Advances in Artificial Intelligence Research (AAIR)
ISSN 2757-7422
www.dergipark.com/aair/

Vol. 3 (No. 1), pp. 36-44, 2023
doi: 10.54569/aair.1200695
Published online: Feb 15, 2023

Prediction of Water Quality with Ensemble Learning Algorithms

Faten Aljarah [1*], Aydin Cetin [2]

[1] Graduate School of Informatics, Gazi University, Ankara, Türkiye
[2] Computer Engineering Department, Faculty of Technology, Gazi University, Ankara, Türkiye

Abstract

Monitoring and controlling the quality of the water is one of the most important issues in the world since only 74% of the world's population use safely managed water where the water is treated well to reach the minimum limit of safety and quality standards. To observe the water potability and take immediate actions to improve the water quality, real-time monitoring and classification process are required. However, monitoring and controlling the water quality is not an easy task since it has many requirements such as the collection and analysis of data and calculations to be made. In this paper, we focus on applying machine learning for the evaluation of the water quality. We have chosen five ensemble learning algorithms namely, Adaptive Boosting (AdaBoost), Random Forest (RF), Extremely randomized Trees (Extra Tree), Gradient Boosting (GB), and Stacking Classifier to evaluate their classification performances in determining the water quality. The Stacking Classifier had achieved the highest accuracy (0.67) and F-score (0.64).

Keywords: *Water Potability; artificial intelligent; production WQ; Machine Learning; ML*

1. Introduction

Human life, in general, depends on the availability of water, and the water quality is one of the important factors affecting the practical improvement of daily life. A specific standard of water quality needs to be reached in order for the water to be considered potable water, which is safe for humans to drink. In contrast, non-potable water is the water that is used for everything except human use. Many large-scale procedures are carried out on non-potable water before use, although it remains unfit for direct human consumption [1]. According to the World Health Organization, in 2020 more than 74% of the world's population (5.8 billion people) use safely managed water where the water is treated well to reach the minimum limit of safety and quality standards. However, more than 2 billion people in the World have access only to polluted or undrinkable water [2, 3]. For example, according to [4] it was reported that approximately 1.8 billion people worldwide use non-potable water sources. As a result, it affects the lives of people especially children, resulting in their death. According to a 2017 report from the World Health Organization, about 525,000 children under the age of five die from diarrhea every year [5]. The process of obtaining fresh water from ground and surface water in the past was easier than now. This is due to the increased dependence of human life on the availability of water. A rise in the issues of water pollution is also a result of the industrial and economic growth that humanity has reached, as well as a lack of knowledge about the right use of wastewater and adequate water use.

It is important to monitor water quality to find out the degree of water pollution, ensure access to clean water resources and apply effective guidelines for the protection of Water Resources [6]. Predicting water quality is a difficult task. Many researchers have made a great effort in determining water quality because of its importance to human life [2]. Therefore, it is an urgent necessity for humans to provide safe drinking water as it maintains the health of the kidneys, and the intestines nourish the muscles and help to maintain body fluid [7]. To ensure the potability of water, it is important to devise new technologies and methods. The water quality index (WQI) is a method used for measuring water quality which reflects the impact of different water standards on its quality. The calculation of WQI is necessary to determine the usability of water and to know the water specifications [8] It converts complex analyses of water properties and huge amounts of data into easy information that can be understood and used by specialists and non-specialists [9]. The water quality measurement index has been evaluated by many international studies as the basis for measuring various water indicators [10].

2. Background

A certain degree of water quality must be attained for water to be considered drinkable and safe for human consumption. In contrast, non-potable water is used for reasons other than drinking. The research [11] separated its data into 84% training and 16% tests and utilized the RF model and other models to estimate the quality of

*Corresponding author
E-mail address: fatina99aljarah@gmail.com

Received: 08/Nov/2022; Received in revised form: 07/Dec/2022; Accepted: 08/Dec/2022.

the water. It obtained favorable results after collecting data with Elsevier's Data in Brief (DiB). The KNN model was used in another study [12] that looked at the following parameters: dissolved oxygen (do), pH, conductivity, biological oxygen demand (bod), nitrates, fecal Escherichia coli, and total Escherichia coli form. This study used synthetically generated data. To achieve optimum water quality, the stacking model was employed by multiple researchers [13-15]. In other research, the KNN, RF, and Adaboost have been utilized and resulted in good evaluation [16, 17]. While GB has been compared with RF, KNN, ANN, and other models in Kelantan River, Malaysia [18].

