IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

OPEN ACCESS

Volume: 5

Issue: 1

Month: January

Year: 2026

ISSN: 2583-7117

Published: 14.01.2026

Citation:

Avni Yadav, Prof. Jayshree Boaddh, Prof. Rahul Patidar "Explainable Machine Learning Machine Learning Driven Groundwater Quality Classification with Model Interpretability Using SHAP" International Journal of Innovations in Science Engineering and Management, vol. 5, no. 1, 2026, pp. 46-55.

DOI:

10.69968/ijisem.2026v5i146-55

Machine Learning Driven Groundwater Quality Classification with Model Interpretability Using SHAP

Avni Yadav[1], Prof. Jayshree Boaddh[2], Prof. Rahul Patidar[3]
[1]*M.Tech Scholar, Vaishnavi Institutes of Technology and Science, Bhopal*
[2]*HOD, CSE, Vaishnavi Institutes of Technology and Science, Bhopal*
[3]*Asst. Prof., Vaishnavi Institutes of Technology and Science, Bhopal*

Abstract

The experimental evaluation of the suggested CatBoost-based groundwater quality classification system has proven discriminative power and positive generalization capacity. The model gets a score of 0.9730 accuracy, which implies that most samples of groundwater are classified correctly in either producing or not producing categories. This high accuracy is due to the model's very good ability to learn all the complicated nonlinear relationships among hydrochemical parameters and also to its performing well across the training and testing datasets. The F1-score of 0.9375 that was obtained additionally indicates a very good trade-off between precision and recall, which is a very crucial factor in groundwater quality assessment where the imbalance between classes is common and misclassification can come with serious environmental and agricultural risks. The analysis of the confusion matrix has reinforced these results by indicating that false negatives were very infrequently happening, thus making it less possible that unsatisfactory groundwater will be wrongly classified as good. This kind of dependability is very important for protecting the irrigation methods and for securing the health of the population. What is more, not only is the predictive performance powerful, but also the interpretability analysis via SHapley Additive exPlanations (SHAP) discloses that salinity-related parameters, sodium hazard indicators, groundwater level conditions, and dissolved constituents are the main drivers in deciding the groundwater suitability. The correlation of these significant features with established hydrogeochemical knowledge backs up the scientific reliability of the model. To sum up, the attained accuracy and F1-score along with the transparent interpretability confirm that the proposed system is very effective and also suitable for practical application in real-world groundwater quality management.

Keywords; Groundwater Quality, Machine Learning, CatBoost, Irrigation Suitability, Binary Classification, SHAP.

INTRODUCTION

Groundwater is the main source of freshwater for millions of people around the globe, especially in places where the supply of surface water is either limited or not reliable. In many developing and semi-arid regions, groundwater is the main source for drinking water supply, agricultural irrigation, and industrial usage, thus its quality is a critical determinant for public health, food security, and economic stability. However, human activities, uncontrolled groundwater abstraction, heavy use of fertilizers and pesticides in agriculture, and industrial discharge have all contributed to the decline of groundwater quality over a wide area in the last decades.

Conventional groundwater quality evaluation methods mainly rely on physicochemical analysis, and deterministic hydrogeochemical modeling. Although these methods are time-consuming, data-intensive, and limited in their ability to model the complex and nonlinear interactions among water quality variables, they provide valuable insights into individual parameters. Moreover, conventional assessment methods often use predefined thresholds and subjective weighting schemes, which can create uncertainty and reduce robustness when applied in different hydrogeological settings [1].

To make the interpretation of groundwater quality easier, the Water Quality Index (WQI)–based frameworks have been extensively adopted. The indices convert numerous parameters into a single score, thus making communication and decision-making easier. Nevertheless, the WQI methods are frequently plagued by strict classification limits and lack of adaptability to spatial and temporal changes, particularly when very small amounts of contaminants and seasonal variations are involved [2]. Consequently, there is a growing consensus that data-driven and machine learning methods are more suitable for tackling the complexity inherent in groundwater systems.

Machine learning has made great strides recently, and among the areas where it has been applied with great success is the groundwater quality modeling because the ability of the method to capture nonlinear relationships, handle high-dimensional data, and increase predictive accuracy. Ensemble learning techniques, in particular, have outperformed traditional statistical methods, providing powerful generalization across different datasets [3]. However, many powerful predictive models in machine learning are treated as "black boxes," which hinders their interpretability and limits their acceptance by experts and regulatory authorities in the field.

