

Evaluation of decision tree algorithms for predictive machine learning modelling of potable water quality

Debabrata Roy

`dranthro_rs@caluniv.ac.in`

University of Calcutta

Abhiman Das

University of Calcutta

Soumendra Nath Talapatra

Seacom Skills University

Subrata Sankar Bagchi

University of Calcutta

Research Article

Keywords: Climate change, Drinking water, Machine learning, Potable water, Predictive accuracy

Posted Date: October 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11085415/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study was attempted to identify the highest performance-classifying ML algorithms on the dataset of water quality and the most reliable results to be used for further applications. The water quality dataset such as pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity and Potability as safe and unsafe were retrieved from Kaggle dataset. For ML-based prediction, the algorithm of seven types such as DecisionStump (DS), Hoeffding Tree (HT), Pruned and unpruned decision tree C4 (J48), Logistic Model Tree (LMT), Random Forest (RF), Random tree (RT) and Fast Decision Tree Learner (REPT) were studied by using WEKA (Waikato Environment for Knowledge Analysis) tool (version, 3.8.5). Overall prediction of algorithm model classification, the correctly classified models were observed the highest values in RF (69.28%). In the case of predictive analysis with different statistical methods of studied ML algorithms, better value of precision (69.2% and 69.5%), F-measure value (77.2% and 53.0%), ROC area under curve (AUC) value (71.7%) and PRC value (76.1% and 65.4%) were obtained for RF on safe and unsafe water quality. This work develops water quality modelling as per ML-based algorithms that may be useful for easy predicting the water quality. Herein, RF algorithm obtained better predictive performance.

INTRODUCTION

The earth as a whole is undergoing increasing weather extremes as a consequence of climate change, which is causing serious damage to ecosystems as well as human beings, especially in many regions of globe (de la Hera-Portillo et al. 2020; Guevara-Ochoa et al., 2020; Salila et al., 2020; Abalasei et al., 2025; Salehi Siavashani et al., 2026) and also in India (Tiwari & Srinivasa Rao, 2025). Changes in global temperatures, precipitation patterns, extended droughts, heavy rainfall, floods, sea-level rise, and an increase in the frequency of extreme weather events may all have a major impact on both groundwater and surface-water resources (Dahal et al., 2025).

It has a major impact on the water quality for drinking purposes in many of water bodies like rivers, lakes, and coastal ecosystems (Delpla et al., 2009; Xia et al., 2016; Dao et al., 2024; Shajedul Islam & Mostafa, 2024;). By inducing the sources, migration, conversion, and ecological processes of contaminants, it directly and indirectly has a potential impact on water quality (Li et al. 2020).

Health depends on clean water, which is also a fundamental human right and a part of effective health protection programs. At the national, regional, and local levels, clean water is crucial to better health and development. In certain places, investments in water supply and sanitation have been found to have a net economic benefit because they reduce unnecessary medical costs and unfavorable health consequences to execute with proper costing (Kaddoura, 2022).

Moreover, in India, around 7 crores of people are found failure to safe potable water in which the state of West Bengal reported the deficiency. It is necessary concern for obtaining safe potable water as basic essential need (Chaudhary, 2015).

A combination of biological, physical, and chemical characteristics, such as pH, turbidity, temperature, conductivity, solids, pesticides, nutrients, etc. are used to evaluate the quality of potable water. These parameters may be changed by climate-related changes either directly or indirectly (Jaffar et al., 2020).

Interestingly, the effects of climate change on the quality of potable water are not just restricted to specific water sources, but rather affect the entire water supply chain, from the safeguarding of source water and its treatment to its distribution and household storage (Anwar et al., 2026).

Treatment plants could require more treatment during times of high turbidity, algae blooms, pollution occurrences, or changing chemical features. Warm circumstances might also impact how some disinfection byproducts develop during the water treatment process (Coffey et al., 2018).

Some studies emphasized that the prediction through machine learning (ML) modelling to know the characters of water quality whether safe or unsafe due to the incidence of climate change (Haghiabi et al., 2018; Lu & Ma, 2020; Kaddoura, 2022; Karthick et al., 2024; Salehi Siavashani et al., 2026).

The objective of the study is to identify the highest performance-classifying ML algorithms on the dataset of water quality and the most reliable results to be used for further applications.

