

Article

Spatial Mapping and Prediction of Groundwater Quality Using Ensemble Learning Models and SHapley Additive exPlanations with Spatial Uncertainty Analysis

Shilong Yang¹, Danyuan Luo¹, Jiayao Tan¹, Shuyi Li¹, Xiaoqing Song^{1,2,3}, Ruihan Xiong⁴, Jinghan Wang⁵, Chuanming Ma^{1,*} and Hanxiang Xiong^{1,*}

¹ School of Environmental Studies, China University of Geosciences, Wuhan 430079, China; yangshilong@cug.edu.cn (S.Y.); luodanyuan@cug.edu.cn (D.L.); tanjy@cug.edu.cn (J.T.); syli@cug.edu.cn (S.L.); songxiaoqing234@163.com (X.S.)

² 111 Geological Party, Guizhou Bureau of Geology and Mineral Exploration & Development, Guiyang 550024, China

³ Geo-Engineering Investigation Institute of Guizhou Province, Guiyang 550008, China

⁴ State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430079, China; xiongrh@cug.edu.cn

⁵ School of Energy Science and Engineering, Central South University, Changsha 410017, China; 223912081@csu.edu.cn

* Correspondence: machuanming@cug.edu.cn (C.M.); hanxiangxiong@cug.edu.cn (H.X.)

Abstract: The spatial mapping and prediction of groundwater quality (GWQ) is important for sustainable groundwater management, but several research gaps remain unexplored, including the inaccuracy of spatial interpolation, limited consideration of the geological environment and human activity effects, limitation to specific pollutants, and unsystematic indicator selection. This study utilized the entropy-weighted water quality index (EWQI), the LightGBM model, the pressure-state-response (PSR) framework and SHapley Additive exPlanations (SHAP) analysis to address the above research gaps. The normalized importance (NI) shows that NO_3^- (0.208), Mg^{2+} (0.143), SO_4^{2-} (0.110), Cr^{6+} (0.109) and Na^+ (0.095) should be prioritized as parameters for remediation, and the skewness EWQI distribution indicates that although most sampled locations have acceptable GWQ, a few areas suffer from severely poor GWQ. The PSR framework identifies 13 indicators from geological environments and human activities for the SMP of GWQ. Despite high AUROCs (0.9074, 0.8981, 0.8885, 0.9043) across four random training and testing sets, it was surprising that significant spatial uncertainty was observed, with Pearson correlation coefficients (PCCs) from 0.5365 to 0.8066. We addressed this issue by using the spatial-grid average probabilities of four maps. Additionally, population and nighttime light are key indicators, while net recharge, land use and cover (LULC), and the degree of urbanization have the lowest importance. SHAP analysis highlights both positive and negative impacts of human activities on GWQ, identifying point-source pollution as the main cause of the poor GWQ in the study area. Due to the limited research on this field, future studies should focus on six key aspects: multi-method GWQ assessment, quantitative relationships between indicators and GWQ, comparisons of various spatial mapping and prediction models, the application of the PSR framework for indicator selection, the development of methods to reduce spatial uncertainty, and the use of explainable machine learning techniques in groundwater management.



Citation: Yang, S.; Luo, D.; Tan, J.; Li, S.; Song, X.; Xiong, R.; Wang, J.; Ma, C.; Xiong, H. Spatial Mapping and Prediction of Groundwater Quality Using Ensemble Learning Models and SHapley Additive exPlanations with Spatial Uncertainty Analysis. *Water* **2024**, *16*, 2375. <https://doi.org/10.3390/w16172375>

Academic Editors: Tianxiang Wang and Alex Neumann

Received: 4 July 2024

Revised: 15 August 2024

Accepted: 21 August 2024

Published: 24 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: groundwater quality; mapping and prediction; EWQI; ensemble learning models; SHAP; spatial uncertainty

1. Introduction

Groundwater is an indispensable freshwater resource across many regions, supporting essential services such as water supply, agricultural irrigation, and industrial development [1]. However, in recent decades, groundwater quality (GWQ) has been significantly

compromised due to factors such as urbanization, climate change, overexploitation and inadequate resource management [2–6]. In many areas, deteriorated groundwater is still used for drinking purposes due to the absence of alternative sources, which leads to significant health risks to local populations [7–9]. Moreover, as a crucial component of the global hydrological cycle, the degradation of GWQ directly impacts ecological stability and regional sustainability [10,11]. Under these circumstances, the precise spatial mapping and prediction of GWQ is essential for identifying pollution sources and informing comprehensive water management strategies for sustainable groundwater management.

Currently, numerous index-based methods are available for assessing GWQ, including the water quality index (WQI) and its adaptations, such as the entropy-weighted WQI (EWQI) [12], principal component analysis (PCA)—WQI [13], the integrated WQI (IWQI) [14], the drinking WQI (DWQI) [15], and the CCME-WQI [16]. Additionally, other methodologies, like the Nemerow index [17] and the comprehensive pollution index (CPI) [18], are also widely employed in GWQ assessment. However, while these methods enable us to indicate the overall GWQ based on samples, they fail to map the spatial distribution of GWQ across larger areas [19–21]. This research gap restricts decision-makers in managing groundwater spatially, thereby obstructing the implementation of effective groundwater management strategies across different regions.

Currently, for the spatial prediction of GWQ, spatial interpolation is primarily used (e.g., kriging and inverse distance weighting), which estimates values at unsampled locations by assuming that closer points are more similar than distant ones based on spatial autocorrelation [22–27]. However, while spatial interpolation offers a convenient means to estimate the spatial distribution of GWQ, its limitations are significant. Firstly, the accuracy of predictions heavily depends on the density and distribution of sampling points, with sparse data leading to unreliable results [28,29]. Secondly, these methods often assume spatial homogeneity, failing to account for complex environmental variability and external factors like human activities, both of which can significantly influence GWQ [30–32]. Lastly, the accuracy of these methods diminishes near the edges of the study area due to fewer data points, known as edge effects [33,34]. Therefore, there is a critical need for new approaches in GWQ mapping and prediction that address the limitations of these traditional interpolation methods, accommodate spatial heterogeneity, and ensure accuracy in data-sparse areas.

Recent studies have increasingly utilized machine learning (ML) models due to their ability to accurately assess and predict GWQ. For example, Singha et al. [35] conducted a comparative analysis of random forest (RF), extreme gradient boosting (XGBoost), artificial neural networks (ANNs), and deep learning (DL) methods in GWQ assessment, concluding that phosphates have a high relative importance. El Bilali et al. [36] found that while adaptive boosting (AdaBoost) and RF models performed better in forecasting GWQ for irrigation purposes, ANNs and support vector regression (SVR) models exhibited greater generalizability. Jeihouni et al. [37] used five decision-tree-based data mining algorithms to identify high-quality groundwater zones, finding RF to be the most accurate for creating reliable GIS-based GWQ maps. However, while these studies primarily focus on the physical and chemical characteristics of GWQ and their interrelations, they often overlook the critical impacts of geological environment settings (e.g., geology, topography and climate) and human activities (e.g., urbanization and pollution source) on GWQ. This oversight leads to the challenges in accurately assessing the factors contributing to groundwater pollution. Meanwhile, research on ML models for single pollutants like nitrate, arsenic, salinity, and fluoride is more detailed, focusing on their predictive accuracy in GWQ prediction. For example, the ML models, such as RF, ANN, XGBoost, CART, BRT, SVR, and KNN, and Bayesian-based methods are employed in mapping and the spatial prediction of nitrate distribution, incorporating geological environmental parameters [19,38–42]. Podgorski et al. [43] employed RF and multivariate logistic regression (MLR) to screen 25 indicators and predict the distribution of fluoride across India. Xia et al. [44] utilized four models (XGBoost, RF, AdaBoost, and SVM) to perform spatial predictions for fluoride, arsenic,

and iodine in the Hetao Basin, China by considering different environmental factors. Tran et al. [45] identified ten environmental indicators to compare the performance of various ML models in predicting groundwater salinity in coastal areas and found that the Cat-Boost regression model provides the highest accuracy. Podgorski and Berg [46] used RF and eleven geological and climatic indicators to carry out spatial predictions of arsenic contamination globally. However, although these studies account for various geological environmental factors for GWQ spatial prediction, they primarily focus on specific pollutants. The analysis of all GWQ parameters and environmental stress indicators in these studies is insufficient, and the application of ML models for the spatial prediction of overall GWQ indexes, such as the EWQI, the IWQI, and the CPI, remains underexplored.

The pressure-state-response (PSR) framework, developed by the Organization for Economic Cooperation and Development (OECD), provides a structured framework to analyze the dynamic interactions between human activities, environmental conditions, and management responses [47–49]. This framework has been ingrained in regulatory approvals and development management plans across Australia, especially for mining applications and environmental assessments. Also, many studies, such as those focused on ecological security [50], ecological vulnerability [51], forest management [52], mine area contamination assessment [53], and urban resilience [54], also demonstrated its wide application. However, its application in GWQ prediction for indicator selection is still limited. In fact, in GWQ prediction, there is a lack of systematic approach in the selection of indicators, especially concerning those related to human activities. The PSR framework aptly provides a structured methodology to identify and analyze various indicators crucial for assessing GWQ. This framework is especially suitable in rapidly urbanizing areas, where the dual impact of human activities on GWQ (both detrimental and beneficial) presents a complex challenge that demands further exploration.

Building upon identified research gaps, this study aims to conduct spatial GWQ mapping and prediction using the Guanzhong Basin as the study area. This research integrates the PSR framework, the EWQI, the LightGBM model, and explainable machine learning techniques (EMLTs). In Guanzhong Basin, studies on GWQ primarily include contamination risk [55], human health risk [56], hydrogeochemical processes [57], and water quality assessment [58]. However, the spatial prediction of GWQ has not yet been explored. Therefore, this study includes three innovations: (1) It pioneers the use of the advanced LightGBM model and the EWQI to perform spatial mapping and carry out predictions of GWQ, as this has not been explored; (2) It utilizes the PSR framework to systematically select indicators for GWQ mapping and prediction, considering geological environment indicators, spatial uncertainty and dual impacts of human activities on GWQ; (3) It incorporates SHapley Additive exPlanations (SHAP), a widely used advanced explainable machine learning technique (EMLT), to visualize the influence of these indicators on GWQ distribution, thereby supporting decision-making in sustainable groundwater management.

2. Study Area

The Guanzhong Basin, located in the central part of Shaanxi Province, China, serves as an essential agricultural and industrial area; it is bordered by the Qinling Mountains to the south and the Bei Mountains to the north [59]. This basin covers an area of approximately 18,955.25 km² and is characterized by its longitudinal range of 107°–110°30' E and latitudinal span of 34°00'–35°40' N (Figure 1). Geologically, the basin is distinguished by a thick layer of Mesozoic sedimentary rocks and is underlain by complex hydrogeological structures formed from tertiary river-lake facies, heavily influenced by historical tectonic activities [60]. The region's climate is classified as temperate, with four distinct seasons and an average annual temperature of 13.3 °C, and annual rainfall varies from 544 to 863 mm, predominantly occurring during the summer months [55]. However, the area is prone to droughts due to its high annual evaporation rate of 800–1200 mm [61]. Hydrologically, the Guanzhong Basin is dominated by the Weihe River, the largest tributary of the Yellow River, which plays a crucial role in the regional water system by linking surface water

interactions with the groundwater [62]. Despite its natural water richness, the basin faces challenges related to water scarcity and the uneven seasonal distribution of rainfall, which can impact both agricultural productivity and urban water supply. Groundwater in the basin is found mainly in unconfined aquifers with thicknesses varying from 5 to 80 m, predominantly recharged by precipitation and lateral flows from adjacent mountainous regions [55]. The infiltration coefficients of the floodplains and terraces further highlight the complex interaction between surface and groundwater systems. The significant human activities, including dense urbanization and industrial operations in cities like Xi'an and Xianyang, intensify the demand for water and place additional pressures on the groundwater systems. Given these factors, this study aims to leverage advanced machine learning models to enhance the prediction and management of GWQ within the Guanzhong Basin, focusing on integrating environmental, climatic, and anthropogenic indicators to provide a comprehensive analysis of the region's groundwater sustainability.

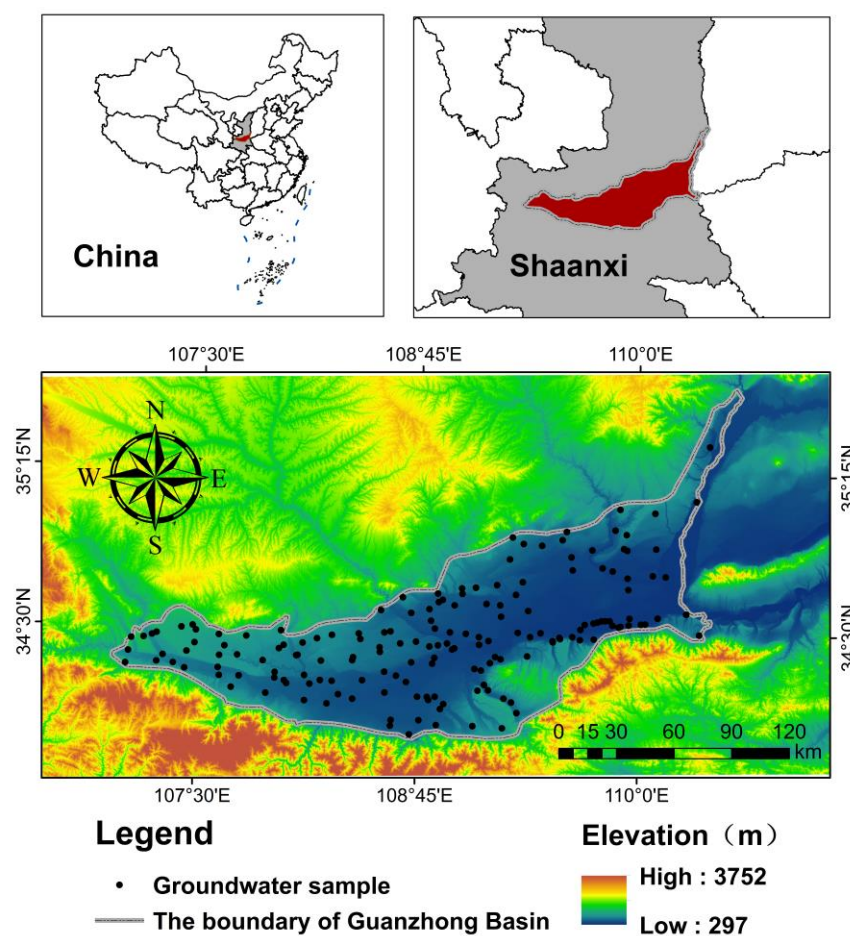


Figure 1. Study area (Guanzhong Basin).

3. Methodology

Figure 2 illustrates the methodological framework of this study, which includes four parts: indicator determination based on the PSR framework; GWQ assessment based on the EWQI; GWQ mapping and spatial prediction based on LightGBM and the TPE; and indicator analysis based on SHAP values. It is crucial to note that GWQ assessment relies on solely groundwater sample analysis using the EWQI to obtain overall GWQ. Conversely, spatial GWQ mapping and prediction extend these EWQI calculations spatially through indicators identified by the PSR framework. Finally, future directions are proposed for sustainable groundwater management.