The water quality index has been predicted in a study by [19] using neural networks at the Tigris River in Baghdad city. In many other studies, the importance of using the water quality index in measuring the potability of water on the Tigris River in the city of Mosul / Iraq is discussed. Also, the factors affecting the water quality index are indicated using the weighted mathematical model [20]. Other studies also aim to implement the Canadian water quality of Environment Ministers (CCME WAI) (CCME WAI) in the Tigris River [21, 22].
Recently, several research articles [23, 24]. have discussed the development of machine learning to assess water quality [25]. Studies have revealed various types of machine learning models applied to water quality, such as fuzzy logic, artificial neural networks, neural inference models, and others [26]. However, there are many variations of machine learning that have not yet been explored in water quality studies [24]. Although machine learning models are common in assessing water quality, they still face some shortcomings, such as the need for human intervention during the modeling process, time-consuming algorithms, and the need for flexible models in solving some environmental problems [27].

From the reviewed literature two major conclusions can be drawn. First, there have been many attempts to improve the performance of machine learning algorithms in water quality assessment. However, the current results still need to be improved. This study focuses on establishing an effective water quality assessment with ensemble methods. Secondly, most studies divide the parameters into chemical, biological, physical, and others for the application of one or more models to monitor the water and predict its quality, but the results are still insufficient. Consequently, this study is trying to fill this gap as well by exploring the ability of five models of machine learning (Adaptive Boosting, Random Forest, Extra trees classifier, Gradient Boosting, and Stacking Classifier) to predict water potability. This paper is organized as follows: After the introductory section, section 2 discusses the methodology and materials for the system that was investigated. Section 3 presents the evaluation metrics and the results of machine learning models considered for determining the potability of water samples and then concluding remarks are given in section 4.

3. Materials and Method

Through the water quality monitoring system and platform, one is able to determine whether the water is fit for a human drink or not. The modeling steps used in this study are presented as a flowchart in Figure 1, with each step being discussed.

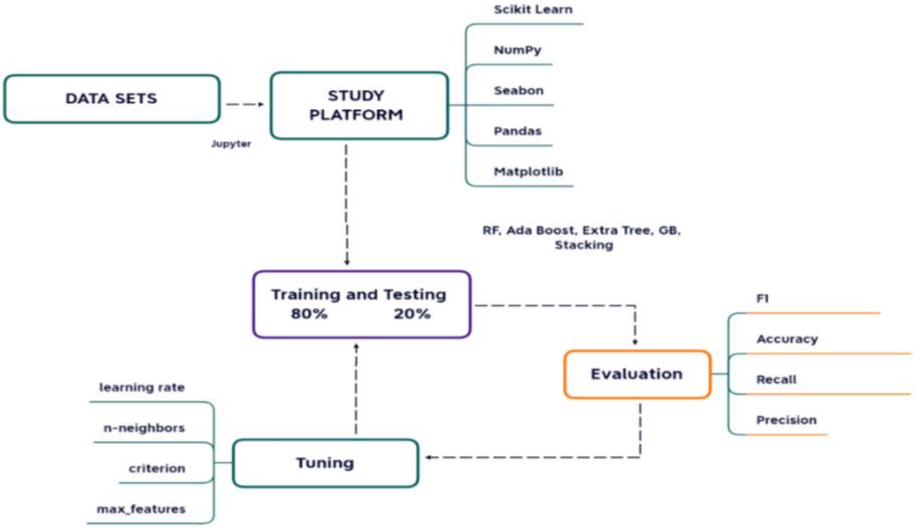


Figure 1. *Flowchart of the proposed model.*

3.1 Data set

The dataset was obtained from Kaggle in a data frame [28]. The information was collected two years ago and contains water quality measurements for 3276 different bodies of water. This dataset includes nine critical parameters: PH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, and Turbidity. Information was laid out in ten columns. This data was obtained from an industrial generator. It represents synthetic data containing information that is created artificially rather than via real-world happenings. Artificial data is made public in order to train machine learning algorithms and validate mathematical principles.

A number of operations and modifications have been performed in order to prepare the data and produce results with fewer errors. To achieve good predictive results, the dataset was divided into training and testing, an 80% training collection, and a 20% test collection. Since the data is not standardized, there is a gap between its values. Thus, the standard numerical is used to measure the data, which varies between 1 and -1 and used to standardize the data. Following a thorough examination of the data, it was discovered that it had empty values (Null) in data characteristics. The null was discovered in the PH, Sulfate, and Triethanolamine values, as illustrated in figure (2). To get rid of noise and unpredictable data, the null is eliminated and corrected using the average data rate.

