

DIFFERENT PERSPECTIVES OF GROUND WATER QUALITY PREDICTION

D.SHRUTHI*, S.HARINATH

Department of Civil Engineering, REVA University, Bangalore, India

(Received 03 September, 2021; accepted 17 September, 2021)

Key words: Quality of water, Prediction, Machine learning technique.

ABSTRACT

Fresh water on the earth is unevenly distributed and is only about 1% of the total water available on the earth. But due to anthropogenic activities of man whole world is facing scarcity of quality water. Therefore knowledge about water, its availability, its quality, weather condition is much necessary. In order to handle enormous data volume, data mining techniques are essential. This paper presents perspectives of different authors for the prediction of water quality using different advanced techniques such as machine learning, artificial neural network, fuzzy logic etc.

INTRODUCTION

Water is one of the essential components necessary for all living beings on the planet earth. But now a days, due to rapid urbanization, industrialization, waste water runoff and many other anthropogenic activities of man leading to heavy contamination of water. Contaminated water is very dangerous for human beings, animals and plants too. When considered for the human beings it may cause problems from minor concerns like diseases related to skin to key ailments such as kidney stones, bone ailments and many more. Therefore it's much essential to assess the eminence of water before it reaches distribution system for the safety of public. But it is very difficult to analyze the water quality regularly since the process of analysis is much time consuming and tedious process. To overcome this issue one can make use of learning methods to predict quality of water before it reaches distribution system by making use of water quality standards provided by World Health Organization (WHO) and Indian Council for Medical Research (ICMR). If the water exceeds these limits one can take necessary action and can improve the quality of water. There is a 50% of contribution from the diseases related to water in patients who are hospitalized as per the announcement from United Nations Development Program (2006). Also, it leads to children death rate to 20% who are under 5-years of age (Nguyen, et al., 2018).

Water Quality Index (WQI) is a method based on mathematical model used to evaluate the quality of water based on chemical and physical characteristics.

It is also used to measure the quality of ground water as well surface water (Kiran, et al., 2019). WQI by thinking about physical, synthetic and organic boundaries, a few scientists utilized cation and anion boundaries to decide water quality record. Notwithstanding figure water quality record a few analysts took the assistance geographic data frameworks like GIS guides to precisely distinguish the dirtied stations, a few specialists surveyed the water quality as well as recognized the dirtying sources, the level of contamination sources evaluated that prompts defiled water, natural, climatic circumstances there by corrupting the nature of water. A few specialists accentuated on recognizing whether the water is nearer to mineral quality or not. A large portion of the examinations that were directed by the analysts depend on measurable and information mining procedures.

Methods of Water Quality Analysis

Quality of any kind of water is of utmost importance when it comes to its suitability or domestic purpose.as for as ground water is concerned it plays a vital role in many of the rural areas as well as urban areas for domestic purposes. Presently water is tested in laboratory by tedious and time consuming processes. Quality of ground water may depend upon precipitation, depth of its availability, geological properties, soil and sediment properties and permeability of the soil. In addition to this ground water quality is also affected by anthropogenic activities of man in a considerable manner. Water is tested in the laboratory is tested by traditional following methods and desirable limits for those parameters set by World Health Organization are shown in the Table 1.

Table 1. Textural classification. actual density (Dr). apparent density (Da) and porosity (P) of studied soils

PARAMETER	UNIT	LIMIT	METHOD
Aluminium	mg Al/l	0.2	AAS
Arsenic	mg As/l	0.05	AAS
Barium	mg Ba/l	0.05	AAS
Beryllium	ug Be/l	0.2	AAS
Cadmium	ug Cd/l	5.0	AAS
Calcium	mg Ca/l	200.0	AAS
Chromium	mg Cr/l	0.05	Colorimetric
Copper	mg Cu/l	1.0	AAS
Iron total	mg Fe/l	0.3	AAS
Lead	mg Pb/l	0.01	AAS
Magnesium	mg Mg/l	150.0	AAS
Manganese	mg Mn/l	0.1	AAS
Mercury	ug Hg/l	1.0	AAS
Selenium	mg Se/l	0.01	AAS
Sodium	mg Na/l	200.0	AAS
Zinc	mg Zn/l	5.0	AAS
Chlorides	mg Cl/l	250.0	Argentometry
Cyanide	mg Cn/l	0.1	Colorimetric
Fluorides	mg F/l	1.5	Colorimetric
Nitrates	mg NO ₃ /l	10.0	Colorimetric
Nitrites	mg NO ₂ /l	-	Colorimetric
Sulphates	mg SO ₄ /l	400.0	Turbidimetric
pH	-	9.2	Electrometric
Total dissolved solids	mg/l	1500	Gravimetric
Total hardness	mg/l	500	Titrometric
Alkalinity	mg/l	500	Titrometric
Total bacteria	Count/ml	100	Pour plate
Coliform	Count/100 ml	0	Filtration
E. Coli	Count/100 ml	0	Filtration
Salmonella	Count/100 ml	0	Filtration

