ISSN 1063-455X, Journal of Water Chemistry and Technology, 2020, Vol. 42, No. 5, pp. 321–328. © Allerton Press, Inc., 2020.

THEORETICAL PRINCIPLES
OF WATER PURIFICATION
AND TREATMENT TECHNOLOGY

A Machine Learning Approach towards Automatic Water Quality Monitoring

Sandeep Bansal[a], * and G. Geetha[b]

[a]*Department of Electronics and Communication Engineering, Lovely Professional University, Phagwara, India*

[b]*Division of Research and Development, Lovely Professional University, Phagwara, India*

**e-mail: er.sandy029@gmail.com*

Received March 26, 2019; revised September 8, 2019; accepted May 14, 2020

Abstract—Increasing rate of water pollution and consequently waterborne diseases are the engrossing evidence towards danger to living organisms. It becomes a great challenge these days to preserve our flora and fauna by controlling various unexpected pollution activities. Although the invention of many schemes and programmes regarding water purification has done a tremendous job, but still there is something that has been lagging. With increase in population, industrialization and global warming situation is getting worse day by day. It becomes very difficult to get safe drinking water and appropriate quality water for other domestic usage and agriculture purpose. Major reasons for water pollution include undesirable increase in impurities. These may cause eutrophication of the water body, change in taste, discolouration and odour of water, water borne diseases, increase in water toxic nature etc. For water to be serviceable it should be aesthetically acceptable, chemically safe, bacteria free; organic substances and radioactive elements should be absent. So, there is an urgent need to look into this situation and take the corrective and necessary actions to overcome this situation. The government is paying an attention to this problem and finding the ways to control the situation. However, major areas are not developed to the point and water quality estimation is totally dependent upon sampling at location and testing in laboratories. Manual sampling and measurements are prone to human errors and these techniques may create ambiguities in predicted output. In this paper we have presented Machine Learning (ML) approach for calculating the Water Quality Index (WQI) and classification of water quality to estimate water characteristics for usage. For analysis, decision tree method is used to estimate water quality information. The standard values of parameters are selected as per guidelines provided by World Health organization (WHO). Results calculated using ML techniques showed prominent accuracy over traditional methods. Accuracy achieved is also significant, i.e. 98 %. Likewise, projection of gathered data was done utilizing web interface and web app to alert the authorities about contamination.

Keywords: machine learning, water quality index, contamination, water quality standards

DOI: 10.3103/S1063455X20050045

INTRODUCTION

Water being the most essential need in present generation, it is as essential as that to keep it clean. Although the invention of many schemes and programs regarding water purification has done a tremendous job, but still there is something that has been lagging. The main causes of water pollution are industrial dumping of waste directly into water, testing of unwanted weapons like bombs by directly throwing them into water to identify their range, oil pollutant etc. [1]. Major reasons for water pollution include undesirable increase in impurities. These may cause eutrophication of the water body, change in taste, discolouration and odour of water, water borne diseases, increase in water toxic nature etc. For water to be serviceable it should be aesthetically acceptable, chemically safe, bacteria free; organic substances, and radioactive elements should be absent. According to statistics WHO reported, approximately 36% of urban and 65% of rural India are without access to safe drinking water [2]. Various studies have been presented-which show that human activities are mostly responsible for water contamination [3].

Water quality can be measured by checking its various physical, chemical and biologic properties, like the contents of inorganic or organic substances, toxic metals, radioactive elements, organic nutrients, oxygen demand, bacterial contents etc. [4]. Chemical parameters include hardness, calcium, magnesium, chloride, sulphate, nitrate etc. Physical parameters include temperature, pH, conductivity, turbidity, taste and odour etc. Among all these parameters, 5 important characteristics have been considered in this work that play an important role while measuring water quality. These parameters are pH, Temperature, Dissolved oxygen (DO), Turbidity, Total Dissolved Solids (TDS). The novelty proposed in this work is using Machine Learning (ML) approach to evaluate water quality with minimal possible error. Prediction of water properties in subsequent years is carried out by training the system using appropriate mechanisms and samples taken from authentic sources. Present work focuses on water quality monitoring by integrating machine learning approach with sensor-based network. This is ready to lend a hand because it is cost effective and easy to operate.