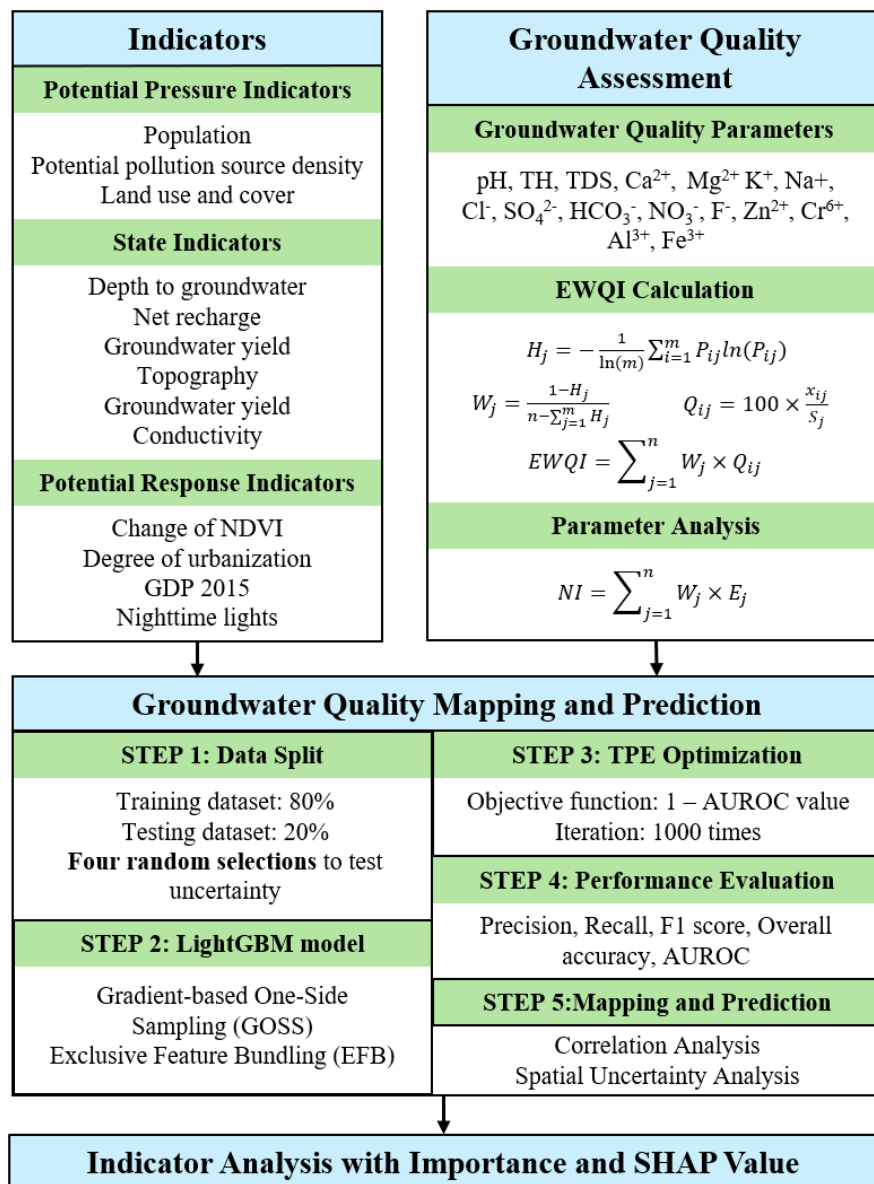


Figure 2. Methodology framework.

3.1. Groundwater Samples Descriptions

The groundwater samples used in this study were derived from the research conducted by Chengzhu et al. [63], which originally comprised 200 groundwater samples. After excluding 10 blank samples and 10 duplicates, a total of 180 samples remained for analysis. The samples were evaluated for various physicochemical parameters and selected for this study if they exceeded the Type III groundwater standards of China (GB/T 14848-2017) [64]. For parameters not covered by the GB/T 14848-2017 standards, the World Health Organization (WHO) drinking water guidelines and a related study [35] were used. A total of 16 indicators were selected, and a description of these data is provided in Table 1.

Table 1. GWQ samples descriptions (mg/L).

Parameters	Min	Max	Mean	SD	Standard
pH	6.96	9.89	7.84	0.32	6.5–8.5
Total Hardness (TH)	9	1885	478.87	291.5	450
Total Dissolved Solids (TDS)	196	10570	1077.61	1098.60	1000
Calcium (Ca ²⁺)	0.56	301	92.58	55.91	75
Magnesium (Mg ²⁺)	1.8	352	60.24	53.36	30
Potassium (K ⁺)	0.13	49.5	3.17	5.82	12
Sodium (Na ⁺)	6.36	1160	140.62	155.88	200
Chloride (Cl ⁻)	3.7	2135	106.54	195.99	250
Sulfate (SO ₄ ²⁻)	1.33	4255	230.9	434.04	250
Bicarbonate (HCO ₃ ³⁻)	117	1349	509.41	189.46	300
Nitrate (NO ₃ ³⁻)	0	373	47.32	53.59	20
Fluoride (F ⁻)	0.12	4.26	0.98	0.77	1
Zinc (Zn ²⁺)	0.001	0.066	0.008	0.008	1
Hexavalent chromium (Cr ⁶⁺)	0.001	0.45	0.033	0.055	0.05
Aluminum (Al ³⁺)	0.003	0.1	0.008	0.01	0.2
Iron (Fe ³⁺)	0	0.35	0.105	0.074	0.3

3.2. GWQ Assessment

3.2.1. EWQI Calculation

The EWQI method employs an entropy-based objective weighting system for each parameter, making it extensively applicable in the assessment of GWQ [12,65–68]. Compared to the traditional WQI, the EWQI provides a more objective evaluation by reducing the subjectivity in parameter weighting, thereby offering a more reliable assessment of GWQ across diverse environments. Given m groundwater samples and n parameters, an $m \times n$ matrix (X) can be constructed, as shown as Equation (1).

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \tag{1}$$

Due to variations in units and magnitudes across different parameters, the normalization step is essential, resulting in a standardized matrix R .

$$R_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{2}$$

where R_{ij} denotes the normalized value of the i^{th} groundwater sample for j^{th} parameter, and $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the j^{th} parameter, respectively.

The entropy value (H_j) is then calculated to determine its relative importance by Equation (3), as lower entropy indicates greater parameter significance.

$$H_j = -\frac{1}{\ln(m)} \sum_{i=1}^m P_{ij} \ln(P_{ij}) \tag{3}$$

where $P_{ij} = \frac{R_{ij}}{\sum_{i=1}^m R_{ij}}$ and $P_{ij} \ln(P_{ij}) = 0$ if $P_{ij} = 0$.

The weight W_j for each parameter is subsequently derived from the entropy values using Equation (4).

$$W_j = \frac{1 - H_j}{n - \sum_{j=1}^m H_j} \tag{4}$$

To evaluate the quality index (Q_{ij}) of groundwater sample i for parameter j , Equation (5) is employed.

$$Q_{ij} = \begin{cases} 100 \times \frac{x_{ij}}{S_j} \\ \frac{C_{pH} - 7}{S_{pH} - 7} \end{cases} \quad (5)$$

where S_j refers to the reference value for parameter j (See Table 1); C_{pH} is the value of pH; and S_{pH} is the permissible limit of pH, which can be either 6.5 or 8.5. When the pH is less than 7, the limit is set at 6.5; when it is greater than 7, the limit is set at 8.5.

Finally, the EWQI is computed the following equation:

$$EWQI = \sum_{j=1}^n W_j \times Q_{ij} \quad (6)$$

3.2.2. Parameter Analysis

Similar to other WQIs, the EWQI calculates the overall GWQ of a sample point but cannot assess the conditions of individual parameters within the area. Moreover, existing methods for single-parameter water quality analysis do not fully consider the importance of the parameters and their exceedance rates. Therefore, this study defines a new term, normalized importance (NI), considering both importance and exceedance rates to reflect the priority level for management of a single GWQ parameter in the study area. Parameters with high NI values should be prioritized for management to reduce the impact of groundwater pollution in the study area. The formula for NI is

$$NI = \sum_{j=1}^n W_j \times E_j \quad (7)$$

where W_j is the weight value for parameter j , and E_j is the exceedance rate of groundwater samples for parameter j compared to the reference value.

3.3. GWQ Mapping and Prediction

3.3.1. Data Split

First, 180 groundwater samples were categorized based on their calculated EWQI values into two groups: 0 (90 samples), representing good GWQ, and 1 (90 samples), representing poor GWQ. To determine the optimal split between training and validation sets, we conducted a preliminary analysis by comparing the average model performance and standard deviation (SD) of six split ratios (65/35, 70/30, 75/25, 80/20, 85/15 and 90/10) based on four random selections. This preliminary analysis has been validated as an effective method for selecting training and validation sets [69–72].

3.3.2. Indicator Selection by PSR Framework

Based on the PSR framework, 13 indicators were selected, including potential pressure indicators, state indicators, and potential response indicators. The reason these are considered “potential” is that it is not yet clear whether the human responses they represent are beneficial or detrimental to GWQ, necessitating further investigation. Table 2 displays information on the data sources, scales, and relevant details of these indicators. To ensure consistency across all indicators during modeling, we standardized the spatial resolution. For vector data such as the point, line and polygon, we utilized the vector-to-raster conversion tool in GIS to achieve a resolution of 30 m (Note: point density was used for PPSD). For raster data of varying resolutions, we applied resampling techniques to uniformize the resolution to 30 m. This standardization ensures uniformity and reliability in our data analysis.

Table 2. Thirteen indicators determined by PSR framework.

Group	Indicators	Sources	Scale	Format
Pressure	Population	SEDAC	250 m	Raster
	PPSD	SPDEE	1:300,000	Point
	LULC	Yang and Huang [73]	30 m	Raster
State	Depth to groundwater	MWRPIC	1:300,000	Line
	Net recharge	Peng et al. [74]	1 km	Raster
	Aquifer water yield capacity	Hydrogeological map	1:300,000	Polygon
	Slope	NASADEM data	30 m	Raster
	Impact of the vadose zone	Zhang et al. [55]	1:300,000	Polygon
Potential response	Conductivity	Zhang et al. [55]	1:300,000	Polygon
	GDP2015	GRDC	1 km	Raster
	Ten years change of NDVI	Yang et al. [75]	30 m	Raster
	Degree of urbanization	SEDAC	250 m	Raster
	Nighttime lights	Elvidge et al. [76]	1 km	Raster

Note(s): SEDAC (Socioeconomic Data and Application Center); SPDEE (Shaanxi Provincial Department of ecology and environment); MWRPIC (Ministry of Water Resource of the People’s Republic of China); GRDC (Geographic Data Sharing Infrastructure, global resources data cloud).

Potential Pressure Indicators

Potential pressure indicators are the factors that many influence GWQ through external forces, such as agricultural activities, industrial emissions, and urban expansion. In this study, population, land use and cover (LULC) and potential pollution-source density (PPSD) are selected as potential pressure indicators for GWQ mapping and prediction (Figure 3). The population size may drive groundwater demand and contribute to wastewater and solid waste production, which can contaminate groundwater [77,78]. Higher population densities typically increase the risk of over-extraction and pollution. In terms of LULC, urbanization increases impervious surfaces, decreasing groundwater recharge and increasing runoff that may carry pollutants [79]. Agricultural practices often use fertilizers and pesticides, risking groundwater contamination through leaching and infiltration [80,81]. PPSD measures the concentration of potential contaminant sources like industrial areas, waste disposal sites, and chemically intensive agriculture [82]. Higher PPSD elevates the contamination risk to groundwater systems.

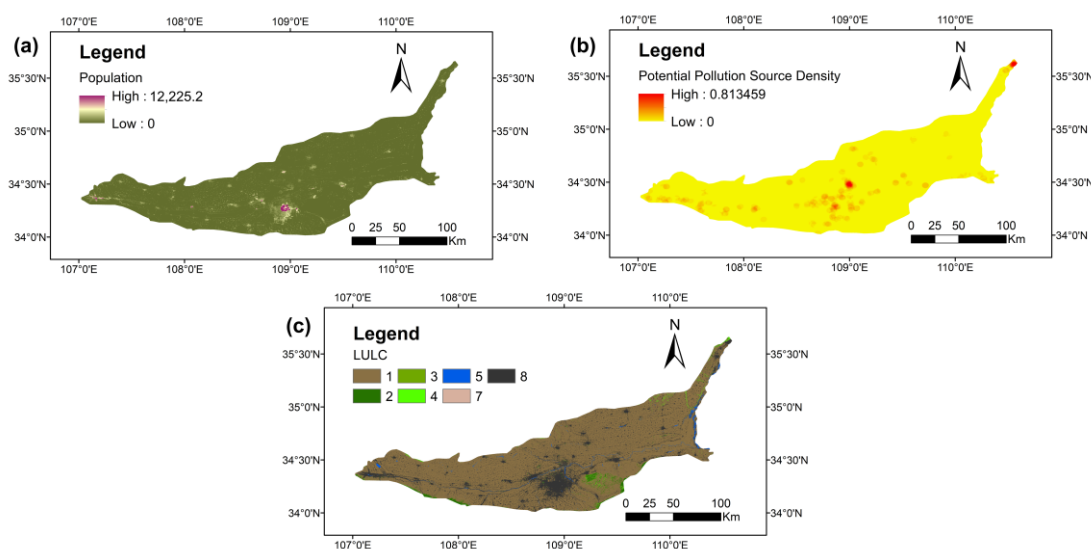


Figure 3. Potential pressure indicators. (a) Population; (b) PPSD; (c) LULC (1: Cropland; 2: Forest; 3: Shrub; 4: Grassland; 5: Water; 7: Barren; 8: Impervious).

State Indicators

State indicators, referencing the DRASTIC model in groundwater vulnerability, are utilized to measure the fundamental hydrogeological conditions that determine GWQ. Groundwater vulnerability is used to measure capacity to resist pollutants entering the groundwater system, which reflects the current state and attributes of the area [83]. In this study, six indicators were selected, including depth to groundwater, net recharge, aquifer water yield capacity, slope, impact of the vadose zone, and conductivity (Figure 4). We removed the indicator of soil media because the net recharge is calculated by multiplying rainfall by an infiltration coefficient, which is determined based on soil type. The influence of these indicators on GWQ and the scores for each indicator can be found in many groundwater vulnerability studies [83–89]. A higher score indicates that the aquifer system in the area is more vulnerable to contamination, while a lower score suggests greater resistance to pollution.

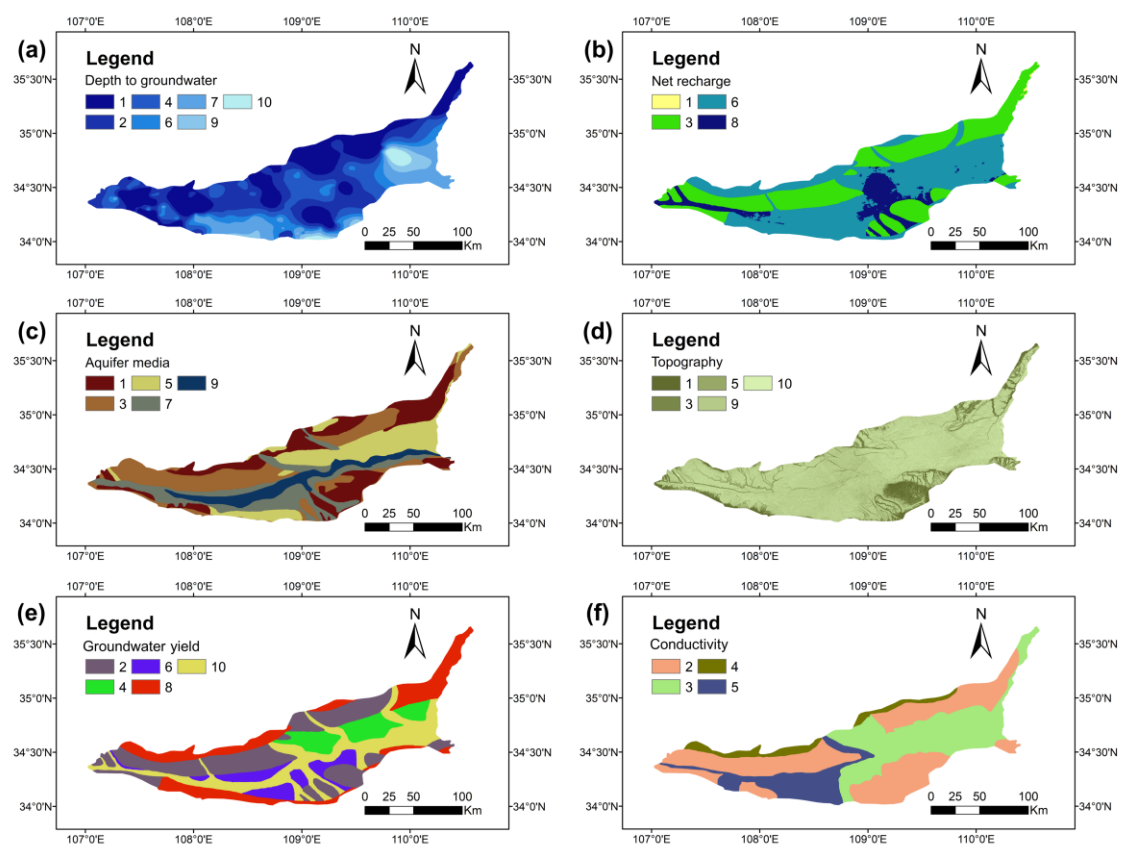


Figure 4. State indicators. (a) Depth to groundwater; (b) Net recharge; (c) Aquifer media; (d) Topography; (e) Groundwater yield; (f) Conductivity.

