

Modeling soil water repellency in loess soils of northern Iran using machine learning

Ali Mohammadian Behbahani^{1*}, Kohzad Heidary², Mohsen Hosseinalizadeh³

¹ Associate Professor, Department of Arid Land Management, Gorgan University of Agricultural Sciences and Natural Resources, Golestan, Iran

² Assistant Research Professor, Soil and Water Conservation Research Department, Khuzestan Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Ahvaz, Iran

³ Associate Professor, Department of Desert Zone Management Faculty of Rangeland and Watershed Management, Gorgan University of Agricultural Sciences and Natural Sciences, Golestan, Iran

Extended Abstract

Introduction

A major hydrological and physical event affecting surface runoff, erosion, and water infiltration is soil water repellency (SWR). Hydrophobic soils reject wetting, therefore causing water droplets to linger on the surface instead of penetrating the soil profile. Particularly in sloping environments and arid ecosystems, this phenomenon causes more overland flow, less water retention, and higher sensitivity to soil loss. SWR is progressively seen in northern Iran, especially in the loess-derived soils of Golestan and Mazandaran provinces, as a result of a mix of climatic conditions and changes in land use. Fine silty to silty-clay textures characterize loess soils in these areas; together with environmental stressors, including drought cycles, agricultural development, and deforestation, they aid in the formation and magnification of SWR. Central to the water repellency are organic compounds, especially hydrophobic plant-derived substances like waxes and lignins. Emphasizing the significance of soil chemical composition, many studies have shown that soil organic carbon is strongly positively linked to SWR intensity. Variations in clay content, pH, and electrical conductivity (EC) can also affect SWR patterns. Although SWR is very important in soil degradation processes, little research has been done employing sophisticated data-driven techniques to forecast its spatial variability. Machine learning (ML) algorithms like Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) provide strong means for modeling complicated soil behavior. Using these algorithms, the current study attempts to forecast SWR in loess soils based on a thorough set of physicochemical parameters, thus helping to justify improved soil management and erosion control measures.

Materials and Methods

Northern Iran served as the location for the study, which looked at specific loess terrains in Golestan and Mazandaran provinces. From many sites, including Gorgan, Maraveh Tappeh, Neka, Sari, and Amol, 45 surface soil samples (depth 0–10 cm) were gathered. While minimizing confounding effects, sampling places were chosen to record changes in topography, land use, and vegetation. Key soil physicochemical characteristics assessed were organic carbon (OC), organic matter (OM), electrical conductivity (EC), pH, mean weight diameter (MWD) of soil aggregates, and particle size distribution (sand, silt, clay). With infiltration time recorded up to 3000 seconds, WDPT tests involved dropping 50 μ L distilled water droplets on air-dried soil surfaces at room temperature. WDPT values were used as the target variable in model development. Procedures used in laboratories complied with the same criteria used in earlier research. R software was used for data preprocessing. Outlier detection based on interquartile range (IQR), Z-score normalization of numerical variables, and multicollinearity analysis using the Variance Inflation Factor (VIF) were included in this. Categorical variables like soil texture classes were converted to dummy variables using one-hot encoding. Three machine learning approaches—Decision Tree (CART approach), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—were applied to the dataset, which was randomly separated into 70% training and 30% testing parts. Models were implemented using R packages rpart, randomForest, and xgboost. Through repeated 10-fold cross-validation, hyperparameter tuning was carried out to enhance prediction accuracy.

Results and Discussion

Initial model performance using default settings revealed limited predictive ability across all algorithms. The Decision Tree (DT) model yielded the weakest results with RMSE = 19.55 and $R^2 = 0.02$, indicating poor capacity to capture the variability in WDPT values. After hyperparameter optimization, both Random Forest (RF) and XGBoost (XGB) showed significant improvements. The RF model achieved RMSE = 15 and $R^2 = 0.42$, while XGB recorded RMSE = 14.7 with the same R^2 , highlighting their comparable predictive power. Feature importance analysis revealed that organic carbon was the most influential predictor of WDPT across all models. Additional influential variables included clay content, sand fraction, EC, OM, and pH, though their relative importance varied by algorithm. In RF, organic matter and sand had high predictive value, whereas in XGB, clay and EC gained prominence. These differences reflect each model's inherent sensitivity to nonlinear interactions. Spatial analysis showed that areas with higher organic carbon content aligned with regions of higher WDPT, confirming the key role of hydrophobic organic compounds in driving soil water repellency. Uncertainty assessment using Bootstrap and Monte Carlo simulations demonstrated that RF was the most stable model, showing the lowest RMSE variability and greater

resilience to noisy input data. Overall, the results confirm that machine learning algorithms, especially RF and XGB, can effectively model and interpret the complex interactions influencing soil water repellency in loess landscapes.

Conclusion

This study demonstrated the applicability of advanced machine learning algorithms for modeling soil water repellency (SWR) in loess-derived soils of northern Iran. Among the three tested models, Random Forest (RF) provided the most reliable and stable predictions, with optimal performance metrics (RMSE = 15, $R^2 = 0.42$) and low sensitivity to data uncertainty. XGBoost (XGB) also yielded competitive results but showed slightly lower stability under uncertain conditions. The Decision Tree (DT) model, while interpretable, lacked sufficient predictive accuracy for complex, nonlinear relationships. The results confirmed that organic carbon is the dominant driver of SWR in the study area, supporting previous findings regarding the hydrophobic nature of plant-derived organic compounds. Other variables, such as clay content, sand fraction, pH, and EC, also played important roles depending on the model structure. Differences in variable importance highlight the benefit of using multiple algorithms to obtain a comprehensive understanding of the underlying mechanisms. Uncertainty analysis showed that RF is less susceptible to overfitting and data noise, making it a more robust choice for environmental modeling. Spatial patterns of WDPT and key soil variables revealed strong regional correlations, suggesting the feasibility of using geospatial ML models for site-specific soil management. Future research should explore hybrid models (e.g., RF-XGB) and deep learning architectures (e.g., CNNs or ActionFormer) to enhance predictive power, particularly in dynamic or post-disturbance soil systems. Moreover, integrating multi-temporal datasets could improve understanding of SWR variability under different environmental and management conditions.