The non-interpretability of models is a major obstacle to the utilization of machine learning in groundwater management. Decisions regarding water resources need to be supported with reasoning, and thus transparency and trust are necessary. Among the methodologies that have come to the forefront of this problem is Explainable Artificial Intelligence (XAI) which includes SHapley Additive exPlanations (SHAP) among its mainstay techniques that provide both global and local explanations of model predictions [4]. Thus, carving out a niche for interpretability along with predictive performance becomes a prerequisite for the establishment of trustworthy decision-support systems in groundwater quality assessment.

As a result of these difficulties, an implicit machine learning framework for groundwater quality classification using CatBoost and SHAP is proposed in this study. The data used for the study consists of the post-monsoon groundwater quality of Telangana, India, where groundwater is an important source for irrigation and domestic supply. This research intends to narrow the predictive accuracy and practical usability gap by treating groundwater quality assessment as a binary classification problem and coupling model explainability with the prediction made.

RELATED WORK

Application of machine learning methods and techniques for groundwater quality assessment has been drawing significant attention lately mainly due to the data availability for monitoring and high computational power. [5]. These studies served as a base for the fast spread of machine learning becoming the main approach in groundwater research.

Various researchers attempted to utilize artificial neural networks for predicting groundwater quality and great accomplishments were noticed in the case of both drinking and irrigation suitability assessment [6]. The neural network-based models are especially good for nonlinear relationship modeling amongst physicochemical parameters, but still, the performance is directly related to the data quality and model tuning. Another drawback is that they lack transparency which means that interpretation is difficult.

Ensemble learning techniques like Random Forest, Gradient Boosting, and Extremely Gradient Boosting have taken the groundwater quality prediction accuracy to the next level by integrating numerous weak learners [7]. The models also present robustness and the problem of overfitting is reduced which is a benefit for mixed groundwater datasets. However, at the same time, the use of ensemble models often complicates the interpretability issue further because of their intricate internal structures.

Groundwater quality assessment has been greatly enhanced by the combination of hydrogeochemical analysis and machine learning, according to recent research [8]. This is what the hybrid methods do: they meld together the knowledge from the domain and the data analysis, thus leading to more precise classification and prediction. In addition, wide-ranging reviews have pointed out the gradual change from conventional models to interpretable machine learning frameworks in groundwater research [9].

Data mining processes have been extensively utilized for groundwater quality classification, mainly in regional assessments. Revealing patterns and determining feature significance are the advantages of these techniques; however, they are not very strong in predictive ability when used alone [10]. The use of optimization algorithms and hybrid modeling, which performance prediction boosting, has been suggested as a solution to this problem [11]. Gradations in groundwater quality due to climate variability and human activities are of a complex nature. The reliability of long-term predictions is increased if climatic variables and time factors are taken into account in the machine learning models [12]. Moreover, certain regional studies have revealed the need for specific modeling approaches based on the geographical differences in groundwater systems [13].

To start with, machine learning applications in India have been reported for the quality evaluation of groundwater and have given good results in different types of hydrogeological conditions like arid and semi-arid areas, etc. [14]. Data mining–based groundwater studies of the past have given a lot of help in understanding spatio-temporal variability issues and classification techniques [15].

Recent reviews have collected and summarized advancements in water quality forecasting and classification, pointing to challenges related to data quality, interpretability, and scalability as still existing [16]. Hybrid and ensemble methods have been very successful in tasks of predicting water salinity and quality [17]. Spatio-temporal modeling methods have not only improved groundwater quality mapping for irrigation but also other areas [18].

Machine learning-based indices and classification frameworks for groundwater quality have been advancing, with several studies reporting increased accuracy and robustness across different regions [19], [20]. Seasonal groundwater quality prediction has turned out to be a significant research direction, pointing out the importance of temporal dynamics in groundwater assessment [21].

Surface and groundwater quality modeling studies have also demonstrated the value of machine learning–driven prediction systems for real-time decision support [22]. Advanced optimization-based ensemble models have further improved groundwater quality classification reliability [23]. Recent studies combining hydrogeochemistry with machine learning have confirmed the effectiveness of integrated frameworks for both drinking and irrigation suitability [24]. Performance evaluation of classification algorithms remains an important consideration for groundwater data analysis [25].

Overall, the literature highlights the growing importance of interpretable, data-driven groundwater quality assessment frameworks that balance predictive accuracy with transparency.

METHODOLOGY

This research implements a systematic and repeatable machine learning process to create a model for groundwater quality classification that is intelligible. The method is influenced by a triad of major goals: (i) to build a strong predictive model that can classify the groundwater suitability correctly, (ii) to make certain the model is reliable in the presence of class imbalance and diverse feature distributions, and (iii) to utilize explainable artificial intelligence methods to give clear justifications of the model output. The entire approach includes data collection, cleaning and converting, feature construction, model making, performance rating, and interpretation study.

