Water Quality Prediction using Artificial Intelligence and Machine learning Algorithms

Jitendra Pandey[#1], Seema Verma[*2]
Banasthali Vidyapeeth, India
jitenndra@gmail.com[#1], seemaverma3@yahoo.com[#2]

Article Info
Page Number: 6114 - 6132
Publication Issue:
Vol 71 No. 4 (2022)

Article History
***Article Received**: 25 March 2022*
***Revised**: 30 April 2022*
***Accepted**: 15 June 2022*
***Publication**: 19 August 2022*

Abstract

Artificial intelligence approaches may significantly lower the cost of water supply and sanitation systems while also ensuring compliance with drinking water and wastewater treatment standards. As a result, there has been a lot of study on modeling and predicting water quality in order to reduce water pollution. The suggested system's innovation is offered in order to establish an effective monitoring system for drinking water in order to maintain a sustainable and environmentally friendly green environment. To estimate the water quality index, the adaptive neuro-fuzzy inference system (ANFIS) method was created in this study (WQI). Water quality was classified using a feed-forward neural network (FFNN) and K-nearest neighbors. Although the dataset contains eight important parameters, only seven were found to have significant values. These statistical factors were used to create the suggested approach. The ANFIS model outperformed the others in terms of predicting WQI values, according to the data. Despite this, the FFNN algorithm has the greatest accuracy (100%) in classifying water quality (WQC). In addition, the ANFIS model correctly predicted WQI, whereas the FFNN model was more robust in identifying the WQC. Furthermore, the ANFIS model performed well in testing, with a regression coefficient of 96.17 percent for predicting WQI, while the FFNN model had the best accuracy (100 percent) for WQC. Water treatment and management might be aided by this suggested technology, which employs superior artificial intelligence.

Keywords: -Adaptive neuro-fuzzy inference system; feed-forward neural network models; water quality index; water quality classification.

Background

Water contamination has worsened as a result of rapid economic expansion and rising urbanization. Understanding the concerns and trends of water quality is also important for reducing and regulating water pollution. To really grasp the condition of the marine ecosystem, most governments throughout the world have begun to build environmental water management programs. Water is the most vital component of life. Although water covers 71 percent of the Earth's surface, the great majority (95 percent) is salt water [1]. As a result, it is critical to preserve the quality of fresh water. Almost one billion people lack access to safe drinking water, and two million people die each year as a result of polluted water, inadequate sanitation, and hygiene [2].

The long-term viability of a diversion plan is dependent on the quality of the water. Predicting water quality entails predicting variation patterns in a water system's quality at a specific period. Predicting water quality is crucial for water preparation and management. Predicting future changes in water safety at various degrees of contamination and establishing logical tactics to avoid and control water contamination may be used to build water contamination prevention and regulation strategies. The overall consistency of water should be assessed in water diversion plans. To address everyday drinking difficulties, a considerable volume of water is carried. As a result, solutions for anticipating water quality in today's civilization should be researched [3].

Poor-quality water may also be costly, because resources must be shifted to repair water delivery infrastructure whenever an issue emerges. The demand for enhanced water treatment and water quality management has been rising for these objectives in order to assure safe drinking water at competitive prices. To address these issues, systematic assessments of raw water, disposal systems, and organizational monitoring issues are necessary [4]. Accurate projections of changes in water quality can greatly increase aquaculture's efficiency. Water quality data is pre-processed in most cases before water quality metrics are anticipated. As a result, there are two stages in this segment. The first stage entails pre-processing water quality data and conducting a correlation study between various water quality metrics.

Water quality modeling has been created utilizing modern computer and artificial intelligence (AI) approaches to handle water quality challenges. By forecasting changes in water quality, artificial neural networks (ANNs) have benefited in the monitoring of water quality systems [5].

They have the potential to greatly increase aquaculture efficiency. The application of the hydrodynamic and water quality model, a relatively new computational technique, presents obstacles and challenges in simulating water quality situations. ANNs are widely used in a variety of fields and give an alternative method for analyzing and monitoring water quality in reservoirs. ANNs have been used to simulate and forecast water quality in bodies of water with great success. ANNs have been employed in a variety of applications, including feed-forward neural networks [6]. To tackle complicated nonlinear systems, the fuzzy logic system was created [7]. The use of ANN applications to calculate and forecast the quality of water bodies has proven effective [8–12]. For designing predictions, ANN models require parameter values [13]. ANNs offer a number of benefits, including the capacity to learn, manage very complicated nonlinear systems, and work in parallel. Shafi et al. [14] utilized data from the Pakistan Council of Research in Water Resources (PCRWR) to identify water quality using support-vector machines, neural networks (NNs), deep NNs, and K-nearest neighbors (KNNs).

