to design a prototype. In strict adherence with Table 1, the appropriate sensors available in market which reflects the attributes mentioned are chosen.

Table 1. Set of attributes for assessing water quality

Attribute	***Normal reading***	***Acceptable variation***	***Observations***
Temperature	13 °C	3% (minimum of 0.2C)	Heavily dependent on the air temperature and depth of measurement
pH	6.5-8.5	0.1	Acid rock drainage and acid soil can lower pH. Algal bloom can increase pH
Conductivity	200-800 μs/cm	3%	Waste waters have conductivity of 10000 s/cm
ReDoX	0.35 mV	100 – 300 mV	Major change can indicate chemical contamination (Indus- trial contamination)
Dissolved Oxygen	79.3%	10%	Low value can indicate organic pollution High value can occur due to water stagnation and excessive growth of water plants
Turbidity	>5 NTU	10%	High turbidity level can indicate a high concentration of metals in the water

2.2. Design of Handheld Device

Upon developing the prototype, the same design is planned to be encapsulated as a handheld device in order to facilitate various use cases like estimating pollutants in stored water used in surface or domestic area of usage as well in running surface water too. The vital components that promote the handy design of IoT prototype is illustrated as Fig. 2.

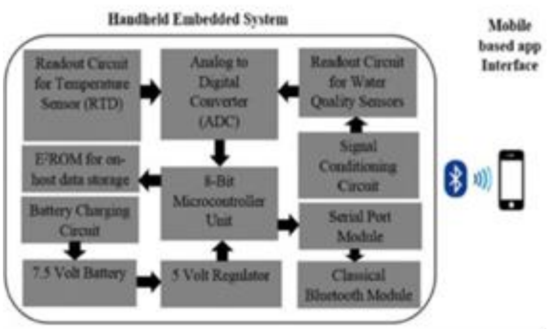

Fig. 2. Schematic diagram of proposed handheld device (Source: Srivastava et al.)

The printed circuit board design is prepared by putting it inside portable mound using 3D printing technology confirms the handy design to be the final circuit and operative one. The handy design of smart water monitor with its demonstration unit is expressed in Fig. 3.

Fig. 3. Proposed design of smart water monitor with demonstration unit (Source: Srivastava et al.)

2.3. Accommodating Deep Learning Algorithm

The previous attempts clearly envisioned that in the process of predicting water pollutants and estimating its quality, the most classical like SVM, decision tree, random forest and ANN based approaches were utilized. In this work, inspired from the past efforts [18-20], an attempt to improvising RF is carried out. RF is capable of handling large dataset in less training time with high accuracy. First, a set of decision trees are built by accompanying best features and then prediction is made upon the classification over the trees built. From the given data set, effective trees are built by ascertaining significant features through bootstrapping. By using bootstrap, each RF tree is formed on a sample of the data as opposed to all of the data during training. The designated segment is referred to as the bag, and the remainder as out of bag (OOB) trials. The results from various trees are pooled once they have all been trained on various bags. The aggregation process aids in minimizing fluctuation or amount of variation. In this way, the random forest is designed to investigate the patterns of interest deeply and faster with less computation time. Increasing the number of trees leads to higher accuracy, in this model built 120 decision trees out of this. The steps in fast forest algorithm proposed for predicting water quality is depicted as Table 2.

Table 2. Algorithmic steps in Deep Fast Forest (DFF)

Algorithm **DeepFastForest**

Input: A training data set Z=(x11,x12,x13,y11), …(xn1,xn2,xn3,yn1), features X and number of trees T

Function DeepFastForest(Z,X)

R $\leftarrow \phi$

loop i $\in 1,..,T\ do$

$Z^{(i)} \leftarrow A\ bootstrap sample\ from\ Z$

$r_i \leftarrow LearnTreeModel(Z^{(i)}, X)$

R $\leftarrow R \cup \{r_i\}$

end loop

return R

end Function

Function *LearnTreeModel*(Z $^{(i)}$, X)

At each node:

a $\leftarrow$ very small subset of X

Split on best feature in a

Return learned tree model

end Function

3. Results and Discussion

In this section, the details about the dataset utilized as an input to evaluate the water quality, development environment on which the proposed machine learning model is implemented and the performance metrics used to assess the proposed model are discussed.