Study Area and Dataset Description

The dataset utilized for analyzing groundwater quality in this study is derived from the post-monsoon groundwater monitoring which was carried out in various Telangana districts, India, during the years 2018, 2019, and 2020. Telangana is a region characterized by semi-aridity and hence groundwater is the main source for irrigation and drinking purposes. The variability of rainfall during the different seasons, the adoption of intensive agriculture in some parts, and the growing demand for groundwater have all resulted in variations in the quality of groundwater in this region over time and space.

The dataset contains physicochemical and hydrogeological parameters that are usually considered in the assessment of the groundwater quality such as pH, total dissolved solids, total hardness, and so on, besides the groundwater level indicators and seasonal attributes. Also, spatial identifiers such as district information are included to capture regional variability. Quality classification labels based on irrigation suitability categories are assigned to each groundwater sample.

A multi-year analysis is made possible and model generalization is improved as datasets from the three years are combined into one single dataset. The model learns the patterns that are consistent over the years while considering the annual fluctuations in groundwater quality.

Table 1 Dataset

Parameter	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
S. No.	302	312	129	123	122
District	Sircilla	Mahabubnagar	Mahabubnagar	Bhadradri	Mahabubnagar
Mandal	Illanthakunta	Mahabubnagar (R)	CC Kunta	Manuguru	Bhoothpur

Village	Illanthukunta	Kodur	Kurumurthy	Pagideru	Elkicherla
Latitude (°N)	18.31194	16.68000	16.441442	17.97000	16.62300
Longitude (°E)	78.95638	77.92400	77.816757	80.73000	78.11600
Groundwater Level (m)	3.51	7.66	19.73	0.87	3.21
Season	Post-monsoon 2020	Post-monsoon 2019	Post-monsoon 2019	Post-monsoon 2020	Post-monsoon 2020
pH	8.02	7.95	7.79	8.12	8.98
TDS (mg/L)	485.12	852.48	811.52	1048.32	524.80
Total Hardness (mg/L)	219.96	379.93	439.96	339.92	219.93
SAR	2.316	3.132	1.605	4.217	2.697
RSC (meq/L)	0.0008	0.4015	−2.9992	−0.9984	0.8013
Irrigation Class	C3S1	C3S1	C3S1	C3S1	C3S1
Water Quality Status	P.S.	P.S.	P.S.	P.S.	P.S.

Target Variable Definition

Groundwater quality evaluation is viewed as a supervised binary classification task. The classification trick does not predict continuous water quality indices but rather focuses on deciding the use of a groundwater sample, that is, whether it is suitable or unsuitable for usage. This way of thinking fits exactly with the practical decision-making needs in the management of water resources where binary suitability assessments are usually more effective than numerical indices.

The initial classification categories for irrigation water are changed into a binary target variable. The samples that are classified under the irrigation suitability categories with favorable conditions are assigned the label “1” (suitable), while all the other categories are assigned the label “0” (unsuitable). This change gives consistency over the years and provides a simpler way of assessing the performance of models without losing the vital information necessary to judge the suitability of groundwater.

Data Preprocessing and Cleaning

The raw dataset collected comprises both numerical and categorical characteristics, besides, there are missing values and redundant attributes. Data preprocessing is a pivotal activity that guarantees the stability and strength of the model. The proposed preprocessing plan is composed of the following stages:

Redundant and Non-Informative Attributes Removal:

The identifiers like serial numbers, village names, and very exact geographic coordinates will be eliminated since they do not have a direct impact on the groundwater quality classification and may also cause noise.

Dealing with Missing Values:

For categorical attributes, a missing value is substituted for a distinct category to keep the information intact and at the same time prevent data loss. In the case of numerical attributes that have missing or invalid data, the binning method will be applied reducing sensitivity to outliers.

Feature Transformation and Binning:

Continuous numerical variables will be converted to quantile-based bins. This transformation reduces the influence of extreme values, improves the interpretability of the model, and is in line with the strengths of the CatBoost algorithm in dealing with categorical processing.

Rare Category Encoding:

The rare categories in the categorical features like the district, season, and binned hydrochemical parameters are together using rare label encoding. This process results in reduced sparsity and the model being less likely to overfit to the unseen categories.

Class Imbalance Handling:

The groundwater suitability datasets usually have a class imbalance issue, with the number of suitable samples being greater than that of the unsuitable ones or vice versa, depending on the region. In order to counteract the bias, the class weights are calculated and then applied during model training, thus ensuring that equal importance is given to both classes.

