ACS Sustainable Resource Management

pubs.acs.org/acssrm

Article

Predicting Uranium in Punjab's Groundwater Using Common Anions: A Machine Learning Approach

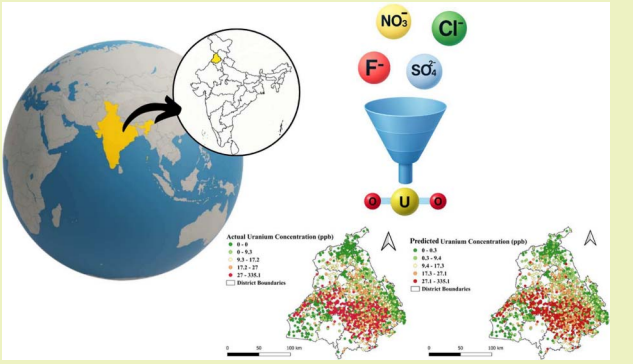
Sankar Sudhir,[Ψ] Vamanie Perumal,[Ψ] Tanmayaa Nayak, and Thalappil Pradeep*

Cite This: https://doi.org/10.1021/acssusresmgt.5c00505

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Accurately predicting uranium concentration in groundwater is vital for protecting public health. Traditional spectroscopic detection methods, although effective and reliable, are time consuming and demand equipment and specialists. In addition, the inclusion of uranium in water quality investigations is rarely a routine practice. We identified Punjab, India as a probable polluted region and measured the risk of uranium contamination using machine learning (ML) models. Uranium was predicted based on water quality parameters that are regularly tested within the DWSS Punjab state drinking water surveillance program, including arsenic, cadmium, mercury, nickel, iron, lead, chromium, nitrate, chloride, fluoride, and sulfate. We employed a large dataset comprising 8735 samples to develop regression models based on Gradient Boosting, Random Forest, and XGBoost algorithms. Furthermore, ensemble methods such as weighted averaging, stacking, and voting provided competitive alternatives and valuable diversity. To account for the variations in water quality characteristics, we introduced K-means and Gaussian mixture clustering and the best regressors were chosen for each cluster. Our results demonstrate that the integration of clustering, regression, and ensemble learning boosted the overall predictive performance. The highest performance (R^2 > 90%) was achieved by the K-means-based XGBoost regressor. Among the available predictors, sulfate, fluoride, nitrate, and chloride emerged as the most informative variables in the ML framework, consistent with their association with hydrochemical regimes linked to elevated uranium. This study contributes to a statistical framework for uranium monitoring and risk reduction, thereby aiding public health and sustainable agriculture in Punjab.



KEYWORDS: *uranium prediction, machine learning, clustering, ensemble learning, predictive modeling*

Downloaded via INDIAN INST OF TECH MADRAS on March 29, 2026 at 17:46:33 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

Groundwater is a major source of drinking water in many countries owing to its ease of treatment as compared to surface water. In addition to anthropogenic sources of pollution of groundwater, it is noteworthy mentioning that the presence of key natural species leads to significant health complications. Countries such as Sweden, Finland, Norway, USA, Canada, and India show significant levels of naturally occurring radioactive elements in groundwater, including uranium. Uranium is a radioactive element with three naturally occurring isotopes: ^{238}U (99.276%), ^{235}U (0.718%), and ^{234}U (0.004%).[1] Its mobility in ground water is attributed to both natural processes as well as anthropogenic interventions. In groundwater, uranium exists predominantly as soluble U(VI) species under oxidizing conditions with a circumneutral pH and as less soluble U(IV) species under reducing conditions. Human factors include nuclear power generation, mining operations, nuclear accidents such as those at Fukushima and Chernobyl, and the application of phosphate fertilizers, as uranium is a naturally occurring impurity in phosphate rocks.[2] Natural processes contributing to its mobilization include volcanic eruptions and weathering.[3]

The presence and distribution of uranyl species (UO_2^{2+}) in aquatic environments are primarily influenced by factors such as redox conditions, pH, CO_2, partial pressure, and the nature of complexing agents. Key agents which influence uranium mobility include carbonates, phosphates, fluorides, sulfates, silicates, and others.[4,5] The formation of complexes by uranium is also driven by pH. For instance, within a pH range of 5 to 7, both cationic species $(UO_2)_3(OH)_5^+$ and $UO_2(OH)^+$, as well as anionic species such as $UO_2(CO_3)_3^{4-}$ and $(UO_2)_2CO_3(OH)_3^-$, are present.[6] Agroecosystems contribute additional pressure on dissolved uranium concentrations in groundwater. In particular, overexploitation of groundwater through agriculture leads to the influx of oxygen into shallow aquifers that favors uranium

Received: September 26, 2025
Revised: March 10, 2026
Accepted: March 10, 2026

ACS Publications

© XXXX American Chemical Society

mobilization in highly alkaline and oxygen-rich environments. Punjab's intense agricultural activities and higher uranium content may be correlated.[7]

The most abundant isotope, ^{238}U with a half-life of 4.5×10^9 years is a long-term risk in aquifers. Similar to cadmium, mercury, and lead, uranium has been recognized as a nephrotoxin.[3] After being released into water and soil, it is absorbed by plants, consumed by grazing animals, and ingested by humans, causing various severe health problems such as impaired bone development, respiratory issues, kidney dysfunction, genetic damage, and reproductive disorders.[3] Researchers observed that elevated uranium intake from groundwater poses a significant risk due to its chemical toxicity, which outweighs concerns related to its radiotoxic effects. Groundwater contamination by uranium is a significant environmental and public health concern, particularly in regions with extensive agricultural activities.[2] The World Health Organization (WHO) has established guidelines for the maximum allowable concentration of uranium in drinking water to minimize health risks.[8] Its occurrence in groundwater necessitates continuous monitoring and treatment to ensure public health safety, especially in regions where natural radioactivity is prevalent.[9]

Conventional methods for uranium measurement include radiochemical, spectroscopic, and photometric techniques.[10−14] While they offer high precision and reliability, they often require the use of expensive instrumentation and involve complex and time-consuming sample preparation processes and demand highly trained personnel.[15] Consequently, the cost and expertise associated with these methods may limit their accessibility for routine water quality monitoring, highlighting the need for more efficient and cost-effective alternatives.[16−18] The development of predictive models to forecast uranium levels in groundwater has several advantages. The primary benefit lies in the proactive management of water resources, to ensure access to safe drinking water for communities and thereby reducing the health risks associated with uranium exposure.[19] A further important implication is that, predictive models enable the tracking of contamination hotspots, which guides effective planning of water treatment and remediation. These findings could also help tackling risk management for agricultural activities and land use planning.

The main sources of geochemistry of regional groundwater have been human activities like the use of fertilizers and the geogenic minerals and the alluvial soil. The less productivity of crops, surface salt accumulation on the ground, high salinity, and total dissolved solids (TDS) were the major factors revealed by researchers to influence the geochemistry of ground water in the semi-arid regions characterized by insufficient rainfall. The highly concentrated use of fertilizers in contemporary farming could be one of the reasons that leads to concentrating harmful substances like uranium and fluoride in groundwater.[20] Previous research has also shown that certain parameters, such as fluoride and sulfate, have higher correlations with uranium content in groundwater.[21,22] Uranyl sulfate species, including UO_2SO_4, $UO_2(SO_4)_2^{2-}$, and $UO_2(SO_4)_3^{4-}$ are highly soluble in water and keep uranium mobile in groundwater in oxidizing environments.[23,24] These uranyl sulfate species are dominant in the acidic pH range ($pH < 6$). Reduced sulfur-bearing minerals, produced by sulfate-reducing bacteria, are now widely acknowledged as a key factor in sustaining uranium-reducing environments in natural groundwater systems with high uranium concentrations.[25−27] Sahoo et al. found that fluoride exhibits a significant positive correlation with uranium, with a stronger association observed in shallow aquifers compared to deep aquifers in groundwater of the alluvial plains of Punjab.[28] Earlier studies have identified elevated electrical conductivity (EC) and TDS, along with alkaline conditions, as key factors contributing to high F^- concentrations in groundwater.[29] The co-occurrence of these elements in groundwater is not merely coincidental but reflects underlying geochemical processes and aquifer characteristics. The importance of chloride and other predictors is further supported by other studies in India. Studies in the Bastar district of Chhattisgarh have shown that uranium concentrations are positively correlated with nitrate, chloride, and sulfate, and they reported that nitrate and chloride act as uranium carriers.[30] Bairwa et al. investigated seasonal variations in the Tonk district of Rajasthan, finding moderate to strong positive correlations between uranium and chloride (correlation coefficients of 0.55 in pre-monsoon and 0.59 in post-monsoon seasons).[31] The study also noted that the order of average anionic settings concentration was $HCO_3^- > Cl^- > SO_4^{2-} > NO_3^-$, indicating the presence of these predictor ions in the groundwater samples.