3.1. Details of Dataset

The proposed model used the dataset named as water_potability.csv for predicting the water quality. The dataset is composed of significant characteristics pertains to water like pH, hardness, turbidity, conductivity, solids, chloramines, sulfate and organic carbon taken from various water bodies accounting 3276 entries. Samples from the dataset are captured in Table 3.

Table 3. Sample values from dataset

pH	*Hardness*	*Solids*	*Chloramines*	*Sulfate*	*Conductivity*	*Organic Carbon*	*Trihalomethanes*	*Turbidity*
3.7	146.5	18360.2	6.6	268.4	399.5	7.94	52.67	5.6
7.8	247.8	23067.2	9.3	316.5	426.7	13.05	45.87	6.5
7.5	256.7	28090.4	8.1	383.2	504.3	9.07	86.74	5.8
8.9	242.5	35670.2	7.5	404.4	570.7	11.34	30.65	7.2
6.7	148.2	44720.5	4.6	399.8	605.3	22.45	72.38	4.5

3.2. Execution Platform

The model is implemented with Colab which allows us to build our own jupyter notebooks over cloud freely. It is a web based integrated development environment (IDE) for experimenting python programs which was released by Google in year of 2017. The platform is renowned as convenient choice of developers in exercising the variety of machine learning and deep learning projects as easy as possible.

3.3. Performance Metrics

In the process of evaluating the performance of the model deployed for water monitoring is achieved through the following measures of interest and it is projected below as Table 4. These measures are rooted from confusion matrix of the model. The confusion matrix, which is frequently shown as a 2*2 matrix, aids in determining whether the model is "confused" while differentiating between the two classes. In this matrix, the row labels stand in for the actual labels, whereas the column labels constitute potential predicted labels. To reflect the two class designations, the two rows' and columns' labels are Positive and Negative. The 4 matrix entries indicate the 4 metrics that track how many accurate and inaccurate predictions the model yielded.

- True Positive (TP): The count of positive predictions
- True Negative (TN): The total negative predictions
- False Positive (FP): The count of false sample predicted as positive
- False Negative (FN): The total count of true samples predicted as negative

Table 4. Assessment Factors

Measure	*Formula*	*Description*
Accuracy	$A = \frac{TP+TN}{TP+TN+FP+FN}$	To indicate correctness
Precision	$P = \frac{TP}{TP+FP}$	To signify overall positive prediction
Recall	$R = \frac{TP}{TP+FN}$	To project negative predictions
F-Score	$F = 2*\frac{P*R}{P+R}$	Harmonic mean of precision and recall

3.4. Evaluation Results

In estimating the model, among the 3276 trials of data, 80 percentage of it used as training and remaining 20 percentage employed as testing the model. The 80:20 ratio is adopted as cross validation strategy and the resultant is visualized in Fig. 4. It gives the visual cues to grasp the performance of the proposed DFF.

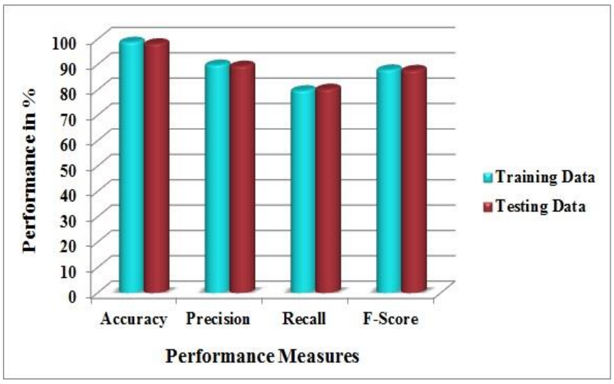

Fig. 4. Performance analysis of proposed DFF model

The proposed DFF model is experimented against the profound efforts encountered in water pollutant detection like ANN, decision tree, random forest and SVM. The performance measures obtained over these machine learning mechanisms is captured in Table 5.

Table 5. Performance matrix of models evaluated

Methods	*Accuracy*	*Precision*	*Recall*	*F-Score*
SVM	96.9	86.2	81.2	83.4
Random Forest	93.7	89.5	85.8	85.8
Decision Tree	93.5	81.7	82.3	81.5
ANN	95.5	88.8	87.3	82.9
DFF (Proposed Model)	98.6	89.7	80.1	87.9

The graphical demonstration of different approaches exercised in predicting water quality is best captured in Fig. 5. It is clearly envisioned that the proposed DFF model yield better results than other methods taken for comparative analysis.