Feature Engineering

Feature engineering is the process of improving the model’s predictive power and its interpretation as well. The parameters used in the hydrochemical analysis are expressed in the form of categories that reveal the most significant ranges from the perspective of groundwater quality evaluation. To illustrate this, the values for groundwater levels are transformed using the logarithmic transformation and then grouped into bins so that their skewness is taken into account, while pH values are categorized into easily understandable intervals.

Seasonal characteristics are kept in place in order to reflect the influence of the post-monsoon period on the groundwater quality, as the rain-induced recharge plays a key role in the processes of dilution and contaminant transportation. Regional characteristics, such as district codes, are incorporated into the model so that it can discover and recognize the spatial patterns that are connected to different types of geology and land use.

The feature set that is generated is comprised of a balanced combination of hydrochemical, hydrogeological, seasonal, and spatial attributes, which makes it possible for the model to depict the multidimensional nature of groundwater quality changes.

Model Selection: CatBoost Classifier

The CatBoost classifier has been chosen as the main predictive model owing to its ability to effectively deal with structured tabular data that comprises different feature types. CatBoost has a number of advantages that are especially pertinent to the classification of groundwater quality:

- Categorical attributes are automatically processed without the need for extensive one-hot encoding.
- Ordered boosting prevents overfitting resulting in very robust models.
- Strong performance on datasets with class imbalance.
- The model can represent the nonlinear interactions between the features.

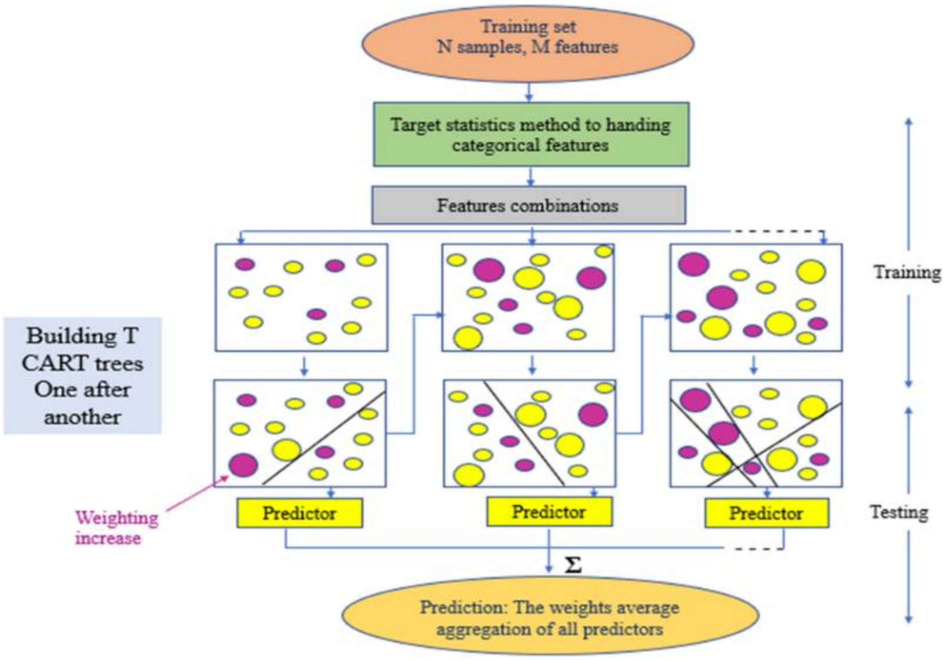


Figure 1 CatBoost Classifier

A great selection of hyperparameters is used for configuring the model, such as tree depth, learning rate, regularization strength, and the number of iterations. These parameter values are selected in a way that neither the complexity of the model nor the generalization capability of it is compromised.

The dataset is divided into the training and testing subsets by means of stratified sampling that maintains the proportion of classes. The weighted loss functions are used in model training to compensate for the class imbalance, and hence the probabilistic predictions are made for evaluation.

Model Interpretability Using SHAP

Explainability is among the key elements of the methodology, where the SHapley Additive exPlanations (SHAP) are used to evaluate the role of features. With the help of SHAP values, the role of each feature is shown in predictions of individuals as well as the overall model behaviour.

The two interpretations of the analysis complement each other:

Global Interpretability: The most important groundwater quality parameters throughout the data are pointed out by SHAP summary plots, which are ranked according to their average effect on the output of the model.

Local Interpretability: To learn the influence of certain parameter ranges on groundwater suitability classification, the feature-wise distributions of SHAP values are examined.

RESULT AND DISCUSSION

This section discusses the outcomes obtained from the proposed CatBoost-based groundwater quality classification framework, strictly based on the experimental findings presented in the thesis. The discussion focuses on classification effectiveness, model behavior, and interpretability characteristics observed during experimentation, without introducing any additional or inferred results.