Keywords: Hydrophobic Organic Compounds; Soil Aggregate Stability; Nonlinear Learning; Sensitivity Analysis; Uncertainty analysis.

Article Type: Research Article

Acknowledgement

We would like to express our sincere gratitude to the Gorgan University of Agricultural Sciences and Natural Resources and Khuzestan Agricultural and Natural Resources Research and Education Center (AREEO) for their financial and logistical support during the implementation of this research project.

Conflicts of interest

The authors of this article declare no conflict of interest regarding the authorship or publication of this article.

Data Availability Statement:

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' contribution

Ali Mohammadian Behbahani: Conceptualization, resources, methodology, visualization, formal analysis and investigation, writing – original draft preparation; **Kohzad Heidary:** Supervision, conceptualization, manuscript editing, formal analysis and investigation; **Mohsen Hosseinalizadeh:** Methodology, conceptualization, resources, formal analysis and investigation, manuscript editing.

*Corresponding Author, E-mail: Ali Mohammadian Behbahani, ali.mohammadian@gau.ac.ir

Citation: Mohammadian Behbahani, A., Heidary, K., Mohsen Hosseinalizadeh, M. (2026). Modeling soil water repellency in loess soils of northern Iran using machine learning. *Water and Soil Management and Modeling*, 6(1), 132-148.

doi: 10.22098/mmws.2025.17919.1633

Received: 26 July 2025, Received in revised form: 06 September 2025, Accepted: 06 September 2025, Published online: 21 March 2026

Water and Soil Management and Modeling, Year 2026, Vol. 6, No. 1, pp. 132-148.

Publisher: University of Mohaghegh Ardabili

© Author(s)





مدل سازی آب‌گریزی خاک‌های لسی شمال ایران با الگوریتم‌های یادگیری ماشین

علی محمدیان بهبهانی^{۱*}، کهزاد حیدری^۲، محسن حسینعلی زاده^۳

^۱ دانشیار، گروه مدیریت مناطق بیابانی، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گلستان، ایران.
^۲ استادیار پژوهشی بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات، آموزش کشاورزی و منابع طبیعی خوزستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، اهواز، خوزستان، ایران.
^۳ دانشیار، گروه مدیریت مناطق بیابانی، دانشکده مرتع و آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گلستان، ایران.

چکیده

یکی از ویژگی‌های مهم خاک که بر چگونگی حرکت آب، نفوذ آن و فرسایش سطحی اثر می‌گذارد، خاصیت آب‌گریزی آن است. این پژوهش باهدف بررسی و تخمین میزان آب‌گریزی خاک بر اساس شاخص WDPT و با استفاده از روش‌های یادگیری ماشینی مثل درخت تصمیم، جنگل تصادفی و XGBoost در خاک‌های لسی شمال ایران انجام شد. به این منظور، از ۴۵ موقعیت مختلف نمونه‌برداری خاک انجام گرفت و ویژگی‌های فیزیکی و شیمیایی آن‌ها، شامل میزان کربن آلی، مواد آلی، هدایت الکتریکی، pH، اندازه ذرات خاک، درصد ماسه، رس و سیلت اندازه‌گیری شد. عملکرد مدل‌ها با استفاده از معیارهای آماری RMSE، MAE و R² و هم‌چنین با تحلیل حساسیت و عدم قطعیت سنجیده شد. در ابتدا، عملکرد مدل‌ها چندان مطلوب نبود، اما با تنظیم دقیق پارامترها، دقت پیش‌بینی به شکل چشم‌گیری بهبود یافت. پس از بهینه‌سازی، مدل جنگل تصادفی با RMSE برابر با ۱۵، MAE برابر با ۹۳/۱۱ و R² برابر با ۴۲ بهترین نتیجه را ارائه داد. مدل XGBoost نیز پس از تنظیم پارامترها با RMSE برابر با ۱۴/۷ و R² برابر با ۴۲ در رتبه دوم قرار گرفت. مدل درخت تصمیم همچنان ضعیف‌ترین عملکرد را داشت. تحلیل عدم قطعیت با روش‌های بوت‌استرپ و مونت کارلو نشان داد که مدل جنگل تصادفی کم‌ترین پراکندگی و نوسان RMSE را دارد و از پایداری بیش‌تری برخوردار است. تحلیل حساسیت مدل‌ها نشان داد که کربن آلی مهم‌ترین عامل در پیش‌بینی WDPT در همه مدل‌ها است. سایر عوامل مانند رس، مواد آلی، هدایت الکتریکی و pH نیز در مدل‌های مختلف نقش مهمی ایفا کردند. نتایج این پژوهش نشان می‌دهد که الگوریتم‌های پیشرفته یادگیری ماشین، به ویژه مدل جنگل تصادفی، می‌توانند ابزارهای قوی و قابل اعتمادی برای مدل‌سازی و مدیریت پدیده‌های پیچیده مانند آب‌گریزی خاک در مناطق حساس مانند خاک‌های لسی شمال ایران باشند. یافته‌های این پژوهش می‌تواند به‌عنوان مبنایی برای ارزیابی و مدیریت آب‌گریزی خاک‌های لسی و بهینه‌سازی راهکارهای کاهش فرسایش و بهبود نفوذپذیری در مناطق مشابه به کار رود.