MACHINE LEARNING

Computers are good at explaining algorithmic and mathematical problems, but frequently the world can't simply be well-defined with a mathematical process. Machine Learning is a thriving field these days, with an increasing credit that ML can play a vital role in variety of applications, such as natural language processing, data mining, image processing, and skilled system [5]. ML is very much helpful in solving problems related to these areas and set to be a scope in near future. ML serves lot of applications including Speech Recognition, Making Predictions and Classifications etc.

Decision tree makes decisions which are more accurate as of regular systems. Decision tree consists of chance nodes, decision node and end nodes. Chance node is represented with circle and commonly used to show probability of certain results. Decision node is represented by square and it is used to make the decisions, and the finally end node represents output [6]. Basically, a decision tree is a tree-like structure which is sub-divided into branches. These branches carry decision node along with them to take a decision for the end node. As classification of water quality based on the sensed parameters has to be done, so decision tree is best fitted model or algorithm to serve the purpose in proposed work.

WATER QUALITY PARAMETERS

Water quality can be measured based upon various parameters defined by standard organizations [7]. Some of the important parameters considered for this work for measuring water quality are as following.

pH represents hydrogen ions content available in water and also known as potential hydrogen. The value of pH in water generally tells about acidic or alkaline characteristics of water. pH level equals to 7 being neutral water, above 7 is more alkaline and below 7 it is more acidic. It has its own importance as human body requires pH level near about 7.4 and also controls the balance of the body [8].

Dissolved oxygen show availability of oxygen amount present in water when air hits directly. It is very much important for water life and aquatic animals. These aquatic animals survive by consuming dissolved oxygen present in water. It is inversely proportional to temperature. Oxygen deficiency occurs when the emission of oxygen through plant respiration does not provide the necessary content in the atmosphere. Occurrence probability of such cases is high in winter times [9].

Temperature is an important parameter because it is the major reason for bio chemical reactions in water life. Survival rate and growth of aquatic life also depends upon water temperature. Higher temperture leads tohigher metabolic rate and vice versa.

Total solids consists of two parts i.e. Total Dissolved Solids and Total Suspended Solids. Dissolved solids refer to the chemicals present in water like lead, mercury, cations, anions etc. Suspended solids states are the elements those are imposed in water and insoluble in nature. Suspended solids can be treated by a filtration method [10].

Turbidity is termed as the physicial parameter of water and represents clarity of water. It carries optical properities, because of that the light that's been imosed on the water is absorbed by the particles of water rather than being refelected.

SYSTEM ARCHITECTURE

Water Quality Index (WQI) tells us the entire quality status of water after multiple parameter evaluations. It is measured over a scale from 0–100. As water being more efficient in nature, along with it water carries several parameters which are divided into 3 sub-categories namely physical, chemical, and biological [11]. Each sub-category is again divided into insignificant categories. Physical parameters include pH, temperature, conductivity etc., chemical parameters include hardness, calcium, magnesium; chloride

Table 1. WQI classification range with effectiveness of water usability

Range	Class	Rating of water quality
0–25	V	Poor
26–50	IV	Below average
51–70	III	Average
71–90	II	Good
91–100	I	Excellent

Table 2. Standards for drinking water

Parameters	Units	WHO	EU	USEPA	CANADA	NESREA	ICMR
pH	–	6.5–8.5	6.5–9.5	6.5–8.5	6.5–7.5	6.5–8.5	6.0–8.5
Temperature	°C				<15	Ambient	
Total solids	mg/L	500	500	500	500	500	500
Turbidity	NTU	<5	4	0.5–1.0	0.1–1.0	<5	2.5–9
Dissolved Oxygen	mg/L						5

levels etc., and biological properties are divided into Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), phenols etc.

Consequently, water quality can be estimated by measuring large number of parameters, those carries some specified range defined by reputed organizations like World Health Organization [2] etc. Also, the limits specified for all parameters may vary from place to place and from sample to sample. Various parameters can be accepted by water research experts, but it is not easily understood by the public or by the scheme makers who are responsible for water improvement schemes. Numerous researches are being done in India as well as United Nations and they finally introduced a term called as Water Quality Index which simplifies the problem to measure efficacy of water quality estimation in standard terms.

