

Groundwater quality assessment and refining vulnerability index through machine learning and pollution index integration: A case study of the Koohpayeh Plain in central Iran

Navideh Najafpour^{1,*} , Shahrokh Soltaninia² 

¹Department of Civil Engineering, Lecturer of Water Science Engineering, Khomeynishahr Branch, Islamic Azad University, Esfahan, Iran.

²Department of Civil Engineering, Faculty of Civil Engineering, Khomeynishahr Branch, Islamic Azad University, Esfahan, Iran.

*Corresponding author: n.najafpour777@gmail.com

Original Research

Received:
13 May 2025
Revised:
24 May 2025
Accepted:
14 June 2025
Published online:
1 July 2025

© 2025 The Author(s). Published by the OICC Press under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract:

Groundwater contamination is a critical issue, especially in arid regions where it serves as the main water supply. This study emphasizes the importance of integrating traditional methods with machine learning techniques to enhance groundwater vulnerability assessments in the Koohpayeh aquifer, central Iran. Groundwater quality was assessed using the Pollution Index of Groundwater (PIG) and Water Quality Index (WQI). WQI values ranged from less than 15 to 55, indicating poor water quality, particularly in central and southeastern regions, due to contaminants from agriculture and urban activities. The PIG index revealed severe pollution in 45% of the area, showing a strong correlation between high PIG and low WQI values. Machine learning methods, including Random Forest (RF), Gradient Boosting (GB), and Support Vector Regression (SVR), were used to refine the DRASTIC model. RF emerged as the most effective model, achieving an MSE of 0.05 and an R^2 of 0.89, outperforming GB, which had a slightly higher R^2 of 0.91 but a greater MSE of 0.08. A strong correlation between nitrate levels and the optimized vulnerability indices validated the improved DRASTIC model's accuracy. The enhanced DRASTIC model provides a more reliable tool for identifying high-risk areas and developing targeted groundwater management strategies. This study underscores the value of combining traditional assessments with machine learning to improve groundwater quality monitoring and protect public health. These findings can support decision-makers in prioritizing pollution control strategies, optimizing groundwater monitoring programs, and designing sustainable management plans in vulnerable arid and semi-arid regions.

Keywords: Water resource; Data-driven approach; Modified DRASTIC; Spatial analysis; Water quality index

1. Introduction

Groundwater is a critical resource worldwide, particularly in arid and semi-arid regions where surface water is scarce. Regional studies have confirmed rapid groundwater depletion in arid zones, such as in Herat Province, Afghanistan, where declining water levels of over 3 meters have been recorded in less than five years (Mostafazadeh et al., 2024). It serves as the primary source of fresh water for drinking, agriculture, and industrial purposes. However, groundwater is susceptible to contamination from various sources, in-

cluding agricultural runoff, industrial discharges, and urban development, which can degrade its quality and pose significant risks to human health and the environment (Davoodi Memar Otagvar et al., 2024). Ensuring groundwater quality is paramount, as contamination can have severe health implications and negatively impact agricultural productivity and ecosystem health (Farhadi et al., 2020). In recent years, the importance of monitoring and maintaining groundwater quality has been underscored by various studies highlighting the prevalence and impact of groundwater pollution (Ram et al., 2021; Fadaei, 2023).

Groundwater vulnerability assessment is a vital component of sustainable water resource management. The DRASTIC index, developed by Aller et al. (1987), is one of the most widely used methods for evaluating groundwater pollution potential. This index incorporates seven hydrogeological factors: Depth to water, Recharge, Aquifer media, Soil media, Topography, Impact of the vadose zone, and Hydraulic conductivity. The lack of knowledge or enough understanding of the degree of vulnerability of groundwater may lead to severe contamination of these resources (Baba and Tayfur, 2011; Khemiri et al., 2013). While the DRASTIC index provides a comprehensive assessment framework, it has limitations, particularly its ability to account for specific contaminants, land use, temporal water quality, and their unique behaviors in the subsurface environment (Al-Adamat et al., 2003). In recent years, researchers have employed a variety of techniques to improve groundwater quality modeling and vulnerability mapping. Traditional index-based models such as DRASTIC, GOD, and SINTACS continue to be utilized due to their interpretability and data availability (Bagherpour et al., 2014; Khemmoudi et al., 2014; Allouche et al., 2017; Baalousha, 2010; Najafpour et al., 2020). To address these limitations, several studies have proposed hybrid approaches by integrating water quality indices like the Water Quality Index (WQI) and the Pollution Index of Groundwater (PIG) into spatial frameworks (Panagopoulos et al., 2005; Jovanovic et al., 2006; Najafpour et al., 2020). These adjustments aimed to provide a more accurate evaluation of groundwater vulnerability, reflecting the specific conditions of the study area. For instance, many studies have enhanced the water quality prediction and DRASTIC index by integrating it with advanced statistical and machine-learning techniques to improve its accuracy and predictive capability (Tayfur et al., 2014; Chitsazan et al., 2015; Asghari Moghadam et al., 2015; Elzain et al., 2023). In this regard, Asghari Moghadam et al. (2015) utilized machine learning models such as Random Forest and Gradient Boosting to optimize the DRASTIC index for groundwater vulnerability assessment. Their results indicated significant improvements in predictive performance compared to the traditional DRASTIC index. Additionally, Hosseini et al. (2018) integrated remote sensing data with machine learning models to evaluate groundwater quality and vulnerability in a semi-arid region of Iran, emphasizing the importance of incorporating diverse data sources for comprehensive assessments. One study incorporated Remote Sensing (RS) data with GIS and machine learning algorithms to assess groundwater vulnerability. They found that integrating remote sensing data significantly enhanced the spatial resolution and accuracy of groundwater vulnerability maps (Lee et al., 2020). Hassanaliipour et al. (2022) investigated the impacts of 30 years of urban expansion in the Ardabil Plain, Iran, on the quantity and quality of surface and groundwater resources. Using satellite imagery and geostatistical analyses, their study revealed a 46% decrease in groundwater reserves and a declining trend in river flows, alongside increasing contamination in urban runoff. Despite groundwater quality remaining within acceptable limits, the authors emphasized

the need for integrated urban runoff management strategies to enhance aquifer recharge and protect water quality. Their findings highlight how unregulated urban growth, coupled with climatic stressors, can significantly degrade hydrological systems in semi-arid basins. Wang et al. (2023) utilized deep learning algorithms to enhance the predictive capabilities of the DRASTIC index for groundwater contamination risk assessment in China, achieving higher accuracy and robustness. Similarly, Motlagh et al. (2023) employed a hybrid model combining DRASTIC and machine learning techniques to assess groundwater vulnerability in Najafabad Plain, Iran, showcasing significant improvements in prediction accuracy (2023).

In terms of water quality assessment, a variety of methods have been employed, with some of the most commonly used approaches including the Water Quality Index (WQI) and the Groundwater Pollution Index (PIG), along with spatial and temporal distribution analyses of groundwater quality parameters. Additionally, various statistical methods, artificial intelligence techniques, and machine learning algorithms have been utilized to improve the prediction of groundwater quality (Nas, 2009; Stigter et al., 2006; Taghizadeh Mehrjardi et al., 2008; Li et al., 2010; Hosseinifard and Mirzaei Aminiyan, 2015; Krishan et al., 2016; Jagannathan et al., 2015; Alam et al., 2024; Najafpour et al., 2024). The incorporation of machine learning algorithms into groundwater quality has become increasingly common. Kulisz et al. (2021) used Artificial Neural Networks (ANNs) to predict groundwater quality parameters based on historical data. Their research highlighted the potential of ANNs in capturing the nonlinear relationships between various hydrogeological factors and groundwater quality. Furthermore, Nasiri Khiavi et al. (2023) compared several machine learning algorithms with the Best–Worst Method (BWM) for groundwater quality modeling in Ardabil, Iran, and found that Random Forest outperformed other models in terms of predictive accuracy. Their results support the integration of MLAs and MCDM tools for improving groundwater management decisions.