واژه‌های کلیدی: ترکیبات آلی آب‌گریز؛ پایداری خاکدانه؛ یادگیری غیرخطی؛ تحلیل حساسیت؛ تحلیل عدم قطعیت.

نوع مقاله: پژوهشی

*مسئول مکاتبات، پست الکترونیکی: علی محمدیان بهبهانی، ali.mohammadian@gau.ac.ir

استناد: محمدیان بهبهانی، علی، حیدری، کهزاد، حسینعلی زاده، محسن (۱۴۰۵). مدل‌سازی آب‌گریزی خاک‌های لسی شمال ایران با الگوریتم‌های یادگیری ماشین. *مدل‌سازی و مدیریت آب و خاک*، ۶(۱)، ۱۳۲-۱۴۸.

doi: 10.22098/mmws.2025.17919.1633

تاریخ دریافت: ۱۴۰۴/۰۵/۰۴، تاریخ بازنگری: ۱۴۰۴/۰۶/۱۵، تاریخ پذیرش: ۱۴۰۴/۰۶/۱۵، تاریخ انتشار: ۱۴۰۵/۰۱/۰۱

مدل‌سازی و مدیریت آب و خاک، سال ۱۴۰۵، دوره ۶، شماره ۱، صفحه ۱۳۲ تا ۱۴۸

© نویسندگان

ناشر: دانشگاه محقق اردبیلی



۱- مقدمه

آب‌گریزی خاک^۱ زمانی رخ می‌دهد که خاک‌ها در برابر جذب آب مقاومت نشان می‌دهند و قطره‌های آب روی سطح آن‌ها باقی می‌مانند. در واقع این خاک‌ها آب را دفع می‌کنند و به خوبی خیس نمی‌شوند (Mao et al., 2022a; Zhang et al., 2025). این پدیده به‌ویژه در مناطق شیب‌دار و خشک از اهمیت زیادی برخوردار است؛ زیرا با کاهش نرخ نفوذ آب به درون پروفیل خاک، حجم آب ذخیره شده در لایه زیرسطحی کاهش یافته و در نتیجه رواناب سطحی افزایش می‌یابد (Hallett et al., 2011; Wang et al., 2024). وقتی که میزان نفوذ آب به درون خاک کاهش می‌یابد، حجم رواناب سطحی افزایش پیدا می‌کند و این موضوع می‌تواند خطر فرسایش را در منطقه به‌طور قابل‌توجهی افزایش دهد. در بخش‌های شمالی ایران، به‌ویژه استان‌های گلستان، مازندران و گیلان، خاک‌هایی با منشأ لسی به‌شکل گسترده‌ای پراکنده‌اند. این نوع خاک‌ها عمدتاً دارای بافت سیلتی یا رسی-سیلتی هستند (Kavian et al., 2014; Heidary et al., 2018; Heidary, et al., 2018). در مطالعه که بر روی این خاک‌ها انجام دادند نشان دادند که هر چقدر میزان درصد سیلت خاک لسی افزایش پیدا کند، مقدار رواناب بیش‌تر خواهد شد؛ از طرف دیگر، پوشش‌های جنگلی شمال (جنگل‌های هیرکانی) معمولاً خاک‌هایی با ماده آلی بیش‌تر و رطوبت کافی دارند که می‌توانند نفوذپذیری خاک را تا حدود زیادی بهبود دهند. با این حال، در مجموع می‌توان گفت که خشکسالی‌های پی‌درپی و پوشش گیاهی خاص شمال ایران، همراه با ترکیب فیزیکی خاک‌های لسی، موجب شده است که پدیده آب‌گریزی خاک در این منطقه اهمیت بالایی پیدا کند و نیازمند مدل‌سازی و بررسی دقیق باشد. عوامل فیزیکی و شیمیایی مختلفی در شدت آب‌گریزی نقش دارند. پژوهش‌های اخیر نشان می‌دهند که ترکیبات آلی خاک به‌ویژه ترین‌ها^۲، لیگنین و موم‌های گیاهی از مهم‌ترین عامل ایجاد آب‌گریزی هستند (Mao et al., 2022b). به عبارت دیگر، هرچه میزان ماده آلی خاک بیش‌تر باشد (به‌ویژه از نوع مشتقات گیاهی آب‌گریزانند موم‌های گیاهی، اسیدهای چرب بلند زنجیره و ترکیبات فنولی)، احتمال و شدت آب‌گریزی نیز بیش‌تر می‌شود. تاکنون رابطه مستقیم بین کربن آلی خاک و میزان آب‌گریزی در پژوهش‌های متعددی تأیید شده است. برای مثال، Hermansen et al. (2021) ، de Blas et al. (2013) و Blaasbjerg et al. (2022) تأکید

کرده‌اند که کربن آلی خاک عامل مهمی در بروز آب‌گریزی خاک^۳ است.