RELATED STUDIES

Several studies have documented the unsafe water quality for drinking due to climate change and ML-based predictive accuracy of water quality. Some available published research articles are as follows:

Drinking water quality due to climate change

Delpla et al. (2009) reviewed the role of the main factors (warming and consequences of extreme events) describing climate change impacts on water quality. They focussed on two main points. Firstly, the impacts on water quality of resources like rivers and lakes, which are altering parameters values (physico-chemical parameters, micropollutants and biological parameters) were considered. Finally, the expected impacts on drinking water production and quality of supplied water were also discussed. They confirmed that a degrading trend of drinking water quality in terms of climate change.

Cool et al. (2019) estimated the impact of future variations in temperature and precipitation, which were related to climate change scenarios based on the probability of total Trihalomethanes (TTHM) concentrations exceeding a threshold in drinking water in the Province of Quebec (Canada). They studied 108 drinking water utilities (DWUs). Temperature and precipitation variations from the period 2006–2009 to three predicted periods (2010–2039, 2040–2069, and 2070–2099) were analyzed using two climate models and three emission scenarios. The probability of TTHM threshold exceedances was calculated using a multilevel logistic regression model based on three variables viz. treatment type, temperature, and precipitation and three hierarchical levels such as TTHM samples, DWUs and source water ecosystem. They found a low but significant increase in the probability of TTHM threshold exceedances

over time (between 1.9% and 4.7%). There was also a significant probability difference between seasons (up to 30%) and between treatment types (between 25% and 40%).

Johnson et al. (2022) reviewed the literature to address climate change impacts on practices commonly used to mitigate the water quality of urban stormwater, agriculture, and forestry. They made qualitative inferences about the resilience of different types of practices. This resilience in the context of two factors such as the sensitivity of practice function to changes in climatic drivers, and the adaptability, or relative ease with which a practice can be reformed as change occurs.

Brkić (2023) reported that the lake ecosystems are one of the important water resources for drinking water supply, which are significantly altered by global warming as increasing temperature in which the evaporation from the free lake surface, lowering the lake level, and changes the water quality in Vrana Lake on Cres-island, Croatia. Their results obtained that the mean annual lake surface water temperature (LSWT) increased by $0.47^{\circ}\text{C decade}^{-1}$ ($p < 0.0001$). The increase in the mean annual lake summer surface water temperature (July–September) was $0.44^{\circ}\text{C decade}^{-1}$ ($p < 0.0001$), and the maximum annual LSWT was $0.56^{\circ}\text{C decade}^{-1}$ ($p < 0.0001$).

Han & Bu (2023) investigated the impacts of climate change on the water quality of Baiyangdian Lake (China) in the past 30 years (1991–2020). The results observed that water quality grade, chemical oxygen demand (COD), total phosphorus (TP) concentrations, and annual average and minimum air temperatures of the lake showed significant differences ($p < 0.05$) during the studied period. The concentration of dissolved oxygen (DO) and TP, annual average and minimum air temperatures, and annual precipitation decreased, while the COD and total nitrogen (TN) concentration, annual maximum temperature, and monthly maximum precipitation increased. The annual average and minimum air temperature affected all water quality variables and explained 12.3–54.5% of variation.

Abalasei et al. (2025) reported that climatic factors, such as rising temperatures rising and altering rainfall patterns, are remarkably important because these regulate the processes, which are fundamental to sustaining life on the Earth. They emphasized that climate change enhances the degradation of drinking water quality and increases public health risks.

Compaoré et al. (2025) evaluated the perceptions, attitudes, and behaviours of Drinking Water System (DWS) staff related to climate change (CC) impacts and adaptation in their roles and responsibilities for managing water quality from the source to the consumer's tap. They also explored the determinants or causes attributed to current practices in the use of decision support systems (DSSs) in the context of CC adaptation and drinking water. They studied 316 public servants (staff) from 263 drinking water utilities of Canadian municipalities. Results indicated that only 32% of DWSs staff perceive CC adaptation as a priority and most of the focus groups participants admitted practicing it in a curative form whereas prevention would be more beneficial to them.

ML prediction accuracy of water quality

Haghiabi et al. (2018) investigated the performance of AI techniques including artificial neural network (ANN), group method of data handling (GMDH) and support vector machine (SVM) for predicting water quality components of Tireh River located in the southwest of Iran. Comparison of predictive outcomes of GMDH model with other algorithms shown that the model has acceptable performance for predicting the components of water quality, its accuracy is slightly less than ANN and SVM.

