Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2020.010202
ISSN 2790-2730 Vol. 1, Issue 2: 11-19

Prediction of Water Quality Monitoring Indicators Based on Random Forest

Erika Ottaviano*

Furtwangen University, 78120 Furtwangen, Germany

**corresponding author*

Keywords: Water Quality Monitoring, Indicator Monitoring, Water Quality Forecast, Random Forest

Abstract: Water resources are not only the material basis to support the current social and economic development, but also an important resource to maintain people's daily life. Therefore, the protection of water resources from being infringed not only requires the cooperation of relevant researchers, but also is closely related to every resident in the society. Everyone needs to pay more attention to it. In the current society, the water environment in different regions is generally protected through multiple processes. Among them, the first line of defense is the monitoring of water quality and the prediction model of indicators. The monitoring and early warning model of water quality indicators is generally based on the real-time monitoring of the content of various substances in the water environment, so as to comprehensively evaluate the water environment and predict the water quality in the future period according to the concentration of different substances. The existing water quality indicator monitoring and prediction model not only is an important technology to control water environmental pollution in the current society, but also can reflect the quality status in the target waters and the future water quality development trend in all aspects. More professional water quality related data can be collected through the existing monitoring models of various indicators of water quality. Based on the in-depth analysis of these data, the main pollutants in the target waters, the current pollution situation and the possible future damage can be fully understood. On the other hand, this water quality indicator monitoring model has also provided assistance for professionals to develop more scientific and reasonable water pollution control plans. In this paper, the existing water quality index monitoring and prediction model was updated by using the stochastic forest algorithm model. The stochastic forest algorithm model generally predicts the indicators and pollution status of the target water area in the future by processing the relevant data of the water quality indicators of the target water area. Finally, the performance of this new water quality monitoring index and prediction model has been improved by about 26% on average.

1. Introduction

Water quality monitoring is one of the key processes in the water resources protection process. The formulation of water resources protection plan and treatment measures also need to analyze the water quality monitoring data before they can be implemented. Although the existing monitoring technology for multiple indicators of water quality has the advantages of high monitoring efficiency and accurate water quality data collection, it can not meet more requirements in the current field of water environmental protection. Therefore, more relevant researchers are needed to explore this, so as to propose a better monitoring and prediction model for water quality indicators.

Some researchers in the field of water resources protection have explored the development of water quality in recent years, and put forward a variety of water resources protection suggestions. Solangi Ghulam Shabir explored the performance of water quality index, comprehensive pollution index and geographical assessment tools in evaluating groundwater quality. Through the comprehensive performance of the three models, he determined that the water quality index had a good performance in the local groundwater quality assessment [1]. Alizadeh Mohamad Javad explored the role of machine learning models in evaluating the water quality of a basin. Through the analysis of machine learning and water quality evaluation model, he determined that it had a good performance in water quality evaluation [2]. Zhi Wei explored the performance of deep learning model in river water quality assessment. Through his research on the deep learning model, he determined that the model could accurately evaluate water quality [3].

Camara Moriken explored the impact of land use rate on water quality in a certain area, and determined that the large area of land use had a great impact on water quality [4]. Son Cao Truong explored the role of water quality and pollution index in the comprehensive evaluation of water area and determined its reliability [5]. Hamid Aadil explored some factors affecting water quality and determined that changes in various factors in the environment could have a great impact on water quality [6]. Muharemi Fitore explored a machine learning method for detecting water quality anomalies and determined the reliability of this method [7]. However, these water resources protection proposals can only slightly improve the existing water resources protection model, and can not directly promote the substantial growth of water resources protection performance.

In addition, some researchers have explored the monitoring and prediction methods of water resources, hoping to get a perfect monitoring and prediction model. Bisht Anil Kumar mainly explored the water quality prediction model in a certain area and determined the feasibility of artificial intelligence in the water quality prediction model [8]. Yang Huanhai explored the water quality prediction model in multi-scale aquaculture and determined the reliability of this multi-scale water quality prediction model [9]. Haghiabi Amir Hamzeh explored the role of a machine learning method in the water quality prediction model, and verified the feasibility of this water quality prediction model through experiments [10]. Barzegar Rahim explored the role of a hybrid algorithm model in water quality prediction and determined the feasibility of this model [11]. Hassan Md Mehedi explored the role of an algorithm model in the prediction of water quality index and determined the reliability of this algorithm model [12]. However, most of these studies can not meet the current social requirements for water resources monitoring and prediction. Therefore, more in-depth research is needed.