از طرف دیگر، بافت و ساختمان خاک نیز اهمیت بالایی در شدت آب‌گریزی دارد. معمولاً خاک‌های ماسه‌ای به‌دلیل سطح ویژه کم‌تری که دارند زمانی که در معرض مواد آلی آب‌گریز قرار می‌گیرند، شدت آب‌گریزی بیش‌تری از خود نشان می‌دهند (Hallett et al., 2011; Weber et al., 2021). در مقابل، خاک‌های سیلتی و رسی-سیلتی به‌دلیل سطح ویژه بالاتر، برای رسیدن به حالت آب‌گریزی شدید، معمولاً به تجمع بیش‌تری از ترکیبات آلی نیاز دارند. علاوه بر این، میزان رطوبت نسبی خاک یکی از عوامل کلیدی و تعیین کننده در بروز پدیده آب‌گریزی است. کاهش رطوبت تا مقادیر پایین باعث می‌شود لایه‌های آب‌گریز اطراف ذرات خاک پایدار مانده و مانع نفوذ آب شوند؛ از اینرو در خاک‌های خشک شدت آب‌گریزی به‌طور چشم‌گیری افزایش می‌یابد. مطالعات نشان داده‌اند که با کاهش رطوبت خاک زاویه تماس قطره آب با خاک افزایش یافته و نفوذپذیری به‌طور چشم‌گیری کاهش می‌یابد (Chen et al., 2020). نقش pH در آب‌گریزی خاک اگرچه تا حدودی پیچیده است، اما نتایج برخی پژوهش‌ها حاکی از آن است که pH خنثی تا حدودی قلبایی (به‌ویژه در خاک‌های آهکی) می‌تواند شرایط را برای آب‌گریزی خاک فراهم کند. (Danielsen et al., 2025) در مطالعه‌ای بر روی یک حوزه آبخیز در دانمارک و با استفاده از مدل مسیر^۴ نشان دادند که pH و EC همراه با کربن آلی حدود نصف تغییرات واریانس آب‌گریزی خاک را تبیین می‌کنند.

در مجموع بررسی عوامل ذکر شده نشان می‌دهد که اثرات متقابل این پارامترها بسیار پیچیده بوده و بسته به شرایط اقلیمی هر منطقه متفاوت است. در سال‌های اخیر، روش‌های یادگیری ماشین جایگاه ویژه‌ای در مدل‌سازی پدیده‌های طبیعی از جمله فرآیندهای تشکیل خاک پیدا کرده‌اند. الگوریتم‌هایی مانند درخت تصمیم^۵ و جنگل تصادفی^۶ به‌دلیل توانایی در شناسایی الگوهای غیرخطی و ترکیب متغیرهای ورودی، در مطالعات خاک مورد توجه قرار گرفته‌اند. (Pahlavanrad et al., 2017) با استفاده از الگوریتم درخت تصمیم‌گیری تصادفی در استان گلستان، نقشه‌بندی رقومی خاک را انجام دادند و قابلیت این الگوریتم را در پیش‌بینی ویژگی‌های خاک در نهشته‌های لسی نشان دادند. (Maleki et al., 2022) نیز

⁴ Path analysis

⁵ Decision Tree

⁶ Random Forest

¹ Hydrophobicity

² Terpenes

³ Soil Water Repellency

هدف اصلی این پژوهش بررسی شدت آب‌گریزی بخشی از رسوبات لسی شمال ایران با استفاده از الگوریتم یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی و XGBoost است. در این مطالعه، برخلاف پژوهش‌های پیشین که بیش‌تر بر مدل‌سازی ویژگی‌های عمومی خاک تمرکز داشتند، برای نخستین‌بار از الگوریتم‌های یادگیری ماشین برای پیش‌بینی شدت آب‌گریزی خاک‌های لسی استفاده شده است. همچنین، تحلیل حساسیت و ارزیابی عدم قطعیت مدل‌ها به‌صورت ترکیبی در چارچوب یک مدل جامع به‌کار رفته است که در مطالعات پیشین مغفول مانده بود. یافته‌های این پژوهش می‌تواند به شناسایی مناطق حساس از منظر آب‌گریزی کمک کرده و منتج به ارائه راهکارهایی برای مدیریت بهینه منابع آب و خاک شود.

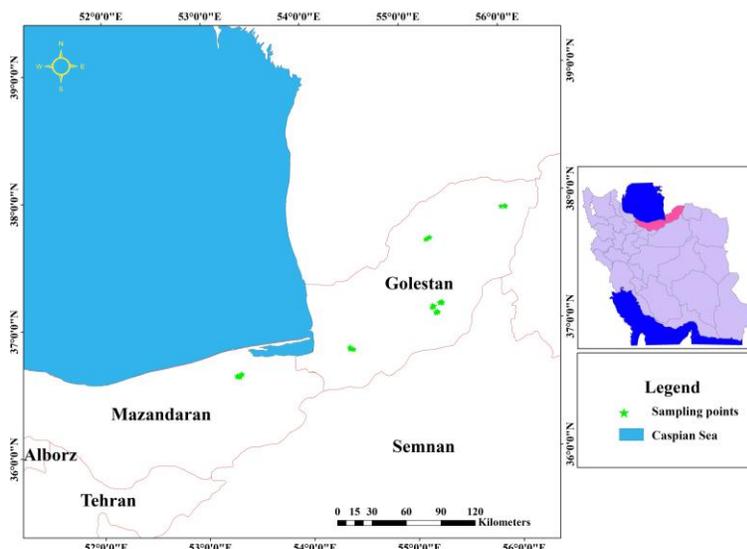
۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

منطقه مورد مطالعه، بخش‌هایی از نواحی شمالی ایران را در بر می‌گیرد که در استان‌های گلستان و مازندران واقع شده‌اند (شکل ۱). نمونه‌برداری از شهرستان‌های گرگان، مراوه‌تپه، نوده، آق‌بند، نکا، ساری و آمل انجام گرفت. این محدوده بخشی از کمربند کوهپایه‌ای و جلگه‌ای جنوب دریای خزر را تشکیل می‌دهد. خاک‌های این منطقه عمدتاً لسی هستند که از رسوبات بادی دوره کواترنر تشکیل شده‌اند (Kariminejad et al., 2022). این رسوبات به‌دلیل لایه‌های متراکم سیلت و رس و نیز میزان بالای کربن آلی، به‌عنوان خاک‌های آسیب‌پذیر به تغییرات کاربری، فرسایش و فرآیندهای هیدرولوژیکی شناخته می‌شوند (Li and Liu, 2024).