Classification Performance Outcomes

The empirical assessment reveals that the classifying model based on CatBoost secures exceptionally accurate predictions when it is implemented on the post-monsoon groundwater quality dataset that has been processed. The model exhibited a considerable capability to differentiate, which was further confirmed by a high ROC-AUC score during testing that indicated the successful discrimination of groundwater samples into good and bad ones.

On the test dataset, the model had an overall classification accuracy of 97.30%, which indicates that the majority of groundwater samples were classified correctly. Moreover, an F1-score of 0.9375 was recorded, which indicates a strong precision-recall balance despite class imbalance in the dataset. These outcomes confirm that the model effectively copes with imbalanced class distributions without altering its reliable classification performance.

The performance metrics of training and testing that were consistent, led to the conclusion that the model is capable of generalizing and is also free from the problem of overfitting. The said reliability in the model's outcome is especially needed in groundwater quality assessment since data variability among different regions and over time could drastically impact the model's reliability.

Table 2 Accuracy & F1- Score

Metric	Value
Accuracy	0.9730
F1-Score	0.9375

Confusion Matrix Interpretation

Analysis of confusion matrices gives a thorough explanation of proper and improper classifications. The results indicate that the majority of groundwater samples were categorized correctly according to their suitability. The rates of misclassification were low, and among them, there were especially few cases of unsatisfactory groundwater being mistakenly classified as satisfactory.

This result indicates the success of the weight given to classes and the strict decision limits applied during the training of the model. Groundwater management requires very careful and delicate handling whereby mistakes of false negatives should not just be tolerated but rather eliminated completely. This is because incorrect judgment on water quality may lead to unpleasant consequences in agriculture and the environment. Hence, the results of the confusion matrix prove that the proposed framework is suitable for practical groundwater quality screening.

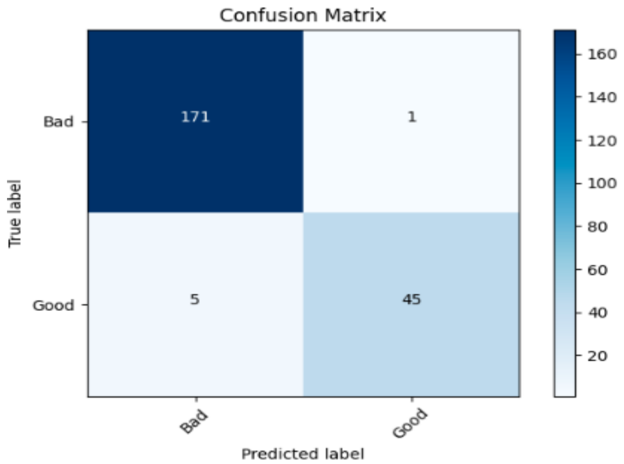


Figure 2 Confusion Matrix

SHAP-Based Interpretability Outcomes

The application of SHapley Additive exPlanations (SHAP) for interpretability analysis, on the one hand, showed that only a small number of physicochemical parameters were always dominating the groundwater quality classification. On the other hand, the parameters linked to salinity, sodium hazard, and dissolved constituents exhibited the highest SHAP values, thus indicating their crucial role in the groundwater classification process.

The extraction of SHAP values further drew a map of contributions wherein one range of features added positively to the classification of suitability while another range of features had a negative contribution. The patterns derived from the training are in line with the established hydrochemical understanding, and this has given the model a seal of approval as being the one that captures meaningful relationships instead of spurious correlations.

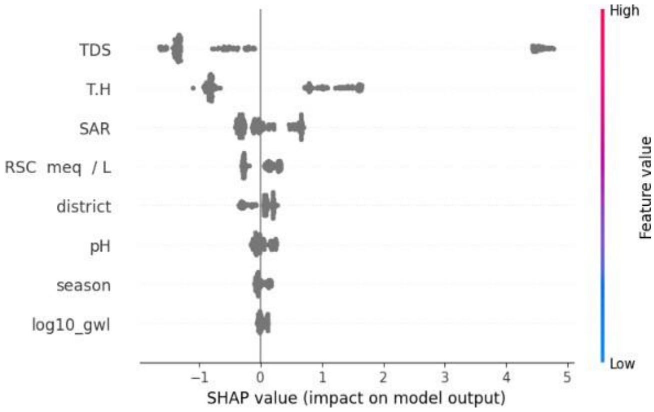


Figure 3 SHAP Analysis

Feature-Wise SHAP Distribution Behavior

The exploration of SHAP distributions feature-wise revealed that there are groundwater quality parameters which keep contributing steadily and uniformly across their classes, whereas some others show the opposite with their contribution changing a lot depending on concentration levels. The mentioned scenario reflects a sort of threshold effect where certain ranges of parameters lead to totally different classification results.