Water quality can be estimated by using Eq.(1) to calculate WQI factor:

$$\mathrm{WQI} = \sum \mathrm{W_Y Q_Y} / \sum \mathrm{W_Y}, \tag{1}$$

where Y are available parameters, Q_Y are values of available parameters, W_Y are weighting factors of available parameters. WQI factor ranges between 0–100 by observing the graphical curves which are designed for water quality parameters. Table 1 shows the WQI classification range [12].

Many authorities (WHO, Bureau of Indian Standards) have set water quality standards for drinking water and other intended uses [2, 13]. These organizations set the permissible limits of various parameters for the consumer usage to reduce consumer health risk and to make sure clean water throughout world. For developing our system, after intensive research we have targeted on some important physiochemical parameters like pH, temperature and turbidity for WQM purpose. Table 2 shows various water quality parameters and their permissible limits for usage described by various organizations [14–18].

As per the finalized parameters to be measured in this work, following sensors have been chosen for taking real time measurements from water resources. Selection of the sensors has been done based on their response time and ability to handle harsh environments for real time applications.

Temperature sensor (DS18B20). This is a waterproof sensor for measuring temperature which is capable to even work in harsh environment conditions. This sensor is one wired which makes easy to connect with many devices. It has capability to measure temperature from negative to positive and its limits are from –55 degrees to +125 degrees Celsius. It is a sensor that does not require power from any external source because it extracts power from data line. Also, this sensor has a physical nature like as that of stainless steel which makes it to bear harsh environment or any wet situations. This sensor will be able to provide readings from 9–12 bits.

pH sensor. This sensor is the new arrangement of calculating pH because the previous sensors are very difficult to connect to any processor board and were very expensive, but this sensor is cost effective, and is specially designed for connection with normal processor boards. The attracting highlights include high accuracy and low-cost range that makes it more popular to use in such applications. It has LED, BNC connector, pH 2.0 sensor interface.

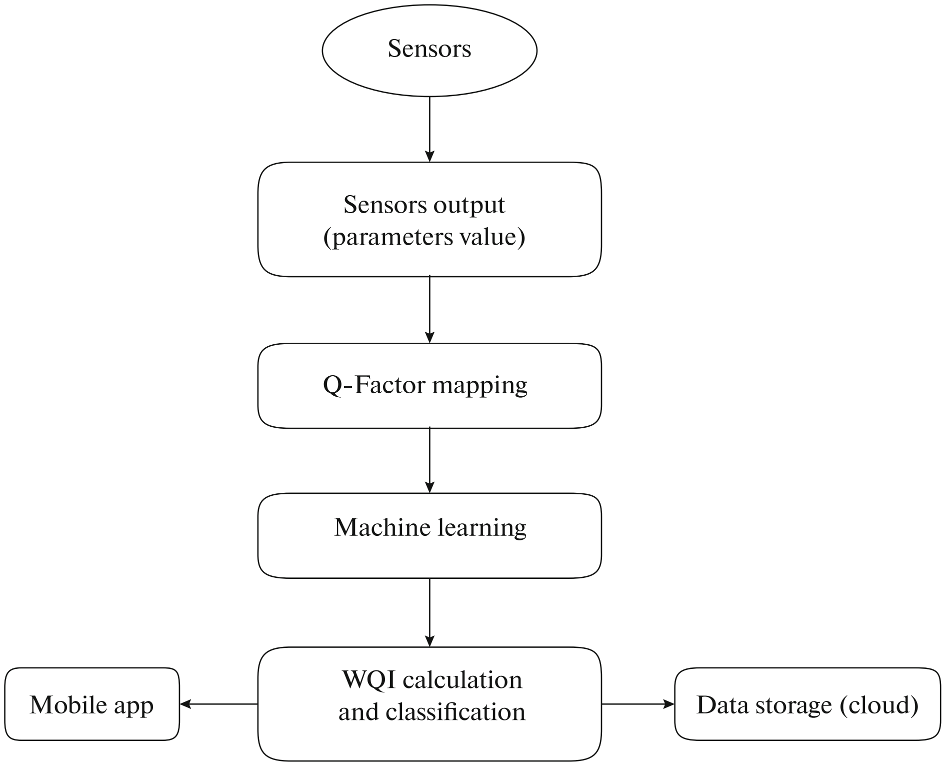

Fig. 1. Flow chart of designed structure.