The main purpose of this paper was to solve the shortcomings of the existing water quality indicator monitoring model that can not analyze and process multiple water quality indicator data and predict the water quality environment. Stochastic forest algorithm model was used to help the existing water quality indicator monitoring model complete in-depth analysis of multiple types of water quality data. The water resources and environment of the target waters were accurately predicted to help the relevant staff formulate more reasonable water quality protection plans.

2. Prediction of Water Quality Monitoring Indicators

Water quality monitoring has been an important basic work in current water quality management. This work is mainly to provide basic data support for the protection and treatment of water areas through real-time monitoring of the pollutants in the target water area and the factors that cause water pollution [13]. However, the early water quality monitoring work is generally only the detection of chemicals in the waters of rivers and lakes. However, with the further acceleration of the process of social industrialization, the damage to water resources in the process of social development is also increasing, which also urges the water quality detection mode to gradually evolve into a professional water quality monitoring department and model to detect the water pollution in most waters. Moreover, with the in-depth research of relevant researchers on the water quality monitoring mode, the projects in the water quality monitoring have been constantly improved, and the monitoring technology has also moved towards informatization and standardization. The general flow of water quality indicator monitoring is shown in Figure 1.

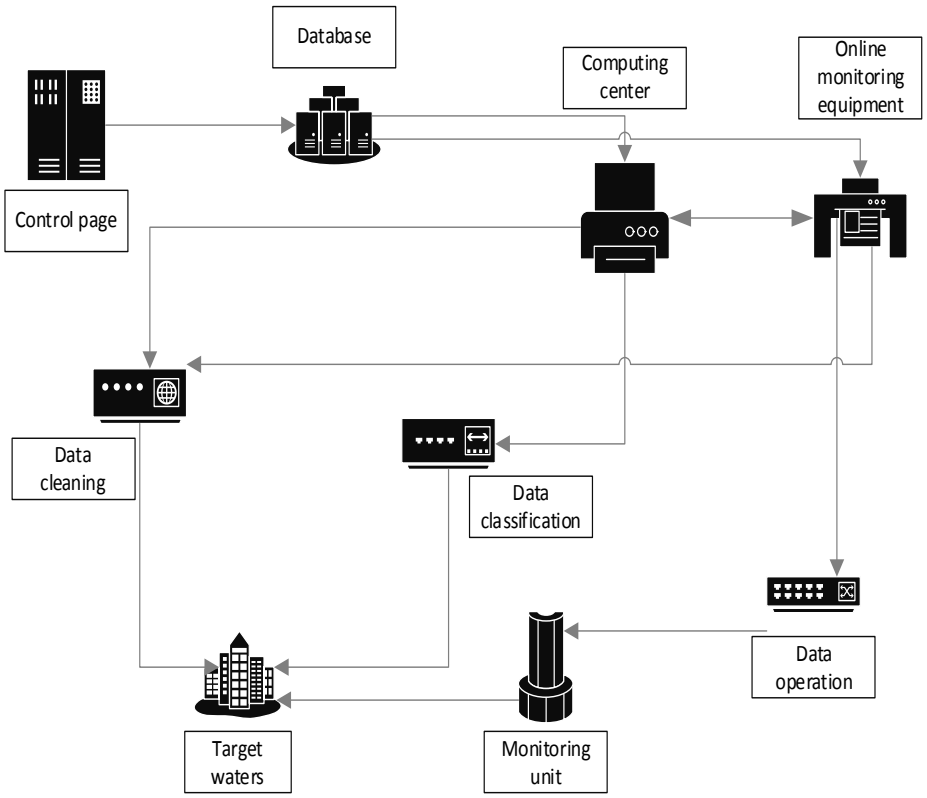


Figure 1. Schematic diagram of the general process of water quality index monitoring