Note: AAS-Atomic Absorption Spectrophotometer

The rest of this paper is structured as: Section III describes various learning methods used to assess the water quality and WQI. It explains the diverse learning methods used to assess the quality of water and concludes the paper with insightful information about water quality assessment/prediction and WQI.

LITERATURE REVIEW

This section describes various challenges posed while assessing water quality and WQI.

Kiran Relangi et al. addressed the issues related to water quality to understand, analyze, and the actions taken to overcome the challenges posed. Also, the authors

discussed about WQI (Kiran, et al., 2019).

Ashwini et al. performed water quality prediction using various techniques. Authors, also reviewed the various learning models and evaluation methods to describe the various categories of water quality (Ashwini, et al., 2019). Various learning methods like Naïve Bayes (NB), Artificial Neural Networks (ANN), Back propagation algorithm, and K Nearest Neighbor (KNN) has been deeply explored in this paper with their advantages and limitations.

Pinter et al. proposed a hybrid method getting to know method to expect the COVID-19 and that they tested capacity of hybrid gadget getting to know the usage of records from Hungary. The hybrid gadget getting to know strategies of Adaptive Network-Primarily based totally Fuzzy Inference System (ANFIS) and Multi-Layered Perceptron-Imperialist Aggressive Algorithm (MLP-ICA) is used to expect time collection of inflamed people and death rate. The fashions expect that through past due, possibly will the outbreak and the whole morality may drop gradually. The justification of the version is carried out up-to 9 days with desirable consequences, which endorses the version accuracy. It is anticipated that the version keeps its accuracy so long as no big interlude occurs. Based on the consequences mentioned, because of the complicated nature of the COVID-19 epidemic and variant in its conduct varies from one kingdom to some other kingdom. This paper presents a preliminary benchmarking to illustrate the capacity of gadget getting to know also can be used as a device in fitness studies context for destiny studies (Pinter, et al., 2020).

Sasikala had taken into consideration the water assets to be had within side the special place of trichy town. All assets are mapped within side the google map the use of Keyhole Mark-up Language (KML) Platform. Studies additionally offers with imparting a graphical vision for the supply of floor water assets of five towns. The groundwater float version for the look at town turned into formulated through the use of enter statistics, inclusive of the region of water assets and suitable boundary conditions. This mission wishes to gather statistics from numerous reasssets and examine the ones statistics with a few statistics mining equipment for prediction or selection making process. After accumulating numerous statistics, important project it to preserve statistics practice transformation and pre-processing of big statistics units for that statistics mining equipment is required. The look at concludes that the use of unsupervised learning, statistics with version may be anticipated at applicable water accuracy rate. Paper famous that ORANGE and WEKA Tool has excessive version as in comparison to the opposite equipment. pH has now no longer plenty version in statistics so, its miles solid in comparison to different and has a bit version for the duration of summer time season because the temperature impacts the water great for the duration

of summer time season (Sasikala, 2020).

Marwan Khan et al. provided a Runoff extent version which receives moisture of soil, irrigation depth, degree of crop and attention time as entry parameters. Some of the learning methods i.e. a couple of Linear Regression (LR), ANN, Decision Trees (DT) and Support Vector machine (SVR) were used for gaining knowledge of and prediction purposes. An evaluation has been made amongst those algorithms to pick high-quality set of rules for irrigation runoff extent prediction. Experimental effects display that a Decision tree offers suitable effects in phrases of maximum R-rectangular value, lowest Mean rectangular error. Whereas LR indicates the more serious bring about phrases of least R-rectangular value, maximum approach rectangular error (Khan, et al., 2019).