Turbidity sensor. It measures the amount of turbidity range present in water by measuring the amount of suspended solids present that are responsible for obstructing light to pass into water. The calculation that is done between scattering value and light transmittance will provide us with the information of how much total suspended solids are available. Total Suspended Solids (TSS) and turbidity are related to each other. Turbidity sensor has both digital and analogue output modes. The output mode can be selected based on which can be either digital or analogue depending on the micro controller unit that we are using. This sensor is not a waterproof sensor.

TDS sensor. Gravity Analog TDS sensor is used in this work. Measuring TDS of water will give us the information about level of cleanliness present in water. It can be applied to all kinds of water resources. TDS pen was once widely used device to measure its value in water. But it possesses some technicality issues related to it, like transmission of data and quality monitoring etc. Keeping this in view we have used a new senor for the TDS checking namely analogue TDS sensor kit which is easy to use and highly compatible with processor board. This will prevent the probe from polarization and make the life of sensor higher. Also, this sensor is a waterproof device that makes it more suitable for such applications.

Dissolved oxygen sensor. Dissolved oxygen present in water is measured with this sensor which is highly compatible with normal processor boards. The probe of this sensor is made of galvanic material. Because of this, polarization chances are also low. Some of the parts of this device are replaceable. It has low maintenance cost.

DESIGN METHODOLOGY

The complete procedure for designing this work is divided into various steps as determined in algorithm 1. Figure 1 shows the step by step methods followed for designed structure.

Algorithm 1

Input: Water quality parameters values with WQI for training, Q-factor mapping values, Water sample to be tested.

Output: Water quality for the sample to be tested.

1. Calculate the Q-factor values for the dataset.

2. Create a report for the Q-factor values of water parameters values with their WQI.

3. Feed the database into WEKA tool to train and generate the decision tree model.

4. Collect the parameters value using designed hardware consisting of various sensors according to the water quality parameters to be tested.

5. Apply the sample values collected at step 4 to the trained decision tree model generated at step 3.

6. Finally, the quality of water sample can be determined after step 5.

Dataset preparation. Calculation of WQI is made possible after calculating Q-value for each parameter. Various researchers have proposed the sample curves for mapping [19]. Some of the relations used for mapping of parameter value with Q-Factor are shown in Fig. 2. Dataset have been prepared for network training. Sample data have been collected from various authenticated sources and used for learning and validation process. Here each parameter is firstly converted into Q-factor. Eqs. (2), (3), (4) are used for conversion from parameter value to Q-Factor of Temperature (T_{Q}), pH ($\mathrm{pH_Q}$) and Total Dissolved Solids ($\mathrm{TDS_Q}$), respectively:

$$T_{\mathrm{Q}} = \left(\left(0.00723232\, t^3\right) - \left(0.261429\, t^2\right) - (0.844661\, t) + 80.6558\right), \tag{2}$$

$$\mathrm{pH_Q} = \left(\left(-0.776289\, \mathrm{ph}^3\right) + \left(11.6597\, \mathrm{ph}^2\right) - (37.9749\, \mathrm{ph}) + 32.1271\right), \tag{3}$$

$$\mathrm{TDS_Q} = \left(\left(1.37374 \times 0.000001 \times \mathrm{tds}^3\right) - \left(0.00108052 \times \mathrm{tds}^2\right) + (0.128167 \times \mathrm{tds}) + 80.0909\right), \tag{4}$$

where t, ph and tds are numerical values of temperature, pH and total dissolved solids, respectively.

For Turbidity, the conversion from NTU to Q-factor is done by mapping using the Fig. 2. Additionally, Dissolved Oxygen Q-factor calculation is done by help of percentage saturation value of DO. The conversion from ppm to Q-factor is done by converting it to percentage saturation first and then to Q-factor by mapping equivalence as given in Fig. 2.

WEKA tool. WEKA is open source software that runs in JAVA platform. It allows you to upload any dataset and apply any algorithm on this dataset to produce required analysis results. We have used decision tree model for training and testing of water samples. WEKA uses an algorithm named J48 for the required purpose. J48 is an improvisation over ID3 algorithm. This algorithm generates a tree based upon the dataset consisting of labelled data. To test any water sample, we need to go through this tree from root node in appropriate direction to reach the leaf node which gives the final water quality estimation.