Previous studies have also demonstrated that the simultaneous presence of uranium with arsenic, and fluoride in the aquifers of the Brahmaputra floodplain, showing that high uranium concentrations are primarily associated with drier, oxidizing aquifers, enhanced rock-water interactions, and the presence of amorphous iron hydroxides like ferrihydrite that act as sinks for uranium.[32] Analysis of water samples in southern Brazil concluded that excessive use of nitrate fertilizer is a major deciding factor for uranium mobility in the non-agricultural catchments.[33] Nitrate has the capability to decrease the reducing conditions of groundwater, which in turn can increase uranium concentration in non-agricultural areas. Uranium mobility has a positive correlation with carbonate, nitrate, and high oxidizing conditions, upon analyzing water samples from the northeastern state of Bihar, India.[34] Analysis of 102 groundwater (bore-well and hand pump) samples from Bathinda and Mansa concluded that the positive redox potential (ORP) and formation of predominant anionic species of mixed hydroxo-carbonate complexes in groundwater may be the cause for high uranium solubility.[22] Therefore, areas with their high concentrations of such species may have elevated uranium levels, which can guide targeted quality monitoring and remediation efforts. Uranium mobility and concentrations are characterized by complicated interactions with many hydrochemical variables which fluctuate concurrently. This warrants for additional research to identify and validate the degree to which hydrochemical data can be relied upon to predict uranium and if predictive models are sufficient to explain the dynamic nature of these interdependencies.

In recent years, machine learning (ML) and artificial intelligence (AI) techniques have been beneficial for the assessment of water quality and modeling of complex non-linear behaviors in water resources, driven by rapid advancements in sensor technologies and the pressing need for more robust methods to address groundwater contamination and sustainability challenges.[35−42] ML-based water quality surrogates from routine analytes have shown high skill in Indian alluvial settings.[43] Arid-region case studies reveal consistent monitoring gaps globally, where limited baseline data hinders predictive risk assessment, motivating ML approaches to upscale uranium screening from routine hydrochemistry.[44] In adjacent Upper Ganga alluvial aquifers, Azmi et al.[45] combined PHREEQC speciation with Monte Carlo-based health-risk modeling, generating full probability distributions of uranium hazard indices and

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX

showing, via sensitivity analysis, that U concentration overwhelmingly controls non-carcinogenic and radiological risk; this supports our use of predictive models to prioritize high-U wells for management. Despite the progress, persistent challenges remain, such as the scarcity of standardized, high-quality groundwater datasets, challenges with model transferability, the need for routine model retraining, handling of model drift, and a lack of integration of physical hydrogeological knowledge. Addressing these gaps, specifically through improved data availability, feature engineering, and interdisciplinary modeling remains key for accurate, early prediction of uranium contamination in diverse hydrogeological settings.[46]

There is limited existing literature that have directly addressed the prediction of uranium concentrations in groundwater using modeling approaches, as shown in Table 1. The first study, by Lopez et al., employed a Random Forest regression model with 558 data points, to establish a relationship between surface soil and water geochemical properties and uranium contamination in groundwater in the Central Valley, California.[47] This approach highlights the potential of mathematical modeling in identifying key geochemical variables that influence uranium distribution. The second study, by Khurelbaatar et al., with 135 data points, utilized a combination of polynomial and multiple linear regression models to predict uranium concentrations in groundwater in Mongolia.[48] This study further integrates Principal Component Analysis (PCA) to identify significant hydrochemical parameters that serve as key predictors in their models. These studies provide substantial evidence on the impact of strong statistical methods for uranium modeling; however, the limited number of investigations entails further development of models across different geographies, with large scale datasets.

Our study examines the correlation between uranium and 11 hydrochemical parameters; and further models uranium using a 3-stage approach consisting of clustering, regression and ensemble. This accounts for the inherent variability in water quality characteristics and allows for more targeted modeling strategies. The overall prediction accuracy is improved by addressing systematic errors in different data subgroups.[46,47] The choice to test multiple ML models for uranium prediction follows prior groundwater studies where ensembles of regression, tree-based, and neural models were jointly evaluated for WQI/GWQI prediction.[49,50] By combining advanced machine learning algorithms with the clustering approach and ensemble methods, we aim to develop a comprehensive and adaptive framework for uranium prediction that can handle the complex and variable nature of water quality data. One promising application of our work is to make groundwater monitoring more accessible, inclusive and sustainable by reducing the dependency on sophisticated and resource-intensive analyses. The approach has immediate practical value, as it enables cost-effective risk assessment and prioritization for detailed testing using advanced facilities, using available water quality datasets. With robust validation across a large area, our study offers a novel, scalable method that directly supports sustainable groundwater governance and provides actionable insights for administrators and resource managers.

2. MATERIALS AND METHODS

2.1. Data Source

In the present study, water samples were collected from multiple locations across Punjab encompassing a wide geographic region (50,362 km^2), as shown in Figure 1. The regions covered included major districts such as Amritsar, Bathinda, Faridkot, Hoshiarpur, Jalandhar, Ludhiana, Mansa, Moga, Patiala, Sangrur, Tarn Taran, Pathankot, Kapurthala, Fazilka, Ferozepur, Gurdaspur, SBS Nagar, SAS Nagar, Barnala, and Roopnagar, with a total population of 31.19 million. The strategic choice of locations ensures that we obtain a comprehensive view of groundwater quality across the state, with a coverage of both rural and urban areas.

Water-quality data used in this study were obtained from the Department of Water Supply and Sanitation (DWSS), Government of Punjab, which maintains a state-wide monitoring database for drinking-water sources. Within Punjab, this surveillance program includes an expanded analyte panel in several high-risk districts, however, the analytical panel are not uniform across all sampling events, resulting in Not Tested (NT) entries for some parameters including uranium in a substantial portion of the database. The DWSS collects and analyzes samples from groundwater-based supplies (e.g., tubewells/handpumps) and canal-based sources used for community distribution. They conduct such studies in their own laboratories, following standard practices. With such an expansive temporal and spatial scale,

Table 1. Summary of Literature for Uranium Prediction in Groundwater Using Machine Learning, Compared to Our Work[a]

Predictors	Location	Techniques	Sample Size	Performance Metrics		Reference
Ca, Mg, Ba, Sr, Na, K, Fe, Mn, DO, SO_4^{2-}, NO^3−N, alkalinity, pH	Central Valley, California	Random Forest	558	RMSE MAE r pseudo R^2	0.59 0.43 NA 0.57	47
pH, EC, ORP, DO, TDS, Cl^-, Mg^{2+}, HCO_3^-, Ca^{2+}	Mongolia	Combination of Polynomial and Multiple Linear Regression	135	RMSE MAE r R^2	2.116 1.402 0.963 NA	48
		PLS Regression		RMSE MAE r R^2	9.932 5.045 0.388 NA	
F^-, Cl^-, NO_3^-, SO_4^{2-}, Fe, Ni, Cd, As, Hg, Cr, Pb	Punjab, India	K-means clustering-XGBoost Regression Cluster 1	8735	RMSE MAE r R^2	4.543 1.039 0.963 0.923	Our work
		K-means clustering-XGBoost Regression Cluster 2		RMSE MAE r R^2	11.344 5.305 0.973 0.917	

[a]DO: dissolved oxygen; PLS: partial least squares; TDS: total dissolved solids; EC: electrical conductivity; ORP: oxidation−reduction potential; RMSE: root mean square error; MAE: mean absolute error; R^2: coefficient of discrimination; r: correlation between actual and predicted values.

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX

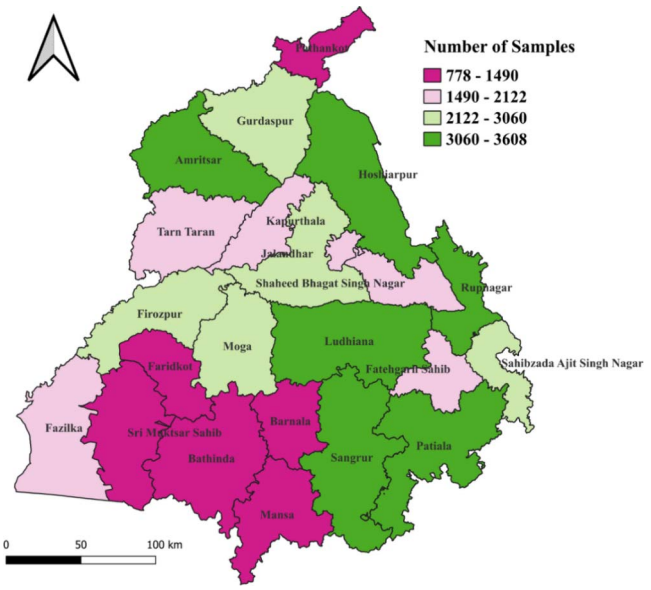


Figure 1. Geographical locations of the samples collected in the districts of Punjab. The number of water samples collected from each district (names are within their geographical boundaries) is indicated.

this analysis gives a fine-grained picture of the quality of groundwater across Punjab for better public health management. The number of samples collected from each source over the last five years is shown in Table S1. Rural communities are dependent on these tubewell/canal/handpumps for community water distribution services.

