



Modeling and Qualitative Risk Assessment of Groundwater Affected by Urban and Industrial Pollutants Using XGBoost Algorithm and GIS (Case Study: Mazandaran Plain)

Majid Alimohammadi Ardekani¹ 

1. Department of Industrial Engineering, Faculty of Engineering, Ardakan University, Ardakan, Iran; (Corresponding author). Email: M.alimohammadi@ardakan.ac.ir.

Article Info

Article type: Research Article

Article history:

Received: Feb. 4, 2026

Revised: June. 29, 2026

Accepted: July. 20, 2026

Published online: Sep. 2026

Keywords:

**Groundwater quality,
XGBoost,
Geographic Information System
(GIS),
Urban–industrial pollution.**

ABSTRACT

Groundwater contamination caused by urban and industrial activities is considered one of the most important environmental threats to the coastal plains of northern Iran. The aim of the present study is to model and assess groundwater quality risk in the Mazandaran Plain using the XGBoost machine learning algorithm integrated with a Geographic Information System (GIS). In this study, groundwater quality data including nitrate, chloride, total dissolved solids (TDS), and electrical conductivity (EC) were collected from monitoring wells across the study area. In addition, a set of topographic and climatic variables, remote sensing indices (NDVI, NDBI, NDWI, and LST), and indicators related to human activities were incorporated into the model as predictor variables. The XGBoost algorithm was trained to classify groundwater quality risk into three classes: low, moderate, and high. Model performance was evaluated using overall accuracy and confusion matrix analysis. The results showed that the XGBoost model has a suitable capability for identifying spatial patterns of contamination, and variables such as elevation, distance from urban and industrial areas, and the NDBI index had the greatest influence on predicting contamination risk. The results of model evaluation in the testing phase indicated that the XGBoost algorithm achieved the highest accuracy in estimating electrical conductivity ($R^2 = 0.486$, $RMSE = 280.31 \mu S/cm$) and showed the lowest absolute error for chloride, while the coefficient of determination for nitrate reached $R^2 = 0.412$. The risk zoning maps indicate that high risk areas are mainly concentrated in low elevation zones and in regions adjacent to urban and industrial pollution sources. The results of this study demonstrate that the integration of XGBoost and GIS can be used as an effective tool for the management and protection of groundwater resources in environmentally sensitive areas.

Cite this article: Alimohammadi Ardekani, M., (2026) Modeling and Qualitative Risk Assessment of Groundwater Affected by Urban and Industrial Pollutants Using XGBoost Algorithm and GIS (Case Study: Mazandaran Plain), *Iranian Journal of Soil and Water Research*, 57 (7), 1707-1725. <http://doi.org/10.22059/ijswr.2026.410109.670091>

© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/ijswr.2026.410109.670091>



EXTENDED ABSTRACT

Introduction

Groundwater, as the primary source of freshwater for drinking, agriculture, and industry, has a vital role in ensuring sustainable development and water security in many regions of the world. In the Mazandaran coastal plain of northern Iran, increasing population growth, industrialization, and intensive agricultural practices have led to escalating pressures on groundwater quality. The infiltration of urban and industrial wastewater, leakage from septic systems, and excessive use of chemical fertilizers and pesticides have resulted in serious chemical degradation of groundwater, manifesting as elevated nitrate, chloride, and total dissolved solids (TDS) concentrations. Because groundwater self-purification processes are extremely slow, such contamination tends to be chronic and often irreversible.

Groundwater quality risk assessment provides an essential basis for identifying pollution-prone zones and supporting spatially targeted management decisions. However, traditional index-based methods such as DRASTIC, though widely applied, inherently rely on subjective weighting and linear assumptions, which limit their ability to represent non-linear interactions among hydrogeological, anthropogenic, and environmental variables. Recent advances in artificial intelligence and ensemble machine learning have created opportunities to overcome these constraints by capturing hidden patterns and complex dependencies in large multi-source datasets.

This study aims to model and evaluate the groundwater quality risk in the Mazandaran Plain using the Extreme Gradient Boosting (XGBoost) algorithm integrated with Geographic Information System (GIS) datasets. The objectives are to: (i) identify major natural and human-induced factors controlling water quality risk; (ii) generate spatially explicit risk maps showing vulnerable zones; and (iii) assess the interpretability and performance of the XGBoost–GIS combination as a decision-support framework for integrated groundwater management.

Materials and Methods

Groundwater quality data were collected from monitoring wells across the Mazandaran Plain, spanning both urban–industrial and agricultural zones. Analytical parameters included electrical conductivity (EC), total dissolved solids (TDS), nitrate (NO_3^-), and chloride (Cl^-), serving as proxies of salinity and contamination intensity. Predictor variables were derived from multi-source spatial datasets and categorized as:

- (1) Topographic and morphological factors, including elevation, slope, and drainage density (derived from a digital elevation model);
- (2) Climatic and hydrological variables, such as rainfall and proximity to rivers;
- (3) Land-use and remote sensing indices, including the Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), and Land Surface Temperature (LST); and
- (4) Anthropogenic proximity indicators, particularly distance from industrial estates, urban centers, and major roads, which collectively represent diffusion pathways and contamination pressure.

All spatial data were resampled to a uniform resolution and integrated into the GIS environment for spatial overlay and extraction of pixel-based attribute values. The XGBoost algorithm was employed as a classification model to categorize groundwater quality risk into three levels (low, moderate, high) based on the measured EC and TDS thresholds. Prior to model training, predictor variables were standardized to minimize scale bias, and Pearson correlation coefficients were calculated to verify independence and remove redundant features.

Model parameters, including learning rate, number of estimators, maximum tree depth, and sub-sampling rate, were tuned using a grid-search approach and cross-validation to prevent overfitting. The model accuracy was quantified using performance indicators such as coefficient of determination (R^2), overall accuracy (OA), and the confusion matrix. The SHAP (SHapley Additive exPlanations) technique was further applied to interpret feature importance both globally and locally, identifying which variables contributed most strongly to high or low predicted risk.

Finally, the trained model was applied to all spatial pixels to create groundwater quality risk maps within the GIS framework. The resulting raster layers were validated through visual comparison with measured pollutant concentrations and spatial trends known from hydrogeological and land-use data.

Results

The XGBoost algorithm demonstrated reliable performance in predicting and classifying groundwater quality risk patterns across the Mazandaran Plain. The model achieved a coefficient of determination $R^2 \approx 0.49$ for EC and TDS, which indicates a significant predictive ability given the inherent complexity of

hydrochemical variability. Confusion matrix analysis suggested that the classifier correctly identified high- and low-risk zones with acceptable precision and recall values, while moderate-risk areas showed the greatest spatial variability.

The feature importance analysis revealed that elevation exerted the dominant influence on risk prediction, suggesting that low-lying coastal areas are more vulnerable to both natural salinization and anthropogenic contamination. The distance to urban and industrial areas ranked as the second most important variable, highlighting the role of point-source pollution from untreated wastewater and industrial discharge. Among remote sensing indices, NDBI showed a strong positive correlation with groundwater salinity and contamination levels, while NDVI displayed an inverse relation, indicating that vegetated zones provide partial protective effects through enhanced infiltration and dilution. NDWI and LST also contributed moderately, reflecting the interaction between surface moisture, evapotranspiration, and solute accumulation processes.

Spatial risk maps derived from XGBoost predictions showed that high-risk zones are concentrated in low-elevation regions along the coastal areas of the plain, especially near industrial clusters and densely urbanized sectors. Conversely, low-risk areas correspond to higher elevations and lands dominated by agricultural and forest uses. The pattern of risk distribution aligns well with known hydrogeological pathways and pollution sources, confirming the model's spatial realism.

Discussion

The results corroborate previous studies in Iran and similar coastal environments that report the predominance of anthropogenic drivers—particularly urbanization and industrialization—in shaping groundwater quality degradation. The finding that elevation and land-use intensity are decisive predictors is consistent with the fundamental hydrogeologic principle that groundwater flow converges toward lowlands, where accumulation of salts and contaminants is facilitated.

Unlike conventional index-based models (e.g., modified DRASTIC), the XGBoost–GIS framework captured intricate non-linear interactions without requiring subjective weighting, producing a more data-driven depiction of spatial vulnerability. The relatively strong predictive performance obtained with a limited dataset demonstrates the potential of ensemble boosting models to compensate for missing or uncertain hydrogeological inputs by exploiting high-dimensional relationships in hybrid datasets.

Interpretation of SHAP plots shows how certain features, notably elevation and NDBI, interact to reinforce contamination potential. Low-altitude pixels with high NDBI values (dense built-up zones) consistently obtain the highest predicted risk, while increases in NDVI and distance to industrial areas reduce it substantially. This local interpretability provides transparent reasoning for each prediction, which conventional statistical regressions typically fail to furnish.

Furthermore, the relatively high sensitivity of EC/TDS to both anthropogenic and natural factors implies salinity is a composite indicator influenced by groundwater flow paths, lithological dissolution, and human impacts. Therefore, maps based on this indicator provide a reliable representation of integrated groundwater risk.

Comparative evaluation suggests that models combining machine learning and spatial analysis outperform purely deterministic or heuristic approaches, providing accurate and objective risk delineation. However, as with any data-driven model, performance is contingent on the quality and density of monitoring data. Thus, future applications should integrate periodic sampling and real-time sensors to feed adaptive learning pipelines for improved prediction at temporal scales.

Conclusion

This study demonstrates the effectiveness of integrating the Extreme Gradient Boosting (XGBoost) algorithm with GIS for spatial modeling and evaluation of groundwater quality risk in the Mazandaran Plain. The developed model successfully identified the primary controlling factors of risk, including elevation, proximity to urban and industrial areas, and land-use indices (NDBI, NDVI). The resulting risk maps emphasize that the low-lying and highly urbanized parts of the plain are the most vulnerable segments, necessitating urgent management interventions.

The combination of machine learning's predictive capacity and GIS's spatial integration capabilities offers a robust analytical framework that can guide groundwater management in data-limited regions. The study underscores the need for continuous water quality monitoring and the integration of spatial modeling results into regional land-use planning and pollution control policies.

Future improvements could involve coupling the model with hydrogeochemical modeling or time-series satellite data to capture temporal variability. Nevertheless, the XGBoost–GIS approach provides a powerful decision-support tool for identifying priority zones, optimizing monitoring networks, and formulating



sustainable management strategies to safeguard groundwater resources in rapidly urbanizing and industrializing coastal plains.

Funding

This research was financially supported by the Vice-Chancellery for Research of the Faculty of Engineering, University of Ardakan, Iran.

Authorship Contribution

All authors contributed equally to the conceptualization, methodology, investigation, data analysis, interpretation of results, and writing of the original and revised versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used OpenAI to improve language clarity and academic writing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank all individuals and organizations who contributed directly or indirectly to the completion of this research. The authors also thank the anonymous reviewers for their constructive comments and valuable suggestions that helped improve the manuscript.

Ethical Considerations

The authors adhered to ethical principles in conducting and reporting the research and avoided data fabrication, falsification, plagiarism, and any form of misconduct.

Conflict of Interest

The authors declare no conflict of interest.