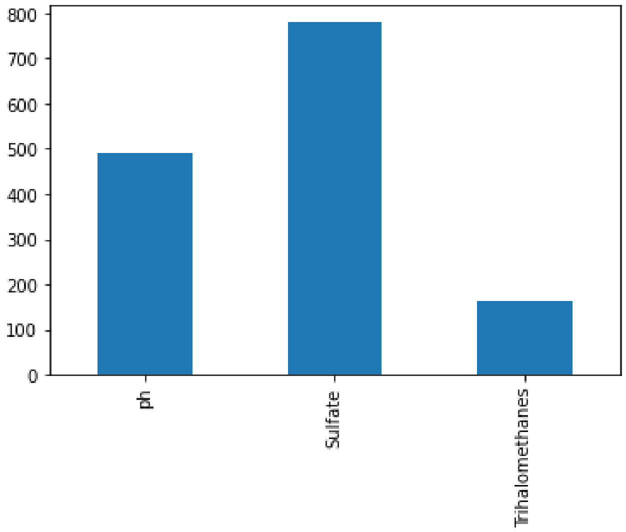


Figure 2. *The number of null in PH, sulfate, and triethanolamine.*

3.2 Development environment

To conduct all the necessary experiments, Jupyter notebook was used as an advanced work environment [29]. The project uses Python language, and as known, the Jupyter environment is one of the most compatible environments with Python.

Scikit_learn, NumPy, Pandas, Seaborn, Matplotlib libraries were used to develop the algorithm. The scikit_learn library includes a set of powerful data extraction and analysis capabilities. It is used to create a collection of machine learning, preprocessing, cross-validation, and visualization algorithms through a uniform interface. [30]. NumPy aims to supply many supporting functions. provides an array object up to 50 times faster than traditional Python lists [31]. Pandas library has been used because it helps to facilitate many time-consuming and repetitive tasks, including data normalization, data visualization, and others [32]. Seaborn library helps in understanding and exploring the data. It is mainly used for making statistical graphics in Python [33]. Matplotlib provides the user with the ability to visualize the data through a set of plots Such as scatter plots, graphs, etc. [34].

3.3 Parameter tuning

The model includes several hyperparameters (external parameters) that can optimize the model by changing its value manually. However, the learning algorithm itself cannot update or change hyperparameters [35]. In this research, we have adjusted the hyperparameters to get the best possible results. The main hyperparameters that contributed to the change in the results are:

- Max_depth: It is applied to determine the depth of the tree. It is the number of nodes from the root to the most distant leaf node. The greater the maximum depth value, the more complex the tree, and thus the greater the likelihood of overfitting. As a result, it is preferable to keep its value low [36]. Practical experiments show that as the maximum depth increases, the training error decreases, while the testing results in very poor accuracy. For example, when the depth is set to 50, the gradient algorithms accuracy is very poor. The best results are achieved by using the rest of the parameters with a depth value equal to 20.

- N_estimators: It is used to control the number of trees used in the ensemble model. This is the number of trees that must be constructed before averaging the final prediction [36]. The value of n-estimators should be set to 100 in order to achieve the greatest results in RF, Ada, Extra, and Gradient. While the algorithm bagging showed different results, the best results were between 220 and 230.

- N_neighbor: determines the number of data points by classifying them into groups. It represents the nearest number of data points that can be found and placed in the same category [36].. When k=1, a high training score shows, but the test result is quite low which results in overfitting. After multiple testing on n-neighbor based on the data of this research, it was discovered that when k=9 in the KNN model, the results are the best.

- L2_regularization: It controls the complexity to get rid of overfitting. This parameter is intended to limit the complexity, noise, and generalization issues that cause overfitting. Because complexity cannot be determined from training data, this parameter solves the complexity problem by determining the right level of complexity in the model [37]. After doing several trials, it was shown that the optimal amount of regulation in the Hist algorithm is 14.9.

- Loss: it is used in the boosting method. and it is a measure of the error that occurs between the output of the algorithm and the target function. This parameter calculates the probability of the expected positive class, to minimize losses [38]. It has been observed that the performance of the Hist model is better when its value is 'auto'. The 'exponential' value in the case of the gradient model is chosen for the best performance. The primary goal is to achieve a balance between under and over-fitting.

- Learning_rate: It can also be called shrinkage, and it works based on taking large steps but after several iterations, it takes smaller steps to reduce the error rate. To reach optimal results, different values of the learning rate have been experimented. The rate of learning is important because it contributes to determining the rates of weight change. It is also used to evaluate the splitting quality [38, 39]. It was found that the Hist and Adaboost algorithms performed better when the learning rate value was larger. On the other hand, the optimal learning value for GB is 0.1, which yields the best outcomes.

- Criterion: This parameter is used to measure the quality of the splitting, so it could stop the algorithm that is running [40]. It is seen that the result in RF and Extra tree is the best when the Criterion is set to “gini”.