The aim of the present study was to optimize the groundwater vulnerability index by combining water quality parameters and machine learning methods, with a focus on a plain in central Iran. By integrating traditional indices like DRASTIC with advanced analytical techniques, we seek to provide a more accurate and comprehensive assessment of groundwater vulnerability. This approach will help in identifying critical areas that require immediate attention and management to ensure the sustainability of groundwater resources. Several previous studies have demonstrated the effectiveness of combining traditional vulnerability indices with modern analytical techniques. Amanah et al. (2019) conducted a study in Semarang, Indonesia, using cluster analysis and principal component analysis to evaluate groundwater quality, revealing significant pollution from industrial and agricultural activities. Similarly, Shihab and Hassan (2006) classified groundwater quality in Mosul, Iraq, using cluster analysis, identifying key factors affecting water quality and highlighting spatial variations. Numerous studies have employed diverse methodological approaches

in assessing groundwater vulnerability. Traditional models such as DRASTIC, SINTACTS, and GOD have been widely used due to their simplicity and regional adaptability. Meanwhile, statistical techniques like PCA, hierarchical clustering, and regression analysis have supported index calibration and parameter selection. In recent years, machine learning algorithms, including Random Forest, Support Vector Machines, and Neural Networks, have emerged as robust tools for predictive modeling. Their integration with spatial tools like GIS and pollution indices (e.g., WQI, PIG) has enabled more accurate mapping of vulnerability zones. However, comprehensive frameworks that combine these diverse approaches remain limited in the literature. Despite the increasing use of machine learning algorithms and pollution indices in groundwater studies, the integration of these methods to refine groundwater vulnerability assessments remains limited. This study aims to bridge this gap by combining the Pollution Index of Groundwater (PIG) with ensemble machine learning models, Random Forest (RF) and Gradient Boosting (GB) within a GIS-based spatial analysis framework. The research not only enhances the predictive capabilities of traditional assessment methods but also contributes to the development of more effective and location-specific management strategies. By applying statistical analyses, machine learning modeling (using Python), and spatial mapping, this hybrid methodological approach offers a comprehensive and novel perspective for assessing groundwater quality and vulnerability in the understudied Koohpayeh Plain, with broader implications for arid and semi-arid regions.

2. Materials and methods

2.1 Study area and data collection

Koohpayeh Plain is part of the Zayandeh Rud basin in central Iran (figure 1). According to the country's division, this area has 6 cities, 7 rural districts, and 25 villages. The north of the study area is along the central Zagros Mountains, and the center and south of the study area are located in the plain. This region has 7 different rain-gauges, climatology, and synoptic stations. According to the statistics of these stations, the average temperature has been recorded as 11.2 degrees Celsius at higher elevations and 14.2 degrees Celsius in the plain. The annual average rain at the mountainous station has been reported to be 350 millimeters, and about 200 millimeters at the station in the plain (Water balance report of Koohpayeh Plain, Ministry of Energy, (Isfahan, 2021)). The studied area is geologically associated with the Sanandaj-Sirjan structural zone (Stocklin, 1968). The region's geomorphological features are shaped by tectonic activity, erosion, and topographic gradients, with general slopes directed toward the southeast and Zayandeh Rud River (Water balance report of Koohpayeh Plain, Ministry of Energy, 2021). Its aquifer is free and alluvial, and according to exploratory logs, its materials are sand and gravel, clay, and rubble, together with conglomerate layers (Stocklin, 1968). Due to the significant return of industrial and agricultural sewage, groundwater resources have been under scrutiny regarding their quality and pollution (Water balance report of Koohpayeh Plain, Ministry of Energy, (Isfahan, 2021)).

Groundwater samples were collected from 48 wells across

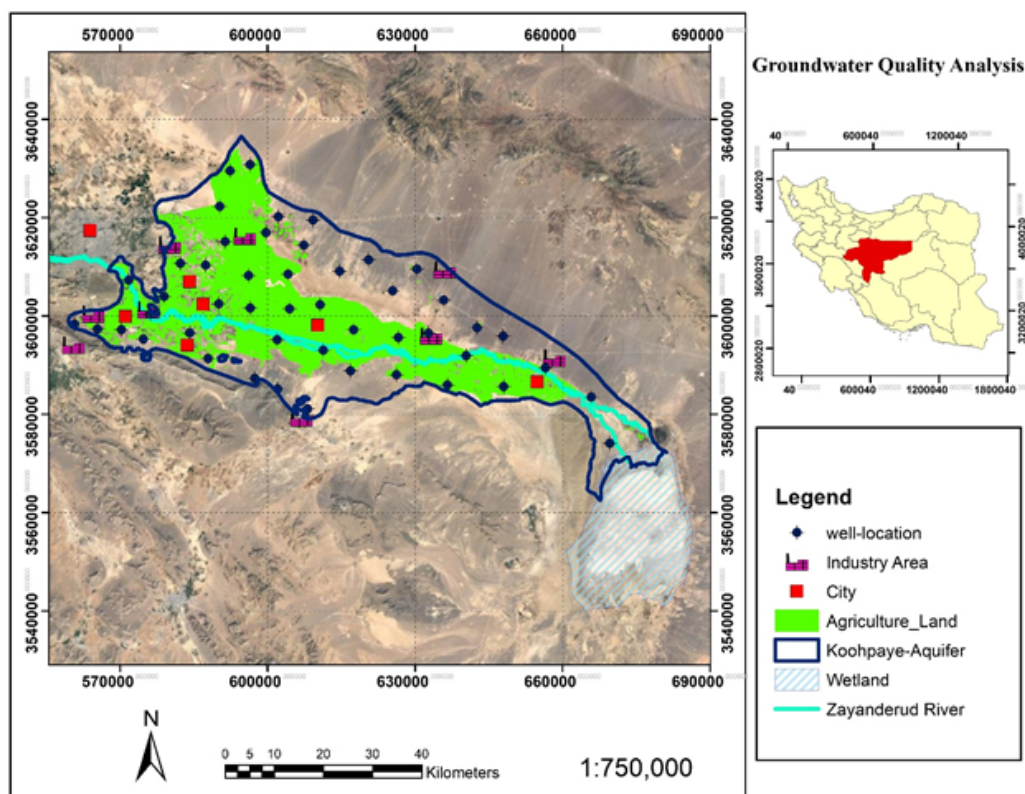


Figure 1. Map of sampling locations in the study area.

the study area during four sampling campaigns designed to capture seasonal variations (figure 1). These field investigations were conducted by the author during both dry and wet seasons from 2021 to 2023. The author was directly responsible for the collection and laboratory analysis of key groundwater quality parameters, ensuring consistency, accuracy, and reliability of the dataset. Wells were selected based on their geographic distribution and potential exposure to different land uses (urban, industrial, agricultural). Samples were collected using pre-cleaned polyethylene bottles. Before sampling, each well was pumped for at least 20 minutes to ensure that the collected water represented the aquifer's current conditions. Samples were collected directly from the well spouts and immediately sealed to prevent contamination. For metal analysis, samples were acidified with nitric acid (95%) to a pH of less than 2. All samples were stored in ice-cooled containers and transported to the laboratory within 24 hours.

The collected groundwater samples were analyzed for a range of physicochemical parameters following standard procedures. pH and electrical conductivity (EC) were measured on-site using calibrated portable meters in accordance with APHA (2012) guidelines. Total dissolved solids (TDS) were estimated from EC values using a standard conversion factor. Major cations and anions, including calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^+), potassium (K^+), chloride (Cl^-), bicarbonate (HCO_3^-), carbonate (CO_3^{2-}), and sulfate (SO_4^{2-}), were quantified using standard titration techniques (APHA, 2012). Nutrients such as nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), and phosphate (PO_4^{3-}) were determined via spectrophotometric methods. Heavy metals were measured using atomic absorption spectrophotometry, following EPA (2007) protocols, to ensure accuracy and compliance with environmental standards. To ensure the reliability and accuracy of the analytical results, strict Quality Assurance (QA) and Quality Control (QC)

protocols were implemented. All laboratory analyses were conducted in an ISO/IEC 17025-certified laboratory. Each test was performed in triplicate to assess repeatability and minimize random error. Instruments were calibrated using certified standard solutions before and during the analysis process. Quality control measures included the use of analytical blanks, duplicate samples, and spiked samples. The relative standard deviation (RSD) for repeated measurements was maintained below 5%, and recovery rates for spiked samples were within the acceptable range of 95 – 105%, in compliance with APHA (2012) and EPA (2007) standards. Statistical analyses were conducted to understand the distribution and variability of the groundwater quality parameters. Descriptive statistics, including mean, standard deviation, minimum, and maximum values, were calculated (Table 1).

The purpose of the studies was to investigate two main parts. Firstly, some applicable indices were determined for the assessment of groundwater quality. Then, the DRASTIC model was used to evaluate groundwater vulnerability, and Machine learning methods improved this model. The trends of these two sections are described below.