Water quality estimation. The values of each parameter are collected through sensors by dipping them into water up to recommended level. The results are fed to the decision tree and finally water quality is estimated.

Alert generation and cloud. Alert generation feature is also introduced in the designed system by setting alarms for any unusual tracing of WQI or class identified, so that action can be taken accordingly. Along with that data is also stored on cloud and a mobile app has been prepared for monitoring of all the activities and dataset collection. Wi-Fi module is used for storage of the data into mobile app and an online database.

RESULTS AND DISCUSSIONS

Decision tree model has been proved very much compatible for identifying water quality in this work. Large database collected from various authenticated resources has also helped in better learning capability of machine and also helped in validation of tested samples. Water quality is classified based upon 5 classes i.e., Class I, Class II, Class III, Class IV and Class V as discussed in Table 1.

The above classification is done in decision tree form by using Weka Software after an adequate training with approximately 500 samples and afterwards 200 samples used for result validation. Figure 3 shows the resultant decision tree graph for testing of water quality based on selected parameters.

Confusion Matrix shown in Table 3 represents the accuracy achieved through decision tree approach. It shows that overall 684 cases were exactly identified as per training data, whereas only 12 instances were incorrectly identified. It leads to 98.28% accuracy achieved while identifying water quality classification through decision tree method. Along with that it shows mean absolute error as 0.0199 and Root Mean Squared error of 0.0996.

After the measurement and analysis through decision tree model all the information is also transmitted to mobile app as well as online database storage for further processing. Graphical analysis is also possible from web data and alarming situations can be noticed and informed to concerned authorities.