To predict total nitrogen, total phosphorous, and total organic carbon in water, researchers used a convolutional neural network (CNN)-long shortterm memory (LSTM) hybrid deep learning model. The long short-term memory (LSTM) network model was used by Liu et al. [16] to estimate the quality of drinking water in the Yangtze River Basin [15]. pH, dissolved oxygen (DO), chemical oxygen demand (COD), and NH3-N were used to create the LSTM model. The LSTM model is said to be promising for monitoring water quality.

Artificial intelligence was proposed by Chen et al. [17] for modeling and predicting water quality. It should be noticed that the ANN model performed better. Singh [18] employed the ANN model to estimate the quality of river water by computing dissolved oxygen (DO) and biological oxygen (BOD) characteristics. To estimate sewage effluent water quality, Zheng et al. [19] used the immune practical swarm optimization (PSO) approach, which used a neural network with a hidden layer. Using the grey correlation analysis approach, Gao [20] improved the back-propagation (BP) neural network to forecast water quality. Zhang et al. [21] used ANNs and a genetic algorithm to estimate water quality utilizing temporal data to improve the forecasting findings' stability. In the south-to-north water diversion (SNWD) Project, Wang et al. [22] suggested a Genetic Regression Neural Network GA-GRNN model to establish an effective technique of forecasting water quality to assure water security. The association between relevant factors was investigated using correlation coefficients. To predict COD and bioche, Abyaneh [23] used ANNs and regression models. The ANN model's kernel function was the radial-basis-function [24,25]. In Small Prespa Lake in Greece, Barzegar et al. [26] constructed a hybrid convolutional neural network (CNN)–LSTM model to predict DO and chlorophyll-a (Chl-a). When compared to the classic support vector regression (SVR) model, the deep learning model outperformed it. Using the ANN model, Maiti et al. [27] estimated dissolved oxygen (DO) values.

Deep learning approaches outperformed classical machine learning techniques in predicting WQI, as did AI techniques like ANN, Bayesian NNs, and adaptive neuro-fuzzy [28]. Piazza et al. [29] compared the numerical optimization technique of the proposed model to the findings of an experimental campaign. The evolutionary algorithm was used in conjunction with a hydraulic simulator to test and assess water quality through monitoring. For anticipating wastewater, Sambito et al. [30] created a smart system based on the Internet of Things and a Bayesian decision network (BDN). The suggested approach was utilized to estimate groundwater water quality WQ using analyses of soluble conservative contaminants such as metals, decision support systems, and auto-regressive moving averages [31].

Currently, water quality is determined through costly and time-consuming laboratory and statistical analyses, which necessitate sample collection, transportation to laboratories, and a significant amount of time and calculation. This approach is ineffective because water is a completely transmissible medium, and time is required if the water is contaminated with disease-causing waste. The disastrous implications of water pollution need a more expedient and cost-effective solution. In this context, we created a real-time system to test an alternate strategy for modeling and predicting water quality based on an advanced artificial intelligence algorithm. These copying models, on the other hand, confront considerable difficulties. They don't take into account elements that alter WQ, for example. The current study's contributions include the deployment of an advanced AI Adaptive neural-fuzzy inference system ANFIS model to forecast the Water Quality Index WQI. For the Water Quality Classification WQC, the feed-forward neural network FFNN and the kernel neural network KNN were utilized. Advanced AI may be generalized and utilized to foresee the water contamination process, assisting decision-makers in planning for timely actions.

Methodology:

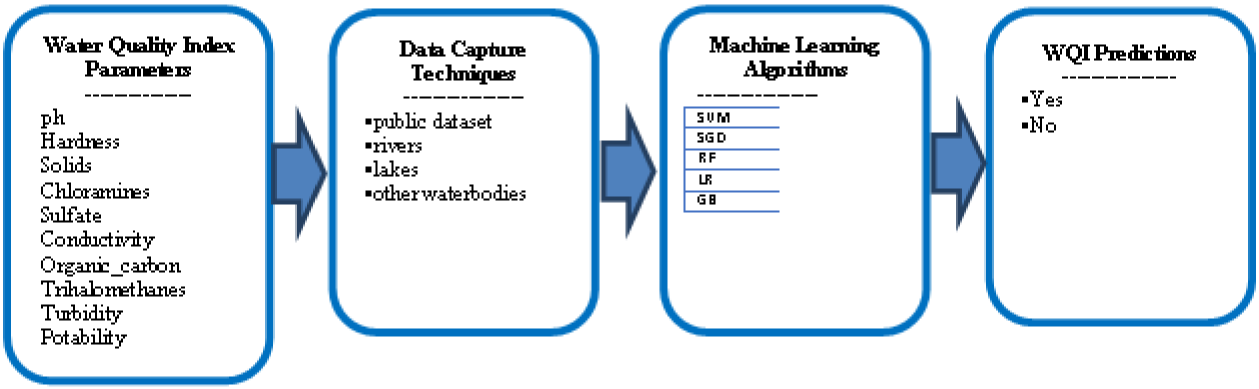


Figure 1 Methodology Used

Algorithms

SVM

The "Support Vector Machine" (SVM) is a supervised machine learning technique that can solve classification and regression problems. It is, however, mostly employed to solve categorization difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space into classes so that additional data points may be readily placed in the proper category in the future. A hyperplane is the name for the optimal choice boundary.