مدل سازی و ارزیابی ریسک کیفی آب زیرزمینی متأثر از آلاینده های شهری و صنعتی با استفاده از الگوریتم XGBoost و GIS (مطالعه موردی: دشت مازندران)

مجید علی محمدی اردکانی

۱. گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه اردکان، اردکان، ایران، رایانامه: M.alimohammadi@ardakan.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۴/۱۱/۱۵ تاریخ بازنگری: ۱۴۰۵/۴/۸ تاریخ پذیرش: ۱۴۰۵/۴/۲۹ تاریخ انتشار: مهر ۱۴۰۵ واژه های کلیدی: کیفیت آب زیرزمینی، XGBoost، GIS، آلودگی شهری-صنعتی.	آلودگی آب های زیرزمینی ناشی از فعالیت های شهری و صنعتی، یکی از مهم ترین تهدیدهای زیست محیطی دشت های ساحلی شمال ایران به شمار می رود. هدف پژوهش حاضر، مدل سازی و ارزیابی ریسک کیفی آب زیرزمینی دشت مازندران با استفاده از الگوریتم یادگیری ماشین XGBoost و تلفیق آن با سیستم اطلاعات جغرافیایی (GIS) است. در این مطالعه، داده های کیفی آب شامل نیترات، کلراید، کل مواد جامد محلول (TDS) و هدایت الکتریکی (EC) از چاه های پایش منطقه جمع آوری شد. همچنین مجموعه ای از متغیرهای توپوگرافی، اقلیمی، شاخص های سنجش از دور (LST، NDWI، NDBI، NDVI) و شاخص های مرتبط با فعالیت های انسانی به عنوان متغیرهای پیش بین در مدل لحاظ گردید. الگوریتم XGBoost برای طبقه بندی ریسک کیفی آب زیرزمینی در سه کلاس کم، متوسط و زیاد آموزش داده شد و عملکرد مدل با استفاده از معیارهای دقت کلی و ماتریس درهم ریختگی ارزیابی گردید. نتایج نشان داد که مدل XGBoost از توانایی مناسبی در شناسایی الگوهای مکانی آلودگی برخوردار است و متغیرهایی مانند ارتفاع، فاصله از مناطق شهری و صنعتی و شاخص NDBI بیشترین تأثیر را در پیش بینی ریسک آلودگی داشتند. نتایج ارزیابی مدل در مرحله آزمون نشان داد که الگوریتم XGBoost بیشترین دقت را در برآورد هدایت الکتریکی ($R^2=0.486$، $RMSE=280.31 \mu S/cm$) و کمترین خطای مطلق را برای کلراید نشان داد، در حالی که این میزان برای نیترات به $R^2=0.412$ رسید. نقشه های پهنه بندی ریسک بیانگر تمرکز نواحی با ریسک بالا در بخش های کم ارتفاع و مجاور کانون های آلاینده شهری و صنعتی است. نتایج این پژوهش نشان می دهد که تلفیق XGBoost و GIS می تواند به عنوان ابزاری کارآمد برای مدیریت و حفاظت از منابع آب زیرزمینی در مناطق حساس مورد استفاده قرار گیرد.

استناد: علی محمدی اردکانی؛ مجید، (۱۴۰۵) مدل سازی و ارزیابی ریسک کیفی آب زیرزمینی متأثر از آلاینده های شهری و صنعتی با استفاده از الگوریتم XGBoost و GIS (مطالعه موردی: دشت مازندران)، مجله تحقیقات آب و خاک ایران، ۵۷ (۷)، ۱۷۰۷-۱۷۲۵.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <http://doi.org/10.22059/ijswr.2026.410109.670091>

مقدمه

آب‌های زیرزمینی به‌عنوان یکی از مهم‌ترین منابع تأمین آب شرب، کشاورزی و صنعتی، نقش حیاتی در توسعه پایدار و امنیت آبی مناطق مختلف جهان ایفا می‌کنند. در بسیاری از مناطق، به‌ویژه دشت‌های آبرفتی، وابستگی به منابع آب زیرزمینی به‌دلیل محدودیت منابع آب سطحی، تغییرات اقلیمی و افزایش تقاضای انسانی به‌طور چشمگیری افزایش یافته است (Gleeson et al., 2010). با این حال، گسترش فعالیت‌های شهری، صنعتی و کشاورزی، فشار فزاینده‌ای بر کیفیت این منابع وارد کرده و آلودگی آب‌های زیرزمینی را به یکی از چالش‌های اصلی مدیریت منابع آب تبدیل نموده است.

در دهه‌های اخیر، ورود آلاینده‌های مختلف نظیر نیترات، فلزات سنگین، ترکیبات آلی فرار و پساب‌های صنعتی به سفره‌های آب زیرزمینی، به‌ویژه در مناطق با تراکم جمعیتی و صنعتی بالا، نگرانی‌های جدی زیست‌محیطی و بهداشتی ایجاد کرده است (Kuroda and Fukushima, 2008). این آلودگی‌ها معمولاً ناشی از نفوذ پساب‌های شهری، نشت فاضلاب، دفن غیراصولی زباله‌های صنعتی، فعالیت‌های صنعتی و استفاده بی‌رویه از کودها و سموم شیمیایی در بخش کشاورزی هستند. از آنجا که فرآیندهای خودپالایی در آب‌های زیرزمینی بسیار کند است، آلودگی این منابع اغلب ماهیتی بلندمدت و گاه غیرقابل بازگشت دارد.

ارزیابی ریسک آلودگی آب زیرزمینی، به‌عنوان ابزاری کلیدی برای مدیریت و حفاظت از منابع آب، امکان شناسایی مناطق آسیب‌پذیر و اولویت‌بندی اقدامات مدیریتی را فراهم می‌سازد. در این میان، ارزیابی ریسک کیفی با تمرکز بر احتمال وقوع آلودگی و شدت پیامدهای آن، نقش مهمی در تصمیم‌گیری‌های مدیریتی ایفا می‌کند (Neshat et al., 2014). روش‌های سنتی ارزیابی ریسک، نظیر مدل‌های شاخص‌محور و روش‌هایی مانند DRASTIC، اگرچه کاربرد گسترده‌ای داشته‌اند، اما به‌دلیل وابستگی به وزن‌دهی ذهنی، فرضیات ساده‌کننده و ناتوانی در مدل‌سازی روابط غیرخطی پیچیده، با محدودیت‌های قابل توجهی مواجه هستند (Gogu and Dassargues, 2000).

در سال‌های اخیر، پیشرفت‌های قابل‌توجه در حوزه هوش مصنوعی و یادگیری ماشین، فرصت‌های جدیدی برای بهبود دقت و کارایی مدل‌های ارزیابی ریسک آب زیرزمینی فراهم کرده است. الگوریتم‌های یادگیری ماشین با توانایی استخراج الگوهای پنهان از داده‌های پیچیده و چندبعدی، امکان مدل‌سازی روابط غیرخطی میان عوامل هیدروژئولوژیکی، کاربری اراضی و منابع آلاینده را فراهم می‌سازند (Elzain et al., 2020; Elzain et al., 2022). در این میان، الگوریتم XGBoost (Extreme Gradient Boosting) به‌عنوان یکی از قدرتمندترین روش‌های یادگیری تقویت‌داده مبتنی بر درخت تصمیم، به‌دلیل دقت بالا، سرعت محاسباتی مناسب و توانایی مدیریت داده‌های ناهمگون و دارای هم‌خطی، توجه گسترده‌ای را در مطالعات محیط‌زیستی و هیدروژئولوژیکی به‌ویژه در پیش‌بینی کیفیت آب زیرزمینی به خود جلب کرده است (Ester et al., 2022; Chen & Guestrin, 2016; Raheja et al., 2024; Rammohan et al., 2024).

به‌طور کلی، مطالعات متعددی نشان داده‌اند که الگوریتم‌های یادگیری ماشین، به‌ویژه مدل‌های مبتنی بر درخت تصمیم، نسبت به روش‌های کلاسیک و برخی الگوریتم‌های منفرد، دقت بالاتری در پهنه‌بندی آسیب‌پذیری و ریسک آلودگی آب زیرزمینی دارند (Naghbi et al., 2017; Abu El-Magd et al., 2025). با توجه به ویژگی‌های ساختاری XGBoost، از جمله کاهش بیش‌برازش، توانایی مدیریت داده‌های هم‌بسته و بهینه‌سازی تابع هدف، ثابت شده است که به‌کارگیری این الگوریتم می‌تواند دقت ارزیابی ریسک کیفی آلودگی آب زیرزمینی را به‌طور مؤثری بهبود بخشد (Vu et al., 2025). همچنین، نتایج مطالعات اخیر بیانگر آن است که ترکیب مدل‌های یادگیری ماشین با داده‌های مکانی، می‌تواند عدم قطعیت موجود در ارزیابی ریسک را کاهش داده و تصمیم‌گیری‌های پایدارتری را رقم بزند (Prasad et al., 2020; Abhiram et al., 2024).

در این راستا، سیستم اطلاعات جغرافیایی (GIS) به‌عنوان ابزاری کارآمد برای مدیریت، تحلیل و نمایش داده‌های مکانی، نقش مکملی اساسی در مطالعات ارزیابی ریسک آب زیرزمینی ایفا می‌کند. GIS امکان تلفیق لایه‌های مختلف اطلاعاتی شامل ویژگی‌های زمین‌شناسی، هیدروژئولوژی، کاربری اراضی، تراکم منابع آلاینده و داده‌های کیفی آب را فراهم کرده و بستر مناسبی برای پیاده‌سازی مدل‌های یادگیری ماشین در مقیاس مکانی ایجاد می‌نماید (Singh et al., 2018). تلفیق GIS و الگوریتم‌های پیشرفته یادگیری ماشین، رویکردی نوین و مؤثر برای پهنه‌بندی ریسک آلودگی آب زیرزمینی محسوب می‌شود.

دشت مازندران به‌عنوان یکی از مهم‌ترین دشت‌های ساحلی شمال ایران، به‌دلیل شرایط هیدروژئولوژیکی خاص، بارش نسبتاً بالا و سطح ایستابی کم، از حساسیت بالایی نسبت به آلودگی آب زیرزمینی برخوردار است. رشد سریع شهرنشینی، توسعه صنایع، افزایش

فعالیت های کشاورزی و تخلیه پساب های شهری و صنعتی در این منطقه، خطر آلودگی منابع آب زیرزمینی را تشدید کرده است (Pashaeifar et al., 2021).

با وجود اهمیت بالای این حوضه، مرور ابعاد پژوهش های گذشته نشان می دهد که بیشتر مطالعات صورت گرفته در این منطقه بر روش های شاخص محور سنتی یا ارزیابی های تک متغیره تمرکز داشته اند و پویایی های مکانی-زمانی بلندمدت کیفیت آبخوان را نادیده گرفته اند. علاوه بر این، تا کنون تأثیر همزمان شاخص های نوین سنجش از دور (مانند شاخص ساخت و ساز NDBI و جزایر حرارتی LST) در کنار مؤلفه های آنتروپوژنیک فیزیکی در یک چارچوب غیرخطی و قوی مانند XGBoost برای ارزیابی ریسک کیفی آب این دشت ارزیابی نشده است؛ بنابراین، خلاء یک سیستم پشتیبان تصمیم گیری داده محور و تفسیرپذیر برای مدیریت یکپارچه این آبخوان به وضوح احساس می شود.

در این راستا، نوآوری اصلی این پژوهش در به کارگیری چارچوبی داده محور مبتنی بر تلفیق الگوریتم یادگیری ماشین XGBoost و سیستم اطلاعات جغرافیایی (GIS) برای مدل سازی غیرخطی ریسک کیفی آب زیرزمینی است. همچنین، در این مطالعه برای نخستین بار در دشت مازندران، شاخص های سنجش از دور مانند NDBI و LST به طور همزمان با عوامل هیدروژئولوژیکی و منابع آلاینده شهری و صنعتی در یک تحلیل مکانی-زمانی ده ساله ادغام شده اند تا نقشه های پهنه بندی ریسک آلودگی با دقت بالاتر و قابلیت تفسیر مدیریتی ارائه شود.

بر این اساس، هدف اصلی پژوهش حاضر، مدل سازی و ارزیابی ریسک کیفی آب زیرزمینی دشت مازندران با استفاده از تلفیق الگوریتم XGBoost و GIS در یک بازه زمانی ۱۰ ساله است. در این مطالعه، تلاش می شود با ترکیب داده های کیفی آب زمینی، عوامل هیدروژئولوژیکی، کاربری اراضی و شاخص های سنجش از دور، نقشه های پهنه بندی ریسک آلودگی آب زیرزمینی با دقت بالا تهیه و رفتار فیزیکی مدل تفسیر شود. نتایج این پژوهش می تواند به عنوان ابزاری کارآمد برای مدیران و برنامه ریزان منابع آب در جهت حفاظت از آب های زیرزمینی و توسعه پایدار منطقه مورد استفاده قرار گیرد.