2.2 Water quality indexes

Several pollution indices are calculated to assess groundwater quality for different uses. The Pollution Index of Groundwater (PIG) and the Water Quality Index (WQI) were calculated for various purposes. In the following, each of these indices is explained.

2.2.1 Pollution Index of Groundwater (PIG)

The PIG (Pollution Index of Groundwater) index is a quantitative tool used to assess the quality of groundwater by evaluating the concentration of various contaminants concerning their permissible limits. The index provides a comprehensive value that reflects the overall pollution status of

Table 1. Descriptive statistics of the groundwater quality parameters.

Parameter	Unit	Count	Mean	STDEV	Min	Max
PH	-	48	7.72	0.42	7.1	8.4
EC	MicMobs/Cm	48	7334.3	40.12	1195	21100
TDS	Mgr/Lit	48	5201.25	4097.36	837	14770
SAR	-	48	11.53	6.91	2.34	25.1
Cl	Mgr/Lit	48	2231.4	2121.44	170	7748
HCO ₃	Mgr/Lit	48	1.5	4.60	0	18
SO ₄	Mgr/Lit	48	311.78	165.69	67.1	701
Ca	Mgr/Lit	48	539.6	455.06	100	1877
Mg	Mgr/Lit	48	281.47	263.57	56.1	1242
Na	Mgr/Lit	48	204.78	181.78	21.88	656
K	Mgr/Lit	48	1108.34	946.94	117	3169
TH	Mgr/LitCaCo ₃	48	9.42	8.06	0.39	31.2
NO ₃	Mgr/Lit	48	1548.35	1324.09	300	5807
PO ₄	Mgr/Lit	48	42.91	10.91208	22.9	63
Total Heavy metals	Mgr/Lit	48	0.39	0.573743	0.07	2.8

a groundwater source. It is particularly useful in identifying areas that require immediate attention for groundwater management and pollution control. The PIG index is calculated as in equation (1).

$$\text{PIG} = \sum_{i=1}^n \left(\frac{C_i}{S_i} \right) \times W_i \quad (1)$$

C_i is the concentration of the i -th parameter in the groundwater sample, S_i is the Permissible limit of the i -th parameter according to regulatory standards, and W_i is the weight assigned to the i -th parameter based on its relative importance to overall groundwater quality (Subba-Rao, 2012). The weight (W_i) assigned to each parameter in the PIG index was determined based on its relative impact on groundwater quality and public health, following the methodology proposed by Subba-Rao (2012). Parameters with greater health implications, such as nitrate and fluoride, were assigned higher weights, while others with lesser impact received lower weights. This approach ensures that the index reflects the true pollution potential of each contaminant.

2.2.2 Water Quality Index (WQI)

The WQI aggregates multiple water quality parameters to provide an overall water quality assessment. The WQI is calculated using equation (2).

$$\text{WQI} = \frac{\sum(q_i \times w_i)}{\sum w_i} \quad (2)$$

where w_i is the weight assigned to each parameter, and q_i is the quality rating of each parameter.

Geographic Information Systems (GIS) were used to create spatial distribution maps for these groundwater quality indexes. These maps provided a visual representation of the spatial variability and helped identify critical areas requiring immediate attention (World Health Organization, 1998). The weights (w_i) used in WQI calculation were adopted from established guidelines such as those of the

World Health Organization (1998) and previous studies (Najafpour et al., 2020; Ram et al., 2021), which consider the relative importance of each parameter to overall water quality. Parameters directly affecting human health, such as nitrate, TDS, and pH, were given higher weights to emphasize their significance in the index formulation.

2.3 DRASTIC and modified DRASTIC index

Groundwater vulnerability assessment aims to pinpoint areas more prone to pollution due to human activities than other regions. Identifying these vulnerable zones allows for implementing targeted measures and land-use restrictions to mitigate groundwater contamination. Groundwater vulnerability denotes the sensitivity of groundwater systems to contamination based on the likelihood of pollutants infiltrating and dispersing within the subsurface environment. This concept offers a relative assessment of the probability of contamination at a specific location rather than providing an absolute measure. The most prevalent method for assessing groundwater vulnerability involves weighting models that assign scores to various parameters based on their influence on groundwater contamination. The DRASTIC model is one of the most commonly applied approaches for evaluating the intrinsic vulnerability of aquifers. The DRASTIC index is a standardized system for evaluating groundwater pollution potential. This index, which indicates the natural potential of the aquifer to become polluted, incorporates seven hydrogeological parameters. As there have been many sources and research works that have used the DRASTIC method to analyze the vulnerability of aquifers, in this study, that is not our main purpose; thus, we will not explain it in detail, and only those methods used in preparing the different layers have been briefly stated. The DRASTIC model is favored for its cost-effectiveness and relatively low data requirements (Aller et al., 1987; Al-Adamat et al., 2003; Babiker et al., 2007). Seven parameters that are used in the evaluation of vulnerability via the DRASTIC method are shown in Table 2 and describe how prepared.

Table 2. Explanation of DRASTIC index parameters.

parameter	description	Data source	Method of measuring
D - Depth to Water	Distance from the ground surface to the water table.	Observation wells (2022 – 2024)	Measured directly using water level indicators in wells.
R - Recharge	Amount of water that infiltrates the ground and reaches the aquifer.	Meteorological data, hydrological studies	Calculated based on rainfall data and infiltration rates.
A - Aquifer Media	The geological formation of the aquifer affects its permeability and porosity.	Geological maps, borehole logs	Determined from geological surveys and borehole logs.
S - Soil Media	The type of soil covering the aquifer impacts water infiltration.	Soil surveys, soil maps	Identified from soil survey reports and soil classification maps.
T - Topography	Slope and shape of the land surface affect runoff and infiltration.	Topographic maps, Digital Elevation Models (DEMs)	Derived from topographic maps and DEMs using GIS software.
I - Impact of Vadose Zone	Characteristics of the unsaturated zone above the water table.	Geological and soil surveys, borehole logs	Evaluated based on soil and geological data from surveys and borehole logs.
C - Hydraulic Conductivity	The ability of the aquifer material to transmit water.	Aquifer tests, hydrological studies	Measured using aquifer pumping tests and data from hydrological studies.

The main goal of vulnerability studies is to create maps that identify areas with a high potential for groundwater contamination. Developing such maps involves choosing a suitable method that considers the availability of hydrogeological data, pollutant distribution, and the environmental characteristics of the region. Geographic Information Systems (GIS) are commonly employed to combine multiple data sources and visualize groundwater vulnerability on a regional level. The spatial analysis features of GIS allow for mapping the seven components of the DRASTIC model, which helps illustrate the spatial connections between land management practices and groundwater vulnerability (Soper, 2006). After preparing each layer of the DRASTIC Index in ArcGIS, these seven layers were weighted according to Table 3 and combined using equation (3).

$$I_{\text{DRASTIC}} = D_r D_w + R_r R_w + A_r A_w + S_r S_w + T_r T_w + I_r I_w + C_r C_w \quad (3)$$

where (r) and (w) represent the ratings and weights assigned to each parameter, respectively.

Table 3. Initial weights of DRASTIC parameters (Aller et al., 1987).

Parameter	D	R	A	S	T	I	C
Weight	5	4	3	2	5	4	3

In this study, advanced methods such as Random Forest (RF), Gradient Boosting (GB), Support Vector Regression (SVR), and Stacking Regressor were employed to enhance the traditional DRASTIC approach by incorporating the Pollution Index of Groundwater (PIG). These models were selected for their ability to handle complex, non-linear relationships between input parameters and groundwater vulnerability, providing a data-driven approach to refine the DRASTIC model's predictions. Below is a detailed explanation of each method and the rationale for its use.

2.3.1 Random Forest (RF)

RF is an ensemble learning method that constructs multiple decision trees during training and merges their results to improve predictive accuracy and control overfitting (Breiman, 2001). In this study, RF was used to refine the DRASTIC index by capturing complex interactions between hydrogeological parameters and observed groundwater contamination levels, such as the PIG index. RF is highly effective in handling large datasets with multiple variables, making it ideal for assessing groundwater vulnerability, where parameter interdependencies significantly impact pollution risk (Barzegar et al., 2021).