The dataset comprises 47,483 unique water samples (one record, per well, per visit of the analyst) collected between 2007 and 2024. Concentrations are measured for two critical categories of contaminants including the heavy metals and anions. Heavy metals considered include arsenic, uranium, lead, cadmium, nickel, iron, and mercury, whilst anions include nitrate, chloride, and sulfate. A total of 15 parameters were measured per sample for 9227 water samples. For 38,255 samples, analysis was not done for some of the parameters.

2.2. Hydrochemical Parameters

Water samples were analyzed by Ion Chromatography (IC), and Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Before ICP-MS analysis, the samples were acidified with 5% HNO_3 and filtered through a 0.22-micron filter. Uranium and other metal concentrations was measured using a NexION 300X ICP-MS, calibrated with relevant reference standards. Major anions (F^-, Cl^-, SO_4^{2-}, NO_3^-) were determined using a conductivity-suppressed ion chromatography system (Metrohm). The instrument was calibrated and standardized using stock solutions of ultra-pure standards for the respective anions.

2.3. Data Preprocessing

The following preprocessing steps were carried out to make the data suitable for analysis. The annotations included in the data set are Not Detected (ND), Below Detection Limit (BDL), and Not Tested (NT). Based on the known procedures and sources of water samples, values of ND were treated as zeros. The Environmental Protection Agency recommends using this method; in it, non-detects or tiny proportions of the detection limits were treated as zeros when they were used in statistical analyses.[51] In the BDL values, we are aware of the detection limits of IC and ICP-MS (0.1 mg/L and 10^{-5} mg/L, for IC and ICP-MS, respectively for most of the species). Therefore, we have taken the BDL values of the parameters tested using IC and ICP-MS as 0.099 mg/L and 10^{-5} mg/L, respectively. NT values present a unique challenge as they indicate missing tests, which can result from various factors, including logistical constraints or the non-requirement of tests at certain locations. For parameters such as aluminum, TDS, selenium, copper, chloride, sulfate, calcium, and magnesium there was a high proportion of NT values (>75%). NT values were removed from the analysis as they cannot be replaced with zeroes. Once data cleaning was complete with these constraints, we ended up with 8735 records, which was used to build our model.

2.4. Machine Learning

A suite of ML models was employed such as Gradient Boosting, Random Forest, and XGBoost for uranium prediction. The methodology comes complete with clustering and ensemble modeling. We present comparison of data subgroups obtained through two approaches including K-means and Gaussian mixture models (GMM). Three ensemble models were incorporated including weighted averaging, stacking, and voting. Descriptions and specifications of all approaches are provided in the Supporting Information attachment in Sections S1, S2, and Table S3. We utilized open-source packages in Python for programming.[52]

The dataset was split for training and testing in the ratio of 0.7:0.3. Data was scaled through *Z*-score normalization prior to modeling. Standalone models were constructed using the 11 analytes as features, from which the residuals are obtained. These residuals were subsequently combined with the 11 analytes to form a new feature space for our clustering analysis.[53]

To visually confirm the presence of distinct groups and show the results of our clustering, two dimensionality reduction techniques were utilized to explore the underlying structure of this combined feature space. PCA was applied to linearly reduce dimensionality while preserving variance, revealing two compact and well-separated groups as shown in Figure 3a. Complementary to this, nonlinear t-Distributed Stochastic Neighbor Embedding (t-SNE) was used to better capture local similarities, also indicating the presence of two distinct clusters as shown in Figure 3b. Similarly, Figure S1 visualizes the clusters obtained using GMM through PCA. After clustering, we compared the hydrochemical characteristics of the resulting clusters by examining the distributions of all input variables stratified by cluster. For each parameter, we generated cluster-wise histograms and applied the Mann−Whitney U test to evaluate whether the distributions differed significantly between clusters ($\alpha = 0.05$) as shown in Figure S20 and Table S4.

3. RESULTS AND DISCUSSION

3.1. Correlation Analysis

We calculated the correlation of uranium with the chosen parameters, wherein the correlation did not exceed 0.5, as shown in Figure 2. Nitrate, chloride, fluoride, and sulfate showed statistically significant correlations with uranium, with correlation coefficients of 0.355, 0.180, 0.269 and 0.425, respectively. Descriptive statistics of the data are presented in Table S2, which includes the mean, median, standard deviation, minimum, and maximum values of the parameters.

3.2. Machine Learning Models

The distinct clusters visualized in Figure 3 for both linear and nonlinear projections, reveal the consistent fact that a standalone global regression model might not be able to reflect on significant differences across the data subgroups. This observation reinforces our integrated approach of detecting data clusters, and training distinct regression models for each cluster. Hence, we can hypothesize and also demonstrate that this cluster-specific modeling technique better represents the underlying data differences in large scale datasets with multiple parameters and therefore provides a better predictive power than a single global model.[51]

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX

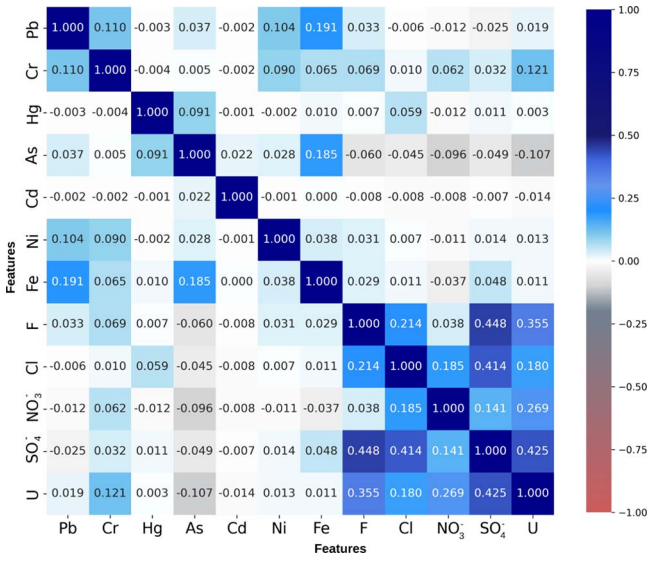


Figure 2. Correlation heatmap of hydrochemical parameters and uranium.

3.3. Model Performance

The performance of all models was evaluated using standard model evaluation metrics, with results for both the training and testing datasets summarized in Table 2. The K-means XGBoost approach emerged as the most effective strategy, demonstrating the lowest RMSE and MAE, coupled with the highest R^2 among all approaches. This was closely followed by the other models with K-means, which showed consistently strong performance across clusters. Most of the K-means based approaches performed better than their counterparts of GMM. This is also illustrated in Figure 3 and Figure S1, where in K-means provides a clearer separation of the data subgroups, while GMM does not confirm a stable separation. The GMM based models reveal a significant variation, with one cluster showing very high accuracy, while the other remains unsatisfactory. It is important to note that all the clustering-based methods were significantly better predictors than individual models. Thus, it can be conceivably hypothesized that the inclusion of data subgroups for the modelling of complex environmental parameters is vital. As presented in Table 2a–c, the model metrics are significantly improved with clustering and ensemble techniques. We consider K-means XGBoost as the best model, although other approaches like K-means Random Forest and K-means ensemble are very close in performance. K-means partitioning yielded two hydrochemical clusters with systematically different composition. Visual inspection of cluster-wise distributions (Figure S20) and Mann–Whitney U tests (Table S4) indicate statistically significant differences for most predictors.

Figure 4 represents the scatter plots for comparing actual versus predicted values for the best models. Similarly, Figure S2 depicts the actual and predicted values for standalone models. Figures S5, S6, S9, and S10 depict the same for standalone and ensemble models for K-means based clusters. Figures S12, S15, S16, and S19 present actual vs. predicted values for standalone and ensemble models for GMM clusters.