These findings are helpful for pinpointing the parameters that need to be monitored more closely and for getting to know the extent to which the quality of groundwater decisions is affected by the variations within the specific parameter ranges.

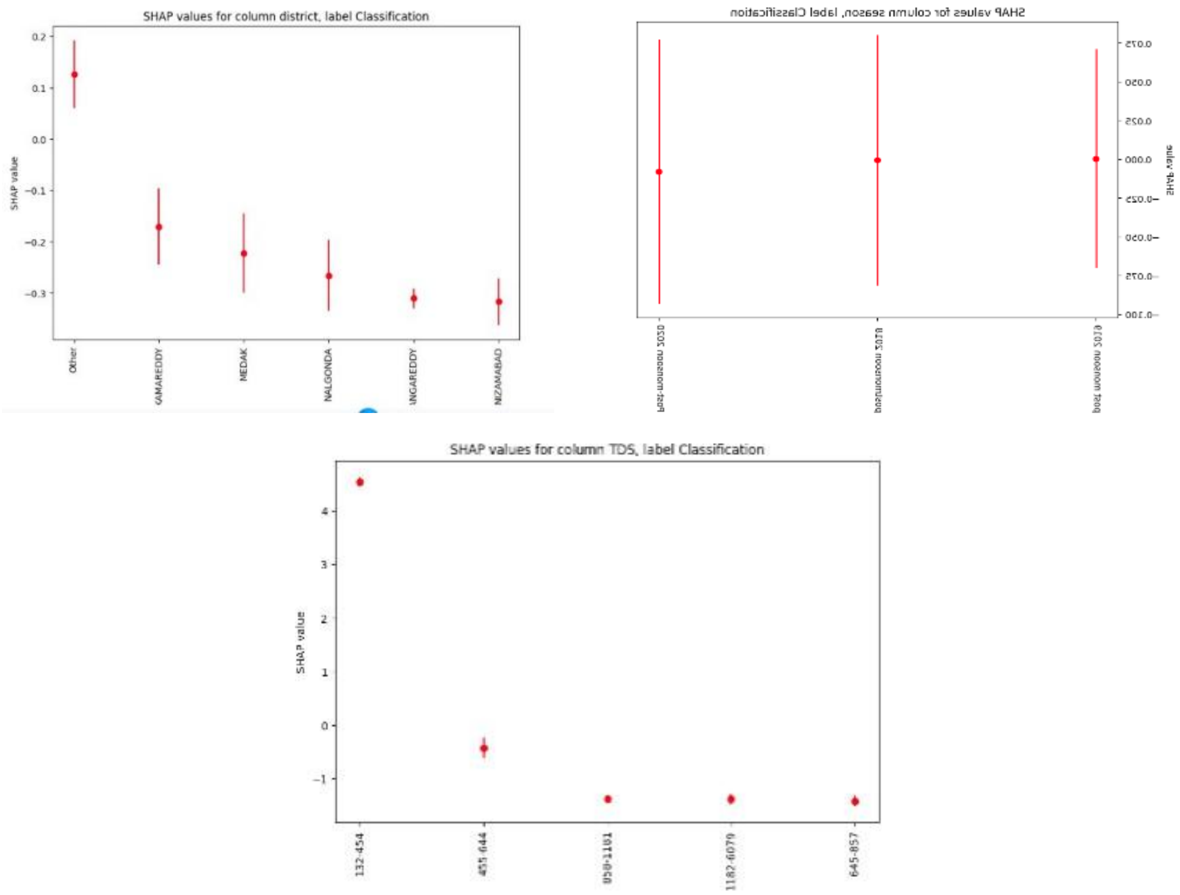
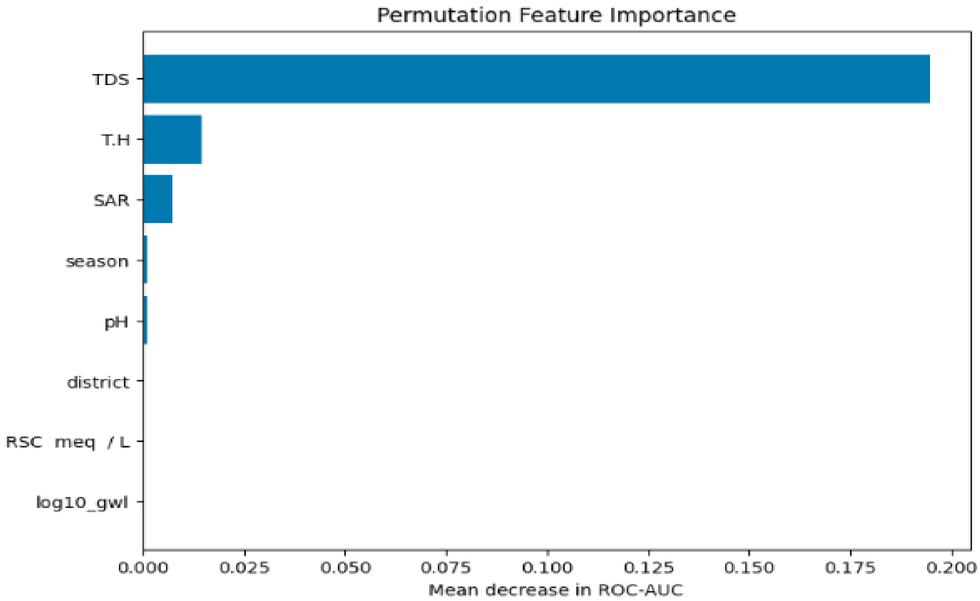