- Max_features: It is used when searching for the best splitting, considering the number of features in the data. If this value is not specified, all features are permitted in each division [41]. When max features are set to ”auto”, the best results are obtained in some models including GB and DT. Nonetheless, the finest outcomes are frequently obtained when this variable is set to “sqrt” as in the Extra tree. Using this value, the model is instructed to choose a specified number of features at random. In this situation, the number of features equals the square root of the entire number of features in the dataset.

- Class_weight: It defines classifications by certain categories (0 and 1). This hyperparameter is used, for models of unbalanced classification [42]. The “balance” value produced the best results in the Extra model. During the typical training process, it modifies the balances of majority and minority group classes.

It is also worth noting that all these hyperparameters’ results change depending on each other. For example, the “gini” option may be the worst or the best depending on different parameters.

3.4 Machine Learning algorithms

Various learning algorithms are employed to generate artificial intelligence. Due to the numerous distinct types of algorithms, 5 ensemble learning methods have been discussed. while the models are constructed according to the data supplied and the type of application. The five models are implemented in this research, and the performance of each model is discussed below. To ensure the accuracy of the findings, 4 different performance metrics were used. Specifically, Precision (Precis), Recall (R), Accuracy (ACC), and F-score (F1 score).

Ada boosting: it mainly works on compiling multiple weak workbooks and gradually learning each of them from the previously wrongly classified objects. We made several manual changes to find out the best suitable parameters for the Ada boosting algorithm. It is noted that the best result is when the parameters of n_estimators= 100, the learning rate= 1.0, and the algorithm= SAMME.R. There is a clear change in the results when the parameters change, and the difference is noticeable.

Random forest: It is an algorithm used for classification, developed from a decision tree. Individual tree decisions are collected to create random forests. After doing many experiments, it is noted that the best results are when using n_estimators = 100 with criterion = gini.

Extra Tree: It is an algorithm that is very similar to the Random Forest algorithm. The difference is that it uses the entire sample instead of the reduction and substitution that the Random Forest algorithm does. It is noted from the extra tree experiment that the change of each of the parameters affects the whole results and the best result is when n_estimater= 100, criterion =gini, max_features= sqrt, and class_weight = balanced.

Gradient boosting: The gradient boosting model is done by building the new model on previous errors and predictions to check if there are any wrong patterns missed from the previous model. The best result is achieved with the parameters: 100, friedman_mse, 0.1, none, 20, exponential. When the parameters are 100, squared_error, 0.1, none, 20, deviance, results have become much worse.

Stacking: the method of action in stacking is to use a different set of models one after the other (sequentially), where the prediction of each of the models is added to produce a new feature. In the end, a final dataset is obtained (new feature), which is fed by the last model called meta-learner as shown in figure (3). The best result is when the meta = LR and the algorithms that are used in the stack are KNN, DT, RF, Adaboost, ET, GB, and Hist Gradient Boost (HGB).

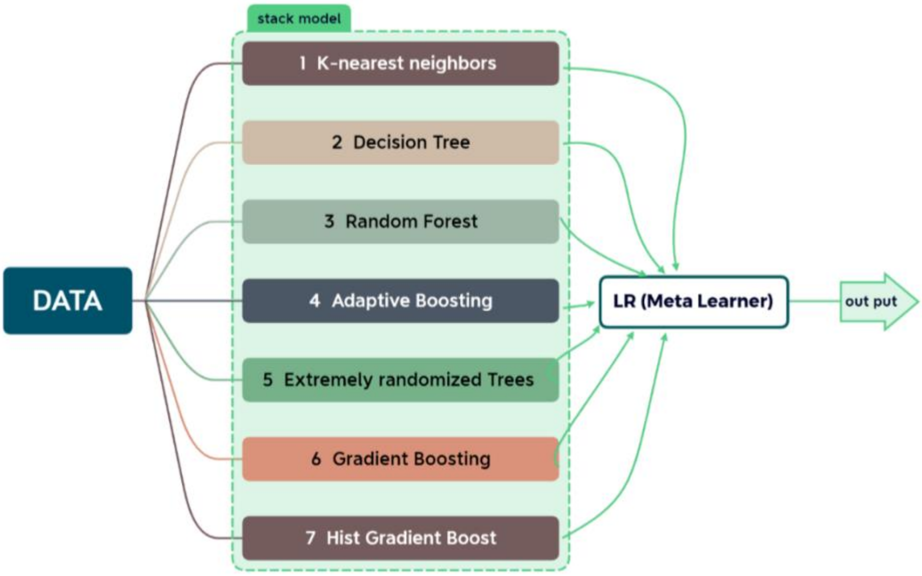


Fig 3: *Stacking in machine learning.*

3.5 Evaluation metrics

To determine whether the forecast results are positive or negative and evaluate the results of the models in predicting water quality, the research [43] used a set of rating performance measures. Confusion matrix was counted based on the next four settings:

1) TP, which means the number of ‘‘true positives’’ predictions.