Potential Response Indicators

Response indicators represent the measures and policies implemented to address or mitigate impacts on GWQ. Due to the dual impact of human activities on GWQ, we selected several potential indicators for the response category. These include GDP2015, ten-year changes in the NDVI, degree of urbanization, and nighttime lights (Figure 5). GDP reflects economic activity levels, where higher values are not only linked to greater environmental impacts from industrial and agricultural runoff affecting GWQ but also indicate increased potential for funding and implementing policies aimed at mitigating these impacts [51,90]. Ten-year changes in the NDVI, the degree of urbanization settlement, and nighttime lights collectively represent the dual impacts of human activities on GWQ. The changes in the NDVI indicate variations in vegetation cover that can either enhance groundwater recharge and pollutant filtration with increased greenery or reduce

these capabilities through land degradation [91–93]. Urbanization increases impervious surfaces and pollution runoff, degrading natural water infiltration and quality, yet it also prompts opportunities for implementing advanced urban planning and sustainable infrastructure to protect groundwater [3,94,95]. Similarly, increased nighttime lights correlate with intensified urban and industrial activities that elevate contamination risks, but they also mark areas where targeted environmental regulations and remediation efforts can effectively mitigate these impacts [96,97]. Each indicator not only reflects the challenges posed by human activities but also underscores the potential for proactive groundwater management responses.

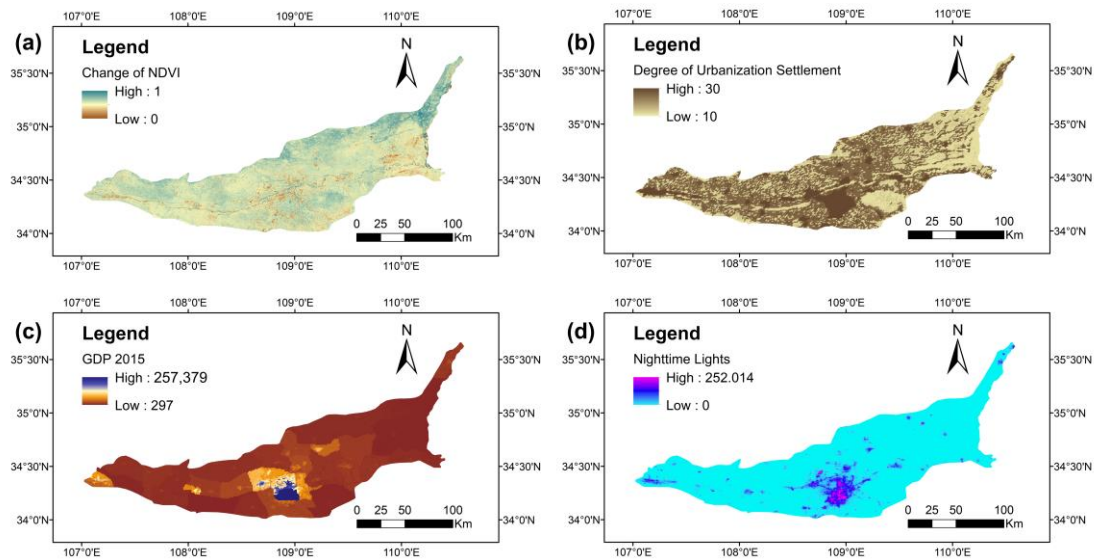


Figure 5. Potential response indicators. (a) Change of NDVI (10 years); (b) Degree of urbanization; (c) GDP 2015; (d) Nighttime lights.

3.3.3. Correlation Analysis

To ensure the accuracy and effectiveness of our model, a correlation analysis was conducted on the dataset variables before modeling. We used the Pearson correlation coefficient (PCC) to identify linear relationships between variables. Any pair of variables with a correlation coefficient exceeding 0.7 was considered strongly correlated, and one variable from each pair was removed to avoid multicollinearity, which can impair model stability and interpretability. The equation of PCC is shown as follows:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (8)$$

where x_i and y_i are the values of the two variables, \bar{x} and \bar{y} are the means of variables, respectively.

3.3.4. LightGBM Model

In our analysis, the LightGBM model, developed by Ke et al. [98], was employed due to its effectiveness in processing large-scale and high-dimensional datasets. This model integrates two key innovations: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) [99]. EFB reduces the dimensionality of the feature space by grouping features that are rarely active at the same time, treating the combination as a graph coloring problem where features are vertices connected by edges when not mutually exclusive. GOSS enhances the training process by focusing on instances with larger gradients by maintaining all instances in the top a percentile (denoted as G_{high}) and

sampling a fraction b from the lower gradients (G_{low}) [100]. The information gain from a feature j at split d is calculated using the formula:

$$\tilde{V}_j(d) = \frac{1}{n} \left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_r^j(d)} \right) \quad (9)$$

where A_l and A_r are subsets of G_{high} , and B_l and B_r are subsets of the sampled G_{low} , with $n_l^j(d)$ and $n_r^j(d)$ representing the number of instances on the left and right sides of the split, respectively.

3.3.5. Hyperparameter Selection and Optimization

Hyperparameter selection and optimization are crucial in ML, as they significantly enhance model performance by fine-tuning settings to align precisely with specific data characteristics and learning objectives [101]. In this study, six hyperparameters and their corresponding parameter spaces are presented in Table 3. The meanings of these hyperparameters are shown in Appendix A.

Table 3. Hyperparameter spaces and optimal hyperparameters.

Hyperparameters	Hyperparameter Spaces
bagging_fraction	hp.uniform('bagging_fraction', 0.5, 0.9)
bagging_freq	hp.choice("bagging_freq", range(4, 7))
boosting_type	hp.choice("boosting_type", ['gbdt', 'dart', 'rf'])
feature_fraction	hp.uniform('feature_fraction', 0.5, 0.9)
learning_rate	hp.uniform('learning_rate', 0.01, 0.5)
num_leaves	hp.choice("num_leaves", range(15, 128))

Also, we used the tree-structured parzen estimator (TPE) for hyperparameter optimization—a method within Bayesian optimization used for its efficiency in refining model parameters by leveraging historical performance data. Specifically, we utilized the “1-area under the receiver operating characteristic curve (AUROC)” as the objective function for minimization, and the iterations were set as 1000. The detailed information about the TPE approach can be found in Xiong et al. [102], Nguyen et al. [103], Rong et al. [104] and Tao et al. [105].

3.3.6. Model Performance Evaluation

Model performance evaluation is crucial in assessing the efficacy and reliability of ML models, providing insights into their predictive accuracy and guiding improvements to ensure robust real-world applications [106,107]. In this study, building on previous research [108,109], we utilized precision (Equation (10)), recall (Equation (11)), F_1 score (Equation (12)), overall accuracy (OA) (Equation (13)), and AUROC (Equation (14)) as metrics to comprehensively evaluate the performance of the model.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F_1 \text{ score} = 2 \times \frac{2TP}{2TP + FP + FN} \quad (12)$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$AUROC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \times \left(\frac{TPR_{i+1} + TPR_i}{2} \right) \quad (14)$$

In this analysis, *TP* denotes instances where the model correctly identified poor GWQ samples, while *FP* indicates cases where good GWQ samples were mistakenly classified as poor. *TN* represents instances where good GWQ samples were correctly recognized, and *FN* refers to cases where the model failed to identify poor GWQ samples. Additionally, FPR_i measures the proportion of good GWQ samples incorrectly classified as poor at the i^{th} threshold. TPR_i quantifies the percentage of actual poor GWQ instances that were correctly identified at the same threshold.

3.4. Spatial Uncertainty Analysis

Some studies have indicated that different sampling methods can affect the uncertainty of the results [110], and similar model performance does not necessarily indicate similar spatial distributions [111]. Therefore, we conducted a spatial uncertainty analysis for four random selections. The spatial uncertainty of four GWQ maps was assessed by two methods. The first approach involves calculating the PCC for all grid cells between each pair of the four GWQ maps to obtain an overall correlation. Stronger correlations (closer to 1) in PCC analysis indicate lower spatial uncertainty

The second method is to spatially visualize the uncertainty between each pair of the four GWQ maps across the study area. For a specific grid cell at position (i, j) in the study area, the spatial uncertainty can be calculated using Equation (15).

$$Uncertainty = \frac{2}{n(n-1)} \sum_{k=1}^{n-1} \sum_{l=k+1}^n |x(k, j) - x(l, j)| \quad (15)$$

where n is the number of GWQ maps, and $x(k, j)$ represents the probability of GWQ from the k^{th} map at grid cell (i, j) . In this study, n equals 4.

3.5. Indicator Importance Analysis and SHAP Analysis

SHAP analysis offers a systematic approach within an EMLT to quantitatively detail the contribution of each feature to a model's predictions; it has been widely used in the groundwater field [112–114]. This approach is crucial for understanding the role of input features (indicators) in determining model outcomes [115]. SHAP analysis is frequently utilized with ensemble models like XGBoost [116], LightGBM [117], CatBoost [118], and RF [119] due to its code compatibility. In SHAP analysis, two key values are included: the SHAP value and the feature value. The former quantifies the impact of each feature (indicator) on the model's prediction. Its positive or negative sign indicates contributions to the binary outcomes of 1 and 0, respectively. The latter refers to the actual value of the indicator itself, which is used as an input in the model. The SHAP method assigns a value to each feature based on its influence, calculated using Equation (16):

$$\varnothing_j = \sum_{S \subseteq N/j} \frac{|S|!(n-|S|-1)!}{n!} [f(S \cup j) - f(S)] \quad (16)$$

where \varnothing_j represents the SHAP value for feature j , derived by summing contributions over all possible subsets of features excluding j ; $|S|$ is the number of features in subset S ; n is the total number of features; $f(S)$ denotes the model's output using subset S without feature j ; and $f(S \cup j)$ is the output when feature j is included.

4. Results

4.1. GWQ Assessment Results

Table 4 presents the weights, exceedance rate and NI of different parameters. The results revealed that the NI of NO_3^- is the highest (0.208), followed by Mg^{2+} (0.143), SO_4^{2-} (0.110), Cr^{6+} (0.109) and Na^+ (0.095). Additionally, while the NI of HCO_3^- , Ca^{2+} , and pH is not high, their exceedance percentages are significantly elevated, at 90%, 60%, and 38.89%, respectively. These findings should prompt significant attention from managers in the study area due to the potential impact on GWQ.

Table 4. Weights, exceedance percentage and NI of parameters.

Parameters	Weights	Exceedance Rate	NI
pH	0.008052	38.89%	0.015521
Total Hardness (TH)	0.021665	34.44%	0.036983
TDS	0.062166	1.67%	0.005146
Calcium (Ca ²⁺)	0.024363	60.00%	0.072454
Magnesium (Mg ²⁺)	0.041843	68.89%	0.142876
Potassium (K ⁺)	0.104982	4.44%	0.023104
Sodium (Na ⁺)	0.068902	27.78%	0.094873
Chloride (Cl ⁻)	0.105335	7.22%	0.037696
Sulfate (SO ₄ ²⁻)	0.107567	20.56%	0.109618
Bicarbonate (HCO ₃ ⁻)	0.014792	90.00%	0.065986
Nitrate (NO ₃ ⁻)	0.067453	62.22%	0.208023
Fluoride (F ⁻)	0.042884	35.56%	0.075585
Zinc (Zn ²⁺)	0.064638	0.00%	0
Hexavalent chromium (Cr ⁶⁺)	0.124156	17.78%	0.109416
Aluminum (Al ³⁺)	0.108349	0.00%	0
Iron (Fe ³⁺)	0.032851	1.67%	0.002719

Figure 6 shows the boxplot and distribution of the EWQI based on 180 groundwater samples. As shown in the boxplot, the median EWQI value is at approximately 64.68, with the interquartile range extending from 48.72 to 100.26. Some outliers can be observed beyond the whiskers, indicating that the distribution is skewed. The histogram further illustrates the distribution of EWQI values, emphasizing a right-skewed trend (skewness = 3.26 and maximum EWQI = 534.72). A majority of the samples cluster towards the lower range of EWQI values (EWQI < 100), which indicates that most of the 180 groundwater samples from the study area are categorized as “Excellent”, “Good” and “Moderate,” using the standards outlined in many studies [67]. As a result, combined with Figure 6, the overall GWQ in the study area is generally good. However, a few areas suffer from severely poor quality, which warrants urgent attention.

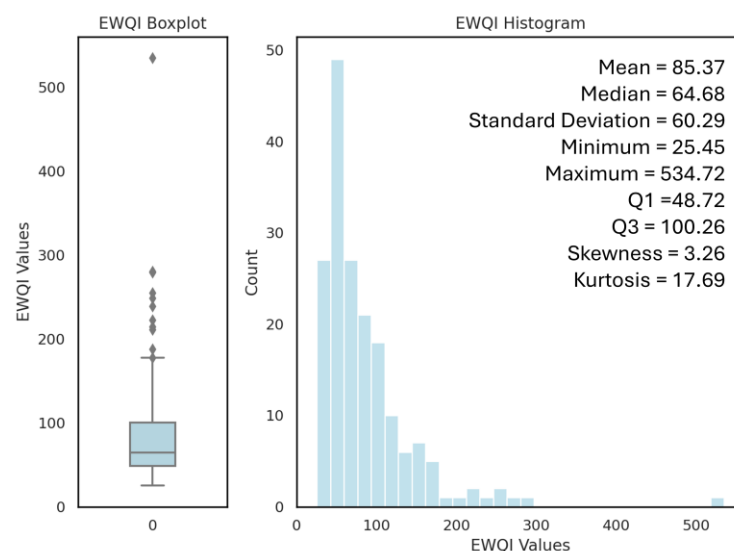


Figure 6. Boxplot and histogram distribution of EWQI.

4.2. Indicator Selection by Correlation Analysis

Figure 7 presents the PCCs between indicators. It is evident that all the PCCs are below 0.7, indicating that multicollinearity among these indicators is not significant. Notably, the coefficient between nighttime light and GDP is 0.63, which can be attributed to the fact that areas with higher economic output often have increased nighttime lighting due

to urbanization and industrial activity [120,121]. Similarly, the PCC between aquifer media and net recharge is 0.66, and this is because the characteristics of aquifer media directly influence the rate and volume of groundwater recharge through their porosity and permeability properties [122]. Despite the correlations among the indicators, their correlations are not strong enough to undermine the models' independence for modeling.