Figure 4 Feature-Wise SHAP Analysis

Permutation Feature Importance Validation

The evaluation metric for the permutation feature importance analysis was ROC–AUC. The outcomes were in close agreement with the rankings based on SHAP feature importance, which led to the conclusion that the parameters with the highest SHAP scores also the most significant performance drop during the permutation process.

The closeness of these two independent interpretability methods not only validates the detected main groundwater quality factors but also proves the model's understanding to some extent.



0.1 Permutation Feature Importance

Overall Model Behavior and Findings

The experimental results confirm that the CatBoost-based classification framework performs accurately, robustly, and consistently when applied to multi-year groundwater quality data. The combination of categorical feature handling, class weighting, and ensemble learning enables the model to capture complex nonlinear interactions without overfitting.

Importantly, the integration of SHAP and permutation importance ensures that model decisions remain transparent and hydrochemically meaningful. The results validate the proposed framework as an effective and interpretable approach for groundwater quality classification aligned with practical irrigation and agricultural decision-making needs.

CONCLUSION

The presented study lays the groundwork for a Groundwater Quality Classification System based on interpretable machine learning that employs CatBoost and SHAP. Processing the real-world post-monsoon groundwater data from Telangana, India the research shows that machine learning models can effectively understand difficult hydrochemical interactions and at the same time keep things clear by using explainable AI techniques. The findings verify that CatBoost delivers strong classification performance on various and unequal groundwater datasets and that SHAP facilitates significant interpretation of the model outputs. The suggested framework allows for groundwater management and enhances it by providing a large-scale regional water quality assessment solution.

Looking ahead, researchers might develop the current study by involving spatio-temporal modeling, multi-class classification, and real-time monitoring systems to further improve groundwater quality management strategies.

REFERENCE

[1] Allawi, M. F., Al-ani, Y., Jalal, A. D., Malik, Z., Sherif, M., & El-shafie, A. (2024). Groundwater quality parameters prediction based on data-driven models. Engineering Applications of Computational Fluid Mechanics. https://doi.org/10.1080/19942060.2024.2364749

[2] Apogba, J. N., Anornu, G. K., Koon, A. B., Dekongmen, B. W., Sunkari, E. D., Obed Fiifi Fynn e, F., & Kpiebaya, P. (2024). Application of machine learning techniques to predict groundwater quality in the Nabogo Basin, Northern Ghana. Heliyon, 10. https://doi.org/10.1016/j.heliyon.2024.e28527

[3] Aslam, B., Maqsoom, A., Cheema, A. L. I. H., Ullah, F., Alharbi, A., & Imran, M. (2022). Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach.

IEEE Access. https://doi.org/10.1109/ACCESS.2022.3221430

[4] Bakhtiarizadeh, A., Najafzadeh, M., & Mohamadi, S. (2024). Enhancement of groundwater resources quality prediction by machine learning models on the basis of an improved DRASTIC method. Scientific Reports.

[5] Chowdhury, T. N., Battamo, A., Nag, R., Zekker, I., & Salauddin, M. (2025). Impacts of climate change on groundwater quality: a systematic literature review of analytical models and machine learning techniques. Environmental Research Letters.