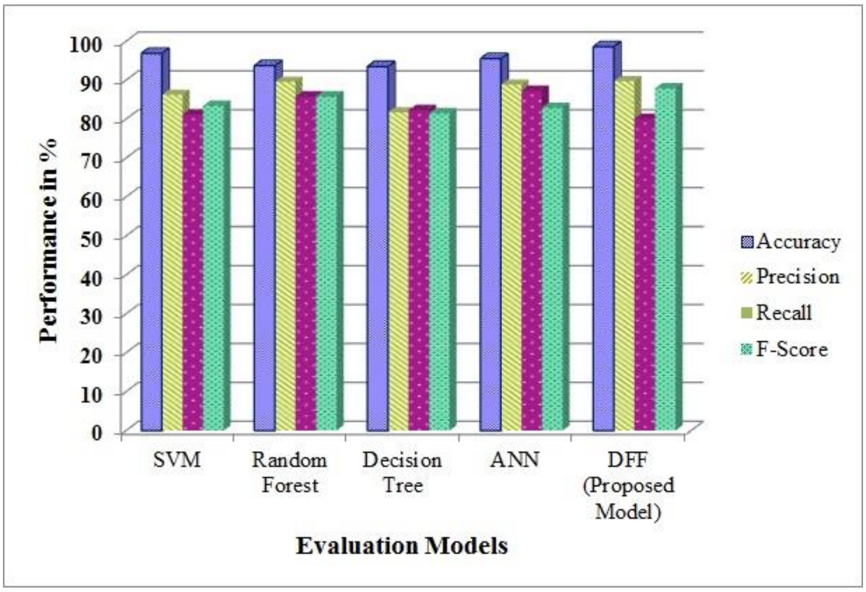

Fig. 5. Graphical Analysis of Different Approaches in Predicting Water Quality

From the graphic, DFF is observed as 1.7% accurate than SVM, 3.1% than ANN, 4.9% than KNN and 5.1% than decision tree. F-Score signifies the overall test accuracy, as F-Score is a harmonic mean of precision and recall, it projects the preciseness and robustness about the classifier or model. The higher F-Score ensures the greater accuracy. The proposed DFF gives its F-Score value as 87.9% which is higher than comparative methods of evaluation.

4. Conclusion

In this work, acute sensors for monitoring surface water pollutants are investigated in deep interest. Moreover an effective IoT based design is also discussed. Further a novel classifier by integrating bootstrapping into RF is built and the performance of the same is evaluated against noteworthy classifiers like ANN, decision tree, random forest and SVM. The experimental results revealed that proposed DFF facilitated promising accuracy 98.6%. The work is much focused on deriving a best possible classifier for predicting water pollutants over publicly viable dataset. In future, the model can be executed over real dataset observed through the IoT device discussed in this work.

Acknowledgements

No funding sources.

Author contributions

All authors are equally contributed in preparing, experimenting and reviewing the article.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] N. S. Kamaruidzaman and Siti Nazahiyah Rahmat, "Water Monitoring System Embedded with Internet of Things (IoT) Device: A Review", *IOP Conf. Series: Earth and Environmental Science*, vol. 498, pp. 1-8, 2020, doi:10.1088/1755-1315/498/1/012068.

[2] Anifatul Faricha, Dimas Adiputra, Isa Hafidz, Moch. Iskandar Riansyah, Lora Khaula Amifia, Moch. Fauzan Rasyid, Moch. Bagus Indrastata and Abdulloh Hamid Nushfi, "Design Configuration of Water Quality Monitoring System in Surabay", *Complete- Journal of Computer, Electronic, and Telecommunication*, vol. 1, no. 1, pp. 1-11, Jul. 2020, doi:10.52435/COMPLETE.V1I1.43.

[3] Vaishnavi V. Daigavane and M.A Gaikwad, "Water Quality Monitoring System Based on IOT", *Advances in Wireless and Mobile Communications*, vol. 10, no. 5, pp. 1107-1116, 2017.

[4] Cho Zin Myint, Lenin Gopal and Yan Lin Aung, "Reconfigurable smart water quality monitoring system in IoT environment", *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, 2017, pp. 435-440, doi: 10.1109/ICIS.2017.7960032.

[5] Fan-Lun Chen , Bo-Chieh Yang , Shu-Yi Peng and Tzu-Chi Lin, "Applying a deployment strategy and data analysis model for water quality continuous monitoring and management", *International Journal of Distributed Sensor Networks*, vol. 16, no. 6, pp. 1-15, Jun.