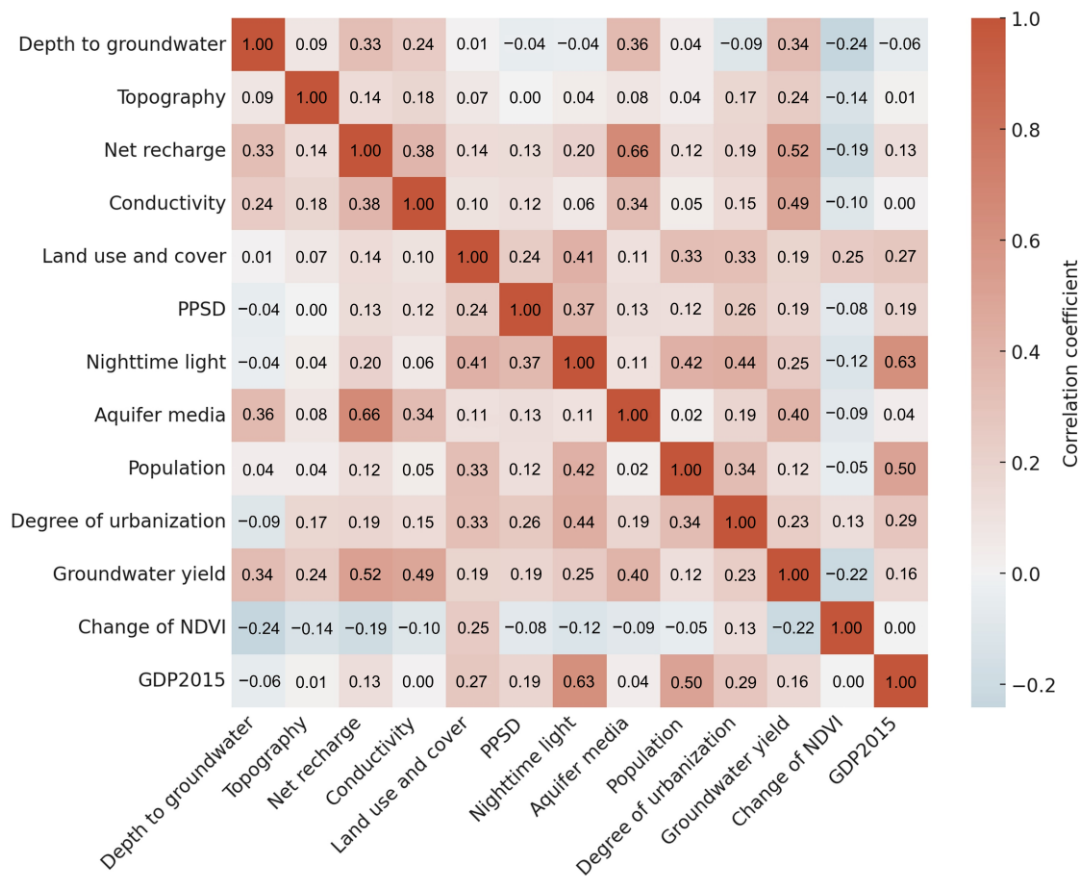


Figure 7. Correlation analysis of indicators.

4.3. Optimal Hyperparameters, Model Performance, and Spatial GWQ Mapping

Based on the preliminary analysis, it was found that the 80/20 split ratio had high model performance (average AUROC = 0.8989) and the highest robustness (SD = 0.0083) (Table 5). Comparing other splits, 80/20 not only maximizes the predictive accuracy but also ensures the model's stability across different test sets. Table 6 illustrates the optimal hyperparameters for four different random sample selections, which highlights that the best optimization results under the TPE approach vary with each sampling method.

Table 5. Data split determination for training and validation dataset by average AUROC and SD.

Selection No.	65/35	70/30	75/25	80/20	85/15	90/10
1	0.8291	0.823	0.8856	0.9074	0.8797	1
2	0.8291	0.8697	0.8573	0.8981	0.8571	1
3	0.7954	0.8011	0.8855	0.8858	0.8546	0.9383
4	0.7808	0.823	0.8601	0.9043	0.8731	0.9877
Average	0.8086	0.8292	0.8721	0.8989	0.8661	0.9815
SD	0.0211	0.0250	0.0135	0.0083	0.0106	0.0254

Table 6. Optimal hyperparameters for different random sample selections.

Hyperparameters	Selection 1	Selection 2	Selection 3	Selection 4
bagging_fraction	0.755129	0.578798	0.628146	0.817471
bagging_freq	5	4	5	5
boosting_type	gbdt	gbdt	gbdt	gbdt
feature_fraction	0.861003	0.561902	0.899878	0.677141
learning_rate	0.310296	0.069457	0.353949	0.314827
num_leaves	102	45	47	21

Figure 8 displays the testing results of the model for four random splits with an 80/20 ratio. It is evident that the model exhibits excellent performance in all four cases, achieving AUROC values of 0.9074, 0.8981, 0.8858, and 0.9043, respectively. Other metrics, including precision, recall, F₁ score, and OA, further corroborate the results. The consistently high performance not only validates the appropriateness of the selected indicators for GWQ mapping and prediction but also reflects the effectiveness of the LightGBM model combined with the TPE, as stated in Li et al. [123], Guo et al. [124] and Li et al. [125]. However, as emphasized by Xiong et al. [102], similar model performances do not necessarily imply similar spatial distributions. We therefore mapped the spatial distribution of GWQ based on the optimal hyperparameters (Figure 9a–d). The natural break method was used to categorize spatial GWQ into five classes (very high, high, moderate, low and very low). The areas of the five classes in the four GWQ maps are displayed in Table 7. It is found that despite similar model performances, there are significant differences in spatial distribution for the four maps. Therefore, we averaged the possibility of each grid cell from the spatial GWQ maps of the four random selections (Figure 9e). This may be a possible strategy to address the spatial uncertainty caused by selecting different datasets.

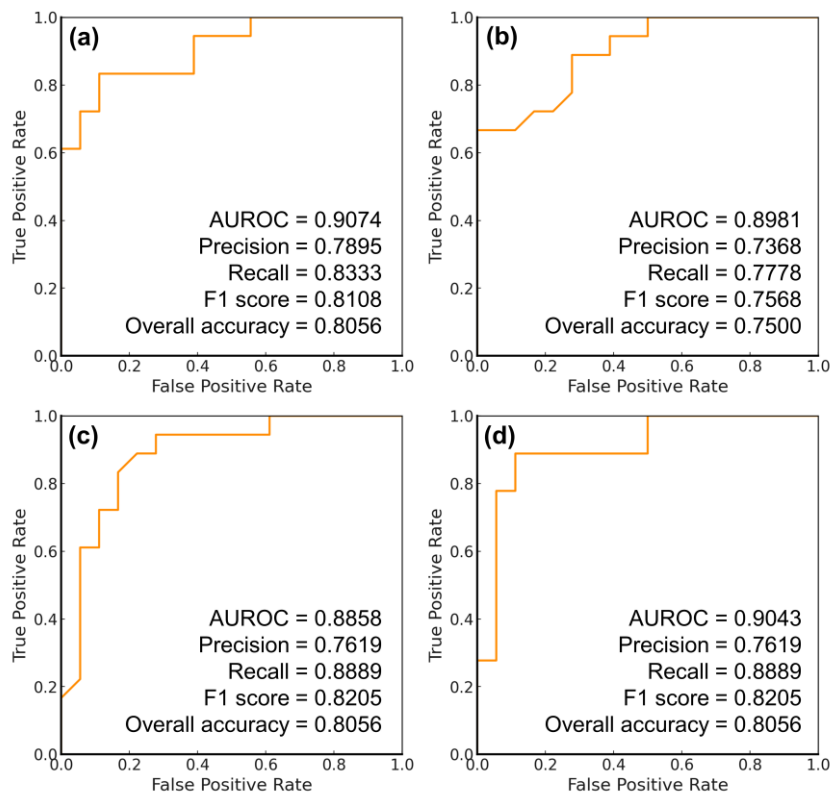


Figure 8. Model performance comparison for four random selections of groundwater samples. (a) Random selection 1; (b) Random selection 2; (c) Random selection 3; (d) Random selection 4.

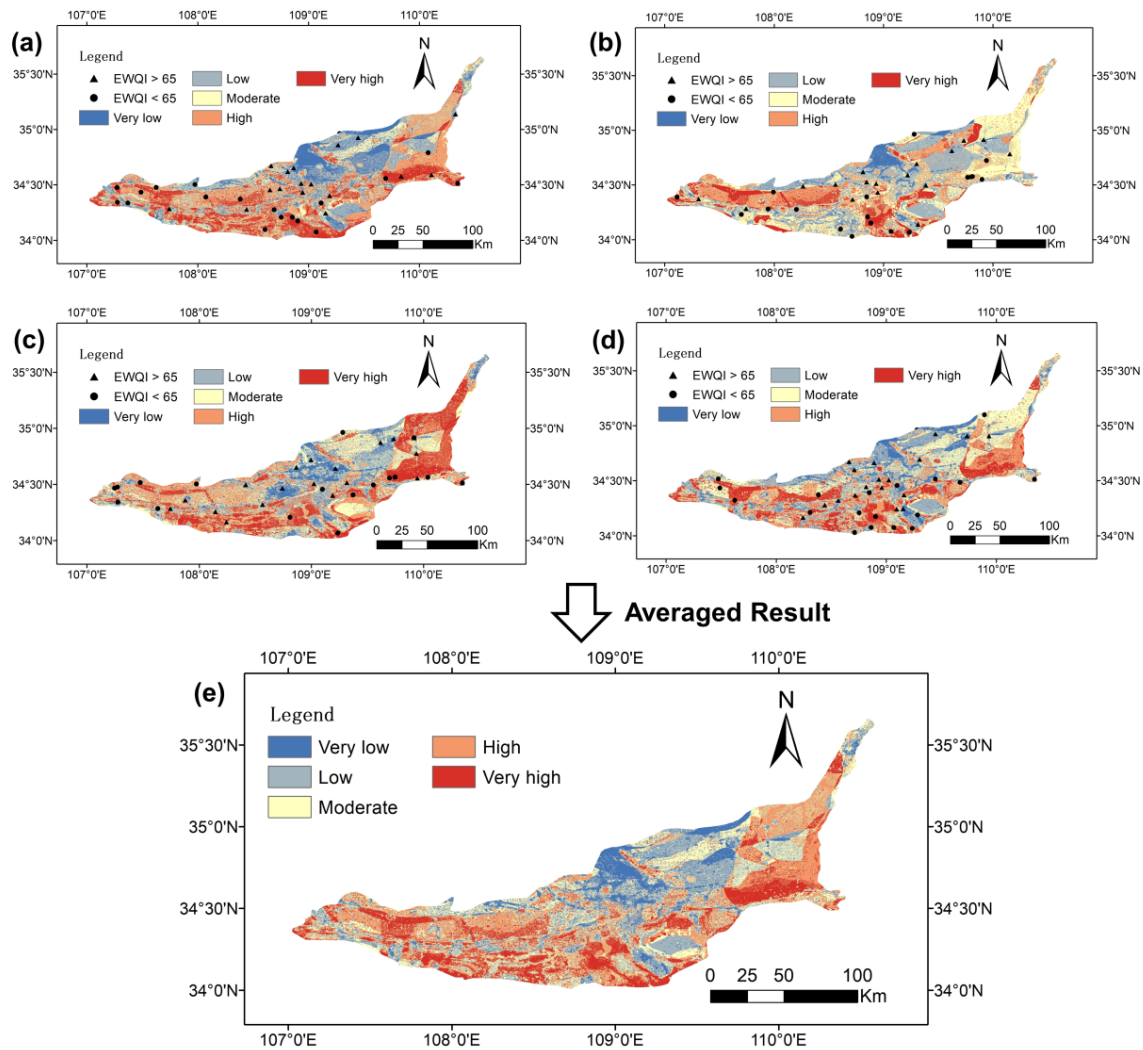


Figure 9. GWQ mapping and prediction with testing data set for 4 random selections. (a) Random selection 1; (b) Random selection 2; (c) Random selection 3; (d) Random selection 4; (e) Average results.

Table 7. Areas of the five classes in the four GWQ maps.

Areas	Selection 1	Selection 2	Selection 3	Selection 4
Very high (km ²)	2625.53	1278.48	1984.58	2261.95
High (km ²)	3736.37	4915.00	3261.57	3981.42
Moderate (km ²)	3559.01	6026.96	4728.50	5181.86
Very low (km ²)	5242.62	4353.83	4695.23	4305.54
Low (km ²)	3791.72	2380.98	4285.36	3224.49

4.4. Spatial Uncertainty in GWQ Mapping

Figure 10 shows the spatial uncertainty analysis results with PCC values. It is found that PCCs among these four maps range from 0.5365 to 0.8066, confirming the observed differences in Figure 9. Particularly notable is the variance between selection 2 and 3, with a PCC of 0.5365. Even for the most similar pair of selections (1 and 4), the PCC is only 0.8066. Figure 10g displays the final results of spatial uncertainty with an average PCC of 0.6707. It indicates that the spatial uncertainty caused by four random selections of training and validation groundwater samples is pronounced. We have highlighted three typical areas with particularly high uncertainty on the map, providing a basis for supplementing groundwater samples in future groundwater management.

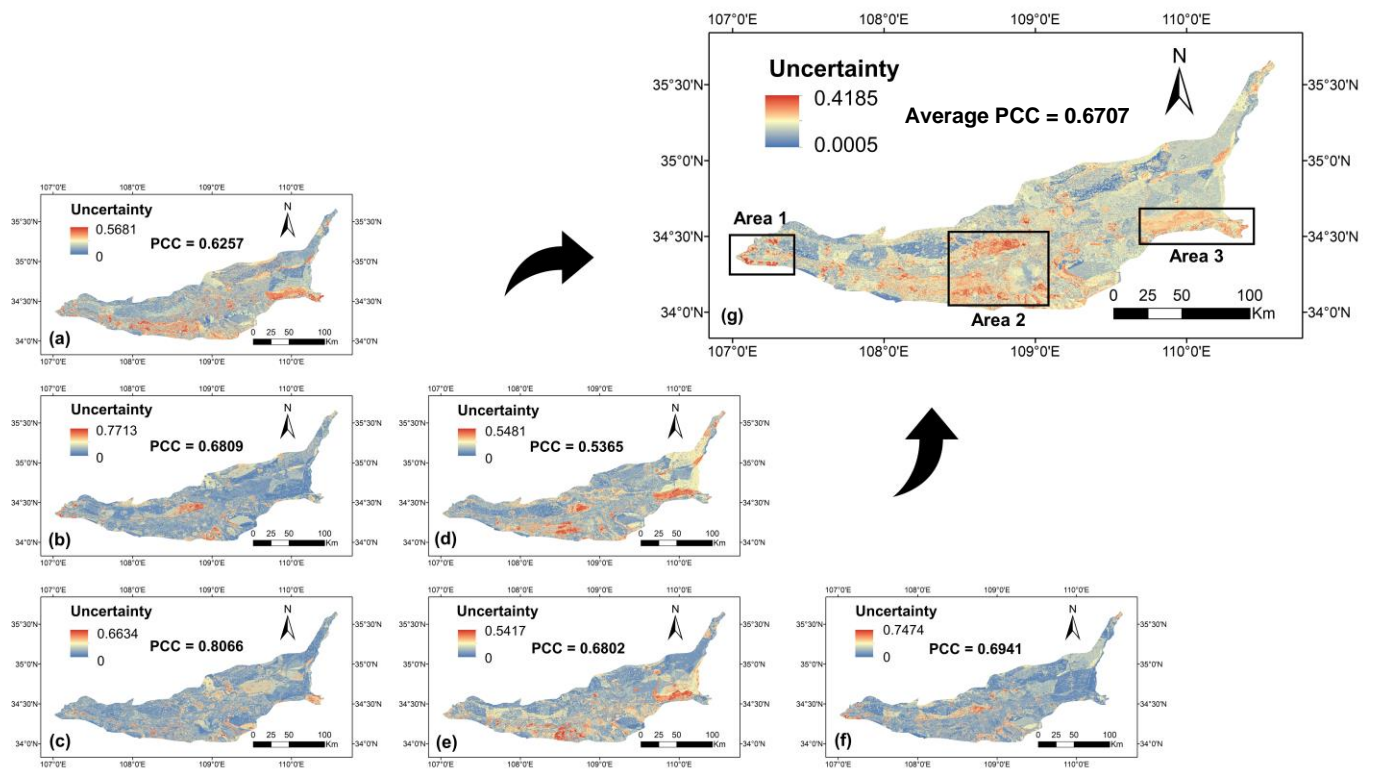


Figure 10. Spatial uncertainty analysis with PCC values. (a) Selection 1 and 2; (b) Selection 1 and 3; (c) Selection 1 and 4; (d) Selection 2 and 3; (e) Selection 2 and 4; (f) Selection 3 and 4; (g) Final average result.