2) FP, which means the number of ‘‘false positives’’ predictions.

3) FN, which means the number of ‘‘false negatives’’ predictions.

4) TN, which means the number of ‘‘true negative’’ predictions.

To separately examine the performance of each category of the model, the model performance has been analyzed into four categories so that it can be compared with other models [44].

a. **Accuracy** is one of the important and main tools in evaluating the model before practical application [45]. It is the arranger of the actual results among the total number of cases tested [46].

$$\text{Accuracy } = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

b. **Recall** is the ratio of correct positives to all actual positives in the data that measures the test's ability to measure the state when the state exists [47].

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

c. **Precision** is the ratio of the correct positive among all expected positives. It refers to how truthful the machine is about the real positives [48].

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

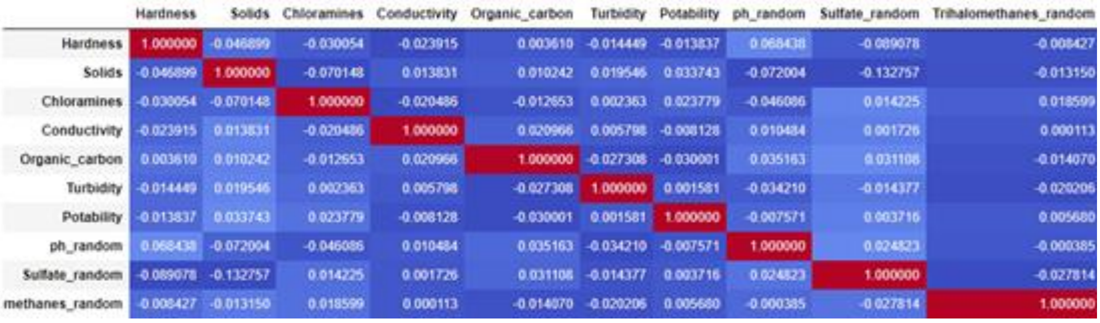
d. **F_score** is defined as an average (harmonic mean) between precision (P) and recall (R) [49] when there is a difference between FP and FN then F1_score will be needed. So, if you have various class distributions F1_score is the best [48].

$$\text{F\_score } = \frac{2 \text{ P R}}{\text{P} + \text{R}} \tag{4}$$

4. Experimental studies

After defining the algorithms to be trained according to the previous section under the methodology, we have performed parameter tuning and then evaluated the results according to evaluation metrics. In this study, the model performance results were evaluated and compared based on 4 criteria including, Precis, R, ACC, and F1 score. The choice of parameters affects obtaining more beneficial results and overall performance which may be more useful than the choice of the model itself. The performance of the model and the selection of parameters are evaluated according to the data set containing various data characteristics, including, PH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, and turbidity.

To apply the methodology, various models and templates were used with the help of Jupiter Notebook as a simple and fast working environment with the help of several important libraries in Python. One of the important steps that will be examined after data collection is the analysis of the relationship between the data. This is done through the correlation matrix, and as shown in Figure (4), there is no correlation between the data available to us, so we do not need to do any additional work to improve the results.



	Hardness	Solids	Chloramines	Conductivity	Organic_carbon	Turbidity	Potability	ph_random	Sulfate_random	Trihalomethanes_random
Hardness	1.000000	-0.046899	-0.030054	-0.023915	0.003610	-0.014449	-0.013837	0.068438	-0.089078	-0.008427
Solids	-0.046899	1.000000	-0.070148	0.013831	0.010242	0.019546	0.033743	-0.072004	-0.132757	-0.013150
Chloramines	-0.030054	-0.070148	1.000000	-0.020486	-0.012653	0.002363	0.023779	-0.046086	0.014225	0.018599
Conductivity	-0.023915	0.013831	-0.020486	1.000000	0.020966	0.005798	-0.008128	0.010484	0.001726	0.000113
Organic_carbon	0.003610	0.010242	-0.012653	0.020966	1.000000	-0.027308	-0.030001	0.035163	0.031108	-0.014070
Turbidity	-0.014449	0.019546	0.002363	0.005798	-0.027308	1.000000	0.001581	-0.034210	-0.014377	-0.020206
Potability	-0.013837	0.033743	0.023779	-0.008128	-0.030001	0.001581	1.000000	-0.007571	0.003716	0.005680
ph_random	0.068438	-0.072004	-0.046086	0.010484	0.035163	-0.034210	-0.007571	1.000000	0.024823	-0.000385
Sulfate_random	-0.089078	-0.132757	0.014225	0.001726	0.031108	-0.014377	0.003716	0.024823	1.000000	-0.027814
methanes_random	-0.008427	-0.013150	0.018599	0.000113	-0.014070	-0.020206	0.005680	-0.000385	-0.027814	1.000000