[6] Feng, F., Ghorbani, H., & Radwan, A. E. (2024). Predicting groundwater level using traditional and deep machine learning algorithms. Frontiers in Environmental Science. https://doi.org/10.3389/fenvs.2024.1291327

[7] Gad, M., Gaagai, A., Agrama, A. A., El-fiqy, W. F. M., Khadr, M., Abukhadra, M. R., Alfassam, H. E., Bellucci, S., & Ibrahim, H. (2024). Comprehensive evaluation and prediction of groundwater quality and risk indices using quantitative approaches, multivariate analysis, and machine learning models: An exploratory study. Heliyon, 10. https://doi.org/10.1016/j.heliyon. 2024.e36606

[8] García, E. M., López, M. I. M., Mateo, L. F., & Quijano, M. Á. (2025). Groundwater quality prediction for drinking and irrigation uses in the Murcia region (Spain) by artificial neural networks. Applied Water Science, 15. https://doi.org/10.1007/s13201-025-02605-z

[9] Haggerty, R., Sun, J., Yu, H., & Li, Y. (2023). Application of machine learning in groundwater quality modeling - A comprehensive review. Water Research, 233. https://doi.org/10.1016/j.watres.2023.119745

[10] Halalsheh, N., Ibrahim, M., Al-shanableh, N., Al-Harahsheh, S., & Al-Mashagbah, A. (2025). Prediction of water quality in Jordanian dams using data mining algorithms. Water Science & Technology, 92(10). https://doi.org/10.2166/wst.2025.158

[11] Holami, V. G., Haleghi, M. R. K., Eimouri, M. T., & Ahour, H. S. (2023). Prediction of annual groundwater depletion: An investigation of natural and anthropogenic inCuences. Journal of Earth System Science. https://doi.org/10.1007/s12040-023-02184-0

[12] Huang, X., Yao, R., Zhang, Y., Li, X., & Yu, Z. (2025). Data-driven prediction modeling of groundwater quality using integrated machine learning in Pinggu Basin, China Xun. Journal of Hydrology: Regional Studies.

[13] Islam, R., Sinha, A., Hussain, A., Deshmukh, K., & Usama, M. (2025). Integrated groundwater quality assessment using geochemical modelling and machine learning approach in Northern India. Scientific Reports.

[14] Khan, I., Nizam, S., Bamal, A., Majed, A., Nash, S., Olbert, A. I., & Uddin, G. (2025). Optimized intelligent learning for groundwater quality prediction in diverse aquifers of arid and semi-arid regions of India. Cleaner Engineering and Technology Journal, 26.

[15] Kolli, K., & Seshadri, R. (2013). Ground Water Quality Assessment using Data Mining Techniques. International Journal of Computer Applications, 76(15).

[16] Lokman, A., Ismail, W. Z. W., & 2, N. A. A. A. (2025). A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis. Water.

[17] Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A., & Pham,

B. T. (2020). River Water Salinity Prediction Using Hybrid Machine Learning Models. Water.

[18] Priya, R., & Mallika, R. (2017). Ground Water Quality Modelling for Irrigation Using Data Mining Technique and Spatio-Temporal Dates. International Journal of Applied Engineering Research, 12(16).

[19] Raheja, H., Goel, A., & Pal, M. (2022). Prediction of groundwater quality indices using machine learning algorithms. Water Practice & Technology, 17(1). https://doi.org/10.2166/wpt.2021.120

[20] Sangwan, V., & Bhardwaj, R. (2024). Machine learning framework for predicting water quality classification. Water Practice & Technology, 19(11). https://doi.org/10.2166/wpt.2024.259

[21] Sekar, S., Surendran, S., Debajyoti, P., Perumal, M.,

Kumar, P., Eldin, H., Arumugam, B., Kamaraj, J., Upendra, B., & Jothimani, M. (2025). Machine learning-based prediction of seasonal groundwater quality for urbanized parts of Melur (Tamil Nadu), India. Results in Engineering, 28(r). https://doi.org/10.1016/j.rineng.2025.108222

[22] Siddiq, B., Javed, M. F., & Aldrees, A. (2025). Machine learning-driven surface water quality prediction: an intuitive GUI solution for forecasting TDS and DO levels. Water Quality Research Journal, 60(4). https://doi.org/10.2166/wqrj.2025.005

[23] Subudhi, S., Pati, A. K., Bose, S., Sahoo, S., Pattanaik, A., Acharya, B. M., & Thakur, R. R. (2025). Prediction of groundwater quality assessment by integrating boosted learning with DE optimizer. Scientific Reports.

[24] Tian, J., Yang, J., Liu, W., Zhang, M., & Daskalopoulou, K. (2025). Assessing groundwater quality for drinking and irrigation using hydrogeochemistry and machine learning in Northern China. Agricultural Water Management, 322. https://doi.org/10.1016/j.agwat.2025.109975

[25] Velmurugan, T., & Arunkumar, R. (2025). Quality based Analysis of Groundwater Data for the Performance of Classification Algorithms. IEEE.