4.5. Indicator Analysis with Importance and SHAP Value

Figure 11 shows the results of feature importance for different randomly selected training and testing sets, based on the Python code “feature importance”. Table 8 shows the accumulated importance of 13 indicators for four random selections. It is evident that population (18.55%) and nighttime light (17.65%) are the most critical indicators, leading over aquifer media, GDP2015, and groundwater yield. After removing population and nighttime light and rerunning the model, a significant decrease in model performance was observed (Appendix B). In contrast, LULC (2.32%), degree of urbanization (1.50%) and net recharge (0.67%) have relatively low importance. Interestingly, both the most and least important indicators include those related to human activities.

Table 8. Accumulated importance of 13 indicators for four random selections.

Indicators	Accumulated Importance	Rank	Proportion
Population	74.22	1	18.55%
Nighttime light	70.59	2	17.65%
Aquifer media	41	3	10.25%
GDP2015	39.49	4	9.87%
Groundwater yield	34.37	5	8.59%
Conductivity	31.23	6	7.81%
Change of NDVI	28.85	7	7.21%
Depth to groundwater	25.21	8	6.30%
PPSD	21.32	9	5.33%
Topography	15.75	10	3.94%
LULC	9.29	11	2.32%
Degree of urbanization	5.99	12	1.50%
Net recharge	2.7	13	0.67%

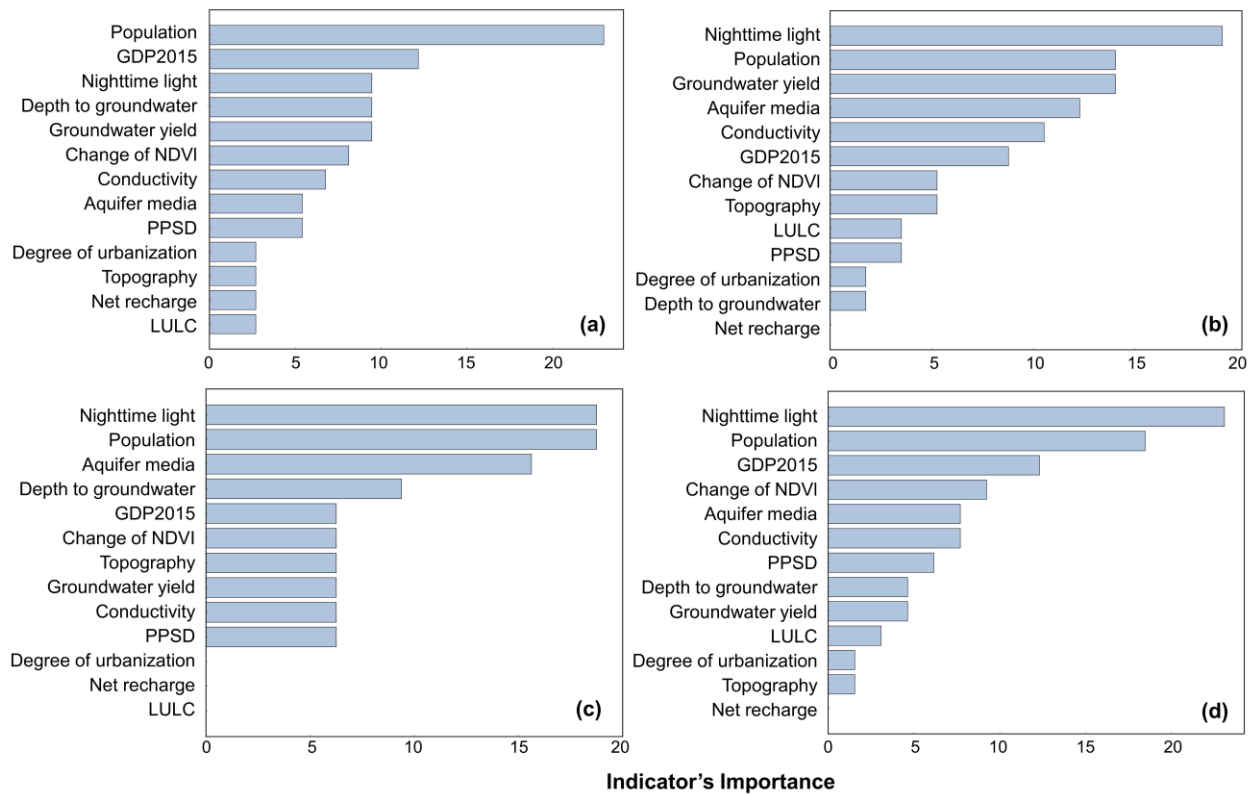


Figure 11. Indicator importance. (a) Selection 1; (b) Selection 2; (c) Selection 3; (d) Selection 4.

Figure 12 shows the SHAP analysis results from four random selections, and high and low feature values (indicator values) are represented by red and blue, respectively. The broader the distribution of an indicator’s feature values, covering a wider range of SHAP values, the more important that indicator is considered. It is found that the indicator importance observed here is consistent with that shown in Figure 11 by “feature importance” code. By analyzing the relationship between feature values and SHAP values, we found that population density, nighttime lights, and GDP, which are theoretically positively correlated, exhibit diverse distributions (Figure 12). Specifically, the contributions of nighttime lights and population to GWQ prediction are opposing, with the areas of poorest GWQ (high SHAP values) being those with higher population density but not necessarily high GDP. Also, the areas with significant NDVI changes and high PPSD exhibit poor GWQ (positive SHAP values), which aligns with the expected outcomes. Additionally, the SHAP analysis results are generally consistent with the scores from the DRASTIC model for groundwater vulnerability, except for the conductivity indicator.

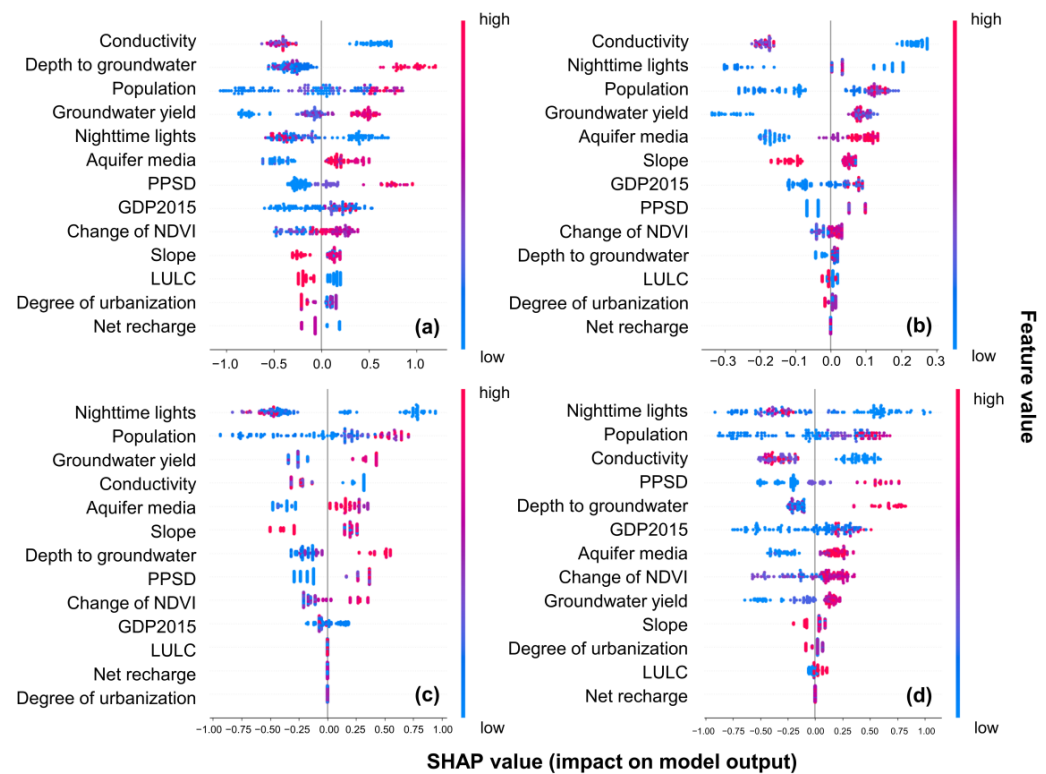


Figure 12. SHAP analysis for training datasets. (a) Selection 1; (b) Selection 2; (c) Selection 3; (d) Selection 4.

5. Discussion

5.1. Discussion on GWQ Assessment

The EWQI offers a refined approach compared to traditional composite GWQ assessment methods like the CPI (GB/T 14848-2017) and the Nemerow index [126]. While the CPI often penalizes overall water quality for a single poor parameter due to its averaging approach, the EWQI mitigates this by using entropy to weight parameters based on their variability and significance. This results in a more balanced and realistic assessment of GWQ [127]. In contrast to the Nemerow index method, which tends to emphasize the worst-case scenario, EWQI provides a broader perspective, integrating various indicators without letting a single outlier skew the overall results [128]. This makes the EWQI particularly useful for creating targeted and effective water management strategies. In fact, if multiple methods are used to evaluate groundwater samples, more comprehensive analysis results can be obtained, and such comparative studies can serve as a future research direction [129,130].

Unlike other studies based on the EWQI [68,126], this paper further considers the quality of individual parameters on the basis of the NI approach. For the single GWQ parameter, the NI value reflects the priority level of management for the parameter. NO_3^- has the highest priority, which is likely highly related to the extensive agricultural land and the use of fertilizers in the study area [56,131,132]. The relatively high NI of Mg^{2+} , SO_4^{2-} , and Na^+ is primarily linked to specific hydrogeochemical processes such as rock weathering, cation exchange, and evaporation [57,133,134]. Additionally, the significant presence of SO_4^{2-} may be attributed to the oxidation of pyrite [135]. It is noted that the NI of Cr^{6+} reached 0.109, highlighting the serious health risks it poses when concentrations exceed safe levels in groundwater. The elevated Cr^{6+} concentrations in groundwater primarily stem from low groundwater velocity in the loess aquifer, cation exchange in alkaline environments, and industrial activities [136]. In terms of other common GWQ parameters such as HCO_3^- , Zn^{2+} and F^- , Fe^{3+} , Al^{3+} , comparing NI values, weights and exceedance rates can provide managers with prioritized information for management. The

managers can develop varying levels of macro-strategic groundwater management plans based on their circumstances, such as economic and policy factors.

5.2. Model Performance and Spatial Uncertainty

In this study, all four selections demonstrated high model performance, with AU-ROC values of around 0.9. However, spatial differences and uncertainties were evident (average PCC = 0.6707). In fact, some studies have demonstrated that similar model performances do not necessarily imply similar spatial distributions in groundwater potential mapping [102] and landslide susceptibility mapping [111]. This is primarily because of the training and testing sets being sourced from specific geographic locations. When these models are generalized spatially, the diversity of indicators and the variation in optimal hyperparameters can lead to inconsistencies in spatial distribution. However, in ML-based spatial predictions, the dataset used for training is inherently limited. It is challenging to guarantee that spatial samples (e.g., groundwater samples) are both sufficient and evenly distributed, which implies that spatial uncertainty cannot be fully eliminated. Therefore, considering that sampling is both time-consuming and expensive, balancing the number of sampling points with spatial uncertainty is an important direction for future research.

Many studies have employed k-fold cross validation to reduce uncertainty in ML models [137–140]. This technique enhances model reliability by ensuring robustness and consistency across different data subsets [141]. However, the effectiveness of k-fold cross-validation in addressing spatial uncertainty is still very limited. This study proposed a possible method that averaged the grid cells from the spatial GWQ maps of four selections. This averaging approach reduces variability and enhances the stability of spatial predictions by mitigating the effects of outliers and random sampling errors. Also, we have highlighted three typical areas with particularly high uncertainty in Figure 10, identifying potential locations for additional groundwater sampling. With more sampling data, the spatial uncertainty of the groundwater quality map is expected to decrease accordingly. However, these methods serve as a starting point, and the current research on spatial uncertainty is notably insufficient. Under this condition, we strongly encourage the development and discussion of more solutions to address spatial uncertainty. Such approaches should be extended beyond GWQ to a broader range of ML-based spatial prediction applications like landslide susceptibility [142], groundwater salinity [143], groundwater potential [72] and nitrate concentrations [19].

5.3. SHAP Observation and Discussion

SHAP analysis is one of the most important EMLTs, and this study further verifies its applicability in spatial GWQ mapping and prediction. In very poor GWQ areas (high SHAP value), the feature value trends of population density and GDP2015 show the variations rather than the expected similarity. The population is depicted in red in poor GWQ areas, while GDP2015 shows the different pattern. The red points for the GDP2015 indicator mainly appear at slightly positive SHAP values, indicating that regions with severe groundwater pollution are not necessarily high-GDP areas. This inconsistency further highlights the importance of considering both the positive and negative impacts of these human related indicators in the analysis. This is why we use the term “potential” in Section 3.3.2 for indicator selection. However, it is found that when it comes to population growth and economic development, many researchers assume their impact on groundwater is primarily groundwater deterioration or some other negative effect [144–147]. Here, we want to emphasize that the positive impacts of these indicators on GWQ should not be ignored. Given the current lack of detailed studies exploring the impacts of economic or population indicators on GWQ, we propose two viewpoints. First, we hypothesize that the impact of these indicators on GWQ may exhibit an inverse U-shape, similar to findings in studies on greenspace and economic growth [148] or population aging and economic growth [149] and carbon emissions and population size [150]. This may be because as economic or population development reaches a certain level, corresponding groundwater

protection and remediation measures are likely to improve, and these indicators may shift from being “pressure” indicators to “response” indicators. Second, we also strongly recommend conducting more research to find evidence that supports the hypothesis or to further investigate the relationship between economic and population indicators and GWQ.

We observed that in areas with poor GWQ (positive SHAP value), the feature value of nighttime light is low while PPSD is high, whereas in areas with good GWQ, the situation is reversed. Nighttime light is often a complex factor that often correlates with GDP, LULC, population, PPSD, urbanization, and other socio-economic indicators (Figure 7). Considering the possible inverse U-shape relationship observed in the analysis of GDP and population, along with the low importance of LULC and degree of urbanization, we have made a reasonable inference regarding the causes of poor GWQ in the study area. Point-source pollution (e.g., industry, farms, mine exploration, hazardous waste disposal sites, and landfills) is the primary cause of GWQ deterioration. The pollution in these areas is characterized by concentrated, localized contamination, often confined to specific sites. Additionally, these areas have low nighttime light intensity and are less influenced by land use and urbanization, which further supports the positive correlation with the PPSD indicator. The impact of point-source pollution on GWQ explains why there is a substantial difference between the spatial distribution map of GWQ created in this study and the groundwater vulnerability map created in current study [55]. This also implies that when using specific pollutant parameter (NO_3^-) to validate groundwater vulnerability models, it is crucial to consider and mitigate the influence of point-source pollution on the results. Therefore, for these potential pollution sources, conducting regular GWQ testing, enhancing wastewater treatment facilities, and providing education on best practices for pollution prevention are top priorities for protecting groundwater in the study area. Additionally, implementing stricter regulations on industrial discharges and monitoring land use changes can further mitigate the risks posed by point-source pollution.

The poorer GWQ in areas with significant NDVI changes indicates that human expansion activities over the past decade have generally degraded GWQ. The low importance of the degree of urbanization further suggests that it is the process of human activity or expansion, rather than the presence of constructed areas themselves, that has led to the deterioration. Human expansion activities often include deforestation, land conversion for agriculture, industrial development, and infrastructure construction [151,152]. These processes contribute to GWQ degradation by disturbing soil and vegetation during land clearing, increasing surface runoff, introducing pollutants before adequate infrastructure is in place, and disrupting natural water recharge areas. Although groundwater protection measures may improve following urban expansion, it is crucial to regulate human activities, particularly by controlling land use changes, implementing sustainable development practices, and enforcing stricter environmental regulations during the human expansion process to address this issue.