Tung & Yaseen (2020) reported that rising trend of contamination in the river due to the climate change and anthropogenic activities, which leads to the deterioration of water quality (WQ). They proposed artificial intelligence (AI) models are the perfect tools for river WQ monitoring, management, sustainability and future policymaking. This research documented that the state of the art of various AI models implemented for river WQ. According to them, this study covers the model structure, input variability, performance metrics, zonal generalisation investigation and comprehensive assessments of AI-based models proceed in river water quality research.

Zhi et al. (2021) used CAMELS-chem, a new dataset with dissolved oxygen (DO) concentrations from 236 nominally disturbed watersheds across the U.S. The model generally learns the theory of DO solubility and captures its decreasing trend with increasing water temperature. It exhibits the potential of predicting DO in “chemically ungauged basins”, defined as basins without any measurements of DO and broadly water quality in general.

Kaddoura (2022) evaluated the efficacy of utilizing ML-based techniques in order to predict the water quality. Thus, a ML classifier model is built to predict the water quality using a real dataset. The results show that support vector machine and k-nearest neighbour are better predictive performance as per F1-score and ROC AUC values. Nevertheless, the LASSO LARS and stochastic gradient descent are suitable based on recall values.

Bordjihene et al. (2026) evaluated and predicted the RF algorithms by using the third ML03 input combination predicted the highest accuracy (99.3%).

MATERIALS AND METHODS

The water quality dataset of nine parameters such as pH, Hardness, Solids, Chloramines, Sulphate, Conductivity, Organic carbon, Trihalomethanes, Turbidity and class effect of Potability as safe and unsafe were retrieved from Kaggle dataset (<https://www.kaggle.com/datasets/adityakadiwal/water-potability/data>). We used only datasets in all rows without missing data. The standard Kaggle water potability dataset contains 3,276 rows (used 2,006 without missing values).

For ML-based prediction, the algorithm of seven types such as DecisionStump (DS), Hoeffding Tree (HT), Pruned and unpruned decision tree C4 (J48), Logistic Model Tree (LMT), Random Forest (RF), Random tree (RT) and Fast Decision Tree Learner (REPT) were studied by using WEKA (Waikato Environment for Knowledge Analysis) tool (version, 3.9.6) (Frank et al., 2016). In this tool, we evaluated the data after pre-processing and classification for these algorithms for predicting different statistical results.

The prediction accuracy of each ML-based algorithm as per correctly and incorrectly classified instances, Kappa (K) statistics, mean absolute error (MAE) and root mean squared error (RMSE) were studied for 10-folds cross validation (CV) test. The summary of results (%) of such as precision value, F-measure value, Matthew's correlation coefficient (MCC), receiver operating characteristic (ROC) curve and precision-recall curve (PRC), respectively as per potability for each algorithm was retrieved from this tool.

RESULTS

Graphical representation after pre-processing step of dataset (Fig. 1). In the case of pH value, the higher 250 instances (6.68–7.11) were obtained. Regarding hardness value, the higher 258 instances (190.90–199.93) were obtained. For solids value, the higher 231 instances (14973.39–17415.47) were obtained. Regarding chloramines value, the higher 248 instances (7.48–7.91) were obtained. Regarding the Sulphate value, the higher 244 instances (333.40–344.76) were obtained. Regarding conductivity value, the higher 208 instances (422.31–444.38) were obtained. For organic carbon value, the higher 218 instances (14.15–15.06) were obtained. For trihalomethanes value, the higher 249 instances (61.85–66.29) were obtained. Regarding turbidity value, the higher 231 instances (3.64–3.86). For potability class, the higher 1194 instances of safe while lower value of 811 instances of safe.

Table 1 describes the summary results of studied algorithms models such as DS, HT, J48, LMT, RF, RT and REPT related to 10 attributes for the prediction of the accuracy as per 10-folds CV test for water quality dataset. Overall prediction of algorithm model classification, the correctly classified models were observed the highest values in RF (69.28%). The statistical values for Kappa statistic (KS), mean absolute error (MAE) and root mean squared error (RMSE) of studied algorithms also predicted better for this algorithm.

In the case of predictive analysis with different statistical methods of studied ML algorithms (Table 2), modest value of precision (69.2% and 69.5%), F-measure value (77.2% and 53.0%), ROC arear under curve (AUC) value (71.7%) and PRC value (76.1% and 65.4%) were obtained for RF on safe and unsafe water quality. The ROC AUC graph is representing in Fig. 2.