2.3.2 Gradient Boosting (GB)

GB is a powerful machine learning technique that builds an ensemble of decision trees in a sequential manner, where each subsequent tree attempts to correct the errors of the previous one (Friedman, 2001). This iterative approach enhances model precision, especially in complex scenarios involving non-linear relationships between variables. GB was utilized in this study to fine-tune the DRASTIC model, effectively integrating the PIG index to adjust vulnerability

scores. The method's ability to minimize errors iteratively makes it particularly valuable for improving the accuracy of groundwater contamination predictions (Subbarayan et al., 2024).

2.3.3 Support Vector Regression (SVR)

SVR, a variant of Support Vector Machines (SVM), was employed for its robustness in regression tasks, particularly where non-linear relationships exist among data points (Subbarayan et al., 2024). SVR works by finding the optimal hyperplane that minimizes prediction error, making it highly suitable for modeling continuous variables such as groundwater vulnerability indices. By integrating SVR, the DRASTIC model was modified to better reflect the actual contamination levels indicated by the PIG index, enhancing its predictive capabilities (Smola and Schölkopf, 2004).

2.3.4 Stacking regressor

Stacking is an ensemble method that combines the predictions of multiple individual models (base learners) to produce a more accurate final prediction (Cortes and Vapnik, 1995). In this study, the Stacking Regressor was used to integrate the outputs of RF, GB, and SVR, leveraging their strengths and compensating for individual weaknesses. This approach provided a comprehensive enhancement of the DRASTIC index, allowing for more precise groundwater vulnerability assessments (Wolpert, 1992).

In this study machine learning models used in this study, including RF, GB, and SVR, were implemented using the Python programming language. The dataset was divided into training and testing subsets using a stratified random sampling strategy, with 80% of the data used for model training and 20% for testing. To ensure model robustness, a 10-fold cross-validation technique was applied during training. Hyperparameters for each model were optimized using grid search with five-fold internal cross-validation. For instance, the number of estimators and maximum tree depth were tuned for RF and GB, while kernel functions and regularization parameters were optimized for SVR.

The traditional DRASTIC model is based on theoretical assessments and expert judgment, which may not fully capture the real-world complexity of groundwater contamination (Asghari Moghadam et al., 2015). By integrating machine learning methods with the DRASTIC index, this study aims to modify and validate the vulnerability scores using observed data such as the Pollution Index of Groundwater (PIG). This approach allows for a more accurate and realistic assessment of groundwater vulnerability, aligning theoretical predictions with observed environmental conditions. Using machine learning not only refines the DRASTIC model but also enables dynamic updates as new data becomes available, making it a powerful tool for sustainable groundwater management. In addition, the PIG index was used to modify and validate the DRASTIC index because it provides a comprehensive assessment of groundwater contamination by integrating various water quality parameters into a single index. Unlike single-parameter approaches, the PIG index captures the cumulative effect of multiple pollutants, making it a more robust indicator of actual groundwater quality. The use of the PIG index allows for the correc-

tion of the DRASTIC model’s vulnerability scores, which are often based on theoretical susceptibility rather than observed contamination levels. By incorporating real-world pollution data from the PIG index, the modified DRASTIC model can better reflect the true state of groundwater vulnerability. In the training phase, DRASTIC parameters were used as inputs, and the corrected vulnerability indices based on PIG values served as the outputs. In the testing phase, the model’s performance was evaluated to ensure that the modified vulnerability indices aligned with actual groundwater quality conditions, providing a more accurate and data-driven vulnerability assessment. Each of these models is briefly explained below. The performance of the models was evaluated using Mean Squared Error (MSE) and R^2 . Then, the best-performing model was selected based on the lowest MSE and highest R^2 values. The optimized DRASTIC index was calculated using the selected model.

3. Results and discussion

The Pearson correlation analysis (Table 4) revealed strong interrelationships among key groundwater quality parameters. Notably, total dissolved solids (TDS) showed strong correlations with chloride (Cl^- ; $r = 0.995$) and calcium (Ca^{2+} ; $r = 0.997$), underscoring the significant influence of these ions on groundwater salinity. Additionally, total hardness (TH) was highly correlated with both calcium and magnesium ($r = 0.997$), reflecting their dominant role in determining water hardness. Chloride further demonstrated a strong association with both EC and TDS ($r = 0.995$), and a slightly lower yet substantial correlation with sodium (Na^+ ; $r = 0.978$), suggesting its anthropogenic and geogenic origins. Moderate correlations between the sodium adsorption ratio (SAR) and both EC ($r = 0.791$) and TDS ($r = 0.792$) indicate a partial dependency, likely modulated by the interplay between Na^+ , Ca^{2+} , and Mg^{2+} . Interestingly, pH was negatively correlated with EC and TDS ($r = -0.688$), suggesting that increased salinity is typically accompanied by more acidic conditions.

3.1 Pollution Index of Groundwater (PIG)

The spatial distribution of the PIG index in the Koohpayeh Plain is shown in figure 3. These values across the groundwater samples in the study area ranged from 0.95 to 11.97,

with a significant spatial variation observed. This map highlights that the central and southeastern parts of the region exhibit the highest pollution levels, with PIG values exceeding 11. These areas correspond to the zones with very high pollution, covering approximately 45% of the study area, as depicted in the pie chart (figure 2). This high pollution is likely driven by intensive agricultural activities, industrial discharges, and natural geochemical processes.

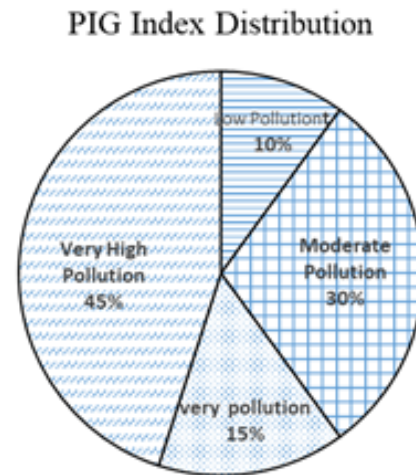


Figure 2. Distribution of PIG index categories in the study area.

The Pollution Index of Groundwater (PIG) values across the Koohpayeh aquifer ranged from 0.95 to 11.97, highlighting substantial spatial variability (figure 3). Approximately 45% of the study area fell within the very high pollution category (figure 2), while only 10% was categorized as low-pollution zones, typically located in peripheral areas with minimal anthropogenic impact.

Based on figure 3 the highest PIG values were predominantly located in the central and southeastern zones, areas subjected to intense agricultural activity, industrial discharges, and wastewater infiltration. Moderate pollution zones, representing 30% of the area, are characterized by slightly elevated levels of electrical conductivity (EC) and total dissolved solids (TDS), while the very high pollution zones are predominantly associated with high concentrations of magnesium (Mg^{2+}) and calcium (Ca^{2+}). The

Table 4. Pearson correlation matrix of the water quality parameters.

	pH	EC	TDS	SAR	Cl	HCO3	SO4	Ca	Mg	Na	K	Total Hardness	NO3	NO2	NH4	PO4
pH	1															
EC	-0.688	1														
TDS	-0.688	0.99**	1													
SAR	-0.374	0.79**	0.79**	1												
Cl	-0.713	0.99**	0.99**	0.74**	1											
HCO3	0.412	-0.811	-0.810	-0.653	-0.805	1										
SO4	0.067	0.328	0.328	0.684*	0.231	-0.300	1									
Ca	-0.707	0.99**	0.99**	0.742**	0.998**	-0.798	0.269	1								
Mg	-0.711	0.99**	0.99**	0.752**	0.997**	-0.811	0.283	0.998**	1							
Na	-0.643	0.99**	0.995**	0.853**	0.978**	-0.802	0.417	0.981**	0.983**	1						
K	-0.354	0.356	0.357	0.398	0.355	-0.324	0.080	0.328	0.347	0.354	1					
Total Hardness	-0.709	0.99**	0.997**	0.746**	0.998**	-0.803	0.274	0.999**	0.999**	0.982**	0.336	1				
NO3	-0.531	0.374	0.373	0.052	0.410	-0.476	-0.213	0.398	0.402	0.323	0.146	0.400	1			
NO2	-0.065	-0.030	-0.029	0.068	-0.049	0.060	0.097	-0.055	-0.040	-0.020	0.762**	-0.050	-0.160	1		
NH4	-0.225	0.132	0.132	0.102	0.125	-0.366	0.084	0.109	0.147	0.110	0.605*	0.123	0.297	0.487	1	
PO4	-0.018	-0.054	-0.053	0.006	-0.065	0.101	0.021	-0.072	-0.059	-0.051	0.713**	-0.067	-0.243	0.958**	0.474	1

*Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level

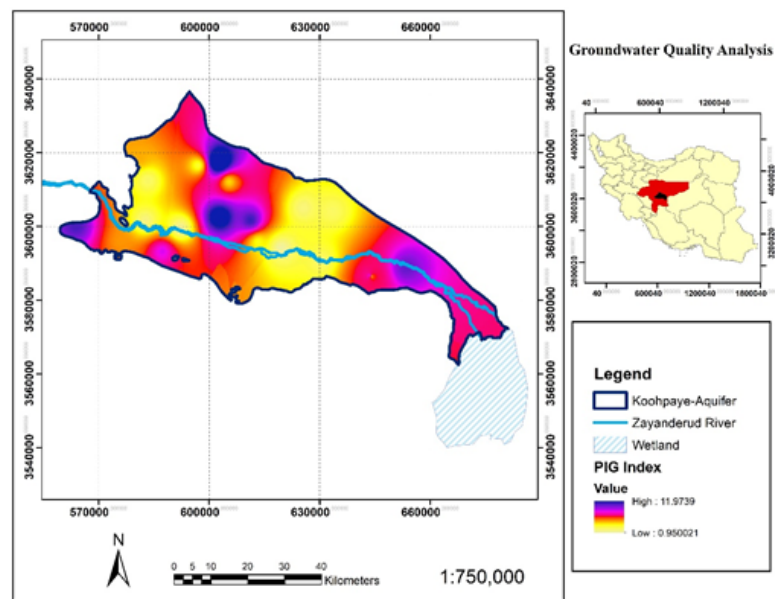


Figure 3. Spatial distribution of PIG index in the Koohpayeh aquifer.

spatial gradient of pollution from low to very high closely follows land use transitions and topographical slopes toward the Zayandeh Rud River. The observed clustering of high PIG values near the river corridor emphasizes the influence of surface-groundwater interactions, leaching from agricultural soils, and unregulated effluent discharge.

These spatial trends are likely influenced by a combination of anthropogenic and natural factors. The southeastern zones coincide with more permeable alluvial formations and shallower groundwater tables, which facilitate the infiltration of contaminants. In addition, this area contains dense agricultural operations, unregulated irrigation practices, and limited wastewater treatment infrastructure, all of which contribute to elevated pollutant loads. The proximity to the river also enhances pollutant transport through sur-

face–subsurface exchange mechanisms. Collectively, these geological and human-induced drivers explain the elevated PIG values in these regions.

3.2 Groundwater Quality Index (WQI)

The Water Quality Index (WQI) ranged from < 15 to 55 across the aquifer (figure 4), indicating considerable heterogeneity in water quality. The poorest water quality (WQI < 15) was observed in the central and southeastern sectors areas overlapping with high PIG zones. Areas around the Zayandeh Rud River have slightly better WQI values (15–30) but still reflect poor water quality. A small central portion shows moderate quality (30–45), and the best quality (45–55) is found in limited yellow zones along the river. Overall, the WQI distribution highlights widespread water quality issues, underscoring the need for targeted manage-

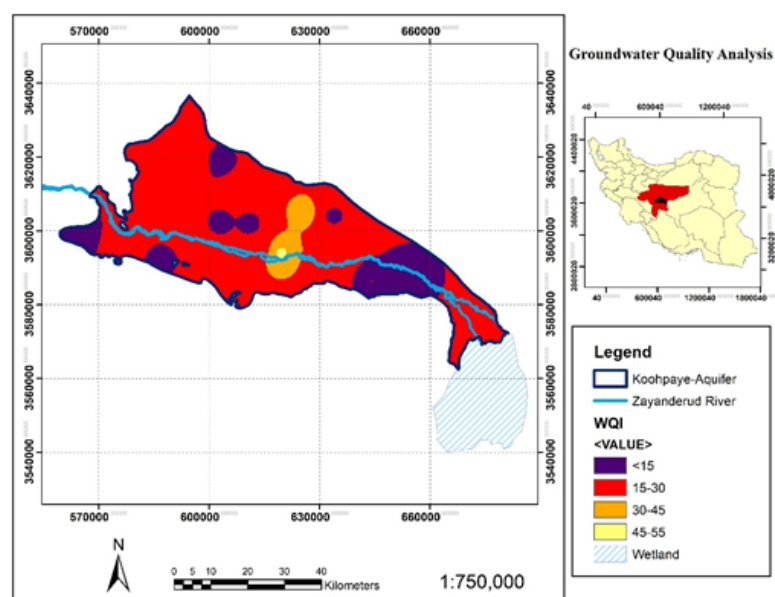


Figure 4. Spatial distribution of water quality index (WQI) in the Koohpayeh aquifer.

ment and pollution mitigation efforts.

The spatial overlap between low WQI and high PIG zones suggests that water quality degradation is primarily driven by cumulative pollution sources. The underlying geology, comprising unconsolidated sediments, combined with flat terrain and poor drainage capacity, amplifies pollutant accumulation. Human activities such as over-application of fertilizers, inefficient irrigation systems, and leakage from septic systems further exacerbate contamination risks. These findings highlight the need for integrated land use and water quality management tailored to the region's hydrogeological characteristics.

3.3 Groundwater vulnerability by modified DRASTIC index

The traditional DRASTIC model, based on the weighted overlay of seven hydrogeological parameters (figure 5), revealed vulnerability scores ranging from 77 to 159 across the Koohpayeh Plain. This figure shows the spatial distribution of groundwater vulnerability using the traditional DRASTIC model, with values ranging from 77 to 159. The highest vulnerability areas are concentrated in the central and southeastern parts of the study area, highlighting zones with increased susceptibility to contamination. These regions are influenced by factors such as high permeability, shallow groundwater depths, and intensive land use, which elevate the risk of pollutants infiltrating the aquifer. However, the conventional model's reliance on expert-assigned weights and generalized assumptions introduces subjectivity and limits predictive accuracy, as reported in previous studies (Aller et al., 1987; Ting and Witten, 1999). To address these limitations, the DRASTIC index was modified using a data-driven approach by integrating the PIG index and applying ensemble machine learning algorithms including Random Forest (RF), Gradient Boosting (GB), and Support Vector Regression (SVR). Figure 6 compares

the performance of various machine learning models, including Linear Regression, Ridge Regression, Random Forest, and Gradient Boosting, in modifying the DRASTIC index using the PIG index. Among the models compared (figure 6), Gradient Boosting (GB) achieved the highest R^2 value (0.91), indicating its strong capacity to capture data variance. However, it also exhibited a higher Mean Squared Error (MSE = 0.15), suggesting that while the model fit the training data well, it may have compromised generalization and introduced overfitting. In contrast, Random Forest (RF) produced slightly lower R^2 (0.86) but significantly outperformed GB in terms of predictive precision with a lower MSE (0.05). Given the trade-off between model complexity and generalizability, RF was selected as the optimal model. Its ensemble structure, which averages predictions across multiple decision trees, tends to reduce variance and improve robustness, especially in environmental datasets with heterogeneous and nonlinear characteristics. Moreover, RF exhibited more stable performance during cross-validation, with smaller fluctuations in R^2 and MSE across folds. This balance between acceptable model fit and lower generalization error highlights RF's capacity to reliably recalibrate the DRASTIC index using observed contamination data such as the PIG index. Thus, despite GB's slightly better fit, RF was favored for its superior predictive consistency, lower error margin, and better resistance to overfitting, making it a more practical and generalizable tool for vulnerability mapping in this study. These findings confirm the robustness of machine learning approaches in refining groundwater vulnerability assessments, aligning with findings from similar research by Ting and Witten (1999), Khan et al. (2022), and Ijlil et al. (2022), where machine learning demonstrated superior predictive capabilities.