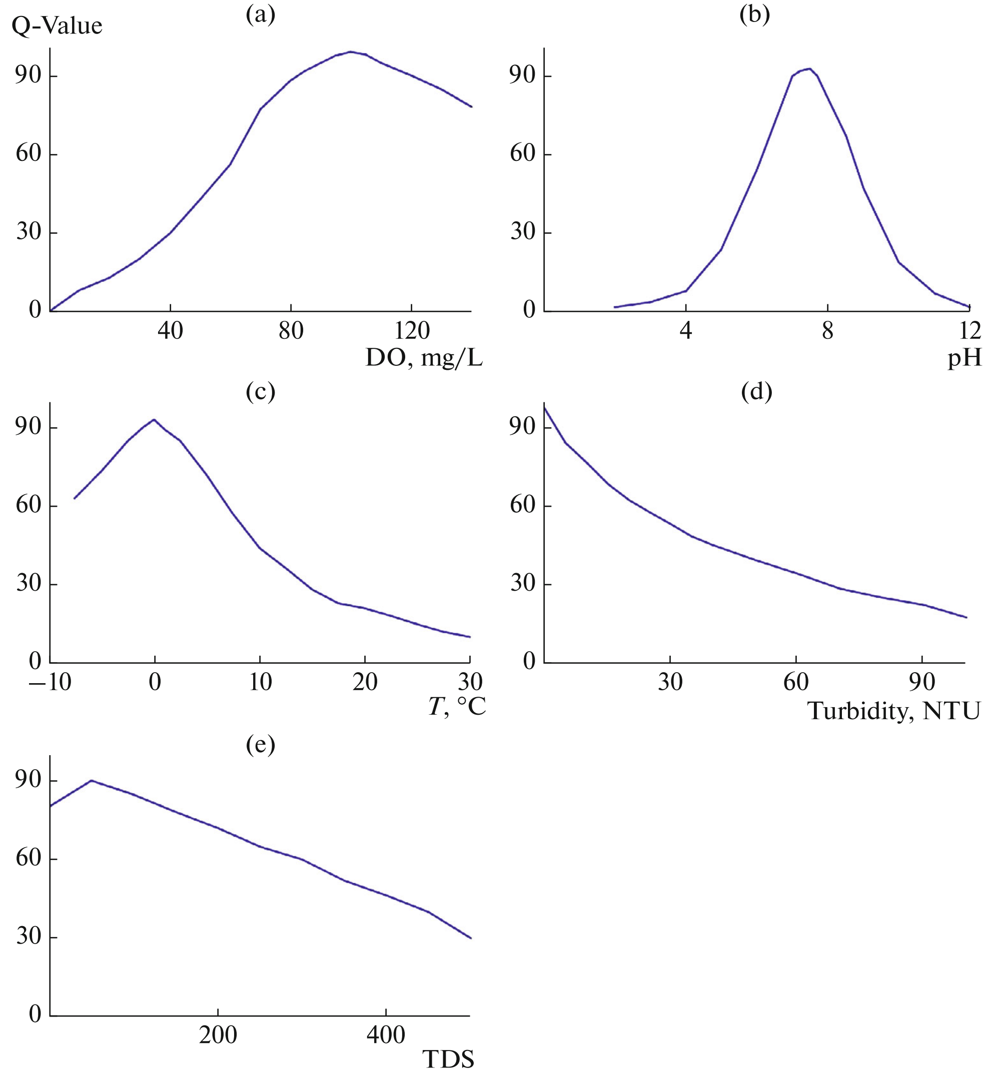

Fig. 2. Relation between DO (a), pH (b), temperature (c), turbidity (d), TDS (e) and Q-value.

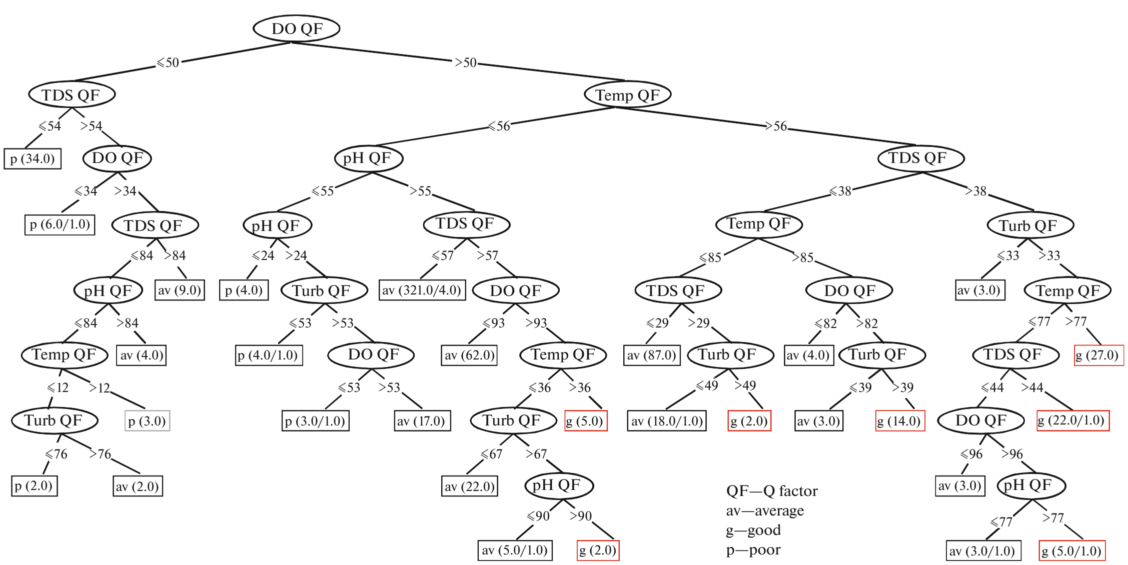

Fig. 3. Decision tree output graph for measuring water quality.

Table 3. Confusion matrix for decision tree output

Class			Classified as	
a	b	c		
75	3	0	Class a	Good
2	556	3	Class b	Average
0	4	53	Class c	Poor

Table 4. Comparative analysis of proposed approach with other traditional methods

Methodology opted for water quality calculation	Accuracy achieved, %
Standard WQI formulae	80.02
Parallel ID3 [20]	95.78
Proposed approach	98.3

Table 4 lists the comparison results of proposed approach with other traditional techniques. We have also used the same dataset for calculating WQI using standard formulae provided by various organizations. The estimated water quality is compared with lab results and the accuracy comes out to be 80% approx. Additionally, Qing and co-authors have proposed parallel ID3 approach for water quality classification and the achieved accuracy was approximately 95% [20]. Hence proposed approach provides significant results over the traditional methods.