2020, https://doi.org/10.1177/1550147720929825.
[6] Hrithik Yadav, Navya Kola, Ashritha Reddy and Kiran Kumar, “Smart Water Quality Monitoring System using IOT and Electronic Sensors”, *International Journal for Research in Applied Science & Engineering Technology (IJRASET*), vol. 10, no. 6, pp., 2873-2876, Jun. 2022, https://doi.org/10.22214/ijraset.2022.44271.
[7] Li Liang, “Water Pollution Prediction Based on Deep Belief Network in Big Data of Water Environment Monitoring”, *Hindawi- Scientific Programming*, pp. 1-11, Dec. 2021, https://doi.org/10.1155/2021/8271950.
[8] Jianhong Li, Ken Cai, Huazhou Chen , Lili Xu, Qinyong Lin and Feng Xu, “Machine Learning Framework for Intelligent Detection of Wastewater Pollution by IoT-Based Spectral Technology”, *Hindawi - Wireless Communications and Mobile Computing*, pp. 1-10, Feb. 2022, https://doi.org/10.1155/2022/9203335.
[9] Farmanullah Jan, Nasro Min-Allah and Dilek Düstegör, “IoT Based SmartWater Quality Monitoring: Recent Techniques, Trends and Challenges for Domestic Applications”, *Water-MDPI,* vol. 13, no. 13, pp. 1-37, Jun. 2021, https://doi.org/10.3390/w13131729
[10] S. Angel Vergina, S. Kayalvizhi, R.M. Bhavadharini and Kalpana Devi, “A Real Time Water Quality Monitoring Using Machine Learning Algorithm*”, European Journal of Molecular & Clinical Medicine*, vol.7, no. 8, pp. 2035-2041, 2020
[11] Sathish Pasika and Sai Teja Gandla, “Smart water quality monitoring system with cost-effective using IoT”, *Heliyon*, vol. 6, no. 7, pp. 1-9, Jul. 2020, DOI: 10.1016/j.heliyon.2020.e04096
[12] Manya Kakkar, Vansh Gupta, Jai Garg and Dr. Surender Dhiman, “Detection of Water Quality using Machine Learning and IoT”, *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 11, pp. 73-75, 2021.
[13] Varsha Lakshmikantha, Anjitha Hiriyannagowda, Akshay Manjunath, Aruna Patted, Jagadeesh Basavaiah and Audre Arlene Anthony, “IoT based smart water quality monitoring system”, *Global Transitions Proceedings*, vol. 2, no. 2, pp. 181-186, Nov. 2021, https://doi.org/10.1016/j.gltp.2021.08.062.
[14] M. Manoj, V. Dhilip Kumar, Muhammad Arif, Elena-Raluca Bulai, Petru Bulai and Oana Geman, “State of the Art Techniques for Water Quality Monitoring Systems for Fish Ponds Using IoT and Underwater Sensors: A Review”, *Sensors*, vol. 22, no. 6, pp. 1-17, Mar. 2022, https://doi.org/10.3390/s22062088.
[15] Duarte Maher, “IoT for fresh water quality monitoring”, Thesis for Degree Project in Information and Communication Technology, KTH University, Stockholm, Sweden, 2018.
[16] T. Uferah Shafi, Rafia Mumtaz, Hirra Anwar, Ali Mustafa Qamar, Hamza Khurshid, Surface Water Pollution Detection using Internet of Things, In Proc. of 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT), pp. 101-126, Pakistan, pp. 1-5, 2018.
[17] Srivastava, S.; Vaddadi, S.; Sadistap, S. “Smartphone-based System for water quality analysis”, Applied Water Science, vol. 8, no. 130, pp. 1-13, 2018.
[18] Sorayya Rezayi, Keivan Maghooli and Soheila Saeedi, “Applying Data Mining Approaches for Chronic Kidney Disease Diagnosis”, International Journal of Intelligent Systems and Applications in Engineering, vol. 9, no. 4, pp. 194-208, 2021.
[19] Anuradha Vashishtha, Anuja Kumar Acharya and Sujata Swain, “A Comparative Study on Various Machine Learning Approaches for the Detection of Alzheimer Disease”, International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 3, pp. 294-304, 2022.
[20] M. Sathya and B. Balakumar, “Insurance Fraud Detection Using Novel Machine Learning Technique”, International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 3, pp. 374-381, 2022.