اقلیم منطقه مورد مطالعه از غرب به شرق، تنوع قابل‌توجهی دارد. نواحی غربی منطقه (مانند شهرستان‌های گرگان و ساری) دارای اقلیم مرطوب تا نیمه‌مرطوب خزری با میانگین بارندگی سالانه حدود ۸۰۰ تا ۱۲۰۰ میلی‌متر هستند، در حالی‌که بخش‌های شرقی‌تر (مانند مراوه‌تپه و نوده) دارای اقلیم نیمه‌خشک تا خشک با بارندگی کم‌تر از ۵۰۰ میلی‌متر در سال هستند (Heidary et al., 2018; Li and Liu, 2024). میانگین دمای سالانه در سطح منطقه حدود ۱۶ درجه سانتی‌گراد برآورد می‌شود (Emami et al., 2024). از نظر توپوگرافی، منطقه شامل ترکیبی از دامنه‌ها، کوهپایه‌ها و جلگه‌های آبرفتی است که این تنوع توپوگرافی در کنار گوناگونی در کاربری اراضی (مانند مرتع، کشاورزی، باغات و جنگل‌های پراکنده) و پوشش گیاهی، منجر به تفاوت در خصوصیات فیزیکی و شیمیایی خاک‌ها شده است (Maleki et al., 2022).

با مدل جنگل تصادفی و در ۲۵۲ نقطه نمونه‌برداری، نقشه‌بندی کربن آلی خاک را در دو عمق خاک (۰-۲۰ و ۲۰-۴۰ سانتی‌متری) در فلات لسی ایران انجام دادند و کارایی بالای مدل را گزارش کردند. (Kariminejad et al., 2022) نیز با استفاده از مدل‌های یادگیری ماشین، به ویژه جنگل تصادفی، به مدل‌سازی وزن مخصوص خاک در مناطق لسی استان گلستان پرداختند. همچنین Emami et al. (2024) با به‌کارگیری مدل جنگل رگرسیونی چندکی به‌همراه تابع اسپیلاین، نقشه سه‌بعدی اجزای بافت خاک را در استان گلستان تهیه کردند. این مطالعات نشان می‌دهند که مدل‌های یادگیری ماشین، از جمله درخت تصمیم، جنگل تصادفی و XGBoost، نه تنها در طبقه‌بندی داده‌ها، بلکه در پیش‌بینی دقیق متغیرهای کمی مانند نفوذپذیری خاک، ظرفیت نگهداری آب و زمان نفوذ قطره آب (WDPT) نیز موفق عمل کرده‌اند. برای مثال، اخیراً با استفاده از الگوریتم‌های بینایی ماشین مبتنی بر ویدئو و شبکه‌های عصبی پیچشی، زمان نفوذ قطره آب در خاک (به‌عنوان شاخص SWR) به‌صورت خودکار اندازه‌گیری شده است (Wang et al., 2024). با اینکه استفاده از الگوهای یادگیری ماشین در بررسی آب‌گریزی خاک هنوز محدود است، اما تجربه موفق مدل‌هایی مانند جنگل تصادفی و درخت تصمیم در پیش‌بینی ویژگی‌های خاک (نظیر کربن آلی و نفوذپذیری) نشان می‌دهد این الگوریتم‌ها می‌توانند ابزارهای کارآمدی برای مدل‌سازی پدیده‌های پیچیده‌ی خاکی باشند (Tarek et al., 2023). مطالعات اخیر نشان داده‌اند که بهره‌گیری از الگوریتم‌های یادگیری ماشین و داده‌کاوی در مدل‌سازی فرآیند نفوذ آب در خاک می‌تواند دقت پیش‌بینی را به‌طور معناداری افزایش دهد. به‌عنوان نمونه، Mehta et al. (2024) با مقایسه مدل‌های ANN، SVM، GMDH و MARS گزارش کردند که شبکه‌های عصبی مصنوعی بالاترین دقت را در برآورد نرخ نفوذ خاک دارند. همچنین، Sepahvand et al. (2022) با استفاده از مدل‌های یادگیری عمیق و ترکیبی CNN، SVR و GMDH بهینه‌سازی شده با الگوریتم‌های فراابتکاری توانستند عملکرد پیش‌بینی نفوذ تجمعی را بهبود بخشند. افزون بر این، Yang and Mei (2022) با به‌کارگیری شبکه‌های عصبی مبتنی بر معادلات فیزیکی (PINNs) فرآیند نفوذ عمودی آب در خاک را با دقت بالاتری نسبت به روش‌های عددی سنتی شبیه‌سازی کردند. این شواهد نشان می‌دهد که ترکیب داده‌کاوی و مدل‌های یادگیری ماشین می‌تواند ابزاری توانمند برای تحلیل و پیش‌بینی پدیده‌های پیچیده خاکی، از جمله نفوذ و آب‌گریزی، فراهم آورد. در همین راستا،



شکل ۱- موقعیت منطقه مورد مطالعه و نقاط نمونه‌برداری
Figure 1- Study area and soil sampling locations

بهتر روابط میان ویژگی‌های خاک و آب‌گریزی کمک می‌کند. تا حد امکان تلاش شد تا خصوصیات توپوگرافی، کاربری و پوشش گیاهی در مکان‌های نمونه‌برداری یکنواخت در نظر گرفته شود. نمونه‌ها پس از برداشت در دمای اتاق خشک شدند و سپس از الک ۲ میلی‌متری عبور داده شدند. در آخر خصوصیات فیزیکی و شیمیایی کلیدی خاک برای آماده‌سازی داده‌ها برای مدل‌سازی، مطابق با روش‌های ذکر شده در جدول ۱ اندازه‌گیری شدند.