In terms of the “state” indicators group, the SHAP analysis results are generally consistent with the scores from the DRASTIC model for groundwater vulnerability [153–155], except for the conductivity indicator. This can also be explained by point-source pollution. In cases of point-source pollution, low conductivity indicates that the polluted groundwater does not easily disperse, thereby affecting regional GWQ. In areas with high conductivity, pollutants from the source diffuse with the water flow, diluting the contaminants and resulting in improved GWQ. Based on the SHAP analysis, the preliminary causes of GWQ condition in the study area have been identified. The next step is for managers to conduct detailed investigations according to these preliminary hypotheses and develop corresponding management strategies. However, it is important to emphasize that SHAP analysis only explains the association between indicators and outcomes from a statistical perspective and does not necessarily imply a definitive causal relationship, particularly for complex indicators like nighttime light. The inferences we make based on SHAP results require further validation through concrete evidence. Causal analysis, as explored by Jia et al. [156], is a direction worth pursuing in future research following SHAP analysis.

5.4. Limitations and Future Research

Although the findings of this study are interesting, there are four limitations to consider. Firstly, for a study area of nearly 20,000 km², the use of 180 sample points is somewhat insufficient for both training and validation sets, which may be a significant reason for spatial uncertainty. However, this issue is common in groundwater studies due to the limited economic and time resources available for extensive sampling. This also highlights the value of our research on spatial uncertainty and provides a foundation for locating where additional groundwater bores need to be drilled. Secondly, although SHAP analysis is an ELMT, the explanation provided is statistical. For some complex socio-economic indicators, the causal relationships between the indicators themselves and between the indicators and the outcomes have yet to be confirmed. Thirdly, this study only used the LightGBM model with an 80/20 split ratio of training and validation datasets for spatial uncertainty analysis. Introducing more models and a broader range of data-split comparisons could increase the stability of the results. Fourth, when determining GWQ standards, we used Chinese, international, and other literature standards, which might introduce some bias to the results. Nonetheless, these limitations have a minimal impact on the findings of this study. Based on this study, future research should include the following six aspects:

- When evaluating GWQ, it is recommended to use multiple methods, including the EWQI, the CPI, and the Nemerow index, and to promote the single parameter analysis method of the NI proposed in this study.
- It is encouraging to confirm the causal relationships between indicators and between the indicators and outcomes, ensuring that the associations identified through SHAP analysis are supported by robust evidence.
- Introduce and compare more models, including deep learning, reinforcement learning, and ensemble learning, to enhance the stability and accuracy of the results.
- Further promote the contribution of the PSR framework in spatial mapping and prediction for indicator selection to ensure the completeness of model construction.
- In addition to calculating spatial average probabilities and supplementing with additional groundwater samples, develop more methods to reduce spatial uncertainty to provide managers with more accurate mapping results.
- Further develop the application of EMLTs in groundwater management.

6. Conclusions

The spatial mapping and prediction of GWQ is essential for identifying pollution sources and informing comprehensive groundwater management strategies. However, this area has not yet been fully explored. The research gaps mainly include the inaccuracy of traditional spatial interpolation for spatial mapping, insufficient consideration of the geological environment and human activities in ML models, the limitation to single pollutants, and the lack of a systematic approach in the selection of indicators. By taking Guanzhong Plain as a case study, this study utilized the EWQI, the LightGBM model, the TPE optimization method, the PSR framework, and SHAP analysis for the spatial mapping and prediction of GWQ, aiming to address the aforementioned research gaps. Through analysis and discussion, we have made several interesting and important findings.

Firstly, according to the NI results for various parameters, NO₃⁻, Mg²⁺, SO₄²⁻, Na⁺ and Cr⁶⁺ should be prioritized for remediation. The skewed distribution of the EWQI indicates that the overall GWQ in the study area is generally good, but a few areas suffer from severely poor quality, which warrants urgent attention. Secondly, based on four randomly selected training and testing sets, although their model performances were high (with AUROC around 0.9), they exhibited spatial uncertainty, with the lowest spatial correlation being only 0.5365 (between selection 2 and 3). This issue is not limited to spatial GWQ mapping and prediction but also extends to other fields. The spatial averaging method and additional groundwater samples may be possible solutions for this issue, but further methods need to be explored. Thirdly, population and nighttime light are the most

critical indicators, while the indicators of net recharge, LULC and degree of urbanization have the lowest importance. Combining SHAP values, we infer that economic development and population have both positive and negative impacts on GWQ, while point-source pollution is the main cause of the decline in GWQ in the study area. Additionally, we speculate that human expansion activities over the past decade have generally had a negative impact on GWQ.

Due to the limited research on the spatial mapping and prediction of GWQ, future studies should include six different aspects in this field, involving multi-method GWQ assessment; causal relationships between indicators and between the indicators and outcomes; the introduction and comparison of more spatial mapping and prediction models; the application of the PSR framework for indicator selection; the development of more methods to reduce spatial uncertainty; and the application of EMLTs in groundwater management. In this way, future research will support the development of the spatial mapping and prediction of GWQ from different perspectives, aiming to further assist groundwater managers in achieving sustainable groundwater management in the future.

Author Contributions: Conceptualization, C.M. and H.X.; methodology, H.X., S.Y. and D.L.; software, S.Y. and S.L.; validation, H.X., D.L., J.T., R.X. and J.W.; formal analysis, H.X., D.L. and J.T.; investigation, J.T., X.S., S.L. and R.X.; resources, X.S.; data curation, S.Y. and J.T.; writing—original draft preparation, H.X. and S.Y.; writing—review and editing, H.X., S.Y., D.L., J.T., S.L., X.S., R.X., J.W. and C.M.; visualization, J.T.; supervision, C.M. and H.X.; project administration, C.M. and X.S.; funding acquisition, X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by Guizhou Provincial Science and Technology Support Plan Project (No. [2022]210); Guizhou Provincial Bureau of Geology and Mineral Resources Research Project (No. [2020]2) and Guiyang Rail Transit Research Project (No. GD3-FW-YJ-03-2020-11-ZB).

Data Availability Statement: Data will be made available on request.

Acknowledgments: We are deeply grateful to Yuzhou Wang for the contributions in generating high-resolution images and for his efforts in enhancing the overall quality of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. The Meaning of Hyperparameters in LightGBM Model

The LightGBM model in this study includes seven hyperparameters, the significance of which is detailed in Table A1. These hyperparameters play a crucial role in controlling the model's complexity, preventing overfitting, and optimizing the trade-off between model accuracy and training efficiency.

Table A1. The meaning of hyperparameters in the LightGBM model.

Hyperparameters	Meanings
bagging_fraction	This parameter specifies the fraction of data to be randomly selected for each iteration, which helps in preventing overfitting.
bagging_freq	This defines how frequently (in terms of iterations) bagging is performed. For instance, setting it to 5 means that bagging is applied every five iterations.
boosting_type	This parameter determines the type of boosting algorithm to use.
feature_fraction	This controls the fraction of features (columns) to be randomly selected for each iteration, helping to improve model generalization.
learning_rate	This is the step size that controls how much the model is adjusted with each iteration, balancing the trade-off between model accuracy and training time.
num_leaves	This specifies the maximum number of leaves in one tree, which directly impacts the complexity and accuracy of the model.

Appendix B. The Results of Model Performance after Removing Population and Nighttime Light

Given that population and nighttime light are important indicators, we removed the two indicators and conducted four additional tests to validate this finding (Table A2). This shows a significant decline in model performance, further confirming the crucial role these indicators play in accurately predicting GWQ.

Table A2. Model performance after removing population and nighttime light.

Performance Metrics	Test 1	Test 2	Test 3	Test 4
AUROC	0.8348	0.8534	0.8673	0.8380
Precision	0.8333	0.9167	0.7778	0.7857
Recall	0.6	0.6111	0.7778	0.6111
F1 score	0.6977	0.7333	0.7778	0.6875
Overall accuracy	0.6389	0.7778	0.7778	0.7222

References

- Gleeson, T.; Wada, Y.; Bierkens, M.F.; Van Beek, L.P. Water balance of global aquifers revealed by groundwater footprint. *Nature* **2012**, *488*, 197–200. [[CrossRef](#)] [[PubMed](#)]
- Belitz, K.; Fram, M.S.; Johnson, T.D. Metrics for assessing the quality of groundwater used for public supply, CA, USA: Equivalent-population and area. *Environ. Sci. Technol.* **2015**, *49*, 8330–8338. [[CrossRef](#)]
- Brindha, K.; Schneider, M. Impact of urbanization on groundwater quality. *GIS Geostat. Tech. Groundw. Sci.* **2019**, *2019*, 179–196.
- Barbieri, M.; Barberio, M.D.; Banzato, F.; Billi, A.; Boschetti, T.; Franchini, S.; Gori, F.; Petitta, M. Climate change and its effect on groundwater quality. *Environ. Geochem. Health* **2023**, *45*, 1133–1144. [[CrossRef](#)]
- Foster, S.; Chilton, J.; Nijsten, G.-J.; Richts, A. Groundwater—A global focus on the ‘local resource’. *Curr. Opin. Environ. Sustain.* **2013**, *5*, 685–695. [[CrossRef](#)]
- Pophare, A.M.; Lamsoge, B.R.; Katpatal, Y.B.; Nawale, V.P. Impact of over-exploitation on groundwater quality: A case study from WR-2 Watershed, India. *J. Earth Syst. Sci.* **2014**, *123*, 1541–1566. [[CrossRef](#)]
- Karangoda, R.; Nanayakkara, K. Use of the water quality index and multivariate analysis to assess groundwater quality for drinking purpose in Ratnapura district, Sri Lanka. *Groundw. Sustain. Dev.* **2023**, *21*, 100910. [[CrossRef](#)]
- Adimalla, N. Groundwater quality for drinking and irrigation purposes and potential health risks assessment: A case study from semi-arid region of South India. *Expo. Health* **2019**, *11*, 109–123. [[CrossRef](#)]
- Li, P.; Li, X.; Meng, X.; Li, M.; Zhang, Y. Appraising groundwater quality and health risks from contamination in a semiarid region of northwest China. *Expo. Health* **2016**, *8*, 361–379. [[CrossRef](#)]
- Güler, C.; Kurt, M.A.; Korkut, R.N. Assessment of groundwater vulnerability to nonpoint source pollution in a Mediterranean coastal zone (Mersin, Turkey) under conflicting land use practices. *Ocean. Coast. Manag.* **2013**, *71*, 141–152. [[CrossRef](#)]
- Wen, X.; Lu, J.; Wu, J.; Lin, Y.; Luo, Y. Influence of coastal groundwater salinization on the distribution and risks of heavy metals. *Sci. Total Environ.* **2019**, *652*, 267–277. [[CrossRef](#)]
- Amiri, V.; Rezaei, M.; Sohrabi, N. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environ. Earth Sci.* **2014**, *72*, 3479–3490. [[CrossRef](#)]
- Hajji, S.; Ayed, B.; Riahi, I.; Allouche, N.; Boughariou, E.; Bouri, S. Assessment and mapping groundwater quality using hybrid PCA-WQI model: Case of the Middle Miocene aquifer of Hajeb Layoun-Jelma basin (Central Tunisia). *Arab. J. Geosci.* **2018**, *11*, 620. [[CrossRef](#)]
- Zhang, Q.; Qian, H.; Xu, P.; Hou, K.; Yang, F. Groundwater quality assessment using a new integrated-weight water quality index (IWQI) and driver analysis in the Jiaokou Irrigation District, China. *Ecotoxicol. Environ. Saf.* **2021**, *212*, 111992. [[CrossRef](#)] [[PubMed](#)]
- Mohebbi, M.R.; Saeedi, R.; Montazeri, A.; Vaghefi, K.A.; Labbafi, S.; Oktaie, S.; Abtahi, M.; Mohagheghian, A. Assessment of water quality in groundwater resources of Iran using a modified drinking water quality index (DWQI). *Ecol. Indic.* **2013**, *30*, 28–34. [[CrossRef](#)]
- Lumb, A.; Halliwell, D.; Sharma, T. Application of CCME Water Quality Index to monitor water quality: A case study of the Mackenzie River basin, Canada. *Environ. Monit. Assess.* **2006**, *113*, 411–429. [[CrossRef](#)] [[PubMed](#)]
- Yang, Q.; Zhang, J.; Hou, Z.; Lei, X.; Tai, W.; Chen, W.; Chen, T. Shallow groundwater quality assessment: Use of the improved Nemerow pollution index, wavelet transform and neural networks. *J. Hydroinformatics* **2017**, *19*, 784–794. [[CrossRef](#)]
- El Mountassir, O.; Bahir, M.; Ouazar, D.; Chehbouni, A.; Carreira, P.M. Temporal and spatial assessment of groundwater contamination with nitrate using nitrate pollution index (NPI), groundwater pollution index (GPI), and GIS (case study: Essaouira basin, Morocco). *Environ. Sci. Pollut. Res.* **2022**, *29*, 17132–17149. [[CrossRef](#)]

19. Knoll, L.; Breuer, L.; Bach, M. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Sci. Total Environ.* **2019**, *668*, 1317–1327. [[CrossRef](#)]
20. Masocha, M.; Dube, T.; Dube, T. Integrating microbiological and physico-chemical parameters for enhanced spatial prediction of groundwater quality in Harare. *Phys. Chem. Earth Parts A/B/C* **2019**, *112*, 125–133. [[CrossRef](#)]
21. Maroufpoor, S.; Jalali, M.; Nikmehr, S.; Shiri, N.; Shiri, J.; Maroufpoor, E. Modeling groundwater quality by using hybrid intelligent and geostatistical methods. *Environ. Sci. Pollut. Res.* **2020**, *27*, 28183–28197. [[CrossRef](#)]
22. Singh, P.; Verma, P. A comparative study of spatial interpolation technique (IDW and Kriging) for determining groundwater quality. *GIS Geostat. Tech. Groundw. Sci.* **2019**, 43–56. [[CrossRef](#)]
23. Pebesma, E.J.; De Kwaadsteniet, J. Mapping groundwater quality in the Netherlands. *J. Hydrol.* **1997**, *200*, 364–386. [[CrossRef](#)]
24. Ahmad, A.Y.; Saleh, I.A.; Balakrishnan, P.; Al-Ghouti, M.A. Comparison GIS-Based interpolation methods for mapping groundwater quality in the state of Qatar. *Groundw. Sustain. Dev.* **2021**, *13*, 100573. [[CrossRef](#)]
25. Belkhiri, L.; Tiri, A.; Mouni, L. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations. *Groundw. Sustain. Dev.* **2020**, *11*, 100473. [[CrossRef](#)]
26. Chakma, A.; Bhowmik, T.; Mallik, S.; Mishra, U. Application of GIS and geostatistical interpolation method for groundwater mapping. In *Advanced Modelling and Innovations in Water Resources Engineering: Select Proceedings of AMIWRE 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 419–428.
27. Lee, K.-J.; Yun, S.-T.; Yu, S.; Kim, K.-H.; Lee, J.-H.; Lee, S.-H. The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan city, South Korea. *J. Hydrol.* **2019**, *569*, 685–697. [[CrossRef](#)]
28. Paiement, A.; Mirmehdi, M.; Xie, X.; Hamilton, M.C. Integrated segmentation and interpolation of sparse data. *IEEE Trans. Image Process.* **2013**, *23*, 110–125. [[CrossRef](#)] [[PubMed](#)]
29. Rivest, M.; Marcotte, D.; Pasquier, P. Sparse data integration for the interpolation of concentration measurements using kriging in natural coordinates. *J. Hydrol.* **2012**, *416*, 72–82. [[CrossRef](#)]
30. Li, J.; Heap, A.D. Spatial interpolation methods applied in the environmental sciences: A review. *Environ. Model. Softw.* **2014**, *53*, 173–189. [[CrossRef](#)]
31. Li, J.; Heap, A.D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecol. Inform.* **2011**, *6*, 228–241. [[CrossRef](#)]
32. Guo, B.; Yang, F.; Wu, H.; Zhang, R.; Zang, W.; Wei, C.; Jiang, G.; Meng, C.; Zhao, H.; Zhen, X. How the variations of terrain factors affect the optimal interpolation methods for multiple types of climatic elements? *Earth Sci. Inform.* **2021**, *14*, 1021–1032. [[CrossRef](#)]
33. Conolly, J. Spatial interpolation. In *Archaeological Spatial Analysis*; Routledge: London, UK, 2020; pp. 118–134.
34. Gharavi, H.; Gao, S. Spatial interpolation algorithm for error concealment. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1153–1156.
35. Singha, S.; Pasupuleti, S.; Singha, S.S.; Singh, R.; Kumar, S. Prediction of groundwater quality using efficient machine learning technique. *Chemosphere* **2021**, *276*, 130265. [[CrossRef](#)]
36. El Bilali, A.; Taleb, A.; Brouziyne, Y. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric. Water Manag.* **2021**, *245*, 106625. [[CrossRef](#)]
37. Jelihouni, M.; Toomanian, A.; Mansourian, A. Decision tree-based data mining and rule induction for identifying high quality groundwater zones to water supply management: A novel hybrid use of data mining and GIS. *Water Resour. Manag.* **2020**, *34*, 139–154. [[CrossRef](#)]
38. Mahboobi, H.; Shakiba, A.; Mirbagheri, B. Improving groundwater nitrate concentration prediction using local ensemble of machine learning models. *J. Environ. Manag.* **2023**, *345*, 118782. [[CrossRef](#)] [[PubMed](#)]
39. Band, S.S.; Janizadeh, S.; Pal, S.C.; Chowdhuri, I.; Siabi, Z.; Norouzi, A.; Melesse, A.M.; Shokri, M.; Mosavi, A. Comparative analysis of artificial intelligence models for accurate estimation of groundwater nitrate concentration. *Sensors* **2020**, *20*, 5763. [[CrossRef](#)]
40. Gholami, V.; Booi, M. Use of machine learning and geographical information system to predict nitrate concentration in an unconfined aquifer in Iran. *J. Clean. Prod.* **2022**, *360*, 131847. [[CrossRef](#)]
41. Alkindi, K.M.; Mukherjee, K.; Pandey, M.; Arora, A.; Janizadeh, S.; Pham, Q.B.; Anh, D.T.; Ahmadi, K. Prediction of groundwater nitrate concentration in a semiarid region using hybrid Bayesian artificial intelligence approaches. *Environ. Sci. Pollut. Res.* **2022**, *29*, 20421–20436. [[CrossRef](#)]
42. Sajedi-Hosseini, F.; Malekian, A.; Choubin, B.; Rahmati, O.; Cipullo, S.; Coulon, F.; Pradhan, B. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci. Total Environ.* **2018**, *644*, 954–962. [[CrossRef](#)]
43. Podgorski, J.E.; Labhasetwar, P.; Saha, D.; Berg, M. Prediction modeling and mapping of groundwater fluoride contamination throughout India. *Environ. Sci. Technol.* **2018**, *52*, 9889–9898. [[CrossRef](#)]
44. Xia, P.; Zhao, Y.; Xie, X.; Li, J.; Qian, K.; You, H.; Zhang, J.; Ge, W.; Pan, H.; Wang, Y. Machine learning prediction of health risk and spatial dependence of geogenic contaminated groundwater from the Hetao Basin, China. *J. Geochem. Explor.* **2024**, *262*, 107497. [[CrossRef](#)]