The extreme points/vectors that assist create the hyperplane are chosen via SVM. Support vectors are the extreme instances, and the method is called a Support Vector Machine. Consider the Figure 2 below, which shows how a decision boundary or hyperplane is used to classify two separate categories:

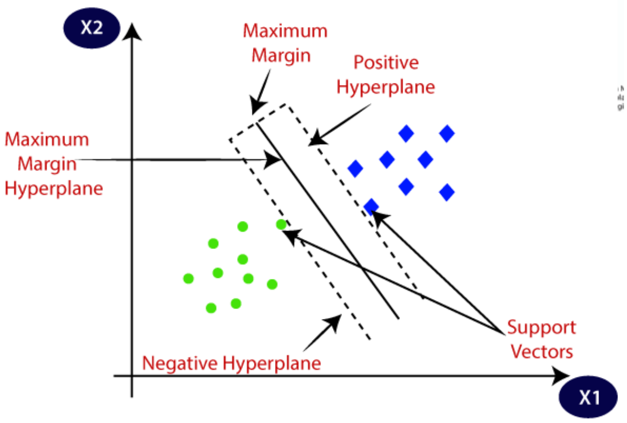


Figure 2 SVM

Simply put, support vectors are the coordinates of each unique observation. The SVM classifier is a frontier that separates the two classes (hyper-plane/line) the most effectively.

SGD

Stochastic gradient descent is a well-known and widely used technique in many Machine Learning algorithms, and it is the foundation of Neural Networks.

Stochastic Gradient Descent (SGD) is a quick and easy way to fit linear classifiers and regressors to convex loss functions like (linear) Support Vector Machines and Logistic Regression. Despite the fact that SGD has been present for a long time in the machine learning field, it has only lately gotten a lot of attention in the context of large-scale learning.

SGD has been used to solve large-scale, sparse machine learning issues that are common in text categorization and natural language processing. Because the data is sparse, the classifiers in this module may readily scale to problems with more than 105 training samples and characteristics.

The class SGDClassifier provides a simple stochastic gradient descent learning method that supports a variety of classification loss functions and penalties. A SGDClassifier trained with the hinge loss, which is identical to a linear SVM, has the decision boundary shown below.

SGD, like other classifiers, requires two arrays: an array X of form (n samples, n features) that has the training samples, and an array y of shape (n samples, n class labels) that holds the target values (class labels) for the training data.

Random Forest

Random Forest is a well-known machine learning algorithm that uses the supervised learning method. In machine learning, it may be utilized for both classification and regression issues. It is based on ensemble learning, which is a method of integrating several classifiers to solve a complicated issue and increase the model's performance.

"Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset," according to the name. Instead than depending on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions.

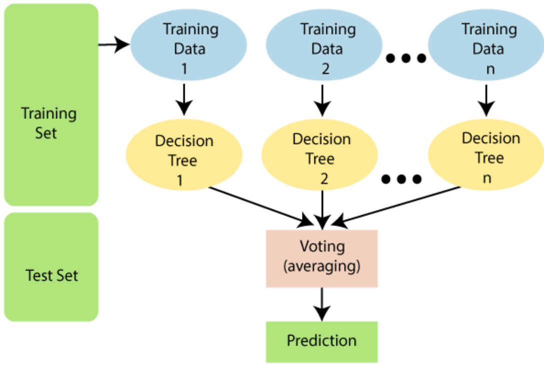


Figure 3Random Forest Classifier

Random Forest Assumptions

Because the random forest mixes numerous trees to forecast the dataset's class, some decision trees may correctly predict the output while others may not. However, when all of the trees are combined, the proper result is predicted. As a result, two assumptions for a better Random forest classifier are as follows:

The dataset's feature variable should have some real values so that the classifier can predict correct outcomes rather than guesses.

- Each tree's predictions must have very low correlations.
- Why would you want to utilize Random Forest?

The following are some reasons why we should utilize the Random Forest algorithm:

- When compared to other algorithms, it takes less time to train.
- It predicts output with good accuracy, and it runs quickly even with a huge dataset.
- When a considerable amount of the data is missing, it can still retain accuracy.