CONCLUSION

Implemented algorithm is easy to understand due to its structure and division. Prediction is the key feature of a decision tree which analyses the present data and further can be used to predict future data. Different physical and chemical parameters are considered for this work to monitor real time water quality information. Five main parameters including pH, temperature, TDS, DO and turbidity are considered to assess water quality. This study reveals that machine learning approach is showing significant improvement in tested values over traditional methods. Water Quality is estimated during this study.

ACKNOWLEDGMENTS

The authors are thankful to management of Lovely Professional University, India for providing their support and assistance to carry out this work.

REFERENCES

1. Rengamani, J. and Venkatraman, V., Study on the safety and security of Indian ports, *Int. J. Manage.*, 2015, vol. 6, no. 10.
2. *Guidelines for Drinking-Water Quality*, Vol. 1: *Rrecommendations*, Geneva: World Health Org., 2004.
3. Rubonis, A.V. and Bickman, L., Psychological impairment in the wake of disaster: The disaster—psychopathology relationship, *Psychol. Bull.*, 1991, vol. 109, no. 3, p. 384.
4. *Water Quality Assessments: A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*, Chapman, D., Ed., London: E&FN Spon, 1996.
5. *Machine Learning: An Artificial Intelligence Approach*, Michalski, R.S., Carbonell, J.G., and Mitchell, T.M., Eds., New York: Springer, 2013.
6. Langley, P. and Simon, H.A., Applications of machine learning and rule induction, *Commun. ACM*, 1995, vol. 38, no. 11, pp. 54–64.
7. *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*, Geneva: World Health Org., 2017.
8. Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., and Kouimtzis, T., Assessment of the surface water quality in Northern Greece, *Water Res.*, 2003, vol. 37, no. 17, pp. 4119–4124.
9. Sánchez, E., Colmenarejo, M.F., Vicente, J., Rubio, A., García, M.G., Travieso, L., and Borja, R., Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution, *Ecol. Indic.*, 2007, vol. 7, no. 2, pp. 315–328.

10. Singh, K.P., Malik, A., Mohan, D., and Sinha, S., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study, *Water Res.*, 2004, vol. 38, no. 18, pp. 3980–3992.
11. Cude, C.G., Oregon water quality index a tool for evaluating water quality management effectiveness, *J. Am. Water Resour. Assoc.*, 2001, vol. 37, no. 1, pp. 125–137.
12. Said, A., Stevens, D.K., and Sehlke, G., An innovative index for evaluating water quality in streams, *Environ. Manage.*, 2004, vol. 34, no. 3, pp. 406–414.
13. *ICS 13.060.20: Indian Standard Drinking Water Specification*, New Delhi: Bureau Indian Stand., 2012, 2nd ed.
14. *Statutory Instruments No. 122: European Union (Drinking Water) Regulations 2014*, Dublin: Stationery Off., 2014, pp. 1–34.
15. Yamamura, S., Drinking water guidelines and standards, in *United Nations Synthesis Report on Arsenic in Drinking Water (Draft Report)*, Geneva: World Health Org., 2001.
16. Spry, D., An overview of Canadian water quality guidelines, *Proc. US Environmental Protection Agency Expert Meeting*, Washington, DC: US Environ. Prot. Agency, 2015.
17. Ladan, M.T., Review of NESREA act 2007 and regulations 2009–2011: A new Dawn in environmental compliance and enforcement in Nigeria, *SSRN Electron. J.*, 2012, vol. 8, p. 116.
18. Gupta, N., Pandey, P., and Hussain, J., Effect of physicochemical and biological parameters on the quality of river water of Narmada, Madhya Pradesh, India, *Water Sci.*, 2017, vol. 31, no. 1, pp. 11–23.
19. Srivastava, G., Water quality index with missing parameters, *Int. J. Res. Eng. Technol.*, 2013, vol. 2, pp. 609–614.
20. He, Q., Dong, Z., Zhuang, F., Shang, T., and Shi, Z., Parallel decision tree with application to water quality data analysis, *Proc. Int. Symp. on Neural Networks*, Berlin: Springer, 2012, pp. 628–637.