45. Tran, D.A.; Tsujimura, M.; Ha, N.T.; Van Binh, D.; Dang, T.D.; Doan, Q.-V.; Bui, D.T.; Ngoc, T.A.; Thuc, P.T.B.; Pham, T.D. Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam. *Ecol. Indic.* **2021**, *127*, 107790. [[CrossRef](#)]
46. Podgorski, J.; Berg, M. Global threat of arsenic in groundwater. *Science* **2020**, *368*, 845–850. [[CrossRef](#)] [[PubMed](#)]
47. Wu, J.; Wang, X.; Zhong, B.; Yang, A.; Jue, K.; Wu, J.; Zhang, L.; Xu, W.; Wu, S.; Zhang, N. Ecological environment assessment for Greater Mekong Subregion based on Pressure-State-Response framework by remote sensing. *Ecol. Indic.* **2020**, *117*, 106521. [[CrossRef](#)]
48. Cheng, H.; Zhu, L.; Meng, J. Fuzzy evaluation of the ecological security of land resources in mainland China based on the Pressure-State-Response framework. *Sci. Total Environ.* **2022**, *804*, 150053. [[CrossRef](#)]
49. Chen, Y.; Xiong, K.; Ren, X.; Cheng, C. An overview of ecological vulnerability: A bibliometric analysis based on the Web of Science database. *Environ. Sci. Pollut. Res.* **2022**, *29*, 12984–12996. [[CrossRef](#)]
50. Lu, T.; Li, C.; Zhou, W.; Liu, Y. Fuzzy Assessment of Ecological Security on the Qinghai–Tibet Plateau Based on Pressure–State–Response Framework. *Remote Sens.* **2023**, *15*, 1293. [[CrossRef](#)]
51. Hu, X.; Ma, C.; Huang, P.; Guo, X. Ecological vulnerability assessment based on AHP-PSR method and analysis of its single parameter sensitivity and spatial autocorrelation for ecological protection—A case of Weifang City, China. *Ecol. Indic.* **2021**, *125*, 107464. [[CrossRef](#)]
52. Wang, Y.-T.; Wang, Y.-S.; Wu, M.-L.; Sun, C.-C.; Gu, J.-D. Assessing ecological health of mangrove ecosystems along South China Coast by the pressure–state–response (PSR) model. *Ecotoxicology* **2021**, *30*, 622–631. [[CrossRef](#)]
53. Weaver, T.; Fridell, P.; Ospina, M.; Brooker, R.; Schenkel, M.; Scrase, A. Contamination assessment of mine infrastructure areas for closure and relinquishment: Hazelwood Coal Mine, Victoria, Australia. In *Proceedings of the Mine Closure 2019: Proceedings of the 13th International Conference on Mine Closure, Crawley, Australia, 3–5 September 2019*; pp. 1491–1496.
54. Chen, M.; Jiang, Y.; Wang, E.; Wang, Y.; Zhang, J. Measuring urban infrastructure resilience via pressure-state-response framework in four Chinese municipalities. *Appl. Sci.* **2022**, *12*, 2819. [[CrossRef](#)]
55. Zhang, Q.; Li, P.; Lyu, Q.; Ren, X.; He, S. Groundwater contamination risk assessment using a modified DRATICL model and pollution loading: A case study in the Guanzhong Basin of China. *Chemosphere* **2022**, *291*, 132695. [[CrossRef](#)] [[PubMed](#)]
56. Wang, Y.; Li, P. Appraisal of shallow groundwater quality with human health risk assessment in different seasons in rural areas of the Guanzhong Plain (China). *Environ. Res.* **2022**, *207*, 112210. [[CrossRef](#)] [[PubMed](#)]
57. Ren, X.; Li, P.; He, X.; Su, F.; Elumalai, V. Hydrogeochemical processes affecting groundwater chemistry in the central part of the Guanzhong Basin, China. *Arch. Environ. Contam. Toxicol.* **2021**, *80*, 74–91. [[CrossRef](#)]
58. Nsabimana, A.; Li, P. Hydrogeochemical characterization and appraisal of groundwater quality for industrial purpose using a novel industrial water quality index (IndWQI) in the Guanzhong Basin, China. *Geochemistry* **2023**, *83*, 125922. [[CrossRef](#)]
59. Dong, M.; Wang, Z.-x.; Dong, H.; Ma, L.-c.; Zhang, L.-y. Characteristics of helium accumulation in the Guanzhong Basin, China. *China Geol.* **2019**, *2*, 218–226. [[CrossRef](#)]
60. Wang, Z.; Wang, J.; Yu, D.; Chen, K. Groundwater potential assessment using GIS-based ensemble learning models in Guanzhong Basin, China. *Environ. Monit. Assess.* **2023**, *195*, 690. [[CrossRef](#)] [[PubMed](#)]
61. Bei, N.; Xiao, B.; Meng, N.; Feng, T. Critical role of meteorological conditions in a persistent haze episode in the Guanzhong basin, China. *Sci. Total Environ.* **2016**, *550*, 273–284. [[CrossRef](#)]
62. Kong, F.; Song, J.; Zhang, Y.; Fu, G.; Cheng, D.; Zhang, G.; Xue, Y. Surface water-groundwater interaction in the Guanzhong section of the Weihe River basin, China. *Groundwater* **2019**, *57*, 647–660. [[CrossRef](#)]
63. Chengzhu, L.; Hongyun, M.; Yaoguo, W. An Inorganic Index dataset of groundwater in the Guanzhong Basin (2015). *Geol. China* **2018**, *45*, 23–29.
64. GB/T 14848-2017; Standard for Groundwater Quality. China Quality and Standards Publishing & Media Co., Ltd.: Beijing, China, 2017.
65. Gao, M.; Qian, J.; Li, X.; Wang, Z.; Hou, X.; Gui, C.; Bai, Z.; Li, J.; Zuo, X.; Zhao, C. Assessment of groundwater quality using Entropy-Weighted Quality Index (EWQI) and multivariate statistical approaches in Heilongdong Spring Basin, Northern China. *Environ. Earth Sci.* **2024**, *83*, 196. [[CrossRef](#)]
66. Ahmad, S.; Umar, R.; Ahmad, I. Assessment of groundwater quality using Entropy-Weighted Quality Index (EWQI) and multivariate statistical techniques in Central Ganga plain, India. *Environ. Dev. Sustain.* **2024**, *26*, 1615–1643. [[CrossRef](#)]
67. Das, C.R.; Das, S. Coastal groundwater quality prediction using objective-weighted WQI and machine learning approach. *Environ. Sci. Pollut. Res.* **2024**, *31*, 19439–19457. [[CrossRef](#)] [[PubMed](#)]
68. Yang, Y.; Li, P.; Elumalai, V.; Ning, J.; Xu, F.; Mu, D. Groundwater quality assessment using EWQI with updated water quality classification criteria: A case study in and around Zhouzhi County, Guanzhong Basin (China). *Expo. Health* **2023**, *15*, 825–840. [[CrossRef](#)]
69. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*, 13. [[CrossRef](#)]
70. Sachdeva, S.; Kumar, B. Comparison of gradient boosted decision trees and random forest for groundwater potential mapping in Dholpur (Rajasthan), India. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 287–306. [[CrossRef](#)]
71. Martínez-Santos, P.; Renard, P. Mapping groundwater potential through an ensemble of big data methods. *Groundwater* **2020**, *58*, 583–597. [[CrossRef](#)]

72. Xiong, H.; Guo, X.; Wang, Y.; Xiong, R.; Gui, X.; Hu, X.; Li, Y.; Qiu, Y.; Tan, J.; Ma, C. Spatial prediction of groundwater potential by various novel boosting-based ensemble learning models in mountainous areas. *Geocarto Int.* **2023**, *38*, 2274870. [[CrossRef](#)]
73. Yang, J.; Huang, X. 30 m annual land cover and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data Discuss.* **2021**, *13*, 3907–3925.
74. Peng, S.; Ding, Y.; Li, Z. High-spatial-resolution monthly temperature and precipitation dataset for China for 1901–2017. *Earth Syst. Sci. Data Discuss.* **2019**, *2019*, 1–23.
75. Yang, J.; Dong, J.; Xiao, X.; Dai, J.; Wu, C.; Xia, J.; Zhao, G.; Zhao, M.; Li, Z.; Zhang, Y. Divergent shifts in peak photosynthesis timing of temperate and alpine grasslands in China. *Remote Sens. Environ.* **2019**, *233*, 111395. [[CrossRef](#)]
76. Elvidge, C.D.; Zhizhin, M.; Ghosh, T.; Hsu, F.-C.; Taneja, J. Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sens.* **2021**, *13*, 922. [[CrossRef](#)]
77. Karak, T.; Bhagat, R.; Bhattacharyya, P. Municipal solid waste generation, composition, and management: The world scenario. *Crit. Rev. Environ. Sci. Technol.* **2012**, *42*, 1509–1630. [[CrossRef](#)]
78. Singh, S.; Raju, N.J.; Gossel, W.; Wycisk, P. Assessment of pollution potential of leachate from the municipal solid waste disposal site and its impact on groundwater quality, Varanasi environs, India. *Arab. J. Geosci.* **2016**, *9*, 131. [[CrossRef](#)]
79. Valtanen, M.; Sillanpää, N.; Setälä, H. The effects of urbanization on runoff pollutant concentrations, loadings and their seasonal patterns under cold climate. *Water Air Soil Pollut.* **2014**, *225*, 1977. [[CrossRef](#)]
80. Srivastav, A.L. Chemical fertilizers and pesticides: Role in groundwater contamination. In *Agrochemicals Detection, Treatment and Remediation*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 143–159.
81. El Alfy, M.; Faraj, T. Spatial distribution and health risk assessment for groundwater contamination from intensive pesticide use in arid areas. *Environ. Geochem. Health* **2017**, *39*, 231–253. [[CrossRef](#)]
82. Li, J.; Shi, Z.; Liu, M.; Wang, G.; Liu, F.; Wang, Y. Identifying anthropogenic sources of groundwater contamination by natural background levels and stable isotope application in Pinggu basin, China. *J. Hydrol.* **2021**, *596*, 126092. [[CrossRef](#)]
83. Xiong, H.; Wang, Y.; Guo, X.; Han, J.; Ma, C.; Zhang, X. Current status and future challenges of groundwater vulnerability assessment: A bibliometric analysis. *J. Hydrol.* **2022**, *615*, 128694. [[CrossRef](#)]
84. Wang, J.; He, J.; Chen, H. Assessment of groundwater contamination risk using hazard quantification, a modified DRASTIC model and groundwater value, Beijing Plain, China. *Sci. Total Environ.* **2012**, *432*, 216–226. [[CrossRef](#)]
85. Hu, X.; Ma, C.; Qi, H.; Guo, X. Groundwater vulnerability assessment using the GALDIT model and the improved DRASTIC model: A case in Weibei Plain, China. *Environ. Sci. Pollut. Res.* **2018**, *25*, 32524–32539. [[CrossRef](#)]
86. Luo, D.; Ma, C.; Qiu, Y.; Zhang, Z.; Wang, L. Groundwater vulnerability assessment using AHP-DRASTIC-GALDIT comprehensive model: A case study of Binhai New Area, Tianjin, China. *Environ. Monit. Assess.* **2023**, *195*, 268. [[CrossRef](#)]
87. Wang, Z.; Xiong, H.; Ma, C.; Zhang, F.; Li, X. Assessment of groundwater vulnerability by applying the improved DRASTIC model: A case in Guyuan City, Ningxia, China. *Environ. Sci. Pollut. Res.* **2023**, *30*, 59062–59075. [[CrossRef](#)]
88. Barbulescu, A. Assessing groundwater vulnerability: DRASTIC and DRASTIC-like methods: A review. *Water* **2020**, *12*, 1356. [[CrossRef](#)]
89. Shirazi, S.M.; Imran, H.; Akib, S. GIS-based DRASTIC method for groundwater vulnerability assessment: A review. *J. Risk Res.* **2012**, *15*, 991–1011. [[CrossRef](#)]
90. Zhou, F.; Su, W.; Zhang, F. Influencing indicators and quantitative assessment of water resources security in karst region based on PSER model—The case of Guizhou. *Sustainability* **2019**, *11*, 5671. [[CrossRef](#)]
91. Parizi, E.; Hosseini, S.M.; Ataie-Ashtiani, B.; Simmons, C.T. Normalized difference vegetation index as the dominant predicting factor of groundwater recharge in phreatic aquifers: Case studies across Iran. *Sci. Rep.* **2020**, *10*, 17473. [[CrossRef](#)]
92. Elbeih, S.F.; El-Zeiny, A.M. Qualitative assessment of groundwater quality based on land use spectral retrieved indices: Case study Sohag Governorate, Egypt. *Remote Sens. Appl. Soc. Environ.* **2018**, *10*, 82–92. [[CrossRef](#)]
93. Rodriguez-Galiano, V.F.; Luque-Espinar, J.A.; Chica-Olmo, M.; Mendes, M.P. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Sci. Total Environ.* **2018**, *624*, 661–672. [[CrossRef](#)]
94. Zhang, F.; Huang, G.; Hou, Q.; Liu, C.; Zhang, Y.; Zhang, Q. Groundwater quality in the Pearl River Delta after the rapid expansion of industrialization and urbanization: Distributions, main impact indicators, and driving forces. *J. Hydrol.* **2019**, *577*, 124004. [[CrossRef](#)]
95. Carlson, M.A.; Lohse, K.A.; McIntosh, J.C.; McLain, J.E. Impacts of urbanization on groundwater quality and recharge in a semi-arid alluvial basin. *J. Hydrol.* **2011**, *409*, 196–211. [[CrossRef](#)]
96. Singh, A.; Srivastav, S.K.; Kumar, S.; Chakrapani, G.J. A modified-DRASTIC model (DRASTICA) for assessment of groundwater vulnerability to pollution in an urbanized environment in Lucknow, India. *Environ. Earth Sci.* **2015**, *74*, 5475–5490. [[CrossRef](#)]
97. Tan, Y.; Xin, Y.; Guo, C.; Lyu, S.; Zhang, G.; Long, Y.; Zhai, Y.; Packham, H.; Zhou, Y.; Tan, H. Impact of urbanization on baseflow characteristics in the central catchment of North China Plain, China. *J. Hydrol. Reg. Stud.* **2023**, *50*, 101527. [[CrossRef](#)]
98. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
99. Hajihosseini, M.; Maghsoudi, A.; Ghezelnash, R. A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm. *Nat. Resour. Res.* **2023**, *32*, 2417–2438. [[CrossRef](#)]