Table 1
Results on different classified instances and statistical data for studied algorithms

SI No.	Algorithms	CCI (%)	ICI (%)	KS	MAE	RMSE
1.	DS	61.15	38.85	0.05	0.47	0.49
2.	HT	61.69	38.30	0.13	0.46	0.48
3.	J48	63.79	36.21	0.17	0.43	0.49
4.	LMT	64.29	35.71	0.44	0.44	0.48
5.	RF	69.28	30.72	0.32	0.41	0.45
6.	RT	58.80	41.20	0.15	0.41	0.64
7.	REPT	62.29	37.71	0.17	0.43	0.50
CCI = Correctly classified instances; ICI = Incorrectly classified instances; DS = DecisionStump; HT = Hoeffding Tree; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; REPT = Fast Decision Tree Learner; KS = Kappa Statistics; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error						

Table 2
Average statistical analysis results of studied algorithms

SI No.	Algorithms	Precision (%)	F-measure (%)	MCC (%)	ROC curve (%)	PRC (%)	Potability Class
1.	DS	60.7	75.1	12.4	52.0	60.5	Safe
		71.6	12.0	12.4	52.0	43.4	Unsafe
2.	HT	63.1	72.8	15.0	60.5	68.2	Safe
		55.7	35.4	15.0	60.5	51.2	Unsafe
3.	J48	63.9	74.8	20.3	61.5	68.4	Safe
		63.4	35.6	20.3	61.5	52.3	Unsafe
4.	LMT	64.0	75.3	21.7	61.7	70.2	Safe
		65.8	35.6	21.7	61.7	55.0	Unsafe
5.	RF	69.2	77.2	34.1	71.7	76.1	Safe
		69.5	53.0	34.1	71.7	65.4	Unsafe
6.	RT	65.9	64.8	15.2	57.6	63.6	Safe
		49.1	50.3	15.2	57.6	44.9	Unsafe
7.	REPT	65.0	71.5	18.1	61.7	67.4	Safe
		55.1	44.2	18.1	61.7	52.5	Unsafe
DS = DecisionStump; HT = Hoeffding Tree; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; REPT = Fast Decision Tree Learner; MCC = Matthew’s correlation coefficient; ROC = Receiver operating characteristic; PRC = Precision-recall curve							

DISCUSSION

Changes in climate variables like rising temperatures, severe hydrological consequences (floods and droughts), evapotranspiration, pollutants load and vegetation fires can cause the water quality alteration of surface water resources viz. rivers, lakes, and reservoirs (Delpla et al., 2009; Tachi et al., 2020; Sequeira et al., 2022; Sukanya & Joseph, 2023; Abalasei et al., 2025; Bordjihene et al., 2026).

Moreover, safe water quality for drinking purposes is matter of great concern because human beings are depending upon safe potable water for their growth and survival (Parween et al., 2022). To obtain better potable water the supply chain should be properly maintained otherwise health hazards may cause.

Therefore, it is essential to establish an effective monitoring and assessment system in order to enhance water quality and to safeguard human and environmental well-being, as well as to empower

administrators to take the requisite steps (Mohd Zebaral Hoque et al., 2022; Lap et al., 2023; Satish et al., 2024).

In the present study, the dataset of two parameters viz. Trihalomethanes and Chloramines were used along with other necessary parameters in which the parameters exceeded the limit the water quality is unsafe to drink. Some studies reported that disinfection by-products (DBP) from chlorination Trihalomethanes and Chloramines can form in the water (How et al., 2016; Villanueva et al., 2023). Excessive disinfectants in drinking water may cause adverse health impacts.

In this context, we used ML-based algorithms for water quality assessment in comparison with safe and unsafe values of different parameters. This study predicted the RF algorithm was modest value of precision (69.2% and 69.5%), F-measure value (77.2% and 53.0%), ROC arear under curve (AUC) value (71.7%) and PRC value (76.1% and 65.4%) were obtained for RF on safe and unsafe water quality.

Many studies indicated that RF algorithm was excellent predictive accuracy regarding water quality assessment (Hussain & Khan, 2020; Lap et al., 2023; Islam et al., 2024; Bordjihene et al., 2026). But we obtained comparatively lower predictive accuracy value from other studies.

The results of this study have important practical implications for selecting models and applying predictive results to dam surface water management. The in-depth comparison of 7 ML models demonstrates the effectiveness of the Tree algorithms. The RF algorithm provides effective support to decision-makers in regions with limited water resources, such as lake water bodies facing stagnation, thermal stratification, and nutrient accumulation. These results facilitate the selection of the most accurate and reliable water quality prediction model, particularly in areas where traditional methods may be less effective and costly.