Figure 5b presents residual diagnostics for K-means XGBoost regressor, showing tight, near-zero-centered densities for cluster 2, indicating minimal bias and variance. Figure 5a shows the same for cluster 1, which exhibits slightly broader tails. The residual plots for other models are shown in Figures S3, S4, S7, S8, S11, S13, S14, S17, and S18. The fact that the residuals were symmetrical about zero leads to the conclusion that the models are not systematic over/under performers of the actual values. This empirical finding helps to assume that the model is based on the homoscedasticity assumption because the degree of variance of the residual is equal in case of different strata of the explanatory variables. These findings, therefore, strengthen the inferential soundness of the models and confirm the fact that the fitted models provide a reliable estimate of the dependent variable assuming the above cited predictors. Overall, the normal distribution of residuals affirms that the regression model is appropriate for the dataset under consideration and supports its use for predictive and inferential purposes.

3.4. Spatial Analysis

Figure 6 provides a comparative visualization of observed and predicted uranium concentrations in groundwater samples across Punjab. Observed and predicted uranium (ppb) were visualized in QGIS using Graduated symbology with a Quantile classifier (6 bins). For this dataset, the first and last class limits are 0–9.3 ppb and 27.1–335.1 ppb, respectively. The coordinates have an uncertainty as they were estimated based on village names.

The spatial patterns align with the geochemical reasoning introduced earlier. Southwestern districts of Punjab show the highest observed and predicted U, consistent with oxic–alkaline alluvial settings where sulfate, nitrate, fluoride, and chloride co-occur and promote uranyl complexation and mobility.

The correlation between uranium and anions such as sulfate is often governed by solubility equilibria of uranium-bearing minerals. When minerals like uraninite, uranyl phosphates, or

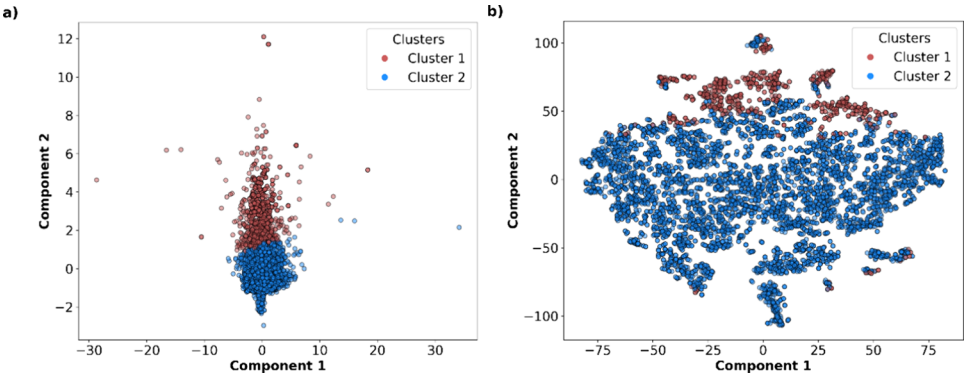


Figure 3. K-means cluster structure visualized by (a) PCA and (b) t-SNE.

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX–XXX

Table 2. Predictive Capabilities of Models for Training and Testing Data[a]

		(a) Standalone models Train				Test			
Clustering	Model	RMSE	R^2	MAE	r	RMSE	R^2	MAE	r
Standalone	Random Forest	4.656	0.951	2.427	0.977	12.346	0.679	7.013	0.825
Standalone	XGBoost	6.568	0.903	4.400	0.952	12.613	0.665	7.851	0.816
Standalone	Gradient Boosting	11.939	0.681	8.908	0.829	14.619	0.550	10.051	0.747
		(b) GMM clustering with models Train				Test			
Clustering	Model	RMSE	R^2	MAE	r	RMSE	R^2	MAE	r
GMM (Cluster 1)	XGBoost	0.509	0.999	0.366	1.000	1.523	0.991	0.842	0.996
GMM (Cluster 1)	Voting	1.034	0.996	0.669	0.998	1.613	0.990	0.948	0.995
GMM (Cluster 1)	Weighted Averaging	1.034	0.996	0.669	0.998	1.613	0.990	0.948	0.995
GMM (Cluster 1)	Stacking	0.843	0.997	0.446	0.999	1.707	0.989	0.848	0.995
GMM (Cluster 1)	Random Forest	1.037	0.996	0.452	0.998	1.983	0.985	1.041	0.993
GMM (Cluster 1)	Gradient Boosting	2.293	0.981	1.551	0.991	2.406	0.978	1.588	0.989
GMM (Cluster 2)	Voting	2.076	0.998	1.098	0.999	26.336	0.544	16.428	0.893
GMM (Cluster 2)	Weighted Averaging	2.076	0.998	1.098	0.999	26.336	0.544	16.428	0.893
GMM (Cluster 2)	Gradient Boosting	2.631	0.997	1.840	0.999	27.589	0.500	17.006	0.891
GMM (Cluster 2)	XGBoost	0.026	1.000	0.017	1.000	28.195	0.478	15.523	0.859
GMM (Cluster 2)	Random Forest	4.766	0.990	1.671	0.996	28.690	0.459	17.602	0.873
GMM (Cluster 2)	Stacking	4.192	0.993	1.573	0.996	29.646	0.423	18.250	0.873
		(c) K-means clustering with models Train				Test			
Clustering	Model	RMSE	R^2	MAE	r	RMSE	R^2	MAE	r
K-means (Cluster 1)	XGBoost	0.096	1.000	0.062	1.000	**11.344**	**0.917**	**5.305**	**0.963**
K-means (Cluster 1)	Voting	2.040	0.997	1.340	0.999	14.161	0.870	7.458	0.943
K-means (Cluster 1)	Weighted Averaging	2.040	0.997	1.340	0.999	14.161	0.870	7.458	0.943
K-means (Cluster 1)	Stacking	2.600	0.996	1.445	0.998	17.648	0.798	8.492	0.915
K-means (Cluster 1)	Gradient Boosting	3.917	0.991	2.887	0.995	19.778	0.747	10.332	0.895
K-means (Cluster 1)	Random Forest	3.057	0.994	1.462	0.997	13.629	0.880	7.472	0.944
K-means (Cluster 2)	XGBoost	0.476	0.999	0.343	0.999	**4.543**	**0.923**	**1.039**	**0.973**
K-means (Cluster 2)	Voting	0.798	0.997	0.545	0.999	3.957	0.942	1.174	0.981
K-means (Cluster 2)	Weighted Averaging	0.798	0.997	0.545	0.999	3.957	0.942	1.174	0.981
K-means (Cluster 2)	Stacking	1.381	0.991	0.421	0.996	5.779	0.876	1.225	0.959
K-means (Cluster 2)	Gradient Boosting	1.668	0.987	1.221	0.993	4.843	0.913	1.832	0.973
K-means (Cluster 2)	Random Forest	0.927	0.996	0.379	0.998	4.200	0.934	1.330	0.976

[a]The best results are in bold face.

sulfates interact with groundwater, uranium concentrations are controlled by the solubility product (K_{sp}) of the respective phases. Elevated sulfate levels promote the formation of soluble uranyl-sulfate complexes, potentially enhancing uranium mobility so long as the product, $[UO_2]^{2+}$ $[SO_4]^{2-}$ does not exceed the K_{sp} value of $[UO_2]\cdot[SO_4](H_2O)_n$. Consequently, environments rich in sulfate, whether from natural sources or human activities, can maintain or even raise uranium levels in groundwater, since

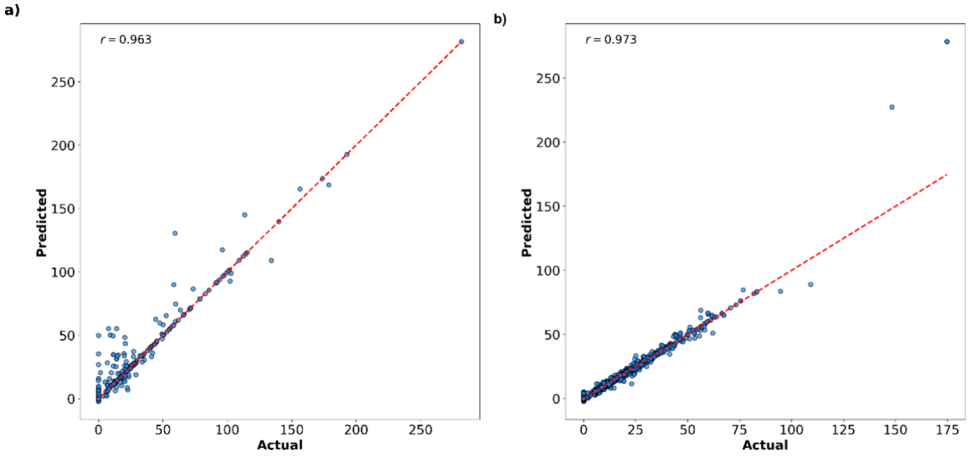


Figure 4. Actual vs predicted values using XGBoost for K-means clusters; (a) Cluster 1 and (b) Cluster 2.