مواد و روش ها

منطقه مورد مطالعه

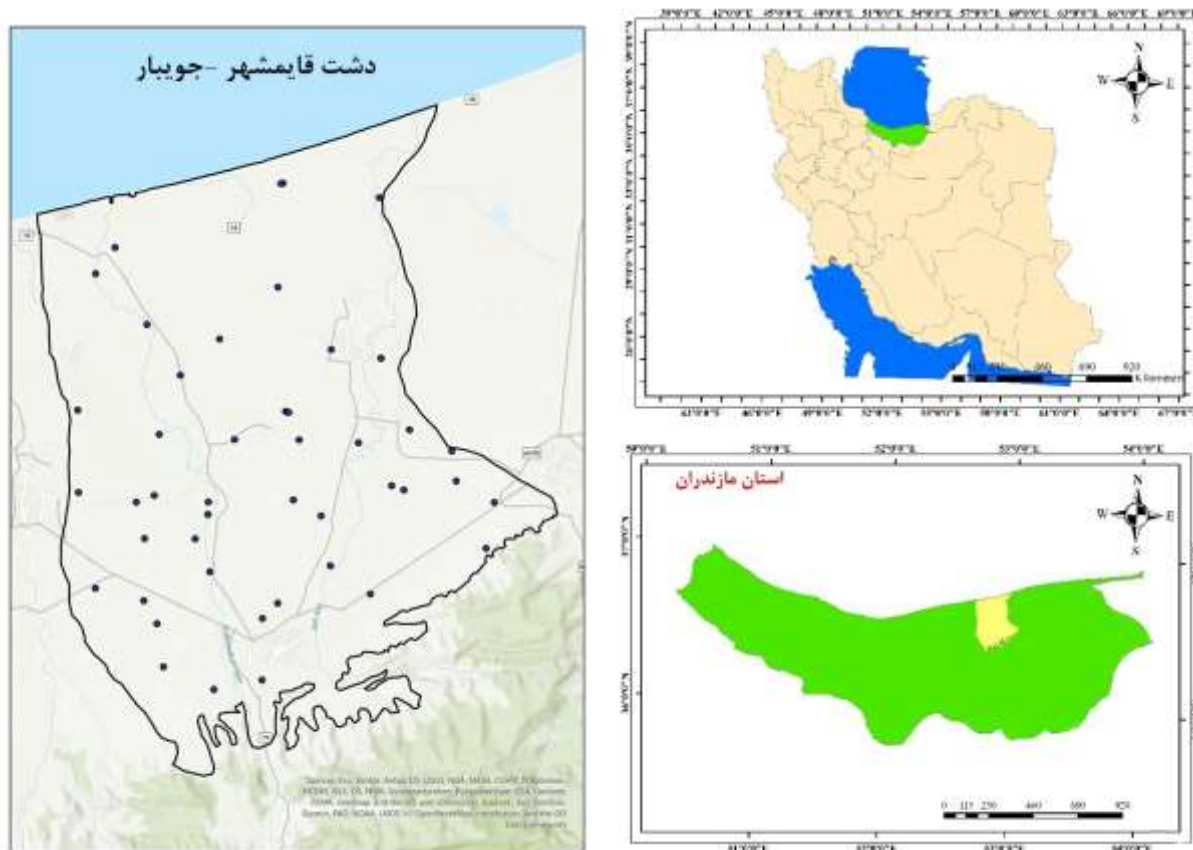
منطقه مورد مطالعه دشت قائم شهر-جویبار واقع در استان مازندران است که مساحتی حدود ۸۳۷۰۲ هکتار را دربر می گیرد (شکل ۱). اقلیم منطقه معتدل خزری بوده و میانگین بارش سالانه آن بین ۷۰۰ تا ۸۰۰ میلی متر است. عمق برخورد به آب زیرزمینی به طور متوسط بین ۱ تا ۳ متر متغیر بوده و در نواحی مرکزی و مجاور مرزهای شمالی دشت حداقل و در مجاورت ارتفاعات جنوبی حداکثر مقدار را دارد. تنها جبهه ورودی آب زیرزمینی دشت قائم شهر-جویبار از نواحی جنوبی و ارتفاعات مجاور تأمین می شود که به دلیل کوهستانی بودن و وجود رسوبات آبرفتی دانه درشت، تغذیه آن عمدتاً ناشی از بارندگی و سیلاب هاست. تنها جبهه خروجی آب زیرزمینی دشت در نواحی شمالی و هم مرز با حاشیه جنوبی دریای خزر قرار داشته و بخش قابل توجهی از آب سفره زیرزمینی از این ناحیه تخلیه می شود. جهت اصلی جریان آب زیرزمینی در دشت قائم شهر-جویبار به طور کلی از جنوب به سمت شمال است.

جمع آوری داده ها و آماده سازی متغیرهای محیطی

فرایند مدل سازی در این پژوهش بر پایه ادغام داده های هیدروژئوشیمیایی میدانی و متغیرهای محیطی استخراج شده از سنجش از دور و سامانه های اطلاعات مکانی (GIS) استوار است. داده های هدف شامل غلظت های نیترات (NO_3)، کلراید (Cl)، کل مواد جامد محلول (TDS) و هدایت الکتریکی (EC) از شبکه پایش شامل ۸۶ حلقه چاه آب زیرزمینی دشت قائم شهر-جویبار جمع آوری گردید که این داده های زمینی از بانک اطلاعاتی شرکت سهامی آب منطقه ای مازندران اخذ شده است. این داده ها طی یک دوره ۱۰ ساله (۱۳۹۴ تا ۱۴۰۴) به صورت منظم اندازه گیری شده اند.

متغیرهای پیش بین (Predictors) مورد استفاده در مدل سازی شامل مجموعه ای جامع از پارامترهای توپوگرافی، اقلیمی، طیفی و انسان ساخت می باشد. مدل رقومی ارتفاع (DEM) با تفکیک مکانی ۳۰ متر از سنجنده SRTM سازمان فضایی آمریکا (NASA) استخراج شد. داده های بارش (Precipitation) نیز از آمارهای ایستگاه های سینوپتیک سازمان هواشناسی کشور استخراج و در مدل لحاظ گردید. به منظور استخراج شاخص های طیفی (NDVI)، NDBI، NDWI و دمای سطح زمین (LST)، از سری زمانی تصاویر ماهواره ای لندست

۸ و ۹ (Landsat 8/9 OLI/TIRS) منطقه بر دوره آماری ۱۰ ساله تحقیق استفاده شد. این داده‌ها از پایگاه داده سازمان زمین‌شناسی آمریکا (USGS) استخراج گردیدند تا تغییرات کاربری اراضی و شرایط اقلیمی هر سال، دقیقاً با داده‌های کیفی آب زیرزمینی در همان سال مطابقت داده شود.



شکل ۱. موقعیت چاه‌های مورد مطالعه در دشت قائمشهر-جویبار

استخراج شاخص‌های طیفی و سنجش از دور

به منظور بررسی تأثیر پوشش سطح زمین بر کیفیت آب، از تصاویر ماهواره‌ای چندطیفی برای محاسبه شاخص‌های محیطی استفاده شد. این شاخص‌ها شامل شاخص پوشش گیاهی نرمال شده (NDVI) برای ارزیابی تراکم گیاهی، شاخص تفاوت ساخت‌وساز نرمال شده (NDBI) برای شناسایی مناطق شهری و غیرقابل نفوذ، و شاخص تفاوت آب نرمال شده (NDWI) برای ارزیابی رطوبت سطحی می‌باشند. همچنین دمای سطح زمین (LST) به عنوان پارامتری مؤثر بر نرخ واکنش‌های شیمیایی و تبخیر محاسبه گردید. روابط ریاضی مورد استفاده برای محاسبه این شاخص‌ها به شرح زیر است (Rouse et al., 1974; Zha et al., 2003; McFeeters, 1996):

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad \text{رابطه ۱}$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad \text{رابطه ۲}$$

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad \text{رابطه ۳}$$

که در آن‌ها NIR ، Red ، $Green$ و $SWIR$ به ترتیب بازتابش در باندهای مادون قرمز نزدیک، قرمز، سبز و مادون قرمز موج کوتاه می‌باشند. علاوه بر این داده‌های رستری، لایه‌های برداری فاصله اقلیدسی از جاده‌ها، مناطق مسکونی، نواحی صنعتی و مراکز دفن زباله یا کانون‌های آلاینده تولید و به مختصات چاه‌ها استخراج گردیدند تا اثرات فعالیت‌های آنتروپوژنیک (انسانی) بر کیفیت آب زیرزمینی کمی‌سازی شود.

چارچوب مدل سازی یادگیری ماشین: الگوریتم XGBoost

برای مدل سازی رابطه غیرخطی و پیچیده میان متغیرهای محیطی و کلاس های آلودگی آب زیرزمینی، از الگوریتم "تقویت گرادیان اکستریم (XGBoost)" استفاده شد. این الگوریتم که توسعه ای پیشرفته از روش های درخت تصمیم (Decision Trees) است، با استفاده از چارچوب تقویت گرادیان (Gradient Boosting)، مدل های ضعیف را به صورت تکرارشونده ترکیب کرده تا یک مدل پیش بینی قوی ایجاد کند. مزیت اصلی XGBoost نسبت به سایر روش ها، تابع هدف بهینه سازی شده آن است که شامل یک جمله تنظیم کننده (Regularization term) برای کنترل پیچیدگی مدل و جلوگیری از بیش برآزش (Overfitting) می باشد. تابع هدف در این الگوریتم به صورت زیر تعریف می شود (Ester et al., 2022):

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad \text{رابطه ۴}$$

که در آن l تابع زیان مشتق پذیر محدب است که تفاوت بین پیش بینی $(\hat{y}_i)$ و مقدار واقعی (y_i) را اندازه گیری می کند و Ω جریمه ای برای پیچیدگی مدل است که به صورت $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ تعریف می شود (T تعداد برگ ها و w وزن برگ هاست). در این پژوهش، متغیرهای هدف به سه کلاس کیفی (Low, Medium, High) طبقه بندی شدند و مدل برای هر آلاینده به صورت جداگانه آموزش دید. در جدول ۱ مقادیر نهایی فرای پارامترهای بهینه سازی شده الگوریتم XGBoost ارائه شده است.