Linear Regression

Linear regression is a basic and quiet statistical regression approach that demonstrates the relationship between continuous variables and is used for predictive analysis. Linear regression, as the name implies, depicts a linear connection between the independent variable (X-axis) and the dependent variable (Y-axis). Simple linear regression is defined as linear regression with only one input variable (x). When there are several input variables, this type of linear regression is referred to as multiple linear regression. A slanted straight line representing the connection between the variables is produced by the linear regression model.

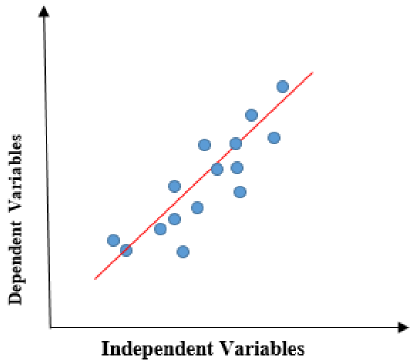


Figure 4Linear regression

$$y = mx+b \implies y = a_0+a_1x$$

y= Dependent Variable.

x= Independent Variable.

a0= intercept of the line.

a1 = Linear regression coefficient.

Linear regression, as previously stated, calculates the connection between a dependent and an independent variable.

Gradient Boosting

Gradient boosting is a strategy that stands out for its predictability and speed, especially when dealing with big and complicated datasets. This algorithm has provided the greatest results in everything from Kaggle contests to machine learning solutions for businesses. We already know that in any machine learning algorithm, mistakes play a significant role. Bias error and variance error are the two most common forms of mistake. The gradient boost approach assists us in reducing the model's bias inaccuracy.

You may have come across the word Boosting when studying machine learning. In the subject of Data Science, it is the most misunderstood word. Boosting algorithms work on the idea of first building a model on the training dataset, then building a second model to correct the faults in the first model.

Boosting is a technique for turning weak students into strong ones. Each new tree in boosting is based on a modified version of the original data set. The gradient boosting algorithm (gbm) is best described by first learning about the AdaBoost Algorithm. The AdaBoost Algorithm starts by training a decision tree with equal weights for each observation. Following the evaluation of the first tree, we raise the weights of the difficult-to-classify data and decrease the weights of the easy-to-classify observations. As a result, the second tree is based on the weighted data. The goal here is to improve on the first tree's forecasts. As a result, our new model is Tree 1 Plus Tree 2. The classification error from this new 2-tree ensemble model is then computed, and a third tree is grown to forecast the updated residuals. This method is repeated for a certain number of iterations. Following trees aid in the classification of observations that were not properly categorised by preceding trees. The weighted total of the predictions provided by the preceding tree models makes up the final ensemble model's predictions.

Dataset:

As outlined by the WWF, only 3% of available water is freshwater. The other 97% relates to salt water which cannot be consumed by humans or animals. With a key resource in limited supply, are there certain areas of the world that have a larger water table supply that can be used by all? The water_potability.csv file contains water quality metrics for 3276 different water bodies. A link to the dataset is given for reference. Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes and Turbidity are all included in the datasets. Seven factors, however, were judged to have significant values, and the resulting models were assessed using various statistical metrics. Temperature settings were used in all of the trials. This data came from Kaggle (https://www.kaggle.com/adityakadiwal/water-potability) (accessed on 3 January 2022).

Finding the missing values for the data cleansing and preparedness.

Parameter	Missing
Hardness	0
Solids	0
Chloramines	0

Sulfate	781
Conductivity	0
Organic_carbon	0
Trihalomethanes	162
Turbidity	0
Potability	0

Table 1- Visualizing the Nullity Matrix(missing values).

Data Processing

In order to increase data quality, the processing step is critical in data analysis. WQI was estimated in this step using the dataset's most critical parameters. The WQI values were then used to classify the water samples. For better accuracy, the z-score approach was utilized as a data normalization tool.

WQI calculation

Water quality was measured using the WQI, which is derived utilizing many characteristics that impact WQ [32]. The suggested system's performance was assessed using a published dataset that included seven significant water quality metrics. The following formula was used to determine the WQI:

$$WQI = \frac{\sum_{i=1}^{N} q_i \times w_i}{\sum_{i=1}^{N} w_i} \quad (1)$$

where N is the total number of parameters in the WQI formula, qi is the quality estimate scale for each parameter I computed by Formula (2), and wi is the unit weight of each parameter calculated by Formula (3)

$$q_i = 100 \times \left(\frac{V_i - V_{Ideal}}{S_i - V_{Ideal}}\right)$$

As stated in Table 1, Vi is a measured value that pertains to the water samples tested, VIdeal is an ideal value that shows clean water (0 for all parameters except OD = 14.6 mg/L and pH = 7.0), and Si is a suggested standard value for parameter i.