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX–XXX

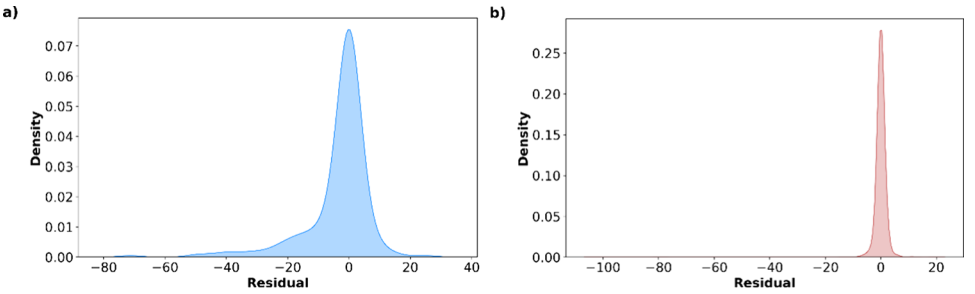


Figure 5. Residual density plots for K-means clustering with XGBoost regressor: (a) Cluster 1 and (b) Cluster 2.

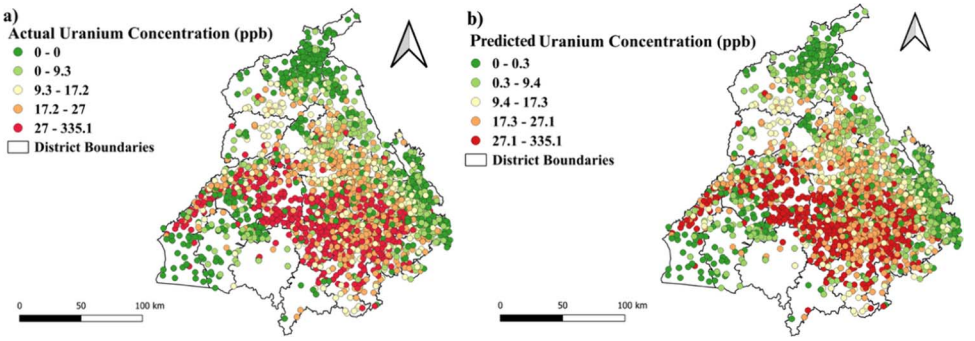


Figure 6. Spatial distribution of uranium in groundwater across Punjab. (a) Observed concentrations; (b) Predicted concentrations.

the formation of these complexes inhibits uranium precipitation, allowing it to stay dissolved in groundwater.[54]

Table 3 summarizes the roles of major anions influencing uranium occurrence and mobility in groundwater.

The data-driven evidence from this study implies that uranium concentration and mobility is governed by key anions and trace metals. The aim of our work was to broaden the current knowledge on synergistic interactions and provoke scientists to question new inter-relationships. The findings reinforce the complexity of uranium geochemistry in aquifers and reiterate the need for a multi-faceted approach towards uranium remediation. The outcomes convey critical information for directing both fundamental research and environmental management policy, while providing real-time insights towards the protection of water resources and citizen health.

Leveraging the strong correlation between specific anions and uranium levels, policies can move from reactive measures to proactive prevention of groundwater contamination. We recommend mandating agricultural practices that restrict fertilizer types known to mobilize uranium, directly addressing the source rather than relying on costly remediation. Ultimately, this framework can form the basis of a "Digital Groundwater Twin" for

Table 3. Hydrogeochemical Role of Major Anions in Uranium Mobilization in Groundwater

Anions	Hydrogeochemical mechanism affecting U mobility in groundwater	Net effect on U(VI) mobility	References
Nitrate (NO_3^-)	• Act as an important control on uranium mobility in groundwater, primarily by promoting oxidizing conditions that convert immobile U(IV) to the more soluble U(VI). • Commonly co-occurs with agriculturally influenced, carbonate-rich recharge waters, which further enhance uranium transport.	Enhances U(VI) mobility	33,55
Fluoride (F^-)	• Alkaline groundwater conditions enhance fluoride enrichment through competitive substitution of F^- from fluoride-bearing minerals by OH^- ions, making fluoride a reliable indicator of alkaline environments. • Fluoride-enriched groundwater is commonly associated with Na−HCO_3 and NaCl hydrochemical types, characterized by high pH and low Ca^{2+} concentrations, conditions that are also conducive to enhanced U(VI) mobility.	Indirect increase in U mobility	20,56−58
Sulfate (SO_4^{2-})	• Elevated SO_4^{2-} concentrations are frequently associated with Na−Cl and Ca−Mg−SO_4 hydrochemical facies, high total dissolved solids (TDS), and long groundwater residence times, conditions that favor uranium enrichment in many arid and semi-arid aquifers. • It participates in coupled sulfur−iron−uranium redox cycling, where sulfate-reducing bacteria can temporarily immobilize uranium by reducing U(VI) to U(IV) and promoting sulfide-mediated precipitation or sorption; however, shifts in redox conditions or microbial activity can re-oxidize U(IV), resulting in uranium remobilization.	System-dependent (immobilization or remobilization)	27,59
Chloride (Cl^-)	• Conservative tracer of salinity and residence time. Does not strongly complex U, but elevated Cl^- indicates prolonged water−rock interaction, evaporative concentration, and association with Na−Cl facies where U(VI) mobility is enhanced.	Indirect increase in U mobility	28,60

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX

Punjab. This digital twin could be a sophisticated, predictive model enabling policymakers to conduct scenario analysis, such as simulating the impact of a shift to less water-intensive crops or changes in fertilizer use on future uranium levels. This would facilitate proactive management practices and data-driven decisions that are also scientifically informed. This directly aligns with long term public-health, sustainability and environmental goals.

4. LIMITATIONS

Although this study offers important insights on the relationships among the given set of analytes and uranium, the following limitations are acknowledged. Real-world systems influencing uranium are considerably more complex, involving numerous other critical factors such as varying pH levels, TDS, conductivity, the presence of other cations (e.g., Ca^{2+}, Mg^{2+}, Fe^{3+}), diverse organic ligands, fluctuations in temperature and redox potential, and interactions with mineral surfaces or microbial communities. Researchers studying groundwater in Bathinda and Mansa districts (Punjab) reported a positive correlation between TDS and uranium concentration, attributing it to reduced ion competition and stronger adsorption on mineral surfaces at lower TDS levels. The absence of TDS data in the present study may limit the accuracy of uranium concentration predictions.[22] Incorporating TDS measurements could enhance the interpretability of uranium mobility and distribution in groundwater. In addition, alkalinity such as carbonate or bicarbonate, a primary control on U(VI) mobility through uranyl-carbonate complexation was not consistently available in the DWSS dataset and therefore could not be included as a predictor in this study. Coyte et al. identified bicarbonate complexation and oxidizing conditions as key geochemical controls on uranium concentrations in groundwaters of Rajasthan and Gujarat.[61] Its inclusion could enhance the understanding of uranium speciation and mobility. The current framework can thus be further enhanced to capture the intricate, often synergistic or antagonistic, effects governing uranium speciation and mobility in diverse environments.

5. FUTURE WORK

Future work will integrate alkalinity/carbonate measurements (and, where available, pH/TDS) and will couple the ML framework with speciation calculations to explicitly link predictive patterns with uranyl-carbonate complexation. Expanding the variables considered is also essential, including not only other critical inorganic ions but also entirely different classes of parameters relevant to specific contexts, such as fertilizer concentrations in agricultural runoff or detailed metrics of soil quality in terrestrial systems, which are known to impact water chemistry and contaminant interactions. While there are many parameters available, achieving high accuracy with the minimum necessary parameters is an area for further optimization. This will allow hand-held sensors to predict uranium in the field using cloud-based analytics. In addition, developing a model with capabilities for real-time considerations and time series analysis is also crucial for understanding and predicting dynamic changes in uranium concentration or behavior over time in fluctuating environments. Extended work should couple the ML−clustering framework with mechanistic tools such as reactive transport or speciation modeling (e.g., PHREEQC) to explicitly simulate geochemical reactions and further disentangle causal pathways. Further investigation into the synergistic and competitive interactions among anions and other chemical species affecting uranium is warranted, potentially employing coupled geochemical transport models to capture nonlinearities and dynamic processes. The modeling framework developed here is not specific to Punjab and can, in principle, be applied to any region where comparable hydrogeochemical settings and similar monitoring panels, with local retraining and validation prior to deployment. Because the pipeline relies on standard clustering and regression algorithms, it can be retrained on local datasets to learn region-specific clusters and predictor−response relationships.