100. Guo, X.; Gui, X.; Xiong, H.; Hu, X.; Li, Y.; Cui, H.; Qiu, Y.; Ma, C. Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms. *J. Hydrol.* **2023**, *621*, 129599. [[CrossRef](#)]
101. Mahmood, J.; Mustafa, G.-e.; Ali, M. Accurate estimation of tool wear levels during milling, drilling and turning operations by designing novel hyperparameter tuned models based on LightGBM and stacking. *Measurement* **2022**, *190*, 110722. [[CrossRef](#)]
102. Xiong, H.; Yang, S.; Tan, J.; Wang, Y.; Guo, X.; Ma, C. Effects of DEM resolution and application of solely DEM-derived indicators on groundwater potential mapping in the mountainous area. *J. Hydrol.* **2024**, *636*, 131349. [[CrossRef](#)]
103. Nguyen, H.-P.; Liu, J.; Zio, E. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. *Appl. Soft Comput.* **2020**, *89*, 106116. [[CrossRef](#)]
104. Rong, G.; Li, K.; Su, Y.; Tong, Z.; Liu, X.; Zhang, J.; Zhang, Y.; Li, T. Comparison of tree-structured parzen estimator optimization in three typical neural network models for landslide susceptibility assessment. *Remote Sens.* **2021**, *13*, 4694. [[CrossRef](#)]
105. Tao, S.; Peng, P.; Li, Y.; Sun, H.; Li, Q.; Wang, H. Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data. *Expert Syst. Appl.* **2024**, *237*, 121294. [[CrossRef](#)]
106. Kumar, R.; Dwivedi, S.B.; Gaur, S. A comparative study of machine learning and Fuzzy-AHP technique to groundwater potential mapping in the data-scarce region. *Comput. Geosci.* **2021**, *155*, 104855. [[CrossRef](#)]
107. Faouzi, E.; Arioua, A.; Namous, M.; Barakat, A.; Mosaid, H.; Ismaili, M.; Eloudi, H.; Houmma, I.H. Spatial mapping of hydrologic soil groups using machine learning in the Mediterranean region. *Catena* **2023**, *232*, 107364. [[CrossRef](#)]
108. Ruidas, D.; Pal, S.C.; Towfiqul Islam, A.R.M.; Saha, A. Hydrogeochemical evaluation of groundwater aquifers and associated health hazard risk mapping using ensemble data driven model in a water scarce plateau region of eastern India. *Expo. Health* **2023**, *15*, 113–131. [[CrossRef](#)]
109. Sarkar, S.K.; Alshehri, F.; Shahfahad; Rahman, A.; Pradhan, B.; Shahab, M. Mapping groundwater potentiality by using hybrid machine learning models under the scenario of climate variability: A national level study of Bangladesh. *Environ. Dev. Sustain.* **2024**, 1–29. [[CrossRef](#)]
110. Nguyen, V.-L.; Shaker, M.H.; Hüllermeier, E. How to measure uncertainty in uncertainty sampling for active learning. *Mach. Learn.* **2022**, *111*, 89–122. [[CrossRef](#)]
111. Vakhshoori, V.; Zare, M. Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomat. Nat. Hazards Risk* **2018**, *9*, 249–266. [[CrossRef](#)]
112. Alshehri, F.; Rahman, A. Coupling machine and deep learning with explainable artificial intelligence for improving prediction of groundwater quality and decision-making in Arid Region, Saudi Arabia. *Water* **2023**, *15*, 2298. [[CrossRef](#)]
113. Niu, X.; Lu, C.; Zhang, Y.; Zhang, Y.; Wu, C.; Saidy, E.; Liu, B.; Shu, L. Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values. *Sci. Total Environ.* **2023**, *904*, 166662. [[CrossRef](#)]
114. Ransom, K.M.; Nolan, B.T.; Stackelberg, P.; Belitz, K.; Fram, M.S. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Sci. Total Environ.* **2022**, *807*, 151065. [[CrossRef](#)]
115. Yang, C.; Chen, M.; Yuan, Q. The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accid. Anal. Prev.* **2021**, *158*, 106153. [[CrossRef](#)]
116. Zhang, J.; Ma, X.; Zhang, J.; Sun, D.; Zhou, X.; Mi, C.; Wen, H. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manag.* **2023**, *332*, 117357. [[CrossRef](#)]
117. Wen, X.; Xie, Y.; Wu, L.; Jiang, L. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP. *Accid. Anal. Prev.* **2021**, *159*, 106261. [[CrossRef](#)] [[PubMed](#)]
118. Joo, C.; Park, H.; Lim, J.; Cho, H.; Kim, J. Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using catboost and shapley additive explanations. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106873. [[CrossRef](#)]
119. Liu, Q.; Gui, D.; Zhang, L.; Niu, J.; Dai, H.; Wei, G.; Hu, B.X. Simulation of regional groundwater levels in arid regions using interpretable machine learning models. *Sci. Total Environ.* **2022**, *831*, 154902. [[CrossRef](#)] [[PubMed](#)]
120. Wu, J.; Wang, Z.; Li, W.; Peng, J. Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. *Remote Sens. Environ.* **2013**, *134*, 111–119. [[CrossRef](#)]
121. Mellander, C.; Lobo, J.; Stolarick, K.; Matheson, Z. Night-time light data: A good proxy measure for economic activity? *PLoS ONE* **2015**, *10*, e0139779. [[CrossRef](#)]
122. Zupanc, V.; Bračič Železnik, B.; Pintar, M.; Čenčur Curk, B. Assessment of groundwater recharge for a coarse-gravel porous aquifer in Slovenia. *Hydrogeol. J.* **2020**, *28*, 1773–1785. [[CrossRef](#)]
123. Li, K.; Xu, H.; Liu, X. Analysis and visualization of accidents severity based on LightGBM-TPE. *Chaos Solitons Fractals* **2022**, *157*, 111987. [[CrossRef](#)]
124. Guo, X.; Xiong, H.; Li, H.; Gui, X.; Hu, X.; Li, Y.; Cui, H.; Qiu, Y.; Zhang, F.; Ma, C. Designing dynamic groundwater management strategies through a composite groundwater vulnerability model: Integrating human-related parameters into the DRASTIC model using LightGBM regression and SHAP analysis. *Environ. Res.* **2023**, *236*, 116871. [[CrossRef](#)]
125. Li, L.; Liu, Z.; Shen, J.; Wang, F.; Qi, W.; Jeon, S. A LightGBM-based strategy to predict tunnel rockmass class from TBM construction data for building control. *Adv. Eng. Inform.* **2023**, *58*, 102130. [[CrossRef](#)]
126. Kumar, P.S.; Augustine, C.M. Entropy-weighted water quality index (EWQI) modeling of groundwater quality and spatial mapping in Uppar Odai Sub-Basin, South India. *Model. Earth Syst. Environ.* **2022**, *8*, 911–924. [[CrossRef](#)]

127. Wang, X.; Liu, B.; He, S.; Yuan, H.; Ji, D.; Li, R.; Song, Y.; Xu, W.; Liu, B.; Xu, Y. Groundwater Environment and Health Risk Assessment in an In Situ Oil Shale Mining Area. *Water* **2024**, *16*, 185. [[CrossRef](#)]
128. Luo, P.; Xu, C.; Kang, S.; Huo, A.; Lyu, J.; Zhou, M.; Nover, D. Heavy metals in water and surface sediments of the Fenghe River Basin, China: Assessment and source analysis. *Water Sci. Technol.* **2021**, *84*, 3072–3090. [[CrossRef](#)] [[PubMed](#)]
129. Patel, P.S.; Pandya, D.M.; Shah, M. A systematic and comparative study of Water Quality Index (WQI) for groundwater quality analysis and assessment. *Environ. Sci. Pollut. Res.* **2023**, *30*, 54303–54323. [[CrossRef](#)] [[PubMed](#)]
130. Zhang, P.; Xiao, M.; Dai, Y.; Zhang, Z.; Liu, G.; Zhao, J. Evaluation of water quality of collected rainwater in the northeastern loess plateau. *Sustainability* **2022**, *14*, 10834. [[CrossRef](#)]
131. Kou, X.; Ding, J.; Li, Y.; Li, Q.; Mao, L.; Xu, C.; Zheng, Q.; Zhuang, S. Tracing nitrate sources in the groundwater of an intensive agricultural region. *Agric. Water Manag.* **2021**, *250*, 106826. [[CrossRef](#)]
132. Serio, F.; Miglietta, P.P.; Lamastra, L.; Ficocelli, S.; Intini, F.; De Leo, F.; De Donno, A. Groundwater nitrate contamination and agricultural land use: A grey water footprint perspective in Southern Apulia Region (Italy). *Sci. Total Environ.* **2018**, *645*, 1425–1431. [[CrossRef](#)]
133. Zhang, Q.; Xu, P.; Qian, H. Assessment of groundwater quality and human health risk (HHR) evaluation of nitrate in the Central-Western Guanzhong Basin, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4246. [[CrossRef](#)]
134. Xu, P.; Feng, W.; Qian, H.; Zhang, Q. Hydrogeochemical characterization and irrigation quality assessment of shallow groundwater in the Central-Western Guanzhong Basin, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1492. [[CrossRef](#)]
135. Kou, X.; Zhao, Z.; Duan, L.; Sun, Y. Hydrogeochemical Behavior of Shallow Groundwater around Hancheng Mining Area, Guanzhong Basin, China. *Water* **2024**, *16*, 660. [[CrossRef](#)]
136. Wang, L.; Li, P.; Duan, R.; He, X. Occurrence, controlling factors and health risks of Cr⁶⁺ in groundwater in the Guanzhong Basin of China. *Expo. Health* **2022**, *14*, 239–251. [[CrossRef](#)]
137. Arabameri, A.; Arora, A.; Pal, S.C.; Mitra, S.; Saha, A.; Nalivan, O.A.; Panahi, S.; Moayedi, H. K-fold and state-of-the-art metaheuristic machine learning approaches for groundwater potential modelling. *Water Resour. Manag.* **2021**, *35*, 1837–1869. [[CrossRef](#)]
138. Xu, T.; Valocchi, A.J.; Choi, J.; Amir, E. Use of machine learning methods to reduce predictive error of groundwater models. *Groundwater* **2014**, *52*, 448–460. [[CrossRef](#)]
139. Roy, J.; Saha, S. Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach. *Artif. Intell. Geosci.* **2022**, *3*, 28–45. [[CrossRef](#)]
140. Mosavi, A.; Sajedi Hosseini, F.; Choubin, B.; Taromideh, F.; Ghodsi, M.; Nazari, B.; Dineva, A.A. Susceptibility mapping of groundwater salinity using machine learning models. *Environ. Sci. Pollut. Res.* **2021**, *28*, 10804–10817. [[CrossRef](#)] [[PubMed](#)]
141. Jiang, G.; Wang, W. Error estimation based on variance analysis of k-fold cross-validation. *Pattern Recognit.* **2017**, *69*, 94–106. [[CrossRef](#)]
142. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Sci. Rev.* **2020**, *207*, 103225. [[CrossRef](#)]
143. Sahour, H.; Gholami, V.; Vazifedan, M. A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. *J. Hydrol.* **2020**, *591*, 125321. [[CrossRef](#)]
144. Ojeda Olivares, E.A.; Sandoval Torres, S.; Belmonte Jiménez, S.I.; Campos Enriquez, J.O.; Zignol, F.; Reygadas, Y.; Tiefenbacher, J.P. Climate change, land use/land cover change, and population growth as drivers of groundwater depletion in the Central Valleys, Oaxaca, Mexico. *Remote Sens.* **2019**, *11*, 1290. [[CrossRef](#)]
145. Bierkens, M.F.; Wada, Y. Non-renewable groundwater use and groundwater depletion: A review. *Environ. Res. Lett.* **2019**, *14*, 063002. [[CrossRef](#)]
146. Dangar, S.; Asoka, A.; Mishra, V. Causes and implications of groundwater depletion in India: A review. *J. Hydrol.* **2021**, *596*, 126103. [[CrossRef](#)]
147. Vaux, H. Groundwater under stress: The importance of management. *Environ. Earth Sci.* **2011**, *62*, 19–23. [[CrossRef](#)]
148. Ai, H.; Zhou, Z. Inhibit or promote: The inverse-U-shape effect of greenspace on economic growth. *Environ. Impact Assess. Rev.* **2023**, *100*, 107094. [[CrossRef](#)]
149. Yang, Y.; Zheng, R.; Zhao, L. Population aging, health investment and economic growth: Based on a cross-country panel data analysis. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1801. [[CrossRef](#)] [[PubMed](#)]
150. Cheng, L.; Mi, Z.; Sudmant, A.; Coffman, D.M. Bigger cities better climate? Results from an analysis of urban areas in China. *Energy Econ.* **2022**, *107*, 105872. [[CrossRef](#)]
151. Li, X.; Chen, D.; Duan, Y.; Ji, H.; Zhang, L.; Chai, Q.; Hu, X. Understanding Land use/Land cover dynamics and impacts of human activities in the Mekong Delta over the last 40 years. *Glob. Ecol. Conserv.* **2020**, *22*, e00991. [[CrossRef](#)]
152. Kumar, R.; Kumar, A.; Saikia, P. Deforestation and forests degradation impacts on the environment. In *Environmental Degradation: Challenges and Strategies for Mitigation*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 19–46.
153. Babiker, I.S.; Mohamed, M.A.; Hiyama, T.; Kato, K. A GIS-based DRASTIC model for assessing aquifer vulnerability in Kakamigahara Heights, Gifu Prefecture, central Japan. *Sci. Total Environ.* **2005**, *345*, 127–140. [[CrossRef](#)]
154. Kang, J.; Zhao, L.; Li, R.; Mo, H.; Li, Y. Groundwater vulnerability assessment based on modified DRASTIC model: A case study in Changli County, China. *Geocarto Int.* **2017**, *32*, 749–758. [[CrossRef](#)]

155. Khosravi, K.; Sartaj, M.; Tsai, F.T.-C.; Singh, V.P.; Kazakis, N.; Melesse, A.M.; Prakash, I.; Bui, D.T.; Pham, B.T. A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. *Sci. Total Environ.* **2018**, *642*, 1032–1049. [[CrossRef](#)]
156. Jia, Y.; Hu, X.; Kang, W.; Dong, X. Unveiling Microbial Nitrogen Metabolism in Rivers using a Machine Learning Approach. *Environ. Sci. Technol.* **2024**, *58*, 6605–6615. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.