In the current monitoring mode of water quality indicators, it is first necessary to clarify the objectives to be monitored. Then, in different water areas, appropriate water quality indicator detection schemes are used to detect multiple types of pollutants in the target water area, thus providing scientific and effective data support for the protection and treatment of water areas. The existing water quality indicator monitoring models generally include a variety of water quality detection methods, including quality analysis, titration analysis, acid-base titration, and so on. These methods also have different functions and detection effects in detecting different waters. Among them, the quality analysis method is generally to separate various substances in the sampled water resources, and then measure the content of various substances. The titration analysis method is to conduct chemical reaction on the sampled water resources by chemical means, and then detect the

substances in the water according to the reaction performance. The working flow of the algorithm in the water quality indicator monitoring and prediction system is shown in Figure 2.

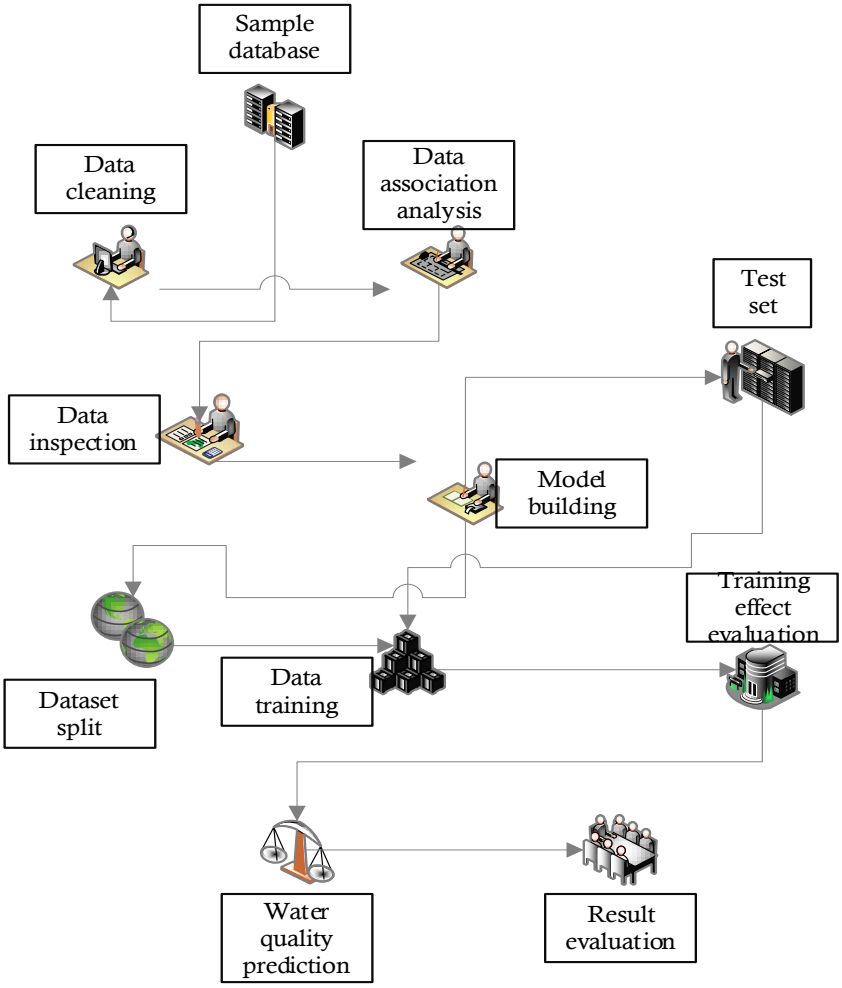


Figure 2. Schematic diagram of the algorithm workflow in the water quality index monitoring and forecasting system