6. CONCLUSIONS

This work demonstrates how machine learning and clustering can be jointly leveraged to move uranium monitoring from sparse, reagent- and instrument-intensive campaigns toward continuous, decision-oriented surveillance using data streams that many regions already collect. By coupling unsupervised partitioning of hydrochemical conditions with non-linear regression, the framework aligns with a broader shift in groundwater science toward models that do not merely interpolate concentrations but actively disentangle how different hydrogeochemical regimes control contaminant occurrence and exceedance risk. By integrating unsupervised clustering (K-means and Gaussian mixture models) with advanced regressors (Gradient Boosting, Random Forest, XGBoost) and ensemble techniques, the pipeline reveals hidden data subgroups that drive predictive superiority over standalone models, achieving robust performance on the largest state-scale dataset to date: 8735 samples from Punjab, India (2007−2024). Beyond technical innovation, this approach underscores the value of subgroup-aware modeling in capturing the inherent complexity of natural systems, where hydrogeochemical interactions defy uniform assumptions. In broader terms, the framework exemplifies how accessible analytes, already collected statewide, can democratize contamination forecasting, fostering proactive public health interventions in arsenic- and uranium-endemic areas like Punjab. Its stability, particularly with K-means and XGBoost, ensures reliability across heterogeneous clusters, while its modular design promotes transferability to heavy metal prediction in diverse global contexts. Ultimately, this work bridges machine learning with environmental geochemistry, empowering data-driven policy, sustainable water management, and equitable access to safe groundwater amid escalating contamination threats. Because our predictors are already measured at scale, this framework enables cost-efficient triage and district-level risk screening prior to confirmatory ICP-MS analyses. Correlation attributions identify SO_4^{2-}, NO_3^-, Cl^-, and F^- as dominant drivers, consistent with known hydrochemistry and provide interpretable guidance for targeted monitoring where laboratory uranium assays are limited. To our knowledge, this is the largest state-scale dataset used to assess U concentrations from routine analytes. This approach can be readily applied to other contaminants and to regions characterized by similar hydrochemical settings. Future work can be carried out to test the applicability of the methodology to other contaminants and other environmental parameters, and this may transform water quality monitoring and management practices. The predictions could be integrated with appropriate hydroinformatics tools and made available to public leading to better public health outcomes.

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX

This framework provides an actionable backbone for reshaping groundwater protection from ad-hoc responses to risk-prioritized management. It enables agencies to use existing chemistry data not just for status reporting, but to dynamically allocate attention, money, and analytical capacity where it will avert the most exposure. The identification of a small set of reliable proxies for uranium risk creates an opportunity to codify simple, threshold-based screening rules in guidelines and standards, so that field laboratories and district engineers can flag emerging hotspots without needing specialized geochemical expertise or full uranium analyses at every site. For practice and policy, the framework offers a concrete pathway to re-prioritize limited ICP-MS capacity, emergency mitigation (e.g., well replacement, blending, or treatment), and infrastructure investments toward locations where modeled risk is highest, rather than distributing effort uniformly or reactively. By distinguishing regions with systematically different risk profiles, the framework supports differentiated regulatory strategies: for example, more stringent permitting, fertilizer management, or well-construction requirements in high-risk hydrogeochemical regimes, and lighter-touch monitoring in demonstrably low-risk zones, thereby improving both the efficiency and the fairness of regulation.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssusresmgt.5c00505.

Modeling methods and model evaluation metrics; figures including visualization of clusters for GMM using PCA and t-SNE, actual concentration versus predicted concentration for standalone models including XGBoost, Gradient Boosting, and Random Forest, residual plots for XGBoost, Gradient Boosting, and Random Forest models, residual plots for standalone models in cluster 1 under K-means clustering, actual versus predicted plots by standalone models in cluster 1 under K-means clustering, actual versus predicted plots by ensemble models in cluster 1 under K-means clustering, residual plots for ensemble models in cluster 1 under K-means clustering, residual plots for standalone models in cluster 2 under K-means clustering, actual versus predicted plots by standalone models in cluster 2 under K-means clustering, actual versus predicted plots by ensemble models in cluster 2 under K-means clustering, residual plots for ensemble models in cluster 2 under K-means clustering, actual versus predicted plots by standalone models in cluster 1 under GMM clustering, residual plots for standalone models in cluster 1 under GMM clustering, residual plots for ensemble models in cluster 1 under GMM clustering, actual versus predicted plots by ensemble models in cluster 1 under GMM clustering, actual versus predicted plots by standalone models in cluster 2 under GMM clustering, residual plots for standalone models in cluster 2 under GMM clustering, residual plots for ensemble models in cluster 2 under GMM clustering, and actual versus predicted plots by ensemble models in cluster 2 under GMM clustering; tables including financial year-wise number of schemes tested for tubewell, canal, and handpump, descriptive statistics of the observed water quality parameters in Punjab with 8735 data points, and comparison of validation metrics for clustering (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Thalappil Pradeep – *DST Unit of Nanoscience (DST UNS) and Thematic Unit of Excellence (TUE), Department of Chemistry, Indian Institute of Technology Madras, Chennai 600 036, India; International Centre for Clean Water, 2nd Floor, B-Block, IIT Madras Research Park, Kanagam Road, Taramani, Chennai 600113, India;*
Email: pradeep@iitm.ac.in; Tel.: +91-44 2257 4208;
Fax: +91-44 2257 0545/0509

Authors

Sankar Sudhir – *DST Unit of Nanoscience (DST UNS) and Thematic Unit of Excellence (TUE), Department of Chemistry, Indian Institute of Technology Madras, Chennai 600 036, India*

Vamanie Perumal – *Department of Engineering Design, Indian Institute of Technology Madras, Chennai 600 036, India*

Tanmayaa Nayak – *DST Unit of Nanoscience (DST UNS) and Thematic Unit of Excellence (TUE), Department of Chemistry, Indian Institute of Technology Madras, Chennai 600 036, India*

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acssusresmgt.5c00505

Author Contributions

[Ψ]S.S. and V.P. have contributed equally to this work. S.S. Initiated the study with T.P., cleaned the data, and, in discussion with T.N., removed variables with many foreign entries of low chemical relevance; ran statistical checks and correlation analyses with uranium and other variables; and wrote the data-processing, results, and discussion sections. V.P. collaborated with S.S. and T.P. to develop the advanced supervised and unsupervised models and drafted the introduction, the models' description, and key parts of the results and discussion. T.N. wrote the introduction, results, and future-work sections and provided the scientific perspective throughout the manuscript. T.P. suggested the problem, obtained the water-quality data from the Department of Water Supply and Sanitation, Punjab, facilitated the collaboration, supervised the work, and reviewed and finalized the manuscript. All authors contributed to the discussion of the results and approved the final manuscript. Initial part of the work in data analytics using this data set was carried out by the International Center for Clean Water (ICCW), IIT Madras. S.S.: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, original draft, and visualization. V.P.: Methodology, software, validation, formal analysis, investigation, data curation, writing - original draft, and visualization. T.N.: writing-original draft. T.P.: conceptualization, resources, supervision, review and editing, project administration, and funding acquisition.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Department of Water Supply and Sanitation, Government of Punjab, for graciously providing us with the water-quality data of Punjab,

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX–XXX

India. Authors also thank Mr. Ragul, an M.S. scholar in the Department of Engineering Design at IIT Madras, for his valuable insights on testing the model in alternative ways. S.S. gratefully acknowledges Dr. Ramya Dwivedi for her thoughtful discussions, which greatly enhanced the quality of this work. T.P. acknowledges financial support from the Centre of Excellence on Molecular Materials and Functions, established under the Institution of Eminence scheme of IIT Madras and the Department of Science and Technology, Government of India, for long years to support. V.P. and T.N. acknowledge funding and research support from the Ministry of Human Resources, Government of India. T.P. is thankful to his former colleagues, especially, Dr. Ganapathy Natarajan, for his initial work on building correlations using water quality data. He also thanks Mrs. Jaspreet Talwar, Mr. Amit Talwar, Mr. S. K. Jain and Mrs. Veenakshi Sharma for their long years of effort in building water quality data of Punjab.

REFERENCES

(1) Skeppström, K.; Olofsson, B. Uranium and Radon in Groundwater. *Eur. Water* **2007**, *17*, 51.