Figure 4. *The Correlation matrix of water potability data parameter.*

Machine learning models are used to predict whether water is safe or unsafe for consumption. This section describes the model's execution of the Adaboost, RF, Extra Tree, GB, and Stacking. Figure (1) shows the step-by-step execution of the models to achieve the desired results. It can be said that each test presented a different method for estimating water potability. Changing the parameters played an important role in improving the results of the models in general. It is noted that the stacking model performed so well, having the best results in F1 Score for estimating potability of water. It is important to mention that the "F1 Score" and "ACC" have importance to demonstrate the credibility of the results. Therefore, these two are given priority since the F1 Score represents the Precis and the R together. When classification ACC is compared, the performance of Random Forest (0.66) Extra Tree (0.66) Gradient Boosting (0.66), and Stacking (0.67) is very close and better than Adaboost (0.61). ACC allows the decision maker to understand the level of accuracy since accuracy is a numerical estimation of the performance of the model. The metrics for WQ predictive models according to ACC and F1 Score are given in detail in Table 1.

Table 1. *Performance metrics for WQ predictive models.*

Models	Accuracy	F_score
Ada boosting	0.61	0.55
Random forest	0.66	0.62
Gradient Boosting	0.66	0.62
Stacking	0.67	0.64
Extra Tree	0.66	0.63

The correlation matrix displays all the levels in the dataset. Since the ACC and F1 Score were discussed, the remainder is the R and Precis. Table (2) shows that the Extra tree represents the best Precis among the remaining models with a value of 0.69.

Table 2. *Performance metrics for WQ predictive models according to recall and precision.*

Models	Precision	Recall
Ada boosting	0.54	0.57
Stacking	0.61	0.69
Gradient Boosting	0.61	0.64
Extra Tree	0.63	0.64
Random forest	0.66	0.60

There have been several research in machine language to analyze water quality. As a result of one of the studies based on the same data that we are working on in the research [50], the model's findings were DT = 0.61 and RF=0.69, It is extremely close to our results. However, the Stacking Classifier model outperforms the other models according to our findings. Other studies [11], using different data and settings, achieved an accuracy of RF= 0.96.

5. Conclusions

Classification prediction methods were used in this study (i.e., Adaptive Boosting, Random Forest, Extra Trees classifier, Gradient Boosting, and Stacking Classifier) to predict water potability. The performances of these five models have been compared using the confusion matrix which includes (TP, FP, TN, and FN) with cross-validation to find the reliability of the models. Our dataset shows recorded values of 9 water parameters such as PH, Hardness, Solids, and turbidity. To eliminate noise and unstable data, null is eliminated and compensated with the average data rate. The parameters were changed continuously, and there was a noticeable difference in the results. From the results of the models obtained, it was observed that the Stacking C lassifier achieved better performance than the other models we have tested in estimating water quality. It is believed that this study can help researchers develop integrated artificial intelligence and machine learning models that will help water system managers in real-time monitoring of the quality of water for future applications.

References

[1] Varila M., “What Is Potable Water? Your Guide to Understanding Types of Water”, viralrang, 2020. [Online]. Available: https://viralrang.com/what-is-potable-water-your-guide-to-understanding-types-of-water/#. [Accessed: Nov 8, 2022]

[2] UNECE, “miyah alshrob,” who, (2022). [Online]. Available: https://www.who.int/ar/news-room/fact-sheets/detail/drinking-water. [Accessed: Oct 19, 2022].

[3] Fluence news team, “What Is Potable Water?”, fluencecorp, 2019. [Online]. Available: https://tinyurl.com/2qj936u9. [Accessed: Nov 8, 2022].

[4] World Health Organization, “Preventing diarrhoea through better water, sanitation and hygiene: exposures and impacts in low- and middle-income countries,” World Health Organization (Report), Villars-sous-Yens, Switzerland, 2014.

[5] World Health Organization, “Diarrhoeal disease,” who, 2017. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease. [Accessed: Dec 3, 2022].

[6] Li D., Liu S., “System and Platform for Water Quality Monitoring Chapter 3,” in Water Quality Monitoring and Management, *China: Academic Press*, 2019, p. 101.

[7] Edition F., Guidelines for Drinking-water Quality - 4th ED., Malta: World Health Organization WHO Library Cataloguing, 2011.

[8] Al safaw Y.A., R. Al Shanouna R.A.A., Messer, N., “Takeem hasaas naweet almeah w hesab muamel WQI le baaz masader almeah fe karyat abo marya kazaa talefar\ muhafazat nainawa,” *Journal of Education and Science*, 27(3), 87, 2018.