3. Random Forest Analysis

Most of the existing water quality indicator monitoring models use manual methods to sample and detect water resources in the waters. With the continuous development of time, various problems have gradually deepened, and can no longer meet the current requirements of water quality monitoring. Therefore, a new automatic water quality indicator monitoring mode is needed to replace this manual mode. This paper completes the construction of this automatic water quality indicator monitoring mode by using random forest and a variety of intelligent sensor equipment. The automatic water quality indicator monitoring mode uses correlation analysis, multiple regression analysis and random forest algorithm model to optimize the operation mode in water quality indicator monitoring. Among them, the random forest algorithm model is mainly a supervised learning algorithm, which can solve multiple classification and regression problems by constructing a set of decision trees [14]. With the continuous development of random forest algorithm, it has been applied in many fields such as medicine, finance, e-commerce, etc. [15]. Benefiting from the powerful data relation mining ability of the current stochastic forest algorithm model, it can quickly analyze and calculate various independent variables in the monitoring mode

of water quality indicators, so as to obtain more accurate data of water quality indicators. This also enables water quality to be monitored more quickly after being polluted, thus helping relevant researchers to propose scientific water quality control plans. In addition, the stochastic forest algorithm model can also make relatively accurate prediction of water pollution in the future through the analysis of existing water quality data. The work flow of random forest in water quality indicator monitoring mode is shown in Figure 3.

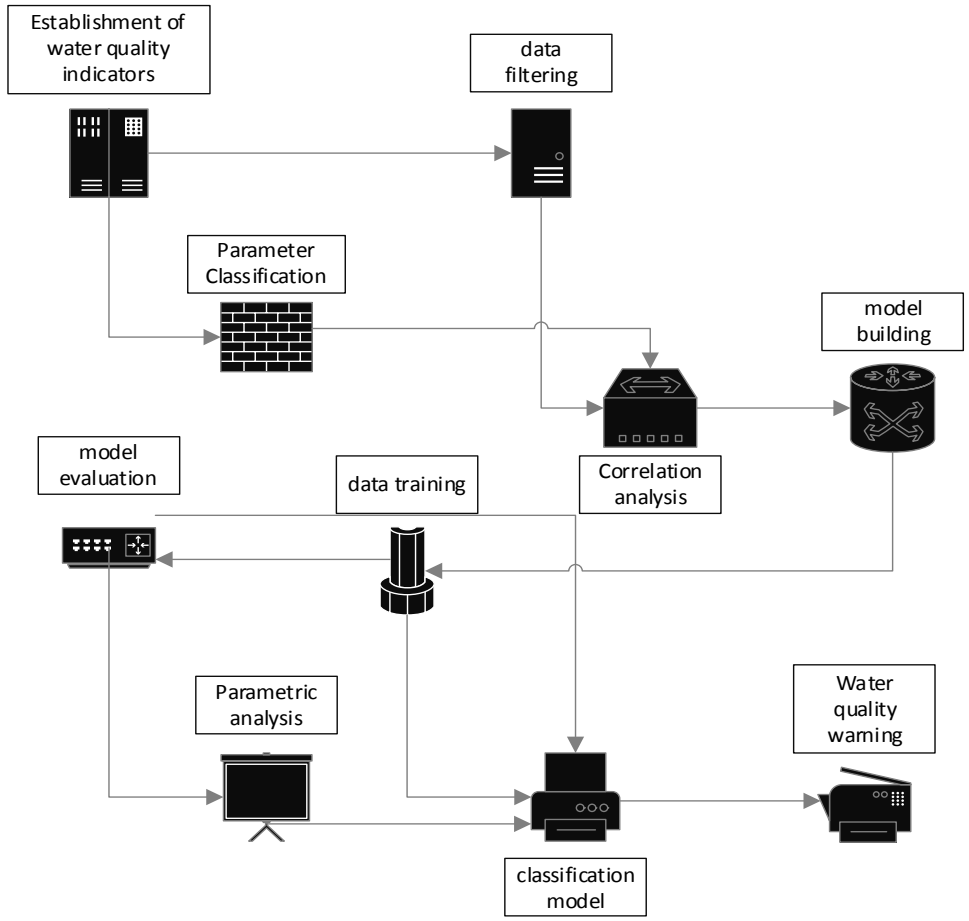


Figure 3. Schematic diagram of the workflow of random forest in the monitoring mode of water quality indicators

4. Random Forest Algorithm

In recent years, with the rapid development of water quality indicator monitoring mode, not only more water quality indicators can be monitored, but also the cost of water quality monitoring is further reduced. However, the development speed of this water quality indicator monitoring mode still cannot meet some requirements of current water resources protection. First of all, the amount of data that needs to be processed in the current water resources protection has increased significantly, and this surge in data volume has also increased the cost of processing data. Therefore, in order to further improve the efficiency of data processing, data are analyzed more effectively, so as to comprehensively reduce the cost of water quality indicator monitoring mode, which requires more algorithm models to be integrated into the water quality indicator monitoring mode. This paper mainly analyzes the correlation between the data of various indicators by using the stochastic forest correlation algorithm model. This analysis can also measure the tightness of the relationship between multiple variable data, so that a data operation model can be obtained by processing these

data to predict the water quality more accurately.