Nguyen et al. addresses the hassle with the aid of using growing a faster, replacement conceptual version primarily based totally at the precise reference fashions. The hydrodynamic statistics and water first-rate method equations from special precise fashions are taken into consideration with inside the evolved version. The version offers idea approximately rivers the use of waterfalls of tanks and piles the advection-diffusion and physico-biochemical processes. They examined the version with the aid of using evaluating its overall performance for the river Molse Neet, Belgium, with famous reference fashions, namely, MIKE eleven and Info Works RS. The outcome display that the theoretical version plays similarly nicely because the reference fashions, however with simulation time 104 instances faster. Successful checking out of this version opens a improvement street closer to hassle fixing within side the discipline of water first-rate manipulate and management (Nguyen, et al., 2019).

Bilali et al. makes use of the parameters which includes Magnesium Adsorption Ratio (MAR), Total Dissolved Solid (TDS), Residual Sodium Carbonate (RSC), Potential Salinity (PS), Exchangeable Sodium Percentage (ESP), Sodium Adsorption Ratio (SAR), and the parameters through Temperature (T), Electrical Conductivity (EC), and pH as inputs. They evolved and evaluated SVR, RF, Adaboost, and ANN approaches. To obtain the results they used 520 samples of information associated with 14 groundwater high-satisfactory features in berrechid aquifer, Morocco. The outcomes acquired indicates that the general prediction performances of adaboost and RF methods are better than the SVR and ANN. Anyhow, the generality capacity and sensitivity to the inputs showed that the ANN and SVR methods are greater generalizable and much less than adaboost and RF. The evolved techniques on this look at were discovered promising in low-price and real-time forecast of groundwater high-satisfactory through the usage of bodily parameters as enter variables (Bilali, et al., 2020).

Di Wu et al. suggested an outline to examine and expect water nice chance. In this work, the entire technique is split into 5 parts (Wu, et al., 2019).

- The uncooked facts are accumulated from the sensor networks and laboratory exams of water supply areas. It covers all of the applicable water nice signs. Here is really well worth to observe that the clustering and declustering strategies are optional.

- After the facts is prepared, we want to discover the important thing elements from more than one dimensions of signs through analysis of correlations, opportunity distribution and generate schooling and trying out facts units.

- The eventual goal of these paintings is to expect water nice chance. In order to discover the chance model, we've got inspected with investigators from water nice control.

- Furthermore, we ought to decluster the consequences and expect correct microorganism signs, each in tendency and values. These values can map to exclusive chance modes in keeping with sensible water supply control requirements in exclusive international locations.

- Future choice help in water remedy flowers can modify to each prediction and chance mode. Also, in practice, the methods want to be advanced with each area know-how facts units growing.

Najah et al. performed a studies with their important goal research is to adapt a algorithmically talented and sturdy approach for the assessment of water first-class features reducing the labour and value for size of these features. This makes a speciality of the Malaysian Johor River located in Johor State in which the water first-class dynamics are notably improved. This paper to begin with evaluated and assessed the correlation most of the features of water first-class on the premise of the experimental records the use of ANN after which to advocate numerous ANN approaches, such as Multi-Layer Perceptron, Neural Network and Radial Basis Function Neural Network which will affirm the effectiveness of those strategies within side the estimation of the features of water first-class. After they may be acquainted with the correctness of the ANFIS within side the prediction of the parameters of water first-class they increase an augmented Wavelet De-Noising approach with the Neuro-Fuzzy Inference System(WDT-ANFIS) to take a look at the effectiveness of the advised version for spatial function via way of means of offering extraordinary situations (Najah, et al., 2019).

Ragi et al. offers quick method to expect unknown parameters together with alkalinity, chloride, sulphate values the use of acknowledged parameters together with electrical conductivity, pH, and TDS. The use of Levenberg-Marquardt set of rules, which facilitates in in addition category of water our bodies for unique application. Curve becoming and sample category strategies have been used and that they discovered that curve becoming is the excellent predicting set of rules. Real time capabilities are vital for evaluation of ingesting water high-satisfactory. Different records set have been used for education; automated water high-satisfactory

category turned into made. New set of rules turned into evolved to optimize neural community weighing mechanism for water high-satisfactory dimension category. Results are as much as an accuracy of 87.9%, 83.94%, 79.48%, 81.736%, in predicting overall-hardness, chloride, overall alkalinity, sulphate, respectively. They proposed that the accuracy may be advanced via way of means of education with very massive quantity of data from unique locations. Increasing greater quantity of entry features for education reason outcomes in predicting even greater water high-satisfactory parameters associated with unique fields of application (Ragi, et al., 2019).