[9] Al Safawi A. Y. T., “Tatbik almuasher alkndy (WQI CCME) le takeem jawdet almeyah le agrad alshrub: dirasat halat jawdet almeyah aljawfeia fe nahiat almehalabia\ muhafazat nainawe,” *Journal of Rafidain Sciences*, 27(4), 199, 2018.

[10] Dilip P.V., Dnyaneshwar, M. S., Rajendra, L. D., Suresh, N. P., “Assessment of Ground Water Quality In Gajanan Colony, Ahmednagar. By Water Quality Index (WQI),” in Second Shri Chhatrapati Shivaji Maharaj QIP Conference on Engineering Innovations, Ahmednagar, India, 105, 2019, ISSN: 2581- 4230.

[11] Ajayi O.O, Bagula A.B, Maluleke H.C., “Water Net: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes”, *IEEE Access*, 10, 48318- 48337. 2022, doi: 10.1109/ACCESS.2022.3172274, 2022.

[12] Aldhyani T.H.H., Al-Yaari M., Al kahtani H., “Water Quality Prediction Using Artificial Intelligence Algorithms,” *Applied Bionics and Biomechanics*, vol.2020, 1-10. doi: 10.1155/2020/6659314, 2020.

[13] Nasir N, Kansal A, Aishalton O, “Water quality classification using machine learning algorithms”, *Journal of Water Process Engineering*, vol.48. doi: 10.1016/j.jwpe.2022.102920, 2022.

[14] Wang L, Zhu Z, Sassoubre L, “improving the robustness of beach water quality modeling using an ensemble machine learning approach”, *Science of the Total Environment*, 765, 1-4, doi: 10.1016/j.scitotenv.2020.142760, 2021.

[15] Rosly R, Makhtar M, Awang M.K, “Comparison of Ensemble Classifiers for Water Quality Dataset,” in Proceedings of the UniSZA Research Conference 2015 (URC ’15), Terengganu, Malaysia, 1-6, 2015

[16] Mogaraju J.K, “Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India,” *Turkish Journal of Engineering*, 7(1), 64 - 72. doi: 10.31127/tuje.1032314, 2023.

[17] El Bilali A, Taleb A, Brouziyne Y, “Groundwater quality forecasting using machine learning algorithms for irrigation purposes”, *Agricultural Water Management*, 245, 106625. doi: 10.1016/j.agwat.2020.106625 , 2021.

[18] Abdul Malek N.H, Wan Yaacob W.F, Md nasir S.A, “Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques”, *Water*, 14(7), 1067. doi: 10.3390/w14071067, 2022

[19] Al-Musawi N, “Prediction and Assessment of Water Quality Index Using Neural Network Model and Gis Case Study: Tigris River in Baghdad City”, *Applied Research Journal*, 3(11), 343-353, 2018.

[20] Talat R.A, Al-Assaf A.Y, Al-Saffawi A.Y.T, “Valuation of water quality for drinking and domestic purposes using WQI: Case study for groundwater of Al-Jameaa and Al-Zeraee quarters in Mosul city/Iraq”, Journal of Physics Conference Series, 1294(7). doi: 10.1088/1742-6596/1294/7/072011, 2019

[21] Safawi A.Y.T.A, “tatbiq al muasher al kanadi(WQI CCME) le taqeem javed almeyah le agrade alshrub”, in The third Scientific Conference of life sciences, Iraq, 27(5), 193-202, 2019.

[22] Mahmood A, “Evaluation of raw water quality in Wassit governorate by Canadian water quality index”, in *Environmental Engineering and Sustainable Development*, Iraq, 162, 1-8. 2018, doi: 10.1051/matecconf/201816205020.

[23] Mosavi A, Ozturk P, Chau K, “Flood Prediction Using Machine Learning Models: Literature Review”, *Water*, 10(11), 1536. doi: 10.3390/w10111536, 2018.

[24] Chen Y, Song L, Liu Y, “A Review of the Artificial Neural Network Models for Water Quality Prediction,” *Applied Sciences*, 10(17), 5776. doi: 10.3390/app10175776, 20 8 2020.

[25] Koranga M., Pant P, Pant D, “SVM Model to Predict the Water Quality Based on Physicochemical Parameters,” *International Journal of Mathematical, Engineering and Management Sciences*, 6(2), 645-659. doi: 10.33889/IJMEMS.2021.6.2.040, 2021

[26] Al-Adhaileh M. H, Alsaade F. W, “Modelling and Prediction of Water Quality by Using Artificial Intelligence,” *Sustainability*, 13(8), 4259. doi: 10.3390/su13084259, 2021

[27] Park S, Jung S, Lee H, “Large-Scale Water Quality Prediction Using Federated Sensing,” *Sensors*, 21(4), 1462. doi: 10.3390/s21041462, 2021.