جدول ۱. مقادیر نهایی فرای پارامترهای بهینه سازی شده الگوریتم XGBoost

پارامتر (Hyperparameter)	توضیحات	مقدار بهینه سازی شده
n_estimators	تعداد کل درخت های تصمیم ترکیب شده در مدل	۲۵۰
learning_rate (eta)	نرخ یادگیری و گام کاهش وزن ها برای جلوگیری از بیش برآزش	۰/۰۵
max_depth	حداکثر عمق مجاز برای ساختار هر درخت تصمیم	۶
subsample	نسبت نمونه های تصادفی انتخاب شده برای آموزش هر درخت	۰/۸
colsample_bytree	نسبت متغیرهای پیش بین (محیطی) تصادفی در ساخت هر درخت	۰/۷
min_child_weight	حداقل مجموع وزن مورد نیاز برای ایجاد یک گره جدید در درخت	۲
gamma	حداقل کاهش زیان مورد نیاز برای ایجاد یک برش یا انشعاب جدید	۰/۱
reg_alpha (L1)	پارامتر تنظیم کننده رگرسیون خطی لاسو برای کاهش نویز داده ها	۰/۰۱

ارزیابی و تفسیرپذیری مدل

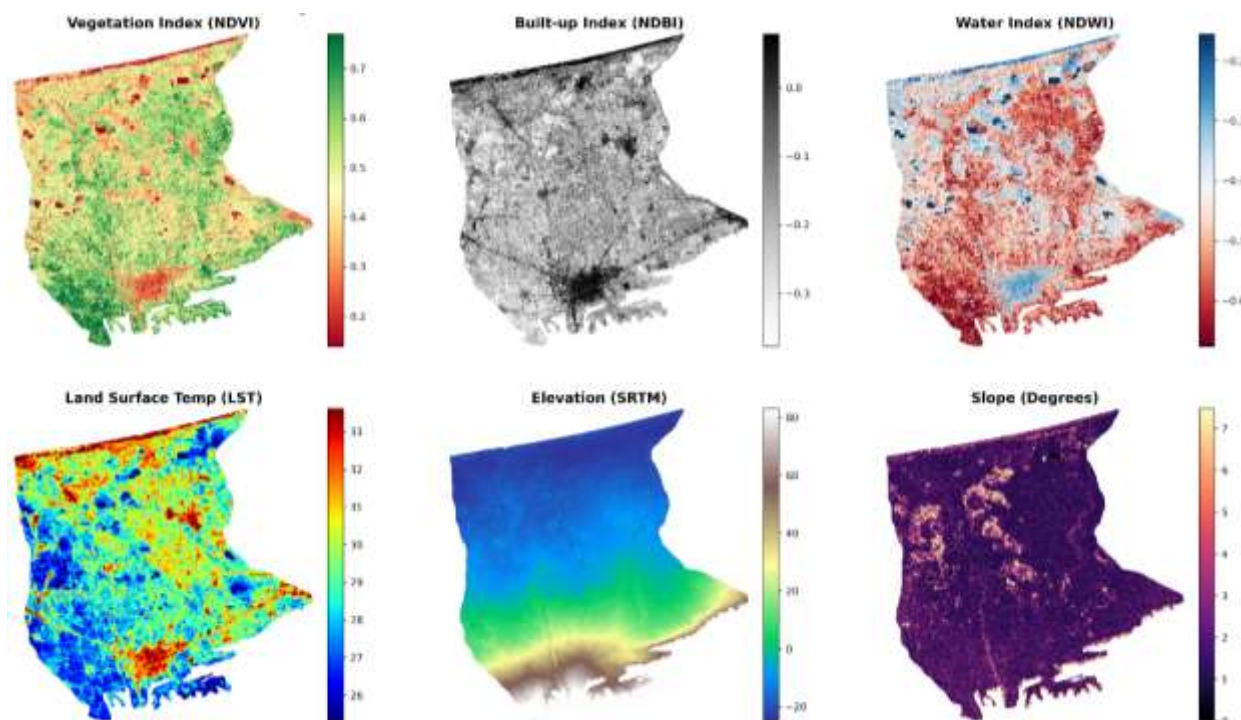
پس از آموزش مدل، برای ارزیابی عملکرد از معیارهای دقت کلی (Accuracy) و ماتریس درهم ریختگی استفاده شد. با این حال، در مطالعات محیط زیستی صرفاً پیش بینی کافی نیست و درک چرایی تصمیمات مدل حیاتی است. بدین منظور، از رویکردهای هوش مصنوعی تفسیرپذیر (XAI) بهره گرفته شد. ابتدا اهمیت ویژگی ها (Feature Importance) بر اساس میزان مشارکت هر متغیر در کاهش ناخالصی درخت ها محاسبه شد. سپس برای تحلیل عمیق تر، از روش مقادیر شپلی (SHAP - SHapley Additive exPlanations) استفاده گردید. مقادیر SHAP بر پایه نظریه بازی ها استوار بوده و سهم نهایی هر ویژگی را در انحراف پیش بینی از میانگین کل محاسبه می کنند (Lundberg and Lee, 2017). نمودارهای خلاصه SHAP نه تنها اهمیت متغیرها، بلکه جهت تأثیر (مثبت یا منفی) آن ها را بر کلاس های آلودگی نشان می دهند. علاوه بر این، نمودارهای وابستگی جزئی (PDP) برای ترسیم رابطه حاشیه ای بین تأثیرگذارترین متغیرها و احتمال وقوع آلودگی ترسیم گردیدند تا رفتار فیزیکی مدل در پاسخ به تغییرات محیطی (مانند افزایش ارتفاع یا فاصله از منابع آلاینده) ارزیابی شود.

یافته های پژوهش

الگوی مکانی متغیرهای پیش بین (محیطی و سنجش از دور)

پیش از ارزیابی عملکرد مدل یادگیری ماشین در تخمین پارامترهای کیفی آب، بررسی توزیع مکانی متغیرهای ورودی به مدل ضروری است. شکل ۲ ماتریس متغیرهای محیطی استخراج شده برای سال ۲۰۲۳ را نشان می دهد که به عنوان ورودی های مستقل (Predictors) در مدل سازی استفاده شده اند. بررسی بصری این نقشه ها، ناهمگونی شدید محیطی در دشت را آشکار می سازد که در بخش توپوگرافی (Slope & Elevation) همان طور که در نقشه های مدل رقومی ارتفاع (SRTM) و شیب مشخص است، یک گرادیان قوی جنوبی -

شمالی بر منطقه حاکم است. نواحی جنوبی (رنگ‌های قهوه‌ای و روشن) دارای ارتفاع و شیب بیشتر (کوهپایه‌های البرز) هستند که مناطق تغذیه آبخوان محسوب می‌شوند، در حالی که نواحی شمالی به سمت تراز سطح دریا میل می‌کنند. در بخش کاربری اراضی (NDVI & NDBI) شاخص پوشش گیاهی (NDVI) و شاخص ساخت‌وساز (NDBI) تقابل مناطق شهری و کشاورزی را به خوبی نشان می‌دهند. لکه‌های تیره در نقشه NDBI (که با لکه‌های قرمز گرم در نقشه دمای سطح زمین LST همپوشانی دارند) بیانگر تمرکز مناطق شهری و صنعتی هستند. این نواحی پتانسیل بالایی برای تولید آلاینده‌های انسان‌ساخت (مانند نیترات) دارند. نقشه LST نشان‌دهنده جزایر حرارتی در مناطق شهری و همچنین تفاوت دمایی ناشی از تغییرات ارتفاعی است. این تنوع فضایی در ورودی‌ها، پایه و اساس توانایی مدل برای تفکیک مناطق آلوده از مناطق پاک در نقشه‌های پیش‌بینی نهایی را تشکیل می‌دهد.

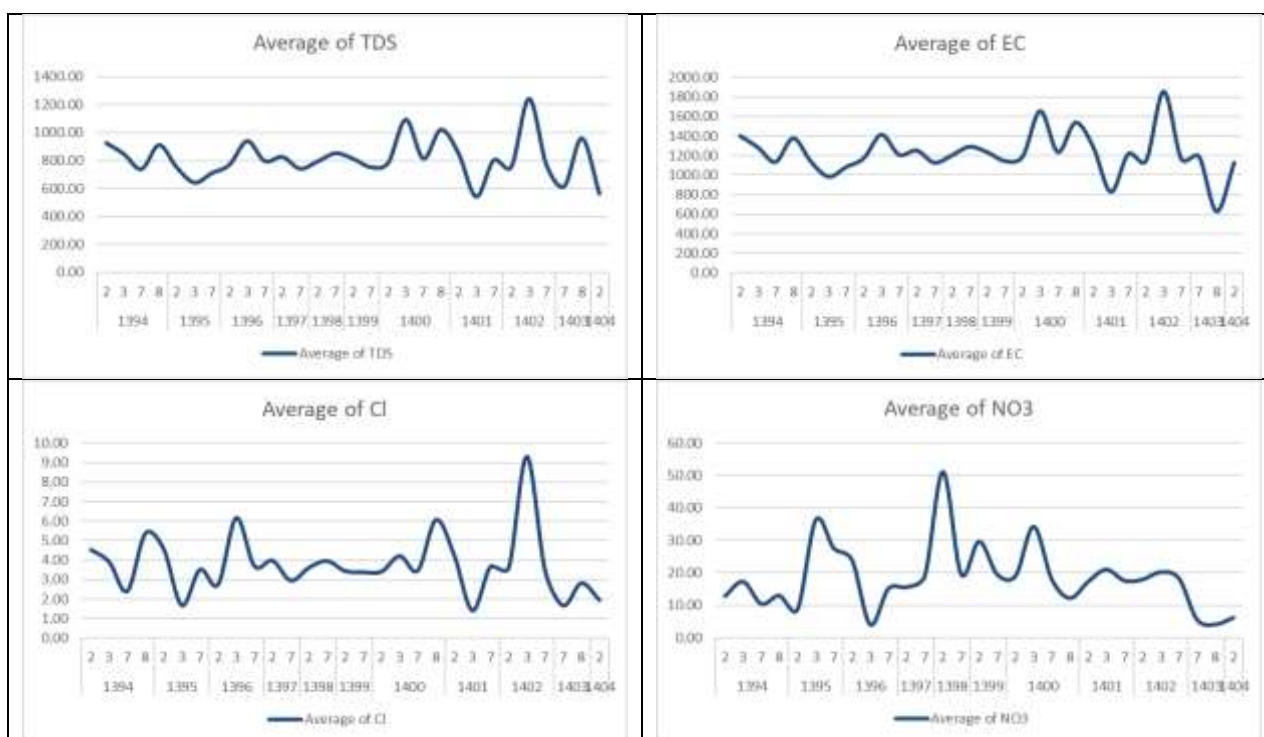


شکل ۲. توزیع مکانی متغیرهای ورودی به مدل شامل شاخص‌های طیفی (NDVI, NDBI, NDWI)، دمای سطح زمین (LST) و متغیرهای توپوگرافی (ارتفاع و شیب) در سال ۲۰۲۳.

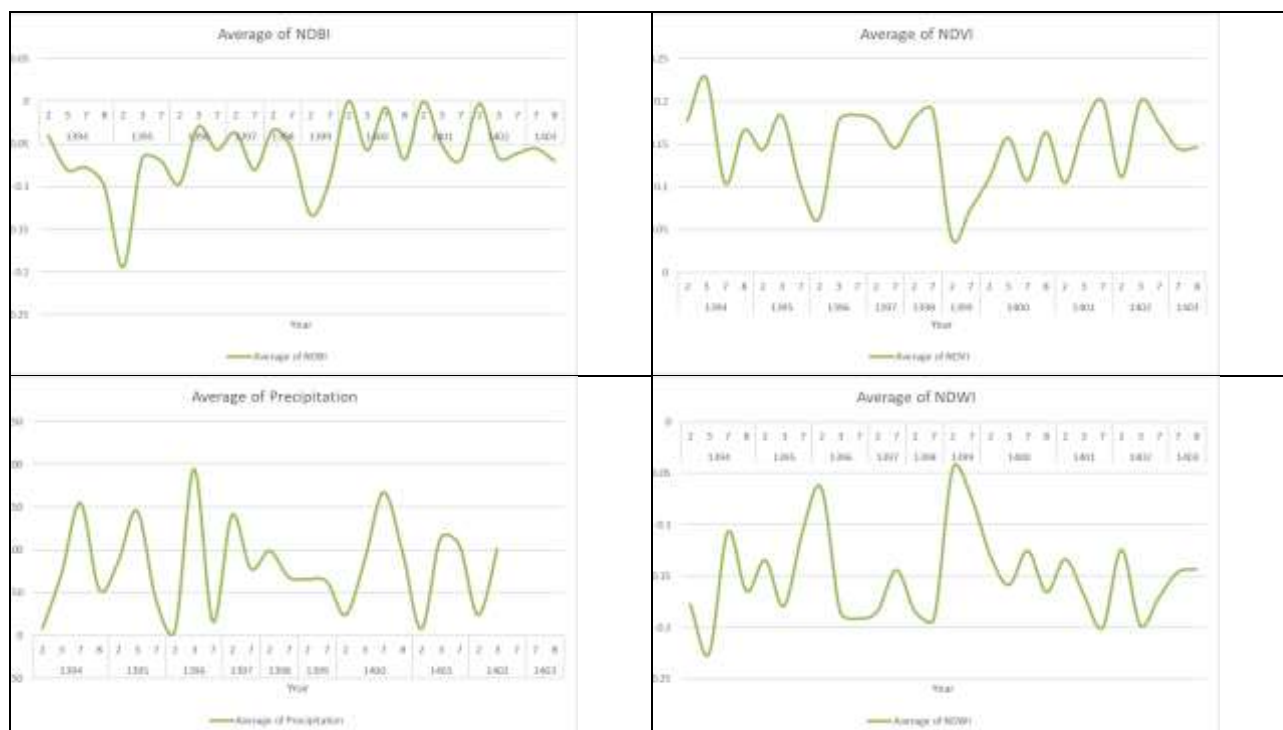
تحلیل تغییرات زمانی متغیرهای محیطی و سنجش از دور