For choosing models and utilizing predictive findings in the management of surface water quality, the findings of this research have significant practical consequences. The extensive comparison of 7 ML models shows the better predictive efficacy of RF algorithms in predicting water quality between the dataset of unsafe and safe, which is supported by the previous study (Bordjihene et al., 2026). Decision-makers in places with scarce water resources, such as stagnant, thermally stratified, and nutrient-accumulating water bodies, may effectively use this model for assistance. In order to choose the most accurate and trustworthy water quality prediction model, especially in locations where traditional methods can be expensive and less effective, these findings are helpful.

CONCLUSION

The United Nations Sustainable Development Goals-6 are associated with water. Access to safe drinking water is a fundamental human right, a crucial element of good health, and a component of better health-protection strategies. Both health and development depend on clean water. In certain regions, expenditures on water supply and sanitation have been demonstrated to have a net economic advantage since they decrease medical expenses and negative health consequences more than they do to

implement them. Overall, this work develops water quality modelling as per ML-based algorithms that combines feature importance methods for easy predicting the water quality. Herein, RF algorithm obtained better predictive performance.

Declarations

Funding

No funding granted in this study.

Author Contribution

D.R and A.D wrote manuscript text, prepared figures. S.T helped data analysis . SSB helped in Data interpretation. All author reviewed the manuscript.

Acknowledgement

Authors are thankful to the developer for this tool used in the present study.

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

Conflict of interest

No conflict of interest.

Data Availability

The water quality dataset of nine parameters such as pH, Hardness, Solids, Chloramines, Sulphate, Conductivity, Organic carbon, Trihalomethanes, Turbidity and class effect of Potability as safe and unsafe were retrieved from Kaggle dataset. This is not a private dataset. We downloaded the dataset as the .csv file from below link: <https://www.kaggle.com/datasets/adityakadiwal/water-potability/data>

References

1. Abalasei ME, Toma D, Dorus M, Teodosiu C. The impact of climate change on water quality: A critical analysis. *Water*. 2025;17:3108.
2. Anwar M, Shah SIA, Akhtar N, Ullah I, Khan D, Khan JA, Shah NS, Han C. (2026). Impact of climate change on water quality, treatment processes, and human health: A comprehensive review. *Environmental Science: Water Research & Technology*.

3. Bordjihene F, Tachi SE, Bouguerra H, Arrar J. Using machine learning models to predict the surface water quality index: a case study of the Cheffia dam (eastern Algeria). *Front Earth Sci.* 2026;14:1761770.
4. Brkić Ž. (2023). Increasing water temperature of the largest freshwater lake on the Mediterranean islands as an indicator of global warming. *Heliyon*, 9(8).
5. Chaudhary A. (2015). Impact of climate change on safe potable water in West Bengal. *Journal of Environment and Sociobiology*, (Spl Iss), 23–24.
6. Compaoré SC, Delpla I, Behmel S, Rodriguez MJ. Perceptions and attitudes of drinking water supply systems staff towards climate change adaptation for drinking water quality management. *Environ Processes.* 2025;12(1):8.
7. Cool G, Delpla I, Gagnon P, Lebel A, Sadiq R, Rodriguez MJ. Climate change and drinking water quality: Predicting high trihalomethane occurrence in water utilities supplied by surface water. *Environ Model Softw.* 2019;120:104479.
8. Dahal D, Bhattarai N, Silwal A, Shrestha S, Shrestha B, Poudel B, Kalra A. A review on climate change impacts on freshwater systems and ecosystem resilience. *Water.* 2025;17(21):3052.
9. Dao PU, Heuzard AG, Le TXH, Zhao J, Yin R, Shang C, Fan C. The impacts of climate change on groundwater quality: A review. *Sci Total Environ.* 2024;912:169241.
10. de La Hera-Portillo Á, López-Gutiérrez J, Zorrilla-Miras P, Mayor B, López-Gunn E. The ecosystem resilience concept applied to hydrogeological systems: A general approach. *Water.* 2020;12(6):1824.
11. Delpla I, Jung AV, Baures E, Clement M, Thomas O. Impacts of climate change on surface water quality in relation to drinking water production. *Environ Int.* 2009;35(8):1225–33.
12. Guevara-Ochoa C, Medina-Sierra A, Vives L. Spatio-temporal effect of climate change on water balance and interactions between groundwater and surface water in plains. *Sci Total Environ.* 2020;722:137886.
13. Haghiabi AH, Nasrolahi AH, Parsaie A. Water quality prediction using machine learning methods. *Water Qual Res J.* 2018;53(1):3–13.
14. Han Y, Bu H. The impact of climate change on the water quality of Baiyangdian Lake (China) in the past 30 years (1991–2020). *Sci Total Environ.* 2023;870:161957.
15. How ZT, Linge KL, Buseti F, Joll CA. Organic chloramines in drinking water: An assessment of formation, stability, reactivity and risk. *Water Res.* 2016;93:65–73.
16. Hussain D, Khan AA. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Sci Inf.* 2020;13(3):939–49.
17. Islam MJ, Salekin SU, Anzum A, Zaman N, Khan AAA, Sarkar D, Hossain MT. Machine learning-driven water quality index prediction: Enhancing accuracy with gradient boosting and explainable AI for sustainable water monitoring. *Appl Agric Sci.* 2024;2(1):1–14.
18. Jaffar A, Thamrin NM, Ali MSAM, Misnan MF, Yassin AIM. The influence of physico-chemical parameters to determine water quality: A review. *J Electr Electron Syst Res.* 2020;17:108–13.