First, the sampling algorithm is used to preliminarily analyze the data in the water quality indicator monitoring model, so as to clean and classify the water quality indicators, as shown in Formula (1).

$$f(n) = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n} \tag{1}$$

Among them, n represents the type data amount of sample data. Then, two data averaging methods are used to calculate the sample data, so as to determine the prediction of water quality through the calculation process of the sample data, as shown in Formula (2).

$$H(x) = \frac{1}{T}\sum_{i=1}^{n} h_i(x) \tag{2}$$

Among them, h_i represents the output value of the sample data, and T represents the vector dimension. Then, the weighted average method is used to fine predict the quality of water quality, as shown in Formula (3).

$$H(x) = \frac{1}{T}\sum_{i=1}^{n} w_i h_i(x) \tag{3}$$

Among them, w_i represents the weight value of the model. The addition of random forest algorithm model not only further reduces the data operation cost of the existing water quality monitoring model, but also greatly improves the efficiency of data operation. This also makes the results of water quality monitoring more timely and plays a greater role in the protection of water resources. On the other hand, the integration of random forest also enables the water quality indicator monitoring model to predict the water quality pollution in the future through the analysis of water quality indicator data.

5. Prediction Experiment of New Water Quality Monitoring Indicators

The high-quality development of social economy has not only promoted the reform of many industries in the society, but also caused more pollution to the water resources and environment in its region. This situation has also attracted the attention of environmental protection staff. In order to protect the existing water resources so that they can be used sustainably, first of all, the water resources in the region need to be better protected and treated. The first and most basic step of the protection and treatment of water resources is the monitoring of water quality. Only through the monitoring data of water quality can reasonable water resources protection and treatment strategies be formulated. On the other hand, with the continuous progress of society, people have higher and higher requirements for the function of water quality indicator monitoring mode. This high requirement is further promoting the high-quality development of water quality indicator monitoring mode and making due contributions to the protection and treatment of water resources. In this paper, the performance of this new water quality index monitoring and prediction model combined with random forest algorithm model was experimentally studied, and the superior performance of this new water quality index monitoring and prediction model combined with random forest was determined. At the same time, its monitoring ability in various water quality indicators was analyzed.

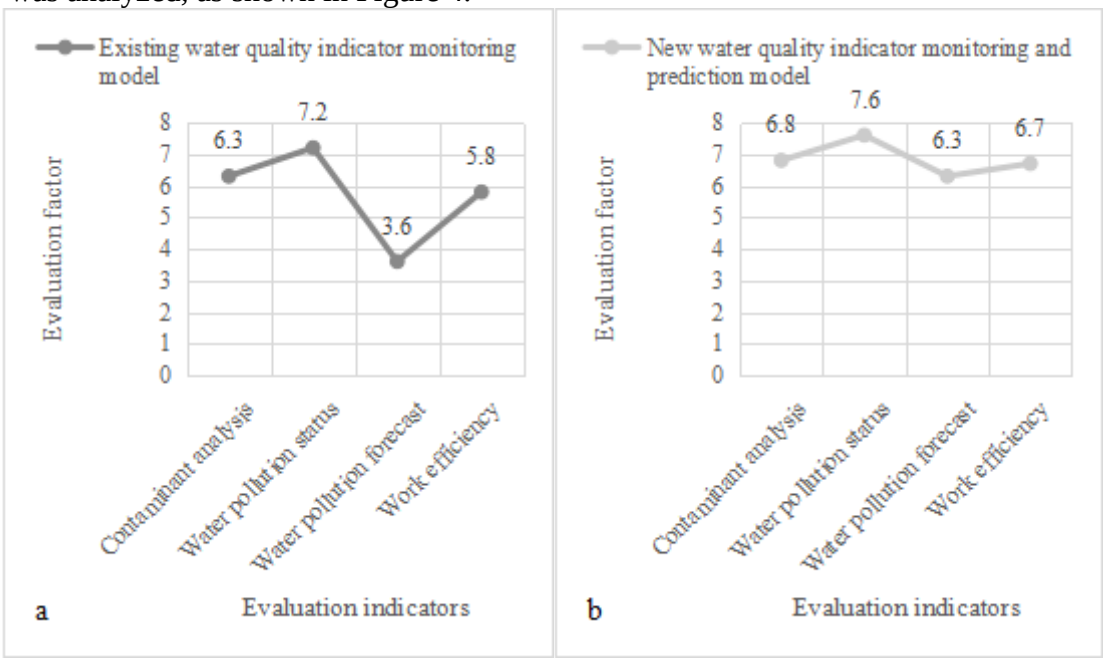
First of all, the concentration of various substances in a section of water and their impact on the water resources in this section of water were analyzed, as shown in Table 1.