Ashwini et al. gives the in your price range method to keep away from infection of water in residential overhead tanks. The nice of water is monitored the usage of IoT gadgets and the destiny prediction of water infection is completed the usage of gadget mastering algorithms. They proposed the device which includes multi sensors linked to NodeMCU to accumulate the water parameters, and the alert message is dispatched to the person earlier than the water receives contaminated. The device allows to store the water from infection and is likewise value effective. The device sends the alert message to person whilst any of the parameters are decrease than the usual values. This allows the person to recognise approximately the infection of water earlier than it reaches person of their residential tanks. This approach cannot handiest be restricted to residential tanks however also can be utilized in water remedy vegetation and industries. The destiny scope for this venture may be prolonged to hit upon the illnesses as a result of distinct parameters and locating the right answer for to smooth the tank. Also biosensors may be used to hit upon the mycobacteria's which might be dangerous for people and to enhance nice of water (Ashwini, et al., 2019).

Sahoo proposed an automatic Hybrid Artificial Neural Network (HANN) version with a brand fresh enter facts handling technique for simulating groundwater stage extrade. They applied, evaluated, and mentioned the version functionality for predicting recurrent groundwater stage extrade at 1124 wells in essential agricultural areas of America. In each area, pumping has produced groundwater stage debilities throughout huge regions and is a developing difficulty for water aid managers. They additionally adopted to check the HANN version on those well-studied aquifers to permit for version outcome assessment. Finally, the HANN version can be maximum beneficial in areas missing acknowledged hydrogeological subsurface parameters. This version makes use of precipitation, temperature, move goes with the drift and weather information as entry features. Adding to this, irrigation call for simulated with the DSSAT version is utilized instead of unreachable groundwater pumping records. They evaluated the overall performance of each our synthetic ANN and traditional regression fashions in specific alignments, one the usage of uncooked enter facts and the

later one the usage of enter facts converted through SSA, mutual information, genetic algorithm, and maximum-correlated lag. The HANN version is outperformed Hybrid Multiple Linear Regression (HMLR) and Hybrid Multiple Nonlinear Regression (HMNLR) methods (Sahoo, 1969).

Bisht et al. advanced River Water Quality Forecasting Model (RWQFM) the usage of the idea of synthetic ANN for the river Ganga. In this work, an attempt were made for growing version for the move Ganga within side the stretch from Devprayag to Roorkee, via way of means of deciding on 5 checking out stations alongside the waterway. Exploratory dataset from one month to any other month for the time association of 2001 to 2015 which includes 4 water high-satisfactory parameters became taken as enter records. Using gifted gadget studying method referred to as ANN an ideal version is advanced via way of means of engaging in numerous experiments in Weka records mining device. The high-satisfactory of water is forecasted for subsequent three hundred and sixty five days and the forecasting accuracy is decided the usage of numerous overall performance measures. The computation of 12-steps in advance WQ indicated that the water comes out to be appropriate for ingesting at some stage in the 12 months 2016 most effective at three places Roorkee, Devprayag, and Rishikesh. The outcomes confirmed that the suggested WQ version is greater green in phrases of the forecasting accuracy. At Rishikesh, the advanced prediction version executed an excellently by achieving accuracy of 100%. Therefore, the prediction of ANN version is appeared as a powerful version similarly to this ANN has demonstrated its importance as a green device within side the forecasting domain. Such fashions will without a doubt be beneficial for the water control our bodies in an effort to manipulate the river pollutants and may totally assist the general public as well (Bisht, et al., 2019).

Some Learning Techniques Used for Prediction of Quality of Water

Artificial neural network (ANN): ANN model is based on deep learning method used for classification which is made of interrelated nodes. This model uses the concept of feed forward, i.e., the output of neurons in one layer will be used as input in the next layer (Bisht, et al., 2019). These models will helps to reveal the interrelationships related to classical information which are not exposed. Hence, predicting and modelling of water quality plays a vital role to control the water pollution.