[28] Kadiwal A., “Water Quality, Drinking water potability,” Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/adityakadiwal/water-potability. [Accessed: March 9, 2022].

[29] Pérez F, Granger B, “jupytercon,” jupyter, 2014. [Online]. Available: https://jupyter.org/. [Accessed: March 5, 2022].

[30] Scikit-learn authors, “1. Supervised learning,” scikit-learn, 2022. [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. [Accessed: April 10, 2022].

[31] Developers, “NumPy 1.23.0 released,” numpy, 2022. [Online]. Available: https://numpy.org. [Accessed: April 10, 2022].

[32] Developers, “pandas: powerful Python data analysis toolkit,” pypi, 2022. [Online]. Available: https://pypi.org/project/pandas/. [Accessed: April 10, 2022].

[33] Developers, “seaborn: statistical data visualization,” seaborn, 2021. [Online]. Available: https://seaborn.pydata.org/. [Accessed: April 10, 2022].

[34] Developers, “Matplotlib: Visualization with Python,” matplotlib, 2022. [Online]. Available: https://matplotlib.org/. [Accessed: April 10, 2022].

[35] Brownlee, J., “What is the Difference Between a Parameter and a Hyperparameter?,” machine learning mastery,

26 6 2017. [Online]. Available: https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/. [Accessed; June 10, 2022].

[36] Yıldırım S, "6 Must-Know Parameters for Machine Learning Algorithms," towards data science, 2022. [Online]. Available: https://towardsdatascience.com/6-must-know-parameters-for-machine-learning-algorithms-ed52964bd7a9. [Accessed: June 10, 2022].

[37] Yıldırım S, "L1 and L2 Regularization — Explained," towardsdatascience, 2020. [Online]. Available: https://towardsdatascience.com/l1-and-l2-regularization-explained-874c3b03f668. [Accessed: June 10, 2022].

[38] Developers, "sklearn.ensemble.HistGradientBoostingClassifier," scikit-learn, 2022. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingClassifier.html. [Accessed: June 11, 2022].

[39] DigitalSreeni, Director, 184 - Scheduling learning rate in keras. [Video]. United States: Site: YouTube, 2020. URL: https://youtu.be/drcagR2zNpw.

[40] Developers, "Sklearn.tree.DecisionTreeClassifier," scikit-learn, 2022. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html [Accessed: June 11, 2022].

[41] Bhatt B, Director, Decision Tree Hyperparameters : max_depth, min_samples_split, min_samples_leaf, max_features. [Video]. India: Site: YouTube, 2019. URL: https://www.youtube.com/watch?v=XABw4Y3GBR4&t=365s.

[42] Paper D, "Scikit-Learn Classifier Tuning from Complex Training Sets," in Hands-on Scikit-Learn for Machine Learning Applications, Logan, UT, USA, Apress, Berkeley, CA, 2020. doi: 10.1007/978-1-4842-5373-1_6.

[43] Alwanas A.A.H, Al-Musawi A.A, Salih S.Q, "Load-carrying capacity and mode failure simulation of beam-column joint connection: Application of self-tuning machine learning model," *Engineering Structures*, 194, 220-229. doi: c10.1016/j.engstruct.2019.05.048, 2019.

[44] Tung T. M, Yaseen Z. M, "A survey on river water quality modelling using artificial intelligence models: 2000--2020", *Journal of Hydrology*, vol. 585, 124670. doi: 10.1016/j.jhydrol.2020.124670, 2020.

[45] QI C, Huang S, Wang X, "Monitoring Water Quality Parameters of Taihu Lake Based on Remote Sensing Images and LSTM-RNN," *IEEE Access*, vol. 8, 188070. doi: 10.1109/ACCESS.2020.3030878, 2020.

[46] Soumik S.K, "How to Calculate Confusion Matrix Manually.", medium, (2020). [Online]. Available: https://medium.com/analytics-vidhya/how-to-calculate-confusion-matrix-manually-14292c802f52. [Accessed: June 22, 2022].

[47] Ho J.Y, Afana H.A, El-Shafie A.H, "Towards a time and cost-effective approach to water quality index class," *Journal of Hydrology*, vol. 575, 148-165. doi: 10.1016/j.jhydrol.2019.05.016, 2019.

[48] Atha R, "Building Classification Model with Python," medium, (2021). [Online]. Available: https://medium.com/analytics-vidhya/building-classification-model-with-python-9bdfc13faa4b. [Accessed: June 22, 2022].

[49] Sasaki Y., "The truth of the F-measure," School of Computer Science, University of Manchester, 2007.

[50] Wiryaseputra M, "Water Quality Prediction Using Machine Learning Classification Algorithm", *International Journal of Scientific & Engineering Research*, 8(9). doi: 10.14299/000000, 2022.