19. Johnson T, Butcher J, Santell S, Schwartz S, Julius S, LeDuc S. A review of climate change effects on practices for mitigating water quality impacts. *J Water Clim Change*. 2022;13(4):1684–705.
20. Kaddoura S. Evaluation of machine learning algorithm on drinking water quality for better sustainability. *Sustainability*. 2022;14(18):11478.
21. Karthick K, Krishnan S, Manikandan R. Water quality prediction: A data-driven approach exploiting advanced machine learning algorithms with data augmentation. *J Water Clim Change*. 2024;15(2):431–52.
22. Lap BQ, Phan TTH, Nguyen HD, Quang LX, Hang PT, Phi NQ, Hoang VT, Linh PG, Hang BTT. Predicting Water Quality Index (WQI) by feature selection and machine learning: a case study of an Kim Hai irrigation system. *Ecol Inf*. 2023;74:101991.
23. Li X, Li Y, Li G. A scientometric review of the research on the impacts of climate change on water quality during 1998–2018. *Environ Sci Pollut Res*. 2020;27(13):14322–41.
24. Lu H, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*. 2020;249:126169.
25. Hoque MZ, Ab J, Aziz NA, Alelyani S, Mohana M, Hosain M. Improving water quality index prediction using regression learning models. *Int J Environ Res Public Health*. 2022;19(20):13702.
26. Parween S, Siddique NA, Diganta MTM, Olbert AI, Uddin MG. Assessment of urban river water quality using modified NSF water quality index model at Siliguri city, West Bengal, India. *Environ Sustain Indic*. 2022;16:100202.
27. Salehi Siavashani N, Pujades-Garnes E, Vizcarra S, Demichele F, Jurado A. A multivariate analysis and machine learning approach to assess the impact of climate change on water quality. *Sustainable Water Resour Manage*. 2026;12:24.
28. Salila E, Sharma MP, Singh R. Climate Change Impacts on Water Quality of Water Bodies-A Review. *Int J Environ Clim Change*. 2020;10(3):26–37.
29. Satish N, Anmala J, Varma MR, Rajitha K. Performance of Machine Learning, Artificial Neural Network (ANN), and stacked ensemble models in predicting Water Quality Index (WQI) from surface water quality parameters, climatic and land use data. *Process Saf Environ Prot*. 2024;192:177–95.
30. Sequeira MD, Castilho A, Tavares AO, Dinis P. The rural fires of 2017 and their influences on water quality: An assessment of causes and effects. *Int J Environ Res Public Health*. 2022;20(1):32.
31. Shajedul Islam M, Mostafa MG. (2024). Impacts of Climate Change on global freshwater quality and availability: A comprehensive review. *J Water Environ Technol* 22 (1): 1–26 Released on J-STAGE February 10, 2024. *Online ISSN*. [https://doi.org/10.2\(9\), 6](https://doi.org/10.2(9), 6).
32. Sukanya S, Joseph S. (2023). Climate change impacts on water resources: An overview. *Visualization techniques Clim change Mach Learn Artif Intell*, 55–76.
33. Tachi SE, Bouguerra H, Derdous O, Djabri L, Benmamar S. Estimating suspended sediment concentration at different time scales in Northeastern Algeria. *Appl Water Sci*. 2020;10(5):118.