Naïve bayes (NB): NB is a supervised learning method based on bayes hypothesis with the postulation of naïve i.e., restrictive liberation among features. This model is used for classification of water quality assessment (Palani, et al., 2008). It works effectively for large datasets. NB presumed that the value of a discrete feature does not depend on the absence or presence of other features. This method works as follows: 1. Data cleaning and extraction.

2. Removes the outsized punctuations. 3. Calculate the probability based on tokens. 4. Adds the probabilities to predict the water quality (Chaunqi, et al., 2009).

Decision tree (DT): Decision tree (DT) is a particular type of tree which uses the concept of probability that enables to make a proper decision. In this method the larger dataset is will be divided into various partitions with smaller size by demonstrating relationship between the child and parent. The nodes which are inside will be referred as inner node and will be labelled with feature as input (Lu, et al., 2009). These models can be used to solve categorical and numerical data, also these are suitable for datasets with large size. These models are capable of optimizing various types of data like numeric and textual. DT is an effective learning method

CONCLUSION

In this paper, an attempt has been made to review previous existing methods for analysis of quality of water and also some learning techniques used in prediction of quality of water. Some researchers used machine learning techniques, some used ANN. Some researchers got good accuracy in Hybrid Artificial Neural Network (HANN) also. These techniques can be used by the management bodies to detect contamination and to alarm the respective concerned authorities for taking safety measures. Accuracy of the techniques can be improved by taking large number of datasets by giving more water quality parameters as inputs. Some of the techniques such as ANN, decision tree, and naïve bayes are also discussed along with their salient features.

REFERENCES

- Nguyen TT, Keupers I and Willems P. 2018. Conceptual river water quality model with flexible model structure. *Environ Model Softw.* 104:102-117.
- Kiran RNDSS, Chaparala A and Sajja R. 2019. Perspectives in water quality assessment. *Int J Recent Technol Eng.* 8:7-11.
- Ashwini K, Vedha JJ and Priya MD. 2019. Intelligent model for predicting water quality. *Int J Adv Res Ideas Innov Technol.* 5:70-75.
- Pinter G, Felde I, Mosavi A, Ghamisi P and Gloaguen R. 2020. COVID-19 pandemic prediction for Hungary: A hybrid machine learning approach. *Mathematics* . 8:890.
- Sasikala R. 2017. A comparative analysis for smart water resource using data mining tools. *Int J Res Granthaalayah.* 5:24-30.
- Khan M and Noor S. 2019. Performance analysis of regression: Machine learning algorithms for predication of runoff time. *Agrotechnology.* 8:1-12.
- Bilali AE, Taleb A and Brouziyne Y. 2020. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric Water Manag.* 245:106625.
- Wu D, Wang H, Mohammed H and Seidu R. 2019. Quality risk analysis for sustainable smart water supply using data perception. *IEEE.* 5:377-388.
- Najah AA, Othman FB, Afan HA, Ibrahim RK, Fai CM, Hossain MS, Ehteram M and Elshafie A. 2019. Machine learning methods for better water quality prediction. *J Hydrol.* 578:124084.
- Ragi NM, Holla R, and Manju G. 2019. Predicting water quality parameters using machine learning. *RTEICT.* 17:1109-1112.
- Ashwini C, Singh UP, Pawar E and Shristi. 2019. Water quality monitoring using machine learning and Iot. *Int J Sci Technol Res.* 8:1046-1048.
- Sahoo S. 1969. Water Resources Research. *J Am Water Resour Assoc.* 5:2-2.
- Bisht AK, Singh R, Bhutiani R and Bhatt A. 2019. Artificial neural network based water quality forecasting model for ganga river. *IJEAT.* 8:2249-8958.
- Palani S, Liong S, Tkalich P. 2008. An ANN application for water quality forecasting. *Mar Pollut Bull.* 56:1586-1597.
- Chaunqi L and Wei W. 2009. Assessment of the water quality near the dam area of three gorges reservoir based on bayes. *IEEE.* 26:26-28.
- Lu J and Huang T. 2009. Data mining on forecast raw water quality from online monitoring station based on decision-making tree. *IEEE.*