The modified DRASTIC index map (figure 7) provides an enhanced spatial depiction of groundwater vulnerability, with index values ranging from 108 to 215. Notably, ar-

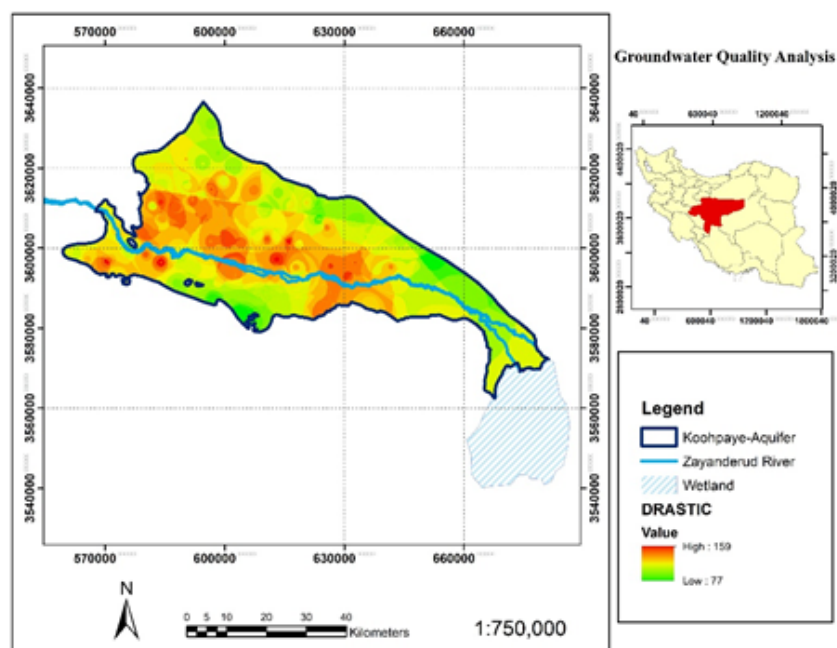


Figure 5. Original DRASTIC index map of groundwater vulnerability.

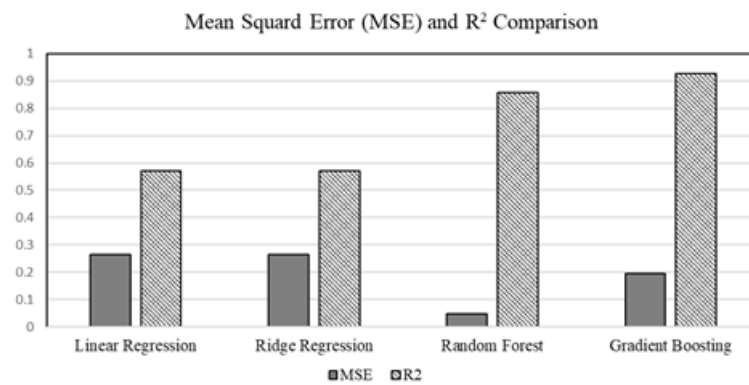


Figure 6. Comparison of model performance for modifying the DRASTIC index.

areas previously identified as moderately vulnerable in the original DRASTIC map exhibited significant increases in vulnerability in the modified version, particularly in the central and southeastern parts of the aquifer. For example, vulnerability scores in urban and agricultural hotspots increased from approximately 130 to 180, reflecting heightened contamination risks in these regions due to intensive human activities. This refined assessment provides critical insights for targeted management strategies and resource allocation, particularly in regions experiencing severe anthropogenic impacts. Moreover, the integration of machine learning allowed for the inclusion of advanced statistical analyses, ensuring the reliability of predictions across diverse hydrogeological settings. By demonstrating the superiority of data-driven approaches, this study highlights the transformative potential of machine learning in refining groundwater vulnerability assessments and underscores the necessity of adopting such methods for sustainable water resource management in arid and semi-arid regions. Nitrate data collected from various points across the aquifer

were used to evaluate and validate the performance of the improved DRASTIC model. The scatter plot (figure 8) illustrates the correlation between the optimized groundwater vulnerability index and nitrate concentrations measured at various locations within the aquifer. The high correlation coefficient ($R^2 = 0.9632$) demonstrates a strong linear relationship, indicating that the optimized vulnerability index effectively predicts areas with higher contamination potential, particularly concerning nitrate levels. This high level of agreement between modeled vulnerability and observed nitrate concentrations validates the accuracy of the modified DRASTIC index and highlights the effectiveness of machine learning methods in refining vulnerability assessments. The robust correlation suggests that the machine learning models successfully adjusted the DRASTIC parameters to align with real-world contamination patterns, specifically for nitrate pollution. While nitrate serves as a key indicator of contamination in this study, it is important to note that the validation process may be extended to include other contaminants for a more comprehensive

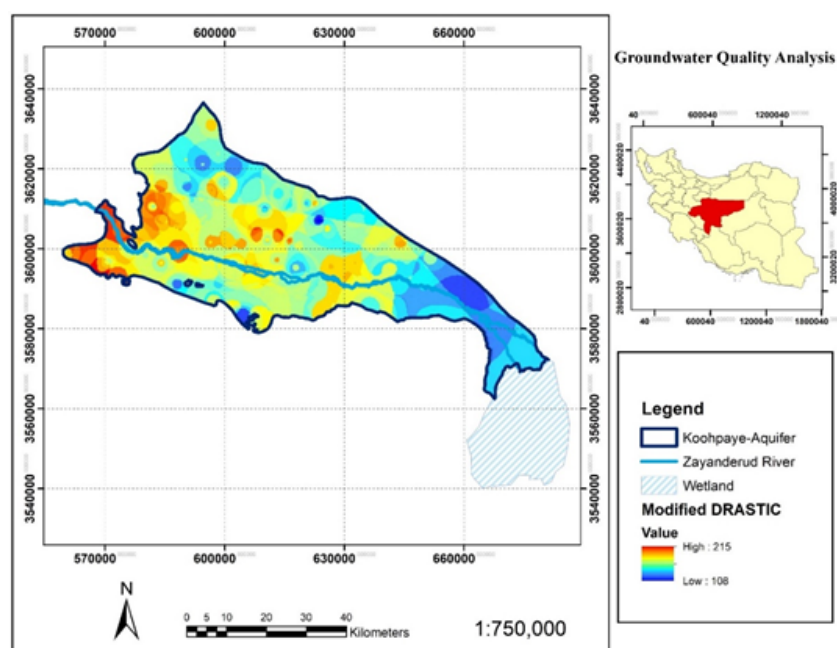


Figure 7. Modified DRASTIC index map incorporating the PIG index and machine learning methods.

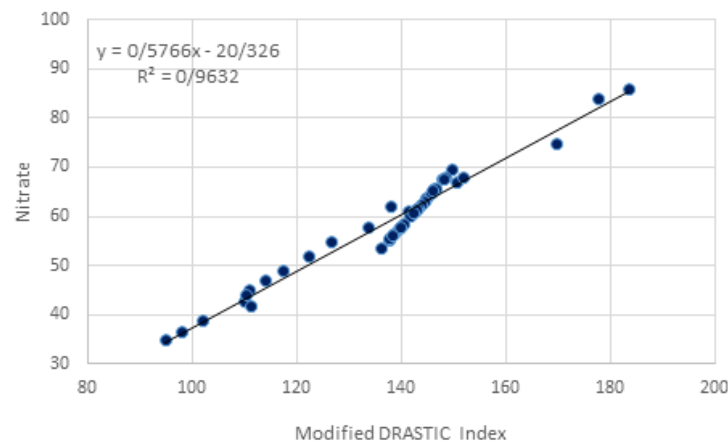


Figure 8. Correlation between optimized vulnerability and nitrate concentrations.

assessment. The validation data used for this analysis were carefully selected from various parts of the aquifer to ensure comprehensive spatial coverage and robust evaluation of the model's performance. This approach offers a transferable framework for other arid and semi-arid regions facing similar water quality challenges. However, this study focused primarily on nitrate as a contamination indicator. Future work should consider validating the models with additional pollutants such as heavy metals or pesticides to broaden applicability. Also, incorporating time-series data and climate variables could further improve model responsiveness to environmental changes.