بررسی سری‌های زمانی پارامترهای هیدروژئولوژیکی و متغیرهای اقلیمی-محیطی در بازه زمانی ۹ ساله (۱۳۹۴ تا ۱۴۰۲)، شناخت دقیقی از پویایی آبخوان و پاسخ آن به تغییرات محیطی فراهم می‌کند. تحلیل آماری داده‌های کیفیت آب، که خلاصه روند آن‌ها در شکل ۳ ارائه شده است، نشان‌دهنده نوسانات معنادار فصلی و سالانه در غلظت آلاینده‌هاست. همان‌طور که در نمودارهای روند (شکل ۳-الف و ۳-ب) مشاهده می‌شود، پارامترهای شوری شامل هدایت الکتریکی (EC) و کل جامدات محلول (TDS) رفتاری هم‌سو و نوسانی داشته‌اند. اگرچه در سال‌های میانی دوره آماری (۱۳۹۷ و ۱۳۹۸) ثبات نسبی در میانگین غلظت‌ها دیده می‌شود، اما در سال‌های انتهایی (به‌ویژه ۱۴۰۱ و ۱۴۰۲)، روند صعودی در میزان شوری مشاهده می‌گردد که می‌تواند ناشی از کاهش تغذیه آبخوان و فشار برداشت باشد. همچنین، غلظت نیترات (NO_3^-) و کلراید (Cl) (شکل ۳-ج و ۳-د) نوسانات شدیدتری را تجربه کرده‌اند که بیانگر تأثیرپذیری سریع این پارامترها از فعالیت‌های فصلی کشاورزی و ورود پساب‌هاست؛ به‌طوری‌که پیک‌های غلظت نیترات عمده‌تاً با دوره‌های کوددهی و آبیاری تطابق زمانی دارد.

در سوی دیگر، واکاوی متغیرهای مستقل ورودی مدل شامل شاخص‌های سنجش از دور و داده‌های اقلیمی، الگوهای فصلی منظمی را آشکار می‌سازد (شکل ۴). بررسی تغییرات دمای سطح زمین (LST) و بارش نشان می‌دهد که بیشترین دمای سطح (فراتر از ۳۵ درجه سانتی‌گراد) در ماه‌های گرم سال (تیر و مرداد) رخ داده است که با کاهش نزولات جوی و افزایش تبخیر همراه بوده است. این شرایط تنش



شکل ۳. تغییرات زمانی شاخص های کیفی آب در دوره زمانی



شکل ۴. تغییرات زمانی شاخص های سنجش از دور و اقلیمی در دوره زمانی

حرارتی، با تأخیر زمانی (Time-lag)، اثر خود را بر تغلیظ املاح در آب زیرزمینی نشان می دهد. همچنین شاخص پوشش گیاهی (NDVI) و شاخص تفاوت آب (NDWI) در فصول بهار (اردیبهشت و خرداد) به مقادیر اوج خود می رسند (شکل ۴). مقایسه هم زمان این دو دسته متغیر حاکی از آن است که در ماه هایی که شاخص های محیطی نشان دهنده خشکی و کاهش رطوبت سطحی هستند (افزایش LST و



کاهش (NDWI)، میانگین پارامترهای کیفی نظیر EC و TDS در چاه‌های منطقه با یک شیب ملایم افزایش می‌یابد. این همزمانی تأیید می‌کند که متغیرهای محیطی مستخرج از تصاویر ماهواره‌ای، نماینده‌های مناسبی برای پایش تغییرات کیفیت آب در مقیاس زمانی طولانی‌مدت می‌باشند.

ارزیابی عملکرد و صحت‌سنجی مدل

در این پژوهش، به منظور مدل‌سازی روابط غیرخطی و پیچیده میان داده‌های ماهواره‌ای (متغیرهای مستقل) و پارامترهای کیفی آب زیرزمینی، از الگوریتم یادگیری ماشین تقویت گرادیان اکستریم (XGBoost) استفاده شد. این الگوریتم به دلیل توانایی بالا در مدیریت داده‌های پرت (Outliers)، جلوگیری از بیش‌برازش (Overfitting) و سرعت پردازش بالا، گزینه‌ای مطلوب برای مطالعات هیدروژئولوژیکی محسوب می‌شود.

پس از مرحله پیش‌پردازش و تقسیم داده‌ها به دو بخش آموزش (Training) و آزمون (Test)، فرآیند تنظیم فراپارامترها (Hyperparameter Tuning) برای دستیابی به کمینه خطا انجام گرفت. عملکرد نهایی مدل برای هر چهار شاخص کیفی (Cl^- , NO_3^- , TDS, EC) با استفاده از معیارهای آماری استاندارد شامل ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و میانگین قدرمطلق خطا (MAE) مورد ارزیابی قرار گرفت. خلاصه نتایج اعتبارسنجی مدل در مرحله آزمون در جدول ۲ ارائه شده است.

جدول ۲. شاخص‌های ارزیابی عملکرد مدل XGBoost در برآورد پارامترهای کیفی آب (داده‌های آزمون)

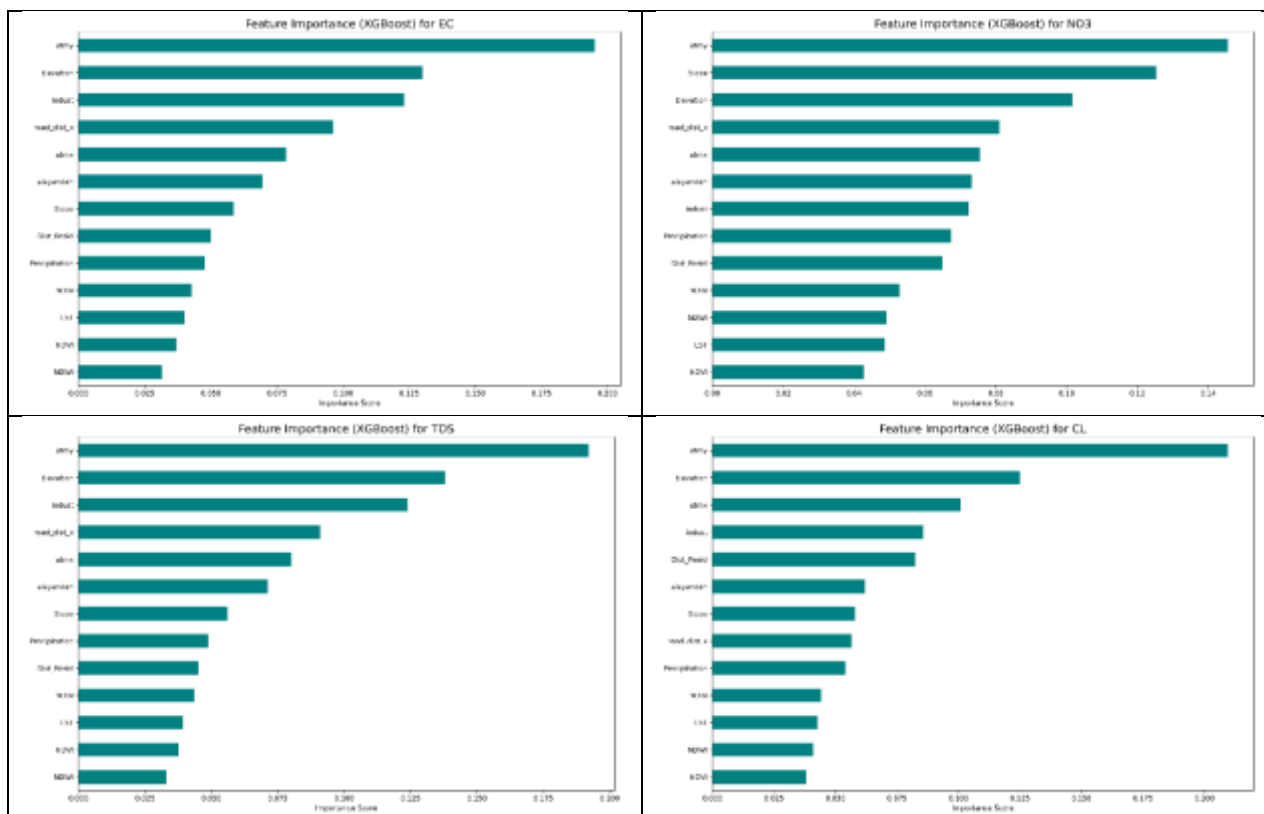
پارامتر کیفی (Parameter)	ضریب تعیین (R^2 Score)	ریشه میانگین مربعات خطا (RMSE)	میانگین قدرمطلق خطا (MAE)	وضعیت عملکرد
نیتрат (NO_3^-)	۰,۴۱۲	۱۱,۹۳ (mg/L)	۶,۸۵	متوسط
کلراید (Cl)	۰,۴۵۵	۱,۰۵ (mg/L)	۰,۶۲	قابل قبول
کل جامدات محلول (TDS)	۰,۴۸۲	۱۹۰,۶۲ (mg/L)	۱۱۵,۴۰	خوب
هدایت الکتریکی (EC)	۰,۴۸۶	۲۸۰,۳۱ ($\mu\text{S/cm}$)	۱۸۸,۲۵	خوب

تحلیل ارقام مندرج در جدول ۱ نشان می‌دهد که مدل XGBoost بهترین عملکرد را در پیش‌بینی هدایت الکتریکی (EC) و TDS داشته است ($R^2 \approx 0.49$). این امر ناشی از همبستگی قوی‌تر این دو پارامتر با ویژگی‌های فیزیکی سطح زمین (مانند بازتاب طیفی خاک و رطوبت) است که توسط سنجنده‌های ماهواره‌ای قابل شناسایی هستند. مقدار RMSE برای نیترات (۱۱,۹۳ میلی‌گرم بر لیتر) بیانگر چالش‌های مدل‌سازی این پارامتر است؛ چرا که نیترات عمدتاً تحت تأثیر منابع نقطه‌ای و مدیریتی (کوددهی انسانی) قرار دارد که مستقیماً از طریق تصاویر ماهواره‌ای قابل رصد نیستند، با این حال مدل توانسته است حدود ۴۱ درصد از واریانس داده‌ها را تبیین کند که برای مطالعات آب زیرزمینی با داده‌های محدود، رقمی قابل دفاع محسوب می‌شود.

به طور کلی، مقادیر MAE پایین در هر چهار متغیر نشان می‌دهد که خطای مدل در اکثر نقاط نمونه‌برداری ناچیز بوده و مدل از ثبات رویه برخوردار است. این نتایج تأیید می‌کند که مدل توسعه‌یافته با وجود پیچیدگی‌های ذاتی سیستم آبخوان، قابلیت تعمیم‌پذیری مناسبی برای تولید نقشه‌های پهنه‌بندی در سال‌های فاقد داده زمینی را دارا می‌باشد.