34. Tiwari VM, Srinivasa Rao N. Impact of climate change on water resources in India. *J Geol Soc India*. 2025;101(6):974–7.
35. Tung TM, Yaseen ZM. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J Hydrol*. 2020;585:124670.
36. Villanueva CM, Evlampidou I, Ibrahim F, Donat-Vargas C, Valentin A, Tugulea AM, Echigo S, Jovanovic D, Lebedev AT, Lemus-Pérez M, Rodriguez-Susa M, Luzati A, de Cássia Dos Santos Nery, Pastén T, Quiñones PA, Regli M, Weisman S, Dong R, Ha S, Phattarapattamawong M, Kogevinas S. M. (2023). Global assessment of chemical quality of drinking water: The case of trihalomethanes. *Water Research*, 230, 119568.
37. Xia X, Wu Q, Mou X, Lai Y. Potential impacts of climate change on the water quality of different water bodies. *J Environ Inf*. 2016;25(2):85–98.
38. Zhi W, Feng D, Tsai WP, Sterle G, Harpold A, Shen C, Li L. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? *Environ Sci Technol*. 2021;55(4):2357–68.

Figures

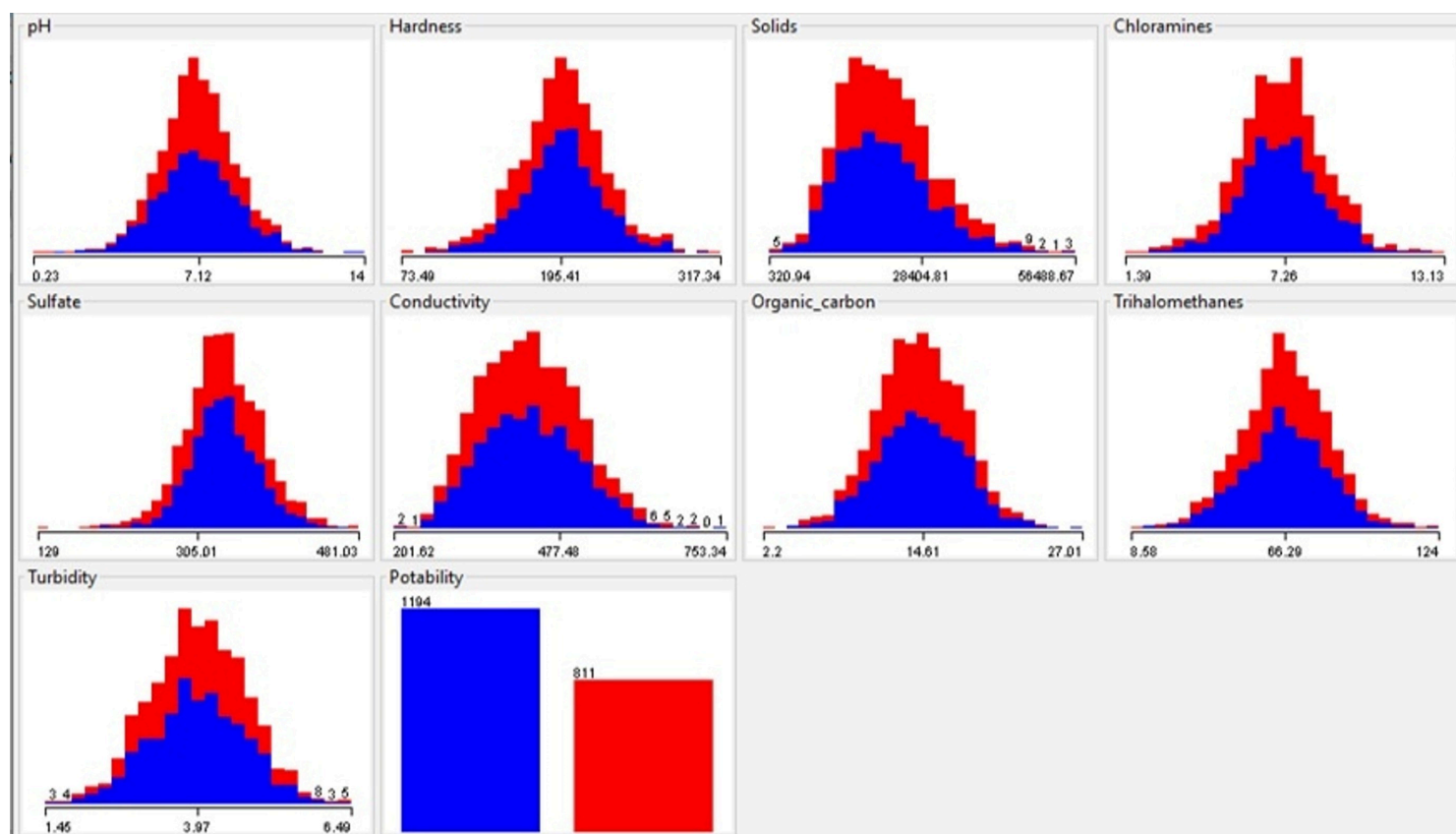


Figure 1

Graphical representation of pre-processing steps of dataset

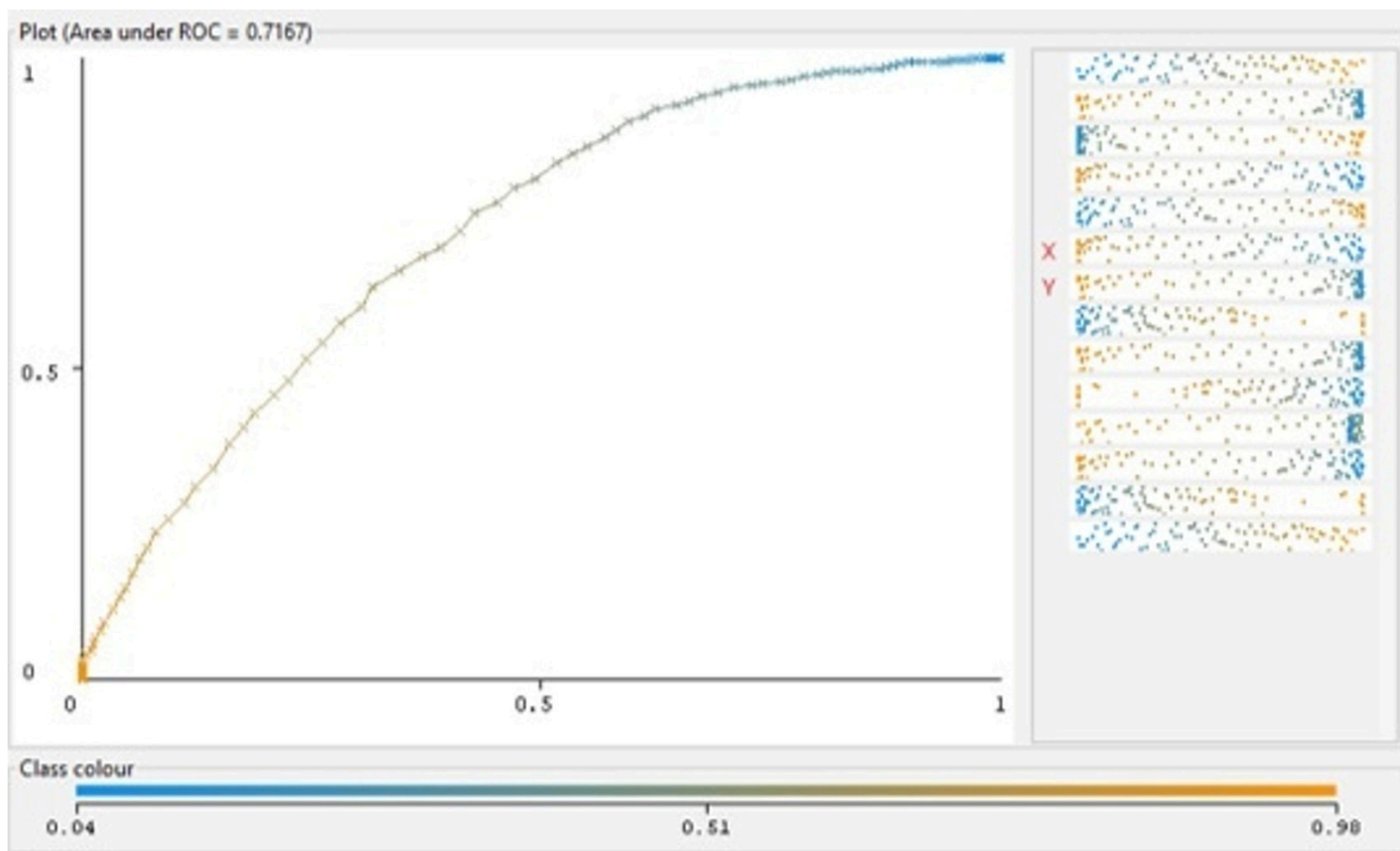


Figure 2

Graphical representation of AUC under ROC curve for RF algorithm