Wi= K/Si

where K is the proportionality constant, which is derived using the formula:

$$K = \frac{1}{\sum_{i=1}^{N} 1/S_i}$$

Parameters	Permissible Limits
Hardness (mg/L)	100
pH	8.5
Conductivity, µS/cm	1000
Solids	500
Chloramines (mg/L)	250
Sulfate (mg/L)	250
Organic_carbon	< 10
Turbidity (NTU)	5
Trihalomethanes	14.36

Table 2. Permissible limits of the parameters used in calculating the WQI [33].

Table 3.

WQI Scale	Category
0–20	C5
21–40	C4
41–60	C3
61–80	C2
81-100	C1

Table 4.

Constraint	Unit Weight (w_i)
Dissolved Oxygen	0.3723
pH	0.219

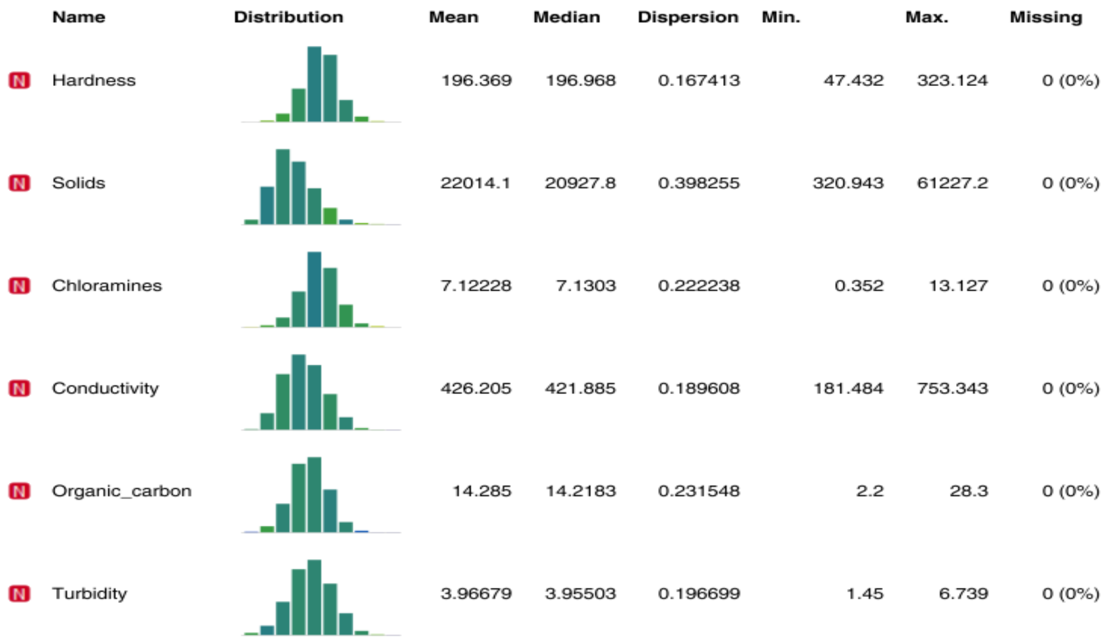
Conductivity	0.371
Biological Oxygen Demand	0.3723
Nitrate	0.0412
Fecal Coliform	0.0221
Total Coliform	0.0022

Tables 3 and 4 represent the parameters of the WQC and unit weight, respectively.

More parameters, including our selection parameters, may be calculated using WQI. The WQI is based on variable data. Any parameter may be tested with any water quality data using the suggested approach.

Discussion

In lab analysis, modeling and prediction of water quality have played a critical and substantial role in reducing time and usage. Artificial intelligence algorithms were investigated as a potential replacement to traditional methods for estimating and forecasting water quality. This study employed experimental data from 3276 samples collected from 3276 distinct rivers and lakes. nine key metrics are included in the dataset. Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes and Turbidity.



Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
Hardness		196.369	196.968	0.167413	47.432	323.124	0 (0%)
Solids		22014.1	20927.8	0.398255	320.943	61227.2	0 (0%)
Chloramines		7.12228	7.1303	0.222238	0.352	13.127	0 (0%)
Conductivity		426.205	421.885	0.189608	181.484	753.343	0 (0%)
Organic_carbon		14.285	14.2183	0.231548	2.2	28.3	0 (0%)
Turbidity		3.96679	3.95503	0.196699	1.45	6.739	0 (0%)

Trihalomethanes	66.3963	66.6225	0.243574	0.738	124	162 (4%)
ph	7.08079	7.03675	0.225121	0.00	14	491 (14%)
Sulfate	333.776	333.074	0.124061	129	481.031	781 (23%)