۲-۲- نمونه‌برداری خاک و آنالیزهای آزمایشگاهی

نمونه‌برداری به صورت سطحی (از عمق ۰ تا ۱۰ سانتی‌متر) در ۴۵ موقعیت از شهرستان‌های گرگان، مراوه‌تپه، نوده، آق‌بند، نکا، ساری و آمل صورت گرفت (شکل ۱). مکان‌های نمونه‌برداری با رویکرد هدفمند و طبقه‌بندی شده انتخاب شدند، به گونه‌ای که تنوع در کاربری اراضی (جنگل، مرتع، کشاورزی) و شرایط اقلیمی (مرطوب تا نیمه‌خشک) در منطقه پوشش داده شود. این روش انتخاب به درک

جدول ۱- مشخصات ویژگی‌های فیزیکی و شیمیایی اندازه‌گیری شده در نمونه‌های خاک و روش‌های آزمایشگاهی مورد استفاده

Property	Measurement method	Reference
Organic carbon (OC)	Walkley-Black method	Enang et al. (2018)
Organic matter (OM)	Calculated as $1.72 \times OC$	Nelson and Sommers (1982)
pH and EC	Soil:water extract (1:2.5)	Kargas et al. (2020)
Soil texture	Bouyoucos hydrometer method	Beretta et al. (2014)
Aggregate stability	Wet sieving method	Almajmaie et al. (2017)
Water drop penetration time (WDPT)	Standard 25 μ L water drop on dry soil surface	Doerr et al. (2009)

جدول ۲ صورت گرفت (Hallett et al., 2011):

جدول ۲- طبقه‌بندی شدت آب‌گریزی خاک بر اساس زمان نفوذ قطره آب (WDPT)

Water repellency class	Water drop penetration time (s)
Not water repellent	$5 >$
Slight	5 – 60
Moderate	60 – 180
Strong	180 – 600

۳-۲- اندازه‌گیری آب‌گریزی خاک

از آزمون زمان نفوذ قطره آب (WDPT) برای اندازه‌گیری شدت آب‌گریزی استفاده شد. در این روش، قطره‌ای از آب مقطر (حجم ۵۰ میکرولیتر) با استفاده از پیست کالیبره شده بر روی سطح خاک خشک‌شده در دمای اتاق ($1C \pm 0.25$) قرار داده می‌شود. سپس زمان نفوذ کامل قطره با کرونومتر اندازه‌گیری شد. هر آزمایش با ۳ تکرار انجام شد و میانگین نتایج ثبت شد (Heidary et al., 2018b). طبقه‌بندی شدت آب‌گریزی مطابق

بافت خاک با استفاده از روش one-hot encoding و به کمک بسته caret به متغیرهای عددی معادل تبدیل شد تا قابلیت استفاده در الگوریتم‌های عددی فراهم شود (Kuhn, 2008). به این ترتیب داده‌ها برای ورود به فرآیند مدل‌سازی آماده شدند.

۲-۵- مدل‌سازی با الگوریتم‌های یادگیری ماشین

در این پژوهش، به‌منظور پیش‌بینی آب‌گریزی خاک با استفاده از شاخص WDPT، سه الگوریتم معتبر یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی و XGBoost (XGB) که توانایی بالایی در مدل‌سازی روابط غیرخطی و پیچیده میان متغیرهای مستقل و وابسته دارند، به کار گرفته شد. در گام نخست، مدل درخت تصمیم با الگوریتم CART و روش رگرسیون ("anova = method) در محیط برنامه‌نویسی R و با بهره‌گیری از بسته rpart پیاده‌سازی شد. در گام دوم، مدل جنگل تصادفی با استفاده از بسته randomForest اجرا شد. در این روش، با ساخت تعداد زیادی درخت تصمیم و ترکیب نتایج حاصل از آن‌ها از طریق میانگین‌گیری، مدل نهایی تولید شد. در گام آخر، الگوریتم XGB با بسته xgboost پیاده‌سازی شد. به‌منظور بهینه‌سازی عملکرد مدل، پارامترهای آن با روش آزمون و خطا^۸ تنظیم شدند.

۲-۶- ارزیابی و اعتبارسنجی مدل‌ها

به‌منظور ارزیابی دقت و کارایی مدل‌های یادگیری ماشین در این پژوهش از سه معیار کمی ریشه میانگین مربعات خطا (RMSE)، میانگین قدر مطلق خطا (MAE) و ضریب تعیین (R^2) برای سنجش میزان انحراف میان مقادیر پیش‌بینی و مقادیر واقعی استفاده شد.