Table 1. Effect of concentration of various substances in waters on the condition of their waters

	Concentration	Water conditions
Turbidity	6.7	5.4
Chloride ion	7.2	6.2
Suspension	6.6	4.8
Electrical conductivity	7.1	6.5

The monitoring mode of water quality indicators is mainly a technology for real-time monitoring of water bodies, which plays a major role in determining the categories and concentrations of pollutants that may exist in a section of water. At the same time, in the monitoring process of water quality indicators, the overall pollution status of the water body can also be evaluated through the classification of pollutants and the change trend of water quality. The current monitoring mode of water quality indicators can generally be divided into two modes: surface water quality monitoring and groundwater monitoring. Although the two water quality indicator monitoring modes are different for the monitored water bodies, their monitoring objectives are mostly the same. After analyzing the concentration of four different substances in the water body and the overall situation of the water body in Table 1, it can be determined that the turbidity and suspended solids concentration of the water body have a great impact on the pollution status of the water body, which also provides a reliable reference for the protection and treatment modes of different water bodies.

Then, the performance of the new water quality indicator monitoring and prediction model proposed by the random forest and the existing water quality indicator monitoring model in many aspects was analyzed, as shown in Figure 4.



a. Existing water quality indicators monitoring model performance diagram

b. New water quality indicators monitoring mode performance schematic

Figure 4. Schematic representation of the performance of the new water quality indicator monitoring and prediction model and the existing water quality indicator monitoring model under random forest

With the development of society, the monitoring of water quality needs to be carried out more

and more frequently, so as to grasp the pollution status of the water area more timely, so as to ensure the sustainability of the water area. Generally speaking, the water quality indicator monitoring mode can fully describe the parameters of the target water quality. With the progress of society, various substances in the water have also begun to increase rapidly. The performance of the current water quality indicator monitoring mode is also beginning to fail to meet the requirements. Through the analysis of the performance of the existing water quality indicator detection mode in the four aspects of pollutant analysis, water pollution status, water pollution prediction and work efficiency in Figure 4a, it is determined that the performance of the existing water quality indicator detection mode in the prediction of water pollution cannot meet the expected goal, and this module needs to be optimized in combination with other technologies. On the other hand, after the analysis of the performance of the new water quality index detection and prediction model combined with random forest in the same four aspects in Figure 4b, it is determined that the new water quality index detection and prediction model combined with random forest has the highest improvement in water pollution prediction, and other aspects have also improved. Finally, in a comprehensive way, the performance of this new water quality index detection and prediction model has improved by about 26% on average.