Linear Regression	0.390031	0.390134	0.0808367	0.278233	0.516533	0 (0%)
Gradient Boosting	0.389945	0.36865	0.351797	-0.012543	1.23716	0 (0%)
SVM	0.325304	0.320632	0.556229	-0.249383	0.935809	0 (0%)
Random Forest	0.396443	0.387083	0.53812	0.00	1	0 (0%)
SGD	0.40434	0.40318	0.274571	-0.000409478	0.925355	0 (0%)

The ultimate purpose of this research is to support and align with Sustainable Development Goal (SDG) 6, which aims to guarantee that everyone has access to safe drinking water. The created model may be used to forecast water quality and index, and hence water quality categorization, with high accuracy, in a simple and low-cost manner. Furthermore, this type of model is reliable and can predict water contamination, allowing authorized governments/agencies to develop effective strategies for improved water sustainability and management, such as removing the contamination source and/or seeking an alternative source of pure water to meet community demand.

Models Compared by MSE

Model	SVM	SGD	RF	LR	GB
SVM		0.844	0.999	0.968	0.997

SGD	0.156		0.945	0.862	0.992
RF	0.001	0.055		0.031	0.927
LR	0.032	0.138	0.969		0.998
GB	0.003	0.008	0.073	0.002	

Result Analysis

Model	MSE	RMSE	MAE	R2
SVM	0.26	0.51	0.449	-0.092
SGD	0.246	0.496	0.474	-0.032
RF	0.228	0.478	0.422	0.04
LR	0.239	0.489	0.476	-0.005
GB	0.219	0.468	0.439	0.078

As we can see that results from various algorithms compared in the study are different. With the given dataset and choice of algorithms, it is clearly visible that SVM out performs all other algorithms with the highest MSE of 0.260.

R 2 = 1 − sum squared regression (SSR) total sum of squares (SST) , = 1 − ∑ (y i − y i ^) 2 ∑ (y i − y ¯) 2

Whereas R2 score for the given dataset is not valid as it is negligent.

Conclusion

The use of AI algorithms to model and predict water quality is critical for environmental preservation. Using data from rivers collected from various sites across India, artificial intelligence algorithms were constructed to forecast and categorize water quality for drinking. nine essential parameters were calculated using WQI: Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes and Turbidity. These were regarded as important water quality indicators. Using powerful AI ANFIS algorithms to develop new techniques can assist ensure a secure environment. Advanced ANFIS algorithms were employed to forecast WQI in this suggested technique. The WQC data was classified using the FFNN method. Statistics were used to analyze and test the suggested technique. Using sophisticated AI, the following conclusions may be derived.

First, the current study looked into an alternative artificial intelligence technique for predicting water quality using minimum and readily available water quality criteria. The datasets used in the study were collected from 3276 different water bodies distinct rivers and lakes. WQI was predicted and classified using artificial intelligence techniques.

Second, by choosing essential factors from a standard dataset, an advanced AI ANFIS model may be created to forecast WQI. The forecast values were strikingly similar to the observation values.

Third, for WQC, machine learning algorithms such as FFNN and KNN can be constructed. In WQC, the FFNN outperformed the KNN. The FFNN method outperformed the KNN approach in classification.

References

[1] Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines. Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), 2017. Licence: CC BY-NC-SA 3.0 IGO.

[2] H. Vicente, et al., Int. J. of Design & Nature and Ecodynamics. Vol. 7, No. 3 (2012)

[3] Sakizadeh, M,(2015), Artificial intelligence for the prediction of water quality index in groundwater systems. Model. Earth Syst. Environ. (2016) 2:8. DOI 10.1007/s40808-015-0063-9

[4] Najah, Al-Mahfoodh et. al (2009). Prediction of Johor River Water Quality Parameters Using Artificial Neural Networks. European Journal of Scientific Research. 28.

[5] H. Vicente et al. | Prediction of public water supply quality using artificial neural networks Journal of Water Supply: Research and Technology—AQUA | 61.7 | 2012

[6] J. Earth Syst. Sci., DOI 10.1007/s12040-016-0708-8, 125, No. 5, July 2016, pp. 985–995

[7] Heddam, Salim. (2016). Multilayer perceptron neural network-based approach for modeling phycocyanin pigment concentrations: case study from lower Charles River buoy, USA. Environmental Science and Pollution Research. 23. 10.1007/s11356-016-6905-9.

[8] Emamgholizadeh, Samad et. al (2013). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. International Journal of Environmental Science and Technology. 11. 645-656. 10.1007/s13762-013-0378-x.