4. Conclusion

This study aimed to enhance groundwater vulnerability assessment in the Koochpayeh aquifer in central Iran by integrating traditional methods with advanced machine learning techniques. This study advances the current understanding of groundwater quality assessment and vulnerability mapping by integrating traditional index-based methods with modern machine learning techniques. The key contribution lies in refining the DRASTIC model using pollution data (PIG index) and validating it against field-measured nitrate concentrations, thereby bridging the gap between theoretical susceptibility and observed contamination patterns. By demonstrating the superiority of ensemble learning models, particularly Random Forest, in enhancing prediction accuracy, the study highlights the potential of data-driven methods for developing dynamic, region-specific groundwater vulnerability frameworks. The coupling of pollution indices (PIG) with spatial analysis also provides a holistic approach for identifying high-risk zones, a crucial advancement for groundwater management in arid and semi-arid regions. This work offers practical implications for policymakers and environmental managers. It supports evidence-based zoning, targeted monitoring, and resource allocation, especially in vulnerable zones influenced by anthropogenic pressures. Moreover, the study sets a precedent for integrating multidisciplinary datasets, including hydrogeology, geochemistry, and land use, with machine learning to improve environmental

decision-making. One notable limitation of this study lies in the validation phase, which primarily relied on nitrate (NO_3^-) concentrations as the indicator of groundwater contamination. While nitrate is a well-established and widely used parameter for assessing pollution, particularly from agricultural sources, this focus may not fully capture the complexity of groundwater contamination arising from other pollutants such as heavy metals, industrial effluents, or emerging contaminants. As a result, the generalizability of the modified DRASTIC model might be constrained to scenarios where nitrate is the dominant contaminant. Future studies should consider incorporating multiple pollutants for model validation to enhance the robustness and applicability of the vulnerability assessment framework across diverse hydrochemical environments, and should expand this approach by incorporating temporal data, a broader range of contaminants, and climate projections to enhance model robustness and applicability. The methodology developed here can be adapted to other regions, offering a replicable template for sustainable groundwater resource protection worldwide. Overall, this study offers a replicable and data-driven approach for sustainable groundwater resource protection, providing a scientific basis for targeted zoning, policy-making, and proactive water quality management across vulnerable aquifers.

Authors contributions

Navideh Najafpour: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources software supervision, validation, visualization, writing, writing-review & editing. Shahrokh Soltaninia: Conceptualization, supervision, validation, verification, visualization, and writing.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Al-Adamat R. A. N., Foster I. D. L., Baban S. M. J. (2003) Groundwater vulnerability and risk mapping for the Basaltic aquifer of the Azraq Basin of Jordan using GIS, remote sensing, and DRASTIC. *Appl Geogr* 23 (4): 303–324. DOI: <https://doi.org/10.1016/j.apgeog.2003.08.007>.
- Alam A., Kumar A., Singh A. (2024) A GIS approach for groundwater quality evaluation with entropy method and fluoride exposure with health risk assessment. *Environ Geochem Health* 46 (47): 1–23. DOI: <https://doi.org/10.1007/s10653-023-01822-2>.
- Aller L., Bennett T., Lehr J. H., Petty R. J., Hackett G. (1987) A standardized system for evaluating groundwater pollution potential using hydrogeologic settings. *U.S. Environmental Protection Agency* (Washington, D.C.) EPA/600/2-87/035
- Allouche N., Maanan M., Gontara M., Rollo N., Jmal I. (2017) A global risk approach to assessing groundwater vulnerability. *J Environ Model Softw* 88:168. DOI: <https://doi.org/10.1016/j.envsoft.2016.11.023>.
- Amanah H., Purwanto S., Hadihardaja I. K. (2019) Application of cluster analysis and principal component analysis for assessment of groundwater quality, A study in Semarang, Central Java, Indonesia. *Earth Environ Sci Conf Ser* 248:012009. DOI: <https://doi.org/10.1088/1755-1315/248/1/012063>.
- APHA (2012) Standard methods for the examination of water and wastewater, 22nd edn. *American Public Health Association* (Washington, D.C.)
- Asghari Moghadam A., Fijani E., Nadiri A. (2015) Optimization of DRASTIC Model by Artificial Intelligence for Groundwater Vulnerability Assessment in Maragheh-Bonab Plain, Optimization of DRASTIC Model by Artificial Intelligence for Groundwater Vulnerability Assessment in Maragheh-Bonab Plain. *Scientific Quarterly Journal of Geoscience* 152:154–166. DOI: <https://doi.org/10.22071/gsj.2015.43279>.
- Baalousha H. (2010) Assessment of a groundwater quality monitoring network using vulnerability mapping and geostatistics, A case study from Heretaunga Plains. *Journal of Agricultural Water Management* 97:240–246. DOI: <https://doi.org/10.1016/j.agwat.2009.09.013>.
- Baba A., Tayfur G. (2011) Groundwater contamination and its effect on health in Turkey. *Environ Monit Assess* 83 (1): 77–86. DOI: <https://doi.org/10.1007/s10661-011-1907-z>.
- Babiker I. S., Mohamed M. A. A., Hiyama T. (2007) Assessing groundwater quality using GIS. *Water Resour Manag* 11:669. DOI: <https://doi.org/10.1007/s11269-006-9059-6>.
- Bagherpour M. A., Talebheydokhti N., Tabatabaee S. H. R., Fadaei Nobandegani A. (2014) Assessment of groundwater nitrate pollution and determination of groundwater protection zones using DRASTIC and Composite DRASTIC (CD) models: The case of Shiraz unconfined aquifer. *J Health Sci Surveill Syst* 2 (2): 65.
- Barzegar R., Razzagh S., Quilty J., Adamowski J., Kheyrollah Pour H., Booi M. J. (2021) Improving GALDIT-based groundwater vulnerability predictive mapping using coupled resampling algorithms and machine learning models. *J Hydrol* 598:126370. DOI: <https://doi.org/10.1016/j.jhydrol.2021.126370>.
- Breiman L. (2001) Random forests. *Mach Learn* 45 (1): 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>.
- Chitsazan N., Nadiri A. A., Tsai F. T. C. (2015) Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging. *J Hydrol* 528:52–62. DOI: <https://doi.org/10.1016/j.jhydrol.2015.06.026>.
- Cortes C., Vapnik V. (1995) Support-vector networks. *Mach Learn* 20 (3): 273–297. DOI: <https://doi.org/10.1007/BF00994018>.
- Davoodi Memar Otagvar L., Fataei E., Tajiabadi M., Naeimi B. (2024) Spatio-temporal assessment of changes in groundwater quality in Qazvin Plain. *Journal of Range and Watershed Management* 77 (3): 353–369. DOI: <https://doi.org/10.22059/jrwm.2024.370962.1744>.
- Elzain E. H., Chung S. Y., Venkatramanan S., Selvam S., Ahmed H. A., Seo K. K. (2023) Novel machine learning algorithms to predict the groundwater vulnerability index to nitrate pollution at two levels of modeling. *Chemosphere* 314:137671. DOI: <https://doi.org/10.1016/j.chemosphere.2022.137671>.
- EPA (2007) Method 7000B: Flame atomic absorption spectrophotometry. *U.S. Environmental Protection Agency* (Washington, D.C.)
- Fadaei A. (2023) Study of contaminants in Iranian drinking water and their health outcomes: A review. *Journal of Pollution* 9 (4): 1450–1474. DOI: <https://doi.org/10.22059/poll.2023.356999.1839>.
- Farhadi H., Fataei E., Kharrat Sadeghi M. (2020) The relationship between nitrate distribution in groundwater and agricultural landuse (case study: Ardabil plain, Iran). *Anthropogenic Pollution* 4 (1): 50–56. DOI: <https://doi.org/10.22034/ap.2020.1885788.1059>.
- Friedman J. H. (2001) Greedy function approximation: A gradient boosting machine. *Ann Stat* 29 (5): 1189–1232. DOI: <https://doi.org/10.1214/AOS/113203451>.
- Hassanalipour Y., Mostafazadeh R., Esmali-Ouri A., Ahmadi M., Imani R. (2022) Evaluation of the effects of urban development on the quantity and quality of surface and groundwater in Ardabil plain. *Journal of Environmental Science Studies* 7 (3): 5374–5385. DOI: <https://doi.org/10.22034/jess.2022.343220.1792>.
- Hosseini F., Malekian A., Choubin B., Rahmati O., Cipullo S., Coulon F., Pradhan B. (2018) A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Sci Total Environ* 644:954–962. DOI: <https://doi.org/10.1016/j.scitotenv.2018.07.017>.
- Hosseinfard S. H., Mirzaei Aminiyam M. (2015) Hydrochemical characterization of groundwater quality for drinking and agricultural purposes: A case study in Rafsanjan Plain, Iran. *J Water Qual Explor Health* 7 (4): 531. DOI: <https://doi.org/10.1007/s12403-015-0173-0>.
- Ijlil S., Essahlaoui A., Mohajane M., Essahlaoui N., Mili E. M., Rompaey A. V. (2022) Machine learning algorithms for modeling and mapping of groundwater pollution risk: A study to reach water security and sustainable development (SDG) goals in a Mediterranean aquifer system. *J Remote Sens* 14 (10): 2379. DOI: <https://doi.org/10.3390/rs14102379>.
- Isfahan Regional Water Authority. (2021) Water Balance Report of Koohpayeh Plain *Ministry of Energy* (Iran), In Persian
- Jagannathan U., Venkatesh A., Sudharsanam A., Muniraj K. (2015) Assessment of groundwater quality using GIS and statistical approaches. *Asian J Earth Sci* 8 (4): 97–113. DOI: <https://doi.org/10.3923/ajes.2015.97.113>.
- Jovanovic N. Z., Adams S., Thomas A., Fey M., Beekman H. E., Campbell R., Saayman I., Conrad J. (2006) Improved DRASTIC method for assessment of groundwater vulnerability to generic aqueous-phase contaminants. *WIT Trans Ecol Environ* 92:393. DOI: <https://doi.org/10.2495/wm060421>.
- Khan Q., Liaqat M. U., Mohamed M. M. (2022) A comparative assessment of modeling groundwater vulnerability using the DRASTIC method from GIS and a novel classification method using machine learning classifiers. *Geocarto International* 37:5832–5850. DOI: <https://doi.org/10.1080/10106049.2022.2049861>.
- Khemiri S., Khnissi A., Ben Alaya M., Saidi S., Zargouni F. (2013) Using GIS for the comparison of intrinsic parametric methods assessment of groundwater vulnerability to pollution in scenarios of semi-arid climate, the case of Foussana groundwater in the central of Tunisia. *Journal of Water Resources and Protection* 5:835–847. DOI: <https://doi.org/10.4236/jwarp.2013.57085>.
- Khemmoudi K., Bendadouche H., Merabet S. (2014) Assessment of the vulnerability of an aquifer by DRASTIC and SINTACS methods: Aquifer of Bazer-Geult Zerga area (northeast Algeria). *Journal of Environmental Research and Management* 5 (9): 0169.