6. Conclusion

With the rapid development of various modern information technologies, the heavy industry and other related industries in society have developed well. However, in this rapid development process, these industries have caused great damage to the local ecological environment in many aspects. This destruction of the ecological environment is also seriously endangering the quality of life of residents in the region, and has a serious negative impact on the health of residents. Therefore, with the development of technology and the improvement of people's awareness of ecological environmental protection, more and more researchers are studying the environmental protection model. Among them, the water quality indicator monitoring model is a key technology in the water resources environmental protection model. The monitoring of multiple indicators in the water quality of different waters can not only help the environmental protection personnel in the region to have a deeper understanding of the water pollution situation, but also analyze the main causes of water pollution. In addition, the analysis of such water quality data can also help the relevant water resources protection staff to build a more scientific and reasonable water resources protection model and better control the current situation of water pollution. Through in-depth analysis of the stochastic forest algorithm model, this paper determined that this algorithm model could be better applied in the current water quality indicator monitoring model, and provided a better role for the water quality prediction model. This water quality monitoring and prediction model based on random forest can predict the pollution status and pollution sources of the target water area in a certain extent through the analysis and calculation of various water quality data, thus helping relevant staff to carry out prevention work in advance.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Solangi Ghulam Shabir. Groundwater quality evaluation using the water quality index (WQI), the synthetic pollution index (SPI), and geospatial tools: a case study of Sujawal district, Pakistan. Human and Ecological Risk Assessment: An International Journal. (2020) 26(6): 1529-1549. https://doi.org/10.1080/10807039.2019.1588099

[2] Alizadeh Mohamad Javad. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Engineering Applications of Computational Fluid Mechanics. (2018) 12(1): 810-823. https://doi.org/10.1080/19942060.2018.1528480

[3] Zhi Wei. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? Environmental science & technology. (2020) 55(4): 2357-2368. https://doi.org/10.1021/acs.est.0c06783

[4] Camara Moriken, Nor Rohaizah Jamil, Ahmad Fikri Bin Abdullah. Impact of land uses on water quality in Malaysia: a review. Ecological Processes. (2019) 8(1): 1-10. https://doi.org/10.1186/s13717-019-0164-x

[5] Son Cao Truong. Assessment of Cau River water quality assessment using a combination of water quality and pollution indices. Journal of Water Supply: Research and Technology-Aqua. (2020) 69(2): 160-172. https://doi.org/10.2166/aqua.2020.122

[6] Hamid Aadil, Sami Ullah Bhat, Arshid Jehangir. Local determinants influencing stream water quality. Applied Water Science. (2020) 10(1): 1-16. https://doi.org/10.1007/s13201-019-1043-4

[7] Muharemi Fitore, Florin Leon. Machine learning approaches for anomaly detection of water quality on a real-world data set. Journal of Information and Telecommunication. (2019) 3(3): 294-307. https://doi.org/10.1080/24751839.2019.1565653

[8] Bisht Anil Kumar. Artificial neural network based water quality forecasting model for ganga river. International Journal of Engineering and Advanced Technology. (2019) 8(6): 2778-2785. https://doi.org/10.35940/ijeat.F8841.088619

[9] HuanhaiYang, Shue Liu. A prediction model of aquaculture water quality based on multiscale decomposition. Mathematical Biosciences and Engineering. (2020) 18(6): 7561-7579.

[10] Haghiabi Amir Hamzeh, Ali Heidar Nasrolahi, Abbas Parsaie. Water quality prediction using machine learning methods. Water Quality Research Journal. (2018) 53(1): 3-13. https://doi.org/10.2166/wqrj.2018.025

[11] Barzegar Rahim, Mohammad Taghi Aalami, Jan Adamowski. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. Stochastic Environmental Research and Risk Assessment. (2020) 34(2): 415-433. https://doi.org/10.1007/s00477-020-01776-2

[12] Hassan Md Mehedi. Efficient prediction of water quality index (WQI) using machine learning algorithms. Human-Centric Intelligent Systems. (2020) 1(3-4): 86-97. https://doi.org/10.2991/hcis.k.211203.001

[13] Ahmed Umair. Water quality monitoring: from conventional to emerging technologies. Water Supply. (2020) 20(1): 28-45. https://doi.org/10.2166/ws.2019.144

[14] Resende, Paulo Angelo Alves, Andre Costa Drummond. A survey of random forest based methods for intrusion detection systems. ACM Computing Surveys (CSUR). (2018) 51(3): 1-36. https://doi.org/10.1145/3178582

[15]]Schonlau Matthias, Rosie Yuyan Zou. The random forest algorithm for statistical learning. The Stata Journal. (2020) 20(1): 3-29. https://doi.org/10.1177/1536867X20909688