[9] Samy I. El-Kowrany et al. Water pollution in the Middle Nile Delta, Egypt: An environmental study, Journal of Advanced Research, Volume 7, Issue 5, 2016, Pages 781-794, ISSN 2090-1232,

[10] Gakii, C (2019), A CLASSIFICATION MODEL FOR WATER QUALITY ANALYSIS USING DECISION TREE. European Journal of Computer Science and Information Technology Vol.7, No.3, pp.1-8, June 2019

[11] Elliott, Grace. (2014). Dissertation: An investigation into the water quality and treatment habits within households in Kanungu, Western Uganda.

[12] Liu, P. et. al (2019). Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. Sustainability (Switzerland). 11. 10.3390/su1102058.

[13] Heddam, S., Sanikhani, H., and Kisi, O., "Application of artificial intelligence to estimate phycocyanin pigment concentration using water quality data: a comparative study", <i>Applied Water Science</i>, vol. 9, no. 7, 2019. doi:10.1007/s13201-019-1044-3.

[14] Ragan, Nahad& Yahia, Hussin. (2018). ASSESSMENT OF DRINKING WATER QUALITY AND ITS IMPACT ON RESIDENTS HEALTH IN MUSCAT CITY, OMAN.

[15] Ahmed, Umair et. al. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. Water. 11. 2210. 10.3390/w11112210.

[16] Muharemi, F et. al (2019) Machine learning approaches for anomaly detection of water quality on a real-world data set, Journal of Information and Telecommunication, 3:3, 294-307, DOI: 10.1080/24751839.2019.1565653

[17] Khan, Yafra& Chai, Soo See. (2016). Predicting and analyzing water quality using Machine Learning: A comprehensive model. 1-6. 10.1109/LISAT.2016.7494106.

[18] Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities. New York: United Nations Children's Fund (UNICEF) and World Health Organization, 2019.

[19] Gajendran, C. (2013). Water quality assessment and prediction modelling of Nambiyar river basin, Tamil Nadu, India.

[20] Verma, Seema & Chaudhary, Prachi. (2012). Wireless Sensor Network application for water quality monitoring in India. 2012 National Conference on Computing and Communication Systems, NCCCS 2012 - Proceeding. 1-5. 10.1109/NCCCS.2012.6412990.

[21] Ali Najah et. al Machine learning methods for better water quality prediction, Journal of Hydrology, Volume 578, 2019, 124084, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2019.124084.

[22] Yuan, Fei et al. (2018). A Biological Sensor System Using Computer Vision for Water Quality Monitoring. IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2876336.

[23] Mehmood, H et. al (2020). Strategic Foresight to Applications of Artificial Intelligence to Achieve Water-related Sustainable Development Goals. UNU-INWEH Report Series. United Nations University Institute for Water, Environment and Health (UNU-INWEH).

[24] Chen, Zhengfa&Guifeng, Liu. (2010). Application of Artificial Intelligence Technology in Water Resouces Planning of River Basin. 322 - 325. 10.1109/ISME.2010.267.

[25] Salah, Hussein &Mocanu, Mariana &Florea, Adina Magda. (2019). Development of a Prototype ArcGIS-Web-Based Decision Application WATERDSS: for Water Pollution Management.

[26] Lal, Alvin & Datta, Bithin. (2018). Genetic Programming and Gaussian Process Regression Models for Groundwater Salinity Prediction: Machine Learning for Sustainable Water Resources Management. 10.1109/SusTech.2018.8671343.

[27] Bejarano, Gissella& Jain, Mayank & Ramesh, Arti &Seetharam, Anand & Mishra, Aditya. (2018). Predictive Analytics for Smart Water Management in Developing Regions. 464-469. 10.1109/SMARTCOMP.2018.00047.

[28] Xiaoping, Wang & Zhang, Fei & Ding, Jianli. (2017). Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China. Scientific Reports. 7. 10.1038/s41598-017-12853-y.

[29] Wang, Yuanyuan & Zhou, Jian & Chen, Kejia& Wang, Yunyun& Liu, Linfeng. (2017). Water quality prediction method based on LSTM neural network. 1-5. 10.1109/ISKE.2017.8258814.

[30] Chin Yik Lin, Mohd. Harun Abdullah, Baba Musta, Ahmad Zaharin Aris&Sarva Mangala Praveena. 2010. Assessment of Selected Chemical and Microbial Parameters in Groundwater of PulauTiga, Sabah, Malaysia Sains Malaysiana 39(3): 337–345

[31] Kamble.P.N. (2008) Physico-Chemical Characteristics of water from Khadakwasala Reservoir, Pune, Maharashtra State,Int. J. Chem. Sci, 6(1), 325-332.

[32] S.R. Warhate, (2012), The evaluation of water quality index around WCL Kolera-Pimpri coal mines,Sci. Revs. Chem. Comm, 2(3), 197-200.