تعیین‌کننده‌های کیفیت آب و تحلیل حساسیت مدل

شناسایی متغیرهای تأثیرگذار بر کیفیت آبخوان، گامی حیاتی در مدیریت منابع آب و شناسایی کانون‌های آلودگی است. در این پژوهش، از معیار «اهمیت ویژگی» (Feature Importance) مبتنی بر شاخص جینی در الگوریتم XGBoost استفاده شد تا وزن نسبی هر یک از متغیرهای محیطی و سنجش از دوری در پیش‌بینی غلظت آلاینده‌ها مشخص گردد. نتایج حاصل از این تحلیل برای چهار پارامتر کیفی در شکل ۵ نمایش داده شده است.



شکل ۵. نمودار میله‌ای اهمیت ویژگی‌ها (Feature Importance) برای پیش‌بینی پارامترهای کیفی.

همان‌طور که در نمودارهای اهمیت ویژگی (شکل ۵) مشهود است، متغیر ارتفاع (Elevation) در هر چهار مدل (EC , TDS , NO_3^- , Cl) به عنوان مهم‌ترین عامل تعیین‌کننده شناخته شده است (با امتیاز اهمیت نسبی بین ۰٫۱۷ تا ۰٫۲۴). این یافته از نظر هیدروژئولوژیکی کاملاً توجیه‌پذیر است؛ زیرا گرادیان هیدرولیکی و جهت جریان آب زیرزمینی تابعی از توپوگرافی سطح و سنگ بستر است. آبخوان دشت قائم‌شهر، یک سیستم جریانی است که در آن آب از ارتفاعات جنوبی (مناطق تغذیه با کیفیت آب بهتر) به سمت دشت و نواحی شمالی (مناطق تخلیه با تجمع املاح بیشتر) حرکت می‌کند. بنابراین، ارتفاع به عنوان یک متغیر جایگزین (Proxy) قدرتمند برای موقعیت مکانی در سیستم جریان عمل کرده و همبستگی قوی با میزان املاح محلول دارد. جالب توجه است که برخلاف ارتفاع، متغیر شیب (Slope) در تمامی مدل‌ها کمترین اهمیت را داشته است (امتیاز حدود ۰٫۰۲)، که نشان می‌دهد تغییرات محلی شیب در مقایسه با تراز مطلق ارتفاع، نقش کمتری در توزیع آلاینده‌ها در مقیاس دشت دارد.

برای پارامترهای مرتبط با شوری، پس از ارتفاع، متغیر منابع آلاینده و دمای سطح زمین (LST) به ترتیب در رتبه‌های دوم و سوم اهمیت قرار دارند. اهمیت بالای متغیر آلاینده (با امتیاز حدود ۰٫۱۴۶) نشان‌دهنده تأثیر مستقیم فعالیت‌های انسانی و خروجی‌های صنعتی/شهری متمرکز بر افزایش بار املاح آبخوان است. همچنین نقش LST بیانگر فرآیندهای تبخیر و تعرق است؛ در مناطق با دمای سطح بالاتر و پوشش گیاهی کمتر، نرخ تبخیر افزایش یافته و منجر به تغلیظ نمک‌ها در ناحیه غیرواشباع و نهایتاً آب زیرزمینی می‌شود.

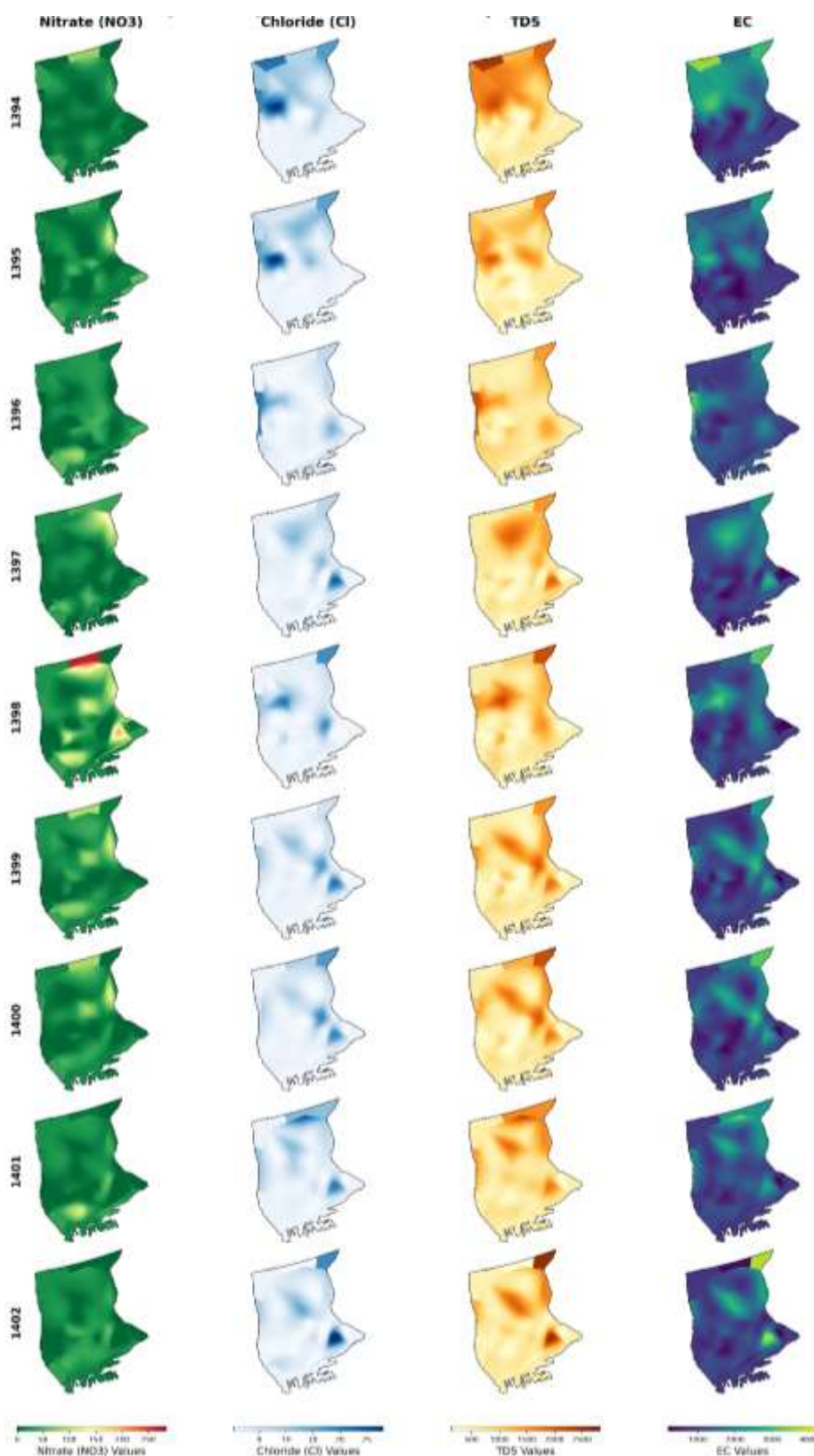
در خصوص نیترات و کلراید، الگوی اهمیت ویژگی‌ها تغییر معناداری به سمت شاخص‌های شهری پیدا می‌کند. برای نیترات (شکل ۵-ج)، شاخص ساخت‌وساز (NDBI) با اهمیت ۰٫۱۴۵ به عنوان دومین عامل مهم شناسایی شده است. این امر تأیید می‌کند که آلودگی نیترات در این منطقه منشأ عمدتاً انسانی (Anthropogenic) دارد و با تراکم مناطق مسکونی و نفوذ فاضلاب‌های شهری همبستگی مستقیم دارد. همچنین متغیر بارش (Precipitation) در مدل نیترات اهمیت بالایی (۰٫۱۱۲) کسب کرده است که مکانیزم آشیوبی (Leaching) کودهای شیمیایی کشاورزی و انتقال آن‌ها به سفره‌های آب زیرزمینی را در فصول بارانی توجیه می‌کند. برای کلراید نیز ترکیبی از عوامل طبیعی (LST) و انسانی (NDBI) و فاصله از مناطق مسکونی) به عنوان محرک‌های اصلی عمل می‌کنند.

تحلیل تغییرات مکانی-زمانی کیفیت آب (۱۳۹۴-۱۴۰۲)

به منظور درک عمیق‌تر پویایی آبخوان و بررسی روند تغییرات آلاینده‌ها در طول زمان، نقشه‌های پهنه‌بندی شده برای چهار پارامتر کیفی



(Cl, TDS, EC, NO_3^-) در بازه زمانی ۹ ساله (۱۳۹۴ تا ۱۴۰۲) تولید و در قالب یک ماتریس مکانی-زمانی در شکل ۶ ارائه شده است. در این ماتریس، ستون‌ها نمایانگر پارامترهای کیفی و سطرها نمایانگر سال‌های هیدرولوژیکی هستند.



شکل ۶. ماتریس تغییرات مکانی-زمانی پارامترهای کیفی آب زیرزمینی (Cl , TDS, EC, NO_3^-) در دشت قائم‌شهر طی دوره آماری ۹ ساله (۱۳۹۴-۱۴۰۲). رنگ‌های گرم‌تر نشان‌دهنده غلظت بالاتر آلاینده‌ها هستند.

با بررسی ستون های مختلف، مشاهده می شود که اگرچه "شدت" آلودگی در سال های مختلف تغییر می کند، اما "مکان" جغرافیایی مناطق بحرانی تقریباً ثابت است. شوری (TDS و EC): بیشترین مقادیر هدایت الکتریکی و کل جامدات محلول همواره در نواحی خاصی (عمدتاً شمال و شمال شرق دشت) متمرکز هستند. این امر تأییدکننده نتایج تحلیل SHAP است که نشان می داد توپوگرافی (ارتفاع پایین) و ساختار زمین شناسی، عوامل اصلی کنترل کننده شوری هستند و این عوامل در طول زمان ثابت اند. نیتрат برخلاف شوری که الگوی پیوسته تری دارد، آلودگی نیترات به صورت لکه ای (Patchy) ظاهر شده است. این لکه ها که نشان دهنده منابع آلاینده نقطه ای (مانند نفوذ فاضلاب شهری یا مصرف متمرکز کود در اراضی کشاورزی) هستند، در برخی سال ها (مانند ۱۳۹۸ و ۱۴۰۰) گسترش یافته و در برخی سال ها محدودتر شده اند.

تغییرات رنگ در ستون ها از بالا به پایین (گذر زمان) نشان دهنده نوسانات غلظت است. نکته قابل توجه این است که روند تخریب کیفیت آب همواره خطی و صعودی نیست. برای مثال، در ستون نیترات، مشاهده می شود که در سال هایی که احتمالاً بارندگی شدیدتر بوده یا فعالیتهای کشاورزی تغییر کرده (مانند سال های میانی بازه)، لکه های قرمز (غلظت بالا) تشدید شده اند. این مشاهده با تحلیل مدل مبنی بر "رفتار دوگانه بارش" (آبشویی نیترات به سمت آبخوان در پی بارش های شدید) همخوانی دارد. در مقابل، پارامترهای Cl و TDS تغییرات آرام تری دارند که ناشی از طبیعت کندتر واکنش آبخوان به فرآیندهای شوری زایی است.

مقایسه ردیف ها نشان می دهد که سال های انتهایی (۱۴۰۱ و ۱۴۰۲) در برخی مناطق شاهد افزایش نسبی مقادیر EC و TDS هستیم که می تواند هشدار برای پیشروی جبهه شور یا کاهش تغذیه آبخوان به دلیل خشکسالی های اخیر باشد. به طور کلی، این نقشه های ماتریسی ابزاری قدرتمند برای مدیریت منابع آب فراهم می کنند، چرا که به تصمیم گیران اجازه می دهند تا به جای تمرکز بر کل دشت، اقدامات اصلاحی خود را دقیقاً بر روی "نواحی جغرافیایی پایدار آلودگی" متمرکز کنند.