(2) Balaram, V.; Rani, A.; Rathore, D. P. S. Uranium in Groundwater in Parts of India and World: A Comprehensive Review of Sources, Impact to the Environment and Human Health, Analytical Techniques, and Mitigation Technologies. *Geosyst. Geoenviron.* **2022**, *1* (2), No. 100043.

(3) Rani, N.; Singh, P.; Kumar, S.; Kumar, P.; Bhankar, V.; Kamra, N.; Kumar, K. Recent Advancement in Nanomaterials for the Detection and Removal of Uranium: A Review. *Environ. Res.* **2023**, *234*, No. 116536.

(4) Cumberland, S. A.; Douglas, G.; Grice, K.; Moreau, J. W. Uranium Mobility in Organic Matter-Rich Sediments: A Review of Geological and Geochemical Processes. *Earth-Sci. Rev.* **2016**, *159*, 160–185.

(5) Thivya, C.; Chidambaram, S.; Keesari, T.; Prasanna, M. V.; Thilagavathi, R.; Adithya, V. S.; Singaraja, C. Lithological and Hydrochemical Controls on Distribution and Speciation of Uranium in Groundwaters of Hard-Rock Granitic Aquifers of Madurai District, Tamil Nadu (India). *Environ. Geochem. Health* **2016**, *38* (2), 497–509.

(6) Nayak, T.; Mukherjee, S.; Kini, A. R.; Islam, M. R.; Nagar, A.; Seth, S.; Pradeep, T. Cellulose-Derived Nanomaterials for Affordable and Rapid Remediation of Uranium in Water. *ACS Sustainable Chem. Eng.* **2025**, *13* (4), 1838–1850.

(7) Nizam, S.; Dutta, S.; Sen, I. S. Geogenic Controls on the High Levels of Uranium in Alluvial Aquifers of the Ganga Basin. *Appl. Geochem.* **2022**, *143*, No. 105374.

(8) Frisbie, S. H.; Mitchell, E. J. World Health Organization Increases Its Drinking-Water Guideline for Uranium. *Environ. Sci.: Processes Impacts* **2013**, *15* (10), 1817.

(9) Waseem, A.; Ullah, H.; Rauf, M. K.; Ahmad, I. Distribution of Natural Uranium in Surface and Groundwater Resources: A Review. *Crit. Rev. Environ. Sci. Technol.* **2015**, *45* (22), 2391–2423.

(10) Tosheva, Z.; Stoyanova, K.; Nikolchev, L. Comparison of Different Methods for Uranium Determination in Water. *J. Environ. Radioact.* **2004**, *72*, 47–55.

(11) Pawlak, Z.; Rabiega, G. Comparison of Inductively Coupled Plasma-Mass Spectrometry and Radiochemical Techniques for Total Uranium in Environmental Water Samples. *Environ. Sci. Technol.* **2002**, *36* (24), 5395–5398.

(12) Yoon, Y. Y.; Cho, S. Y.; Lee, K. Y.; Ko, K. S.; Ha, K. Radiochemical Determination of Uranium and Radium Isotope in Natural Water Using Liquid Scintillation Counter. *J. Radioanal. Nucl. Chem.* **2013**, *296* (1), 397–402.

(13) Jung, E. C.; Cho, H.-R.; Cha, W.; Park, J.-H.; Baik, M. H. Uranium Determination in Groundwater Using Laser Spectroscopy. *Rev. Anal. Chem.* **2014**, *33* (4), 245.

(14) Brina, R.; Miller, A. G. Direct Detection of Trace Levels of Uranium by Laser-Induced Kinetic Phosphorimetry. *Anal. Chem.* **1992**, *64* (13), 1413–1418.

(15) Yu, X.; Su, X.; Wang, Z.; Hou, Z.; Li, B. A Review of Uranium (U) Elemental Detection Methods. *Anal. Methods* **2025**, *17* (8), 1683–1697.

(16) Rathore, D. P. S. Advances in Technologies for the Measurement of Uranium in Diverse Matrices. *Talanta* **2008**, *77* (1), 9–20.

(17) Santos, J. S.; Teixeira, L. S. G.; Dos Santos, W. N. L.; Lemos, V. A.; Godoy, J. M.; Ferreira, S. L. C. Uranium Determination Using Atomic Spectrometric Techniques: An Overview. *Anal. Chim. Acta* **2010**, *674* (2), 143–156.

(18) Wang, T.; Tao, B.; Zuo, B.; Yan, G.; Liu, S.; Wang, R.; Zhao, Z.; Chu, F.; Li, Z.; Yamauchi, Y.; Xu, X. Challenges and Opportunities of Uranium Extraction From Seawater: A Systematic Roadmap From Laboratory to Industry. *Small Methods* **2025**, *9* (3), No. 2401598.

(19) Singh, A. K.; Hasnain, S. I.; Banerjee, D. K. Grain Size and Geochemical Partitioning of Heavy Metals in Sediments of the Damodar River - a Tributary of the Lower Ganga, India. *Environ. Geol.* **1999**, *39* (1), 90–98.

(20) Paikaray, S.; Chander, S. Geochemical Variations in Uranium and Fluoride Enriched Saline Groundwater around a Semi-Arid Region of SW Punjab, India. *Appl. Geochem.* **2022**, *136*, No. 105167.

(21) Makubalo, S. S.; Diamond, R. E. Hydrochemical Evolution of High Uranium, Fluoride and Nitrate Groundwaters of Namakwaland, South Africa. *J. Afr. Earth Sci.* **2020**, *172*, No. 104002.

(22) Kumar, A.; Tripathi, R. M.; Rout, S.; Mishra, M. K.; Ravi, P. M.; Ghosh, A. K. Characterization of Groundwater Composition in Punjab State with Special Emphasis on Uranium Content, Speciation and Mobility. *Radiochim. Acta* **2014**, *102* (3), 239–254.

(23) Lu, G.; Haes, A. J.; Forbes, T. Z. Detection and Identification of Solids, Surfaces, and Solutions of Uranium Using Vibrational Spectroscopy. *Coord. Chem. Rev.* **2018**, *374*, 314–344.

(24) Alcorn, C. D.; Cox, J. S.; Applegarth, L. M. S. G. A.; Tremaine, P. R. Investigation of Uranyl Sulfate Complexation under Hydrothermal Conditions by Quantitative Raman Spectroscopy and Density Functional Theory. *J. Phys. Chem. B* **2019**, *123* (34), 7385–7409.

(25) Noseck, U.; Tullborg, E.-L.; Suksi, J.; Laaksoharju, M.; Havlová, V.; Denecke, M. A.; Buckau, G. Real System Analyses/Natural Analogues. *Appl. Geochem.* **2012**, *27* (2), 490–500.

(26) Skierszkan, E. K.; Dockrey, J. W.; Helsen, J.; Findlater, L.-L.; Bataille, C. P.; De Laplante, G.; McBeth, J. M.; Mayer, K. U.; Beckie, R. D. Persistence of Uranium in Old and Cold Subpermafrost Groundwater Indicated by Linking^{234} U-^{235} U-^{238} U, Groundwater Ages, and Hydrogeochemistry. *ACS Earth Space Chem.* **2021**, *5* (12), 3474–3487.

(27) Sitte, J.; Akob, D. M.; Kaufmann, C.; Finster, K.; Banerjee, D.; Burkhardt, E.-M.; Kostka, J. E.; Scheinost, A. C.; Büchel, G.; Küsel, K. Microbial Links between Sulfate Reduction and Metal Retention in Uranium- and Heavy Metal-Contaminated Soil. *Appl. Environ. Microbiol.* **2010**, *76* (10), 3143–3152.

(28) Sahoo, P. K.; Virk, H. S.; Powell, M. A.; Kumar, R.; Pattanaik, J. K.; Salomão, G. N.; Mittal, S.; Chouhan, L.; Nandabalan, Y. K.; Tiwari, R. P. Meta-Analysis of Uranium Contamination in Groundwater of the Alluvial Plains of Punjab, Northwest India: Status, Health Risk, and Hydrogeochemical Processes. *Sci. Total Environ.* **2022**, *807*, No. 151753.

(29) Abiye, T.; Shaduka, I. Radioactive Seepage through Groundwater Flow from the Uranium Mines, Namibia. *Hydrology* **2017**, *4* (1), 11.

(30) Singh, M.; Sahu, P.; Tapadia, K.; Jhariya, D. Assessment of the Groundwater Quality by Using Multivariate Approach and Non-Carcinogenic Risk of Uranium in the Inhabitants of the Bastar District, Chhattisgarh, Central India. *Water Supply* **2022**, *22* (4), 3863–3878.