۲-۷- تحلیل اهمیت متغیرها در مدل‌های یادگیری ماشین

به‌منظور بررسی دقیق‌تر نقش و سهم هر یک از ویژگی‌های فیزیکی و شیمیایی خاک در پیش‌بینی شاخص WDPT و نیز شناسایی متغیرهای کلیدی مؤثر بر پاسخ مدل، تحلیل اهمیت متغیرها برای هر سه الگوریتم یادگیری ماشین مورد استفاده در این پژوهش صورت گرفت. در گام نخست برای مدل درخت تصمیم اهمیت هر متغیر براساس میزان کاهش مجموع مربعات خطا در هر گره انشعاب محاسبه شد. این روش نخستین‌بار توسط Liaw and Wiener (2002) معرفی شده است. در مدل جنگل تصادفی، از شاخص درصد افزایش خطای میانگین مربعات^۹ استفاده شد تا تا حساسیت مدل

Very strong

> 600

در این پژوهش، مقدار زمان نفوذ قطره (WDPT) به‌عنوان متغیر وابسته برای مدل‌سازی در نظر گرفته شد و مقدار عددی آن در کنار سایر متغیرهای فیزیکی و شیمیایی خاک، مبنای تحلیل‌های یادگیری ماشین قرار گرفت.

۲-۴- پردازش داده‌ها و آماده‌سازی برای مدل‌سازی

برای آماده‌سازی داده‌ها برای مدل‌های یادگیری ماشین، ابتدا مجموعه‌ای از داده‌های خام در نرم‌افزار Excel بازبینی و پس از آن وارد محیط برنامه‌نویسی R شدند تا امکان پردازش بهتر فراهم شود. در این مرحله، متغیرهای فاقد داده و ستون‌های غیرضروری نظیر مختصات جغرافیایی حذف شدند؛ در ادامه برای شناسایی داده‌های پرت از روش فاصله بین چارکی (IQR) استفاده شد. تحلیل داده‌ها نشان داد که مقادیر داده‌های شناسایی شده ناشی از تغییرات طبیعی در ویژگی‌های خاک (مانند تغییرات موضعی در مواد آلی یا بافت خاک) هستند که در مناطق لسی با تنوع توپوگرافی و کاربری اراضی امری متداول محسوب می‌شود. به‌منظور حفظ جامعیت داده‌ها و جلوگیری از حذف اطلاعات ارزشمند، داده‌های پرت از مجموعه داده حذف نشدند. به جای آن، مدل‌هایی که نسبت به داده‌های پرت حساسیت کم‌تری دارند از جمله جنگل تصادفی انتخاب شدند (Ng et al., 2019). در گام بعدی از روش Z-score برای نرمال‌سازی داده‌ها به‌منظور کاهش اثرگذاری متغیرهایی با دامنه مقیاس مختلف استفاده شد. این اقدام به ویژه برای الگوریتم‌هایی نظیر XGBoost که نسبت به مقیاس متغیرها حساس هستند، اهمیت بالایی دارد. در این مطالعه، ۴۵ نمونه خاک با ۸ ویژگی اصلی برای هر نمونه جمع‌آوری شد. نسبت نمونه به متغیر (۱: ۵/۶) و استفاده از روش‌های اعتبارسنجی متقاطع، کفایت داده‌ها برای مدل‌سازی را تضمین می‌کند. سپس داده‌ها به‌صورت تصادفی به دو گروه آموزش (۳۱ نمونه، ۷۰ درصد) و آزمون (۱۴ نمونه، ۳۰ درصد) تقسیم شدند که این نسبت با استانداردهای متداول در مطالعات یادگیری ماشین خاک مطابقت دارد (Brungard et al., 2015). در مرحله بعدی، برای بررسی هم‌خطی میان متغیرهای مستقل، ضریب تورم واریانس (VIF)^۷ محاسبه شد. در مواردی که مقادیر VIF بیش از ۱۰ بود، مدل هم با آن متغیر و هم بدون آن متغیر اجرا شد تا تأثیر حذف یا نگهداری آن بر دقت مدل مورد ارزیابی قرار گیرد. در نهایت، متغیر

⁹ Increase in %MSE

⁷ Variance Inflation Factor

⁸ Trial-and-error

(WDPT) بر اساس خصوصیات فیزیکی و شیمیایی خاک‌های لسی منطقه مورد مطالعه در شمال ایران است. همچنین در ادامه، چارچوب‌هایی برای تحلیل‌های آبی شامل تحلیل حساسیت و عدم قطعیت نیز در نظر گرفته شده است.

۳-۱- بررسی آماری اولیه داده‌ها

به منظور شناخت اولیه از داده‌های جمع‌آوری شده، تحلیل توصیفی ویژگی‌های فیزیکی و شیمیایی خاک و نیز شاخص WDPT انجام شد. نتایج محاسبه آماره‌هایی نظیر میانگین، میانه، انحراف معیار، کمینه و بیشینه (جدول ۳)، نشان داد که متغیرهای مورد بررسی از تنوع قابل توجهی برخوردار هستند. نزدیکی میانگین و میانه در بسیاری از ویژگی‌های مورد بررسی حاکی از توزیع نسبتاً نرمال داده‌ها بود، هر چند در برخی متغیرها مانند EC، MWD و WDPT چولگی مثبت و کشیدگی بالا مشاهده شد که بیانگر وجود داده‌های پرت و توزیع‌های غیرنرمال است. برای نمونه، چولگی EC برابر با ۳/۶ و کشیدگی ۱۳/۹۸ بود که نشان‌دهنده وجود مقادیر بزرگ‌تر از میانگین و فاصله گرفتن داده‌ها از توزیع نرمال است. بررسی بافت خاک نیز بر تنوع ساختاری خاک‌های منطقه تأکید دارد؛ به‌طور خاص، شن با چولگی ۱/۰۸ و کشیدگی ۰/۶۴ احتمال غلبه برخی بافت‌های خاص خاک را مطرح می‌کند. نتایج باکس پلات متغیرها (شکل ۲) نیز حاکی از آن است که متغیرهایی مانند EC، MWD، درصد شن و WDPT دارای داده‌های پرت قابل توجهی هستند. این داده‌های پرت می‌توانند ناشی از شرایط طبیعی خاک، نحوی نمونه‌برداری یا حتی خطاهای آزمایشگاهی باشند.