تحلیل معناداری آماری خروجی های مدل

به منظور اطمینان از صحت عملکرد الگوریتم در تفکیک زون های کیفی و بررسی اینکه آیا تفاوت غلظت آلاینده ها در کلاس های ریسک پیش بینی شده (کم، متوسط و زیاد) از نظر آماری معنادار است یا خیر، از آزمون ناپارامتری کروسکال-والیس (Kruskal-Wallis) استفاده گردید. همان طور که در جدول X مشاهده می شود، نتایج این آزمون حاکی از آن است که برای تمامی پارامترهای هیدروژئوشیمیایی (NO_3^- ، Cl^- ، TDS و EC)، تفاوت میانگین رتبه ها در سطح اطمینان فراتر از ۹۹ درصد ($P < 0.001$) کاملاً معنادار است. بیشترین آماره آزمون (H) متعلق به نیترات ($H = 672.04$) و شاخص های شوری ($H = 672.00$) بود که نشان دهنده قدرت تفکیک بسیار بالای مدل در جداسازی پهنه های بحرانی از مناطق ایمن است. همچنین نتایج حاصل از آزمون تعقیبی دان (Dunn's post-hoc test) با تصحیح بونفرونی تأیید نمود که اختلاف غلظت بین تمامی جفت کلاس ها (به ویژه میان کلاس های با ریسک کم و زیاد) برای هر چهار پارامتر دارای بالاترین سطح تمایز آماری می باشد. این یافته ها ثابت می کند که پهنه بندی های ارائه شده توسط مدل XGBoost، بازتابی از تفاوت های واقعی و فیزیکی در ساختار داده های کیفی آبخوان است و ناشی از خطاهای تصادفی نیست.

جدول ۳. نتایج آزمون آماری کروسکال-والیس برای بررسی معناداری کلاس های ریسک پیش بینی شده

پارامتر کیفی	آماره آزمون کروسکال-والیس (H)	مقدار پی-بی (P-value)	نتیجه آزمون تعقیبی (تمایز دو به دو)	وضعیت معناداری
نیترات (NO_3^-)	۶۷۲/۰۴	$P < 0.001$	معنادار در تمام جفت کلاس ها	کاملاً معنادار
کلراید (Cl^-)	۶۶۵/۴۶	$P < 0.001$	معنادار در تمام جفت کلاس ها	کاملاً معنادار
هدایت الکتریکی (EC)	۶۷۲/۰۰	$P < 0.001$	معنادار در تمام جفت کلاس ها	کاملاً معنادار
کل جامدات محلول (TDS)	۶۷۲/۰۰	$P < 0.001$	معنادار در تمام جفت کلاس ها	کاملاً معنادار

بحث

تفسیر الگوهای مکانی ریسک آلودگی آب زیرزمینی

نتایج حاصل از پهنه بندی ریسک کیفی آب زیرزمینی در دشت مازندران نشان داد که توزیع مکانی آلاینده ها یکنواخت نبوده و الگوهای مشخصی را در ارتباط با شرایط طبیعی و فعالیتهای انسان زاد دنبال می کند. بیشترین سطوح ریسک برای شاخص های شوری شامل EC

TDS عمدتاً در نواحی شمالی و شمال شرقی دشت مشاهده شد. پایداری مکانی این نواحی در طول دوره مطالعه بیانگر آن است که عوامل ساختاری نظیر توپوگرافی، جهت جریان آب زیرزمینی و موقعیت هیدرولیکی آبخوان نقش تعیین کننده‌ای در تجمع املاح محلول ایفا می‌کنند.

برتری عامل ارتفاع به عنوان مهم‌ترین متغیر کنترل کننده شوری، نشان دهنده انتقال تدریجی املاح به سمت نواحی پایین دست و مناطق تخلیه آبخوان است. این الگو با مدل مفهومی جریان آب زیرزمینی در دشت‌های آبرفتی همخوانی دارد که طبق آن، آب‌های نفوذی از ارتفاعات جنوبی به سمت نواحی کم ارتفاع شمالی حرکت کرده و در مسیر خود املاح محلول را تغلیظ می‌کنند. یافته‌های این پژوهش با نتایج مطالعه Pashaeifar و همکاران (۲۰۲۱) در دشت‌های آبرفتی شمال ایران هم‌راستا است که نشان دادند پارامترهای توپوگرافی، به‌ویژه ارتفاع و موقعیت هیدرولیکی، نقش غالبی در کنترل الگوی مکانی کیفیت آب زیرزمینی و تجمع آلاینده‌ها در نواحی پایین دست آبخوان دارند. در مقابل، توزیع مکانی نیترات الگوی متفاوتی را نشان داد و به صورت لکه‌ای و ناپیوسته در سطح دشت گسترش یافته است. این رفتار ناهمگن بیانگر غلبه منابع آلاینده نقطه‌ای و خطی نظیر نواحی شهری، صنعتی و اراضی کشاورزی بر رفتار این آلاینده است. تمرکز لکه‌های پرنسبک نیترات در مجاورت مناطق با NDBI بالا، تأییدی بر نقش نفوذ فاضلاب‌های شهری، نشت شبکه‌های جمع‌آوری فاضلاب و رواناب‌های آلوده در افزایش غلظت نیترات آب زیرزمینی است که با یافته‌های Fukushi و Kuroda (۲۰۰۸) در محیط‌های شهری آسیا مطابقت دارد.

تحلیل تغییرات زمانی و نقش عوامل اقلیمی

بررسی تغییرات زمانی پارامترهای کیفی آب زیرزمینی طی دوره ۱۳۹۴ تا ۱۴۰۲ نشان داد که شاخص‌های شوری به‌ویژه EC و TDS در سال‌های اخیر، به‌خصوص ۱۴۰۱ و ۱۴۰۲، روند افزایشی قابل توجهی داشته‌اند. این افزایش همزمان با کاهش نسبی بارش و افزایش دمای سطح زمین (LST) بوده است که می‌تواند منجر به کاهش تغذیه مؤثر آبخوان و افزایش فرآیندهای تبخیر و تغلیظ املاح شود. نقش LST به عنوان یکی از متغیرهای مهم مدل برای شوری، بیانگر اثرات غیرمستقیم تغییرات اقلیمی بر کیفیت آب زیرزمینی است. افزایش دمای سطح زمین، علاوه بر تشدید تبخیر، می‌تواند سرعت واکنش‌های ژئوشیمیایی را نیز افزایش داده و به آزادسازی بیشتر یون‌ها در محیط آبخوان منجر شود. این یافته با گزارش‌های IPCC و مطالعات منطقه‌ای در خصوص تأثیر گرمایش بر منابع آب زیرزمینی همخوانی دارد.

رفتار نیترات در بعد زمانی متفاوت بوده و نوسانات بیشتری را نشان می‌دهد. کاهش مقطعی غلظت نیترات در سال‌های پربارش می‌تواند ناشی از رقیق‌سازی آبخوان باشد، در حالی که افزایش آن پس از دوره‌های بارش شدید احتمالاً به دلیل شست‌وشوی کودهای شیمیایی و انتقال آن‌ها به زون اشباع است. چنین رفتار دوگانه‌ای بیانگر پیچیدگی واکنش نیترات نسبت به تغییرات اقلیمی و فرآیندهای هیدرولوژیکی است؛ به گونه‌ای که بارش می‌تواند در برخی دوره‌ها موجب رقیق‌سازی آبخوان و کاهش غلظت نیترات شده و در عین حال، در دوره‌های بارش شدید، از طریق آیشویی کودهای شیمیایی و انتقال آن‌ها به زون اشباع، سبب افزایش غلظت نیترات در آب زیرزمینی شود که با یافته‌های Elzain و همکاران (۲۰۲۲) هم‌راستا است.

ارزیابی عملکرد الگوریتم XGBoost در مدل‌سازی ریسک کیفی

نتایج ارزیابی عملکرد مدل XGBoost نشان داد که این الگوریتم توانایی مناسبی در مدل‌سازی ریسک کیفی آب زیرزمینی دارد. مقادیر ضریب تعیین (R^2) در بازه ۰٫۴۹ تا ۰٫۴۹ و خطای RMSE نسبتاً پایین برای پارامترهای شوری، بیانگر دقت قابل قبول مدل در بازتولید الگوهای مشاهده شده است. عملکرد بهتر مدل برای EC و TDS نسبت به نیترات، نشان دهنده وابستگی بیشتر شاخص‌های شوری به متغیرهای پیوسته و قابل استخراج از داده‌های مکانی و سنجش از دور است.

در مقابل، دقت کمتر مدل در پیش‌بینی نیترات (RMSE حدود ۱۱٫۹۳ mg/L) را می‌توان به ماهیت انسان‌زاد، ناهمگن و تا حدی تصادفی منابع این آلاینده نسبت داد. این نتیجه با یافته‌های Elzain و همکاران (۲۰۲۲) هم‌راستا است که نشان دادند حتی پیشرفته‌ترین الگوریتم‌های یادگیری ماشین نیز در مدل‌سازی نیترات با عدم قطعیت بیشتری مواجه‌اند.

مقایسه نتایج این پژوهش با مطالعات Chen و Guestrin (۲۰۱۶) و Naghibi و همکاران (۲۰۱۷) نشان می‌دهد که XGBoost نسبت به بسیاری از روش‌های کلاسیک و حتی برخی الگوریتم‌های یادگیری ماشین منفرد، عملکرد پایدارتر و مقاوم‌تری در برابر هم‌خطی متغیرها و نویز داده‌ها دارد. این ویژگی به‌ویژه در محیط‌های پیچیده‌ای مانند دشت مازندران که همزمان تحت تأثیر عوامل طبیعی و انسانی

است، اهمیت بالایی دارد.

تحلیل اهمیت متغیرها و تفسیر فیزیکی نتایج SHAP

تحلیل اهمیت ویژگی ها و مقادیر SHAP نشان داد که ارتفاع (Elevation) مهم ترین متغیر پیش بین برای تمامی پارامترهای کیفی آب زیرزمینی است. این یافته تأیید می کند که گرادیان هیدرولیکی و موقعیت مکانی چاه ها نسبت به مسیر جریان آب زیرزمینی، نقش اساسی در تعیین سطح ریسک آلودگی دارند. افزایش احتمال ریسک در ارتفاعات پایین تر، نشان دهنده تجمع تدریجی آلاینده ها در نواحی تخلیه آبخوان است.

برای پارامترهای شوری، پس از ارتفاع، متغیرهای LST و فاصله از منابع آلاینده بیشترین سهم را در پیش بینی ریسک داشته اند. این امر بیانگر تأثیر همزمان فرآیندهای اقلیمی و فعالیت های صنعتی-شهری بر افزایش شوری آب زیرزمینی است. در مقابل، برای نیترات، شاخص NDBI و بارندگی از مهم ترین متغیرها بودند که نقش مستقیم توسعه شهری و فرآیندهای آبشویی را برجسته می سازند. استفاده از SHAP در این پژوهش امکان تفسیرپذیری مدل را فراهم کرده و نشان داده است که چگونه ترکیب متغیرهای طبیعی و انسان زاد می تواند به صورت غیرخطی بر ریسک آلودگی اثر بگذارد. این رویکرد با پیشنهاد Lundberg و همکاران (۲۰۱۷) مبنی بر لزوم شفافیت در مدل های یادگیری ماشین محیط زیستی همخوانی دارد.

پیامدهای مدیریتی و کاربردی نتایج

ترکیب XGBoost با GIS و داده های سنجش از دور در این پژوهش نشان داد که این رویکرد می تواند به عنوان ابزاری کارآمد و کم هزینه برای پایش ریسک کیفی آب زیرزمینی مورد استفاده قرار گیرد. با این وجود، باید تأکید کرد که عملکرد بهینه این چارچوب وابستگی زیادی به شرایط هیدروژئولوژیکی بومی دارد؛ لذا تعمیم نتایج این مدل عمدتاً برای دشت های ساحلی با شرایط مشابه (سطح ایستابی بالا و اقلیم معتدل) معتبر است و کاربرد آن در دشت های خشک، نیمه خشک و آبخوان های عمیق، نیازمند واسنجی مجدد (Recalibration) مدل و بررسی محدودیت های اقلیمی خواهد بود.