(31) Bairwa, A.; Gupta, A.; Mishra, V.; Sahoo, S.; Tiwar, S.; Menaria, T.; Gupta, K. Seasonal Variation and Spatial Distribution of Uranium in Sources of Water in Tonk District of Rajasthan India. *Pollution* **2024**, *10* (1).

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX–XXX

(32) Das, N.; Das, A.; Sarma, K. P.; Kumar, M. Provenance, Prevalence and Health Perspective of Co-Occurrences of Arsenic, Fluoride and Uranium in the Aquifers of the Brahmaputra River Floodplain. *Chemosphere* **2018**, *194*, 755–772.
(33) Bonotto, D. M.; Wijesiri, B.; Goonetilleke, A. Nitrate-Dependent Uranium Mobilisation in Groundwater. *Sci. Total Environ.* **2019**, *693*, No. 133655.
(34) Richards, L. A.; Kumar, A.; Shankar, P.; Gaurav, A.; Ghosh, A.; Polya, D. A. Distribution and Geochemical Controls of Arsenic and Uranium in Groundwater-Derived Drinking Water in Bihar, India. *Int. J. Environ. Res. Public Health* **2020**, *17* (7), No. 2500.
(35) Drogkoula, M.; Kokkinos, K.; Samaras, N. A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management. *Appl. Sci.* **2023**, *13* (22), No. 12147.
(36) Ghobadi, F.; Kang, D. Application of Machine Learning in Water Resources Management: A Systematic Literature Review. *Water* **2023**, *15* (4), No. 620.
(37) Tao, H.; Hameed, M. M.; Marhoon, H. A.; Zounemat-Kermani, M.; Heddam, S.; Kim, S.; Sulaiman, S. O.; Tan, M. L.; Sa'adi, Z.; Mehr, A. D.; Allawi, M. F.; Abba, S. I.; Zain, J. M.; Falah, M. W.; Jamei, M.; Bokde, N. D.; Bayatvarkeshi, M.; Al-Mukhtar, M.; Bhagat, S. K.; Tiyasha, T.; Khedher, K. M.; Al-Ansari, N.; Shahid, S.; Yaseen, Z. M. Groundwater Level Prediction Using Machine Learning Models: A Comprehensive Review. *Neurocomputing* **2022**, *489*, 271–308.
(38) Hanoon, M. S.; Ahmed, A. N.; Fai, C. M.; Birima, A. H.; Razzaq, A.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Application of Artificial Intelligence Models for Modeling Water Quality in Groundwater: Comprehensive Review, Evaluation and Future Trends. *Water Air Soil Pollut.* **2021**, *232* (10), 411.
(39) Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* **2022**, *14* (9), No. 1384.
(40) Yan, X.; Zhang, T.; Du, W.; Meng, Q.; Xu, X.; Zhao, X. A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years. *J. Mar. Sci. Eng.* **2024**, *12* (1), No. 159.
(41) Podgorski, J.; Berg, M. Global Threat of Arsenic in Groundwater. *Science* **2020**, *368* (6493), 845–850.
(42) Dharma, V. L.; Nurtanio, N. K.; Nugroho, F. S.; Anggreainy, M. S.; Kurniawan, A. A Review on Machine Learning Methods for Water Quality Prediction. In *2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*; IEEE: IPOH, Malaysia, **2023**, pp 131−138.
(43) Khan, I.; Umar, R. Machine Learning-Driven Optimization of Water Quality Index: A Synergistic ENTROPY-CRITIC Approach Using Spatio-Temporal Data. *Earth Syst. Environ.* **2024**, *8* (4), 1453–1475.
(44) Afrikaner, L. C.; Mapani, B.; Amwele, H. Groundwater in Arid Environments: A Review of Uranium Occurrence and Impacts. *J. Hydrol. Reg. Stud.* **2025**, *62*, No. 102814.
(45) Azmi, A.; Umar, R.; Khan, I. Occurrence, Geochemical Characteristics, and Monte Carlo Simulations−Based Health Risk Assessment of Uranium in Groundwater in Parts of Upper Ganga Basin, India. *Environ. Geochem. Health* **2025**, *47* (10), 404.
(46) Dhapre, M.; Jadhav, S.; Das, D.; Khan, J.; Kim, Y.; Chiao, S.; Danielson, T. A Systematic Review of Machine Learning in Groundwater Monitoring. *Environ. Modell. Software* **2025**, *192*, No. 106549.
(47) Lopez, A. M.; Wells, A.; Fendorf, S. Soil and Aquifer Properties Combine as Predictors of Groundwater Uranium Concentrations within the Central Valley, California. *Environ. Sci. Technol.* **2021**, *55* (1), 352–361.
(48) Khurelbaatar, L.; Batdelger, A.; Khinayat, T.; Oyuntsetseg, B. Pattern Recognition Method from Hydrochemical Parameters to Predict Uranium Concentrations in Groundwater. *Chemometr. Intell. Lab. Syst.* **2022**, *222*, No. 104509.
(49) Niazkar, M.; Piraei, R.; Goodarzi, M. R.; Abedi, M. J. Comparative Assessment of Machine Learning Models for Groundwater Quality Prediction Using Various Parameters. *Environ. Processes* **2025**, *12* (1), 10.
(50) Goodarzi, M. R.; Niknam, A. R.; Barzkar, A.; Niazkar, M.; Zare Mehrjerdi, Y.; Abedi, M. J.; Heydari Pour, M. Water Quality Index Estimations Using Machine Learning Algorithms: A Case Study of Yazd-Ardakan Plain, Iran. *Water* **2023**, *15* (10), No. 1876.
(51) *Guidance for Data Quality Assessment*. https://www.epa.gov/sites/default/files/2015-06/documents/g9-final.pdf (accessed 2025-10-09).
(52) Morita, S. Chemometrics and Related Fields in Python. *Anal. Sci.* **2020**, *36* (1), 107–111.
(53) Terza, J. V.; Basu, A.; Rathouz, P. J. Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *J. Health Econ.* **2008**, *27* (3), 531–543.
(54) Gorman-Lewis, D.; Burns, P. C.; Fein, J. B. Review of Uranyl Mineral Solubility Measurements. *J. Chem. Thermodyn.* **2008**, *40* (3), 335–352.
(55) Van Berk, W.; Fu, Y. Redox Roll-Front Mobilization of Geogenic Uranium by Nitrate Input into Aquifers: Risks for Groundwater Resources. *Environ. Sci. Technol.* **2017**, *51* (1), 337–345.
(56) Chaudhari, U.; Mehta, M.; Sahoo, P. K.; Mittal, S.; Tiwari, R. P. Co-Occurrence of Geogenic Uranium and Fluoride in a Semiarid Belt of the Punjab Plains, India. *Groundwater Sustainable Dev.* **2023**, *23*, No. 101019.
(57) Li, J.; Wang, Y.; Zhu, C.; Xue, X.; Qian, K.; Xie, X.; Wang, Y. Hydrogeochemical Processes Controlling the Mobilization and Enrichment of Fluoride in Groundwater of the North China Plain. *Sci. Total Environ.* **2020**, *730*, No. 138877.
(58) Alam, N.; Kumar, A.; Singh, D. K.; Kumar, S.; Husain, M. A.; Neidhardt, H.; Eiche, E.; Marks, M.; Biswas, A. Testing the Hypothesis of Fluoride and Uranium Co-Mobilization into Groundwater by Competitive Ion Exchange in Alluvial Aquifers of Southern Punjab, India. *J. Hazard. Mater.* **2025**, *492*, No. 138267.
(59) Pandit, P.; Saini, A.; Chidambaram, S.; Kumar, V.; Panda, B.; Ramanathan, A. L.; Sahu, N.; Singh, A. K.; Mehra, R. Tracing Geochemical Sources and Health Risk Assessment of Uranium in Groundwater of Arid Zone of India. *Sci. Rep.* **2022**, *12* (1), 2286.
(60) Sahu, M.; Sar, S. K.; Baghel, T.; Dewangan, R. Seasonal and Geochemical Variation of Uranium and Major Ions in Groundwater at Kanker District of Chhattisgarh, Central India. *Groundwater Sustainable Dev.* **2020**, *10*, No. 100330.
(61) Coyte, R. M.; Jain, R. C.; Srivastava, S. K.; Sharma, K. C.; Khalil, A.; Ma, L.; Vengosh, A. Large-Scale Uranium Contamination of Groundwater Resources in India. *Environ. Sci. Technol. Lett.* **2018**, *5* (6), 341–347.

https://doi.org/10.1021/acssusresmgt.5c00505
ACS Sustainable Resour. Manage. XXXX, XXX, XXX−XXX