نسبت به حذف هر ویژگی ارزیابی شود. این شاخص در مطالعات (Cutler et al. (2007) به کار گرفته شده است. در نهایت در مدل XGB، اهمیت ویژگی‌ها با استفاده از شاخص Gain تعیین شد که سهم هر متغیر در بهبود عملکرد مدل در فرآیند بوستینگ را نشان می‌دهد.

۳-۸- تحلیل عدم قطعیت مدل‌ها

برای ارزیابی پایداری و قابلیت اطمینان سه مدل یادگیری ماشین مورد استفاده در این پژوهش در پیش‌بینی شاخص WDPT، از دو روش تحلیل عدم قطعیت شامل بوت‌استرپ^{۱۰} و مونت‌کارلو^{۱۱} استفاده شد. در روش بوت‌استرپ، با انجام ۱۰۰ بار نمونه‌گیری تصادفی با جایگزینی^{۱۲}، میزان تغییرپذیری خطای پیش‌بینی مدل‌ها بر اساس شاخص RMSE مورد بررسی قرار گرفت. همچنین در روش مونت‌کارلو، با افزودن اختلال^{۱۳} تصادفی به ویژگی‌های خاک، ۱۰۰ مجموعه داده مصنوعی ایجاد شد و مدل‌ها بر روی داده‌ها آموزش داده شدند تا پایداری عملکرد آن‌ها تحت شرایط مختلف ارزیابی شود. نهایتاً، مقایسه بصری نتایج حاصل از این دو رویکرد در قالب نمودار باکس پلات برای هر سه مدل انجام شده و تفاوت‌ها و پایداری نتایج تحلیل شد.

۳- نتایج و بحث

در این بخش، یافته‌های به‌دست‌آمده از تحلیل‌های آماری و مدل‌سازی با استفاده از الگوریتم‌های یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی و XGB ارائه و بررسی شده‌اند. هدف اصلی، ارزیابی دقت این مدل‌ها در پیش‌بینی شاخص آب‌گریزی خاک

جدول ۳- شاخص‌های آماری توصیفی متغیرهای اندازه‌گیری شده در نمونه‌های خاک منطقه مطالعه

Table 3: Descriptive statistical indicators of measured soil variables in the study area

Variable	Mean	Median	Standard deviation	Minimum	Maximum	Skewness	Kurtosis
OC (%)	1/51	1/17	1/10	0/16	4/68	0/88	0/09
OM (%)	2/60	2/01	1/89	0/27	8/05	0/88	-0/09
EC (dS/m)	2/27	1/15	3/32	0/59	19/35	3/60	13/98
pH	7/34	7/40	0/19	6/89	7/70	-0/47	-0/58
MWD (mm)	8/15	5/69	6/47	0/26	26/82	1/23	0/91
Sand (%)	28/67	25/80	12/34	13/80	65/80	1/08	0/64
Clay (%)	17/54	16/80	6/22	6/80	30/80	0/27	-0/68
Silt (%)	53/79	53/40	10/35	27/40	71/40	-0/45	-0/57
WDPT (s)	21/40	19/00	15/49	3/00	64/00	1/11	0/51

تصادفی استفاده شد که به دلیل ساختار غیر خطی و انعطاف‌پذیر خود، تأثیر ناهنجاری‌های داده را به حداقل می‌رساند (Kuhn and

در این پژوهش به جای حذف این داده‌ها از الگوریتم‌های یادگیری ماشین مقاوم به داده‌های پرت مثل درخت تصمیم و جنگل

¹² Sampling with replacement

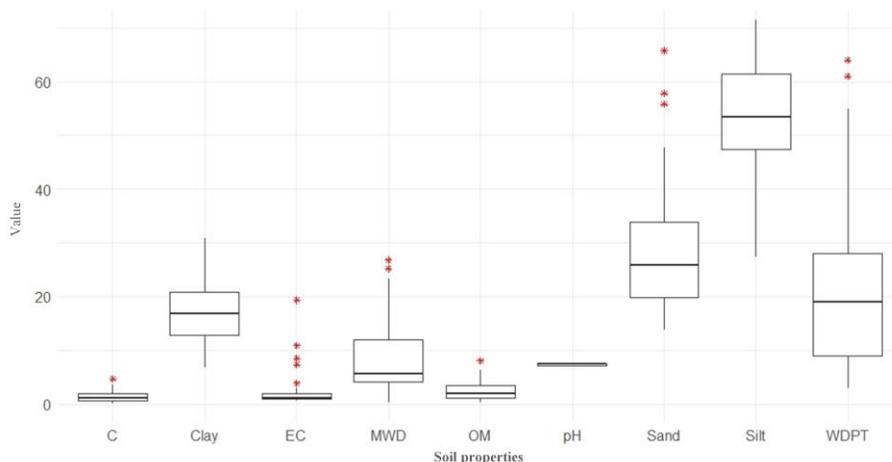
¹³ Noise

¹⁰ Bootstrap

¹¹ Monte Carlo

یافته‌ها نشان‌دهنده همبستگی مثبت میان کربن آلی، ماده آلی و MWD با شاخص WDPT بود. موضوعی که نقش بالقوه این متغیرها را در پیش‌بینی شدت آب‌گریزی خاک تقویت می‌کند.