[33] Sunanda S. A, (2012), Drinking water quality of coal mine surrounded lime industry area, Sci. Revs. Chem. Comm.: 2(3), 188-191.

[34] D. P. Gupta, Sunita & J. P. Saharan. 2009. Physiochemical Analysis of Ground Water of Selected Area of Kaithal City (Haryana) India Researcher, 1(2),

[35] Ajit M. Kalwale1, Padmakar A. Savale. 2012. Determination of Physico-Chemical Parameters of DeoliBhorus Dam water. Advances in Applied Science Research, 3(1):273-279

[36] Hemant Pathak. 2012. Assessment of physico-chemical quality of municipal water samples of Makronia sub-urban area of Bundel khand region, India Analele University Nii din Oradea – SeriaGeografie Year XXII, no. 1: 122 – 127

[37] Aremu M.O. Gav B.L., Opaluwa O.D., Atolaiye B.O., Madu P.C. and Sangari D.U. 2011. Assessment of Physicochemical Contaminants in Waters and Fishes from Selected Rivers in Nasarawa State, Nigeria. Research Journal of Chemical Sciences. 1(4): 6-17

[38] Huneeus C. 2019. Digital Water Transformation: The Promise of Artificial Intelligence. [omline] <https://medium.com/datadriveninvestor/digital-water-transformation-the-promise-of-artificial-intelligence-7d88fb07e79b> accessed on 14-jul-2020.

[39] Shi Chun-yi, Chang-Ning Huang, Principles of Artificial Intelligence, Beijing: Tsinghua University Press, 1993.

[40] Hussein Ali Salah, M. Mocanu and A. Florea, "Development of a prototype ArcGIS- web-based decision application WATERDSS: For water pollution management," 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, 2015, pp. E-1-E-6, doi: 10.1109/ECAI.2015.7301143.

[41] Huneeus C 2019, Digital Water Transformation: The Promise of Artificial Intelligence, Data Driven Investor, [online] <https://medium.com/datadriveninvestor/digital-water-transformation-the-promise-of-artificial-intelligence-7d88fb07e79b> accessed on 19-Jul-2020

[42] A. Lal and B. Datta, "Genetic Programming and Gaussian Process Regression Models for Groundwater Salinity Prediction: Machine Learning for Sustainable Water Resources Management," 2018 IEEE Conference on Technologies for Sustainability (SusTech), Long Beach, CA, USA, 2018, pp. 1-7, doi: 10.1109/SusTech.2018.8671343.

[43] R. K. Bhattacharjya, B. Datta, and M. G. Satish, "Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers," Journal of Hydrologic Engineering, vol. 12, pp. 273-282, 2007.

[44] A. Lal and B. Datta, "Development and Implementation of Support Vector Machine Regression Surrogate Models for Predicting Groundwater Pumping-Induced Saltwater Intrusion into Coastal Aquifers," Water Resources Management, pp. 1-15, 2018.

[45] A. Lal and B. Datta, "Multiple objective management strategies for coastal aquifers utilizing new surrogate models," International Journal of Geomate, vol. 15, pp. 79-85, 2018.

[46] D. K. Roy and B. Datta, "Multivariate Adaptive Regression Spline Ensembles for Management of Multilayered Coastal Aquifers," Journal of Hydrologic Engineering, vol. 22, p. 04017031, 2017.

[47] D. K. Roy and B. Datta, "Fuzzy C-Mean Clustering Based Inference System for Saltwater Intrusion Processes Prediction in Coastal Aquifers," Water Resources Management, pp. 1-22, 2016.

[48] G. Bejarano, M. Jain, A. Ramesh, A. Seetharam and A. Mishra, "Predictive Analytics for Smart Water Management in Developing Regions," 2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, 2018, pp. 464-469, doi: 10.1109/SMARTCOMP.2018.00047.

[49] Wang, X., Zhang, F. & Ding, J. Evaluation of water quality based on a machine learning algorithm and water quality index for the Ebinur Lake Watershed, China. Sci Rep 7, 12858 (2017). https://doi.org/10.1038/s41598-017-12853-y

[50] Y. Wang, J. Zhou, K. Chen, Y. Wang and L. Liu, "Water quality prediction method based on LSTM neural network," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-5, doi: 10.1109/ISKE.2017.8258814.

[51] Wiki A.I. A Beginner’s Guide to Important Topics in AI, Machine Learning, and Deep Learning, [online] < https://pathmind.com/wiki/multilayer-perceptron#one > Accessed on 12-Jul-20

[52] Simpson-Hébert et al. (1996); Sawyer et al. (1998); and Brikké (2000), WHO Guidelines for Drinking-water Quality, second edition, Volume 3; the supporting document Water Safety Plans (section 1.3)