یافته های این پژوهش، فراتر از ارزیابی های تئوریک، می تواند به عنوان یک نقشه راه عملیاتی برای مدیران آب و برنامه ریزان شهری دشت مازندران مورد استفاده قرار گیرد. با توجه به خروجی های مدل XGBoost و نقشه های پهنه بندی ریسک، اقدامات عملیاتی زیر پیشنهاد می شود:

۱. مدیریت منابع شوری (EC و TDS) بر اساس زون بندی توپوگرافی: با توجه به اهمیت غالب متغیر ارتفاع، مشخص شد که زون های شمالی و شمال شرقی دشت (مناطق با کمترین ارتفاع) به عنوان زون های تخلیه، بالاترین پتانسیل انباشت شوری را دارند. اقدام مدیریتی مشخص برای این مناطق، «ممنوعیت صدور مجوز حفر چاه های بهره برداری صنعتی با حجم پمپاژ بالا» است تا از افت بیشتر سطح آب و پیشروی جبهه آب شور دریا (Seawater Intrusion) جلوگیری شود.

۲. کنترل بار نیترات مبتنی بر کاربری اراضی: ارتباط قوی میان شاخص ساخت و ساز (NDBI) و بارندگی با افزایش نیترات، نشان دهنده نشت فاضلاب و شستشوی کودهای کشاورزی است. اقدام عملیاتی برای تصمیم گیران، «اجبار شهرک های صنعتی و مجتمع های مسکونی در شعاع ۲ کیلومتری از پهنه های پرخطر نیترات، به احداث سیستم های سپتیک ایزوله و تصفیه خانه های غیرمتمرکز» است. همچنین جهاد کشاورزی باید در مناطق هم جوار با این پهنه ها، الگوی «کوددهی زمان بندی شده (پرهیز از کوددهی در فصول پربارش)» را به کشاورزان دیکته کند.

۳. بهینه سازی شبکه پایش: نقشه های ماتریسی (شکل ۶) نشان داد که مناطق بحرانی از نظر شوری دارای پایداری مکانی هستند. اقدام عملیاتی در این بخش، «بازطراحی شبکه پایش شرکت سهامی آب منطقه ای» است؛ به گونه ای که چاه های پایش در زون های کم خطر (ارتفاعات جنوبی) کاهش یافته و بودجه تخصیص یافته صرف احداث چاه های پایش برخط (Real-time Sensors) در زون های پرخطر شمال و مجاور کانون های NDBI بالا گردد.

نتیجه گیری

پژوهش حاضر با هدف ارتقای رویکردهای ارزیابی ریسک کیفی آب زیرزمینی در مناطق تحت فشار فعالیت های انسانی، چارچوبی یکپارچه مبتنی بر الگوریتم یادگیری ماشین XGBoost و سیستم اطلاعات جغرافیایی (GIS) را برای دشت مازندران ارائه نمود. تمرکز اصلی این



مطالعه نه صرفاً بر پیش‌بینی آلودگی، بلکه بر درک سازوکارهای کنترل‌کننده الگوهای مکانی و زمانی کیفیت آب زیرزمینی و افزایش قابلیت تفسیرپذیری مدل‌های داده‌محور در مطالعات هیدروژئولوژیکی بود.

نتایج کلی پژوهش نشان می‌دهد که ادغام داده‌های کیفی آب با متغیرهای توپوگرافی، اقلیمی، کاربری اراضی و شاخص‌های سنجش از دور، امکان بازنمایی واقع‌بینانه‌تری از پیچیدگی فرآیندهای حاکم بر آلودگی آبخوان‌های ساحلی را فراهم می‌سازد. به‌ویژه در دشت مازندران، شرایط خاص هیدروژئولوژیکی نظیر سطح ایستابی کم، بارش نسبتاً بالا و تمرکز فعالیت‌های شهری و صنعتی، موجب شکل‌گیری الگوهای ناهمگن و پویا در توزیع آلاینده‌ها شده است که مدل‌های کلاسیک مبتنی بر وزن‌دهی ثابت، توانایی کافی برای شناسایی آن‌ها را ندارند.

به‌کارگیری الگوریتم XGBoost به دلیل توانایی آن در مدل‌سازی روابط غیرخطی، مدیریت هم‌خطی میان متغیرها و کنترل بیش‌برازش، نشان داد که مدل‌های یادگیری تقویتی مبتنی بر درخت تصمیم می‌توانند جایگزین مناسبی برای رویکردهای شاخص‌محور سنتی در ارزیابی ریسک کیفی آب زیرزمینی باشند. افزون بر این، استفاده از روش‌های هوش مصنوعی تفسیرپذیر، به‌ویژه تحلیل مقادیر SHAP، امکان فراتر رفتن از پیش‌بینی صرف و دستیابی به درک مفهومی از نقش نسبی عوامل طبیعی و انسان‌ساخت در کنترل ریسک آلودگی را فراهم ساخت. این ویژگی، مدل ارائه‌شده را به ابزاری قابل اعتمادتر برای پشتیبانی تصمیم‌گیری مدیریتی تبدیل می‌کند.

از منظر مدیریتی، نقشه‌های پهنه‌بندی ریسک کیفی استخراج‌شده در این پژوهش می‌توانند به‌عنوان مبنایی علمی برای اولویت‌بندی اقدامات حفاظتی، کنترل منابع آلاینده شهری و صنعتی، و بهینه‌سازی مکان‌یابی فعالیت‌های توسعه‌ای در دشت مازندران مورد استفاده قرار گیرند. تمرکز بر مناطق با ریسک بالا می‌تواند به کاهش هزینه‌های پایش، افزایش اثربخشی مداخلات مدیریتی و جلوگیری از تشدید آلودگی‌های بلندمدت و بعضاً غیرقابل بازگشت آب زیرزمینی منجر شود.

در نهایت، اگرچه چارچوب پیشنهادی این مطالعه پتانسیل بالایی برای مدل‌سازی مکانی ریسک دارد، اما تعمیم آن به تمامی دشت‌ها با محدودیت‌هایی روبه‌روست. این مدل به‌طور خاص متأثر از ویژگی‌های دشت مازندران (از جمله عمق کم برخورد به آب و تأثیرپذیری سریع از کاربری‌های انسان‌ساخت) توسعه یافته است. بنابراین، کاربرد این رویکرد در سایر حوضه‌های آبریز مستلزم بومی‌سازی متغیرهای ورودی، توجه به رژیم تغذیه آبخوان و ارزیابی مجدد آستانه‌های آلودگی است. با در نظر گرفتن این محدودیت‌ها، رویکرد XGBoost-GIS گامی مؤثر در جهت مدیریت پایدار منابع آب زیرزمینی در مناطق حساس و در حال توسعه ساحلی محسوب می‌شود.

اصول اخلاقی

حامی مالی

این پژوهش با حمایت مالی و معنوی معاونت پژوهشی دانشکده مهندسی، دانشگاه اردکان انجام شده است.

مشارکت نویسندگان

تمامی نویسندگان به‌طور مساوی در مفهوم‌سازی پژوهش، طراحی روش تحقیق، انجام پژوهش، گردآوری داده‌ها، تجزیه و تحلیل و تفسیر نتایج، نگارش پیش‌نویس اولیه و بازبینی و نهایی‌سازی مقاله مشارکت داشته‌اند.

اعلامیه هوش مصنوعی مولد و فناوری‌های مبتنی بر هوش مصنوعی در فرایند نگارش

در طول آماده‌سازی این اثر استفاده کرده‌اند. سپس از هوش مصنوعی به‌منظور بهبود شیوایی زبانی و نگارش علمی متن نویسندگان محتوای مقاله بررسی و در صورت نیاز ویرایش شده و مسئولیت کامل محتوای منتشرشده بر عهده نویسندگان است.

بیانیه دسترسی به داده‌ها

داده‌های پژوهش حاضر بنا به درخواست از نویسندگان قابل دسترسی است.

سپاسگزاری

نویسندگان از تمامی افرادی که به‌صورت مستقیم یا غیرمستقیم در انجام این پژوهش همکاری داشته‌اند تشکر و قدردانی می‌نمایند. همچنین از داوران محترم مجله به‌خاطر ارائه دیدگاه‌ها و پیشنهادهای علمی و سازنده سپاسگزاری می‌شود.

ملاحظات اخلاقی

نویسندگان اصول اخلاق پژوهش را در تمامی مراحل انجام و انتشار این تحقیق رعایت نموده و از هرگونه جعل داده، دست کاری، سرقت علمی و رفتار غیراخلاقی اجتناب کرده اند.

تعارض منافع

بنا بر اظهار نویسندگان، این مقاله فاقد هرگونه تعارض منافع است.

REFERENCES

- Abhiram, K., et al. (2024). Groundwater quality prediction and risk assessment in Kerala, India: A machine-learning approach. *Science of The Total Environment*.
- Abu El-Magd, S., Masoud, A. M., Brink, H. G., et al. (2025). Groundwater vulnerability under climate change: A machine learning framework. *Earth Systems and Environment*, 1-15.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
- Elzain, H. E., Chung, S. Y., Senapathi, V., Sekar, S., Lee, S. Y., Roy, P. D. & Sabarathinam, C. (2022). Comparative study of machine learning models for evaluating groundwater vulnerability to nitrate contamination. *Ecotoxicology and Environmental Safety*, 229, 113061.
- Ester, M., Kriegel, H. P., & Xu, X. J. G. A. (2022). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (vol, pg 785, 2016). *GEOGRAPHICAL ANALYSIS*.
- Gleeson, T., VanderSteen, J., Sophocleous, M. A., Taniguchi, M., Alley, W. M., Allen, D. M., & Zhou, Y. (2010). Groundwater sustainability strategies. *Nature geoscience*, 3(6), 378-379.
- Gogu, R. C., & Dassargues, A. (2000). Sensitivity analysis for the EPIK method of vulnerability assessment in a small karstic aquifer, southern Belgium. *Hydrogeology Journal*, 8(3), 337-345.
- Kuroda, K., & Fukushima, T. (2008). Groundwater contamination in urban areas. In *Groundwater management in Asian cities: technology and policy for sustainability* (pp. 125-149). Tokyo: Springer Japan.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), 1425-1432.
- Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, 31(9), 2761-2775.
- Neshat, A., Pradhan, B., Pirasteh, S., & Shafri, H. Z. M. (2014). Estimating groundwater vulnerability to pollution using a modified DRASTIC model in the Kerman agricultural area, Iran. *Environmental earth sciences*, 71(7), 3119-3131.
- Pashaeifar, M., Dehghanzadeh, R., Ramazani, M. E., Rafieyan, O., & Nejaei, A. (2021). Spatial and temporal assessment of groundwater quality and hydrogeochemical processes in Urmia Lake Basin, Iran. *Water Supply*, 21(8), 4328-4342.
- Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience & Remote Sensing*, 57(6), 735-752.
- Raheja, H., Goel, A., & Pal, M. (2024). A novel approach for prediction of groundwater quality using gradient boosting-based algorithms. *ISH Journal of Hydraulic Engineering*, 30(5), 1-12.
- Rammohan, B., Partheeban, P., Ranganathan, R., & Balaraman, S. (2024). Groundwater Quality Prediction and Analysis Using Machine Learning Models and Geospatial Technology. *Sustainability*, 16(22), 9848.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351, 309.
- Singh, L. K., Jha, M. K., & Chowdary, V. M. (2018). Assessing the accuracy of GIS-based multi-criteria decision analysis approaches for mapping groundwater potential. *Ecological Indicators*, 91, 24-37.
- Vu, P. T., et al. (2025). Evaluating groundwater quality in an arsenic-contaminated aquifer in the Red River Delta using machine learning: a case study in Van Phuc, Hanoi, Vietnam. *Environmental Science: Advances*.
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International journal of remote sensing*, 24(3), 583-594.