- Krishan G., Singh S., Khanna A. (2016) Water quality index of groundwater in Haridwar district, Uttarakhand. *Journal of Water Energy International* 8 (2): 54–60.
- Kulisz M., Kujawska J., Przysucha B., Cel W. (2021) Forecasting water quality index in groundwater using artificial neural network. *Energies* 14 (18): 5875. DOI: <https://doi.org/10.3390/en14185875>.
- Lee S., Hyun Y., Lee S., Lee M. J. (2020) Groundwater potential mapping using remote sensing and GIS-based machine learning techniques. *Remote Sensing* 12 (8): 1200. DOI: <https://doi.org/10.3390/rs12081200>.
- Li P. Y., Qian H., Wu J. H. (2010) Groundwater quality assessment based on improved water quality index in Pengyang County, Ningxia, Northwest China. *Journal of Chemistry* 7 (S1): S209.
- Mostafazadeh R., Malaekah M., Hama P. (2024) Quantitative assessment of groundwater resources depletion in Herat area, Afghanistan. *Civil Engineering and Environmental Science* 1 (1): 38–47. DOI: <https://doi.org/10.22034/aces.2024.457658.1003>.
- Motlagh Z. K., Derakhshani R., Sayadi M. H. (2023) Groundwater vulnerability assessment in central Iran: Integration of GIS-based DRASTIC model and a machine learning approach. *Groundwater Sustainable Development* 23:101037. DOI: <https://doi.org/10.1016/j.gsd.2023.101037>.
- Najafpour N., Torabi Poode H., Yoonesi H. (2020) Evaluation of groundwater quantity and quality on aquifer pollutant (Case study: Lenjanat Plain, Isfahan). *Journal of Environmental Studies* 11 (21): 233–248. DOI: <https://doi.org/20.1001.1.20089597.1399.11.21.20.9>.
- Najafpour N., Vahdatpour N., Aghababaei E. (2024) Evaluating the capability of artificial intelligence in predicting the amount of electrical conductivity and nitrate in underground water resources (a case study of artificial neural methods ANN and ANFIS). *Journal of Quality Engineering and Management* 13 (2): 207–230. DOI: <https://doi.org/10.48313/jqem.2023.192560>.
- Nas B. (2009) Geostatistical approach to assessment of spatial distribution of groundwater quality. *Journal of Environmental* 18 (4): 1073.
- Nasiri Khiavi A., Mostafazadeh R., Adhami M. (2023) Groundwater quality modeling and determining critical points: a comparison of machine learning to Best-Worst Method. *Environ Sci Pollut Res Int.* 30 (54): 115758–115775. DOI: <https://doi.org/10.1007/s11356-023-30530-8>.
- Panagopoulos G., Antonakos A., Lambrakis N. (2005) Optimization of the DRASTIC model for groundwater vulnerability assessment by the use of simple statistical methods and GIS. *Hydrogeology Journal* 13 (3): 922–933. DOI: <https://doi.org/10.1007/s10040-004-0394-7>.
- Ram A., Tiwari S. K., Pandey H. K., et al. (2021) Groundwater quality assessment using water quality index (WQI) under the GIS framework. *Applied Water Science* 11 (46) DOI: <https://doi.org/10.1007/S13201-021-01376-7>.
- Shihab A. S., Hassan Z. A. (2006) Classification of groundwater quality in Mosul, Iraq, using cluster analysis. *Environmental Monitoring and Assessment* 120 (1-3): 179–196. DOI: <https://doi.org/10.1007/s10661-005-9050-5>.
- Smola A. J., Schölkopf B. (2004) A tutorial on support vector regression. *Statistics and Computing* 14 (2): 199–222. DOI: <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- Soper D. S. (2006) Mapping groundwater vulnerability using the DRAS-TIC index: A spatial approach integrating GIS. *Journal of Hydrology* 333 (1-2): 234–245. DOI: <https://doi.org/10.1016/j.jhydrol.2006.08.001>.
- Stigter T. Y., Ribeiro L., Carvalho Dill A. M. M. (2006) Application of a groundwater quality index as an assessment and communication tool in agro-environmental policies: Two Portuguese case studies. *Journal of Hydrology* 327 (3-4): 578. DOI: <https://doi.org/10.1016/j.jhydrol.2005.12.001>.
- Stocklin J. (1968) Structural history and tectonic of Iran: A review. *AAPG Bulletin* 52 (7): 1229–1258. DOI: <https://doi.org/10.1306/5D25C4E3-16C1-11D7-8645000102C1865D>.
- Subba-Rao N. (2012) PIG: A numerical index for dissemination of groundwater contamination zones. *Hydrological Processes* 26:3344–3350. DOI: <https://doi.org/10.1002/hyp.8456>.
- Subbarayan S., Thiyagarajan S., Karuppanan S., Panneerselvam B. (2024) Enhancing groundwater vulnerability assessment: Comparative study of three machine learning models and five classification schemes for Cuddalore district. *Environmental Research*, 117769. DOI: <https://doi.org/10.1016/j.envres.2023.117769>.
- Taghizadeh Mehrjardi R., Akbarzadeh S., Mahmoodi S. h., Heidari A., Sarmadian F. (2008) Application of geostatistical methods for mapping groundwater quality in Azarbaijan Province, Iran. *American Eurasian Journal of Agriculture and Environmental Science* 3 (5): 726. DOI: <https://doi.org/20.1001.1.20089554.1387.2.5.7.4>.
- Tayfur G., Nadiri A. A., Asghari Moghaddam A. (2014) Supervised intelligent committee machine method for hydraulic conductivity estimation. *Water Resources Management* 28 (4): 1173–1184. DOI: <https://doi.org/10.1007/S11269-014-0553-y>.
- Ting K. M., Witten I. H. (1999) Issues in stacked generalization. *Journal of Artificial Intelligence Research* 10:271–289. DOI: <https://doi.org/10.1613/jair.594>.
- Wang C., Wang X., Zhang G., Zhang F., Li J., Chen S., Sauvage S., Sánchez-Pérez J., Han Y., Liu J. (2023) Identification of critical effect factors for prediction of spatial and interannual variability of shallow groundwater nitrate in agricultural areas. *Science of the Total Environment* 891:164342. DOI: <https://doi.org/10.1016/j.scitotenv.2023.164342>.
- Wolpert D. H. (1992) Stacked generalization. *Neural Networks* 5 (2): 241–259. DOI: [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- World Health Organization (1998) Guidelines for drinking water quality, 2nd Ed. WHO (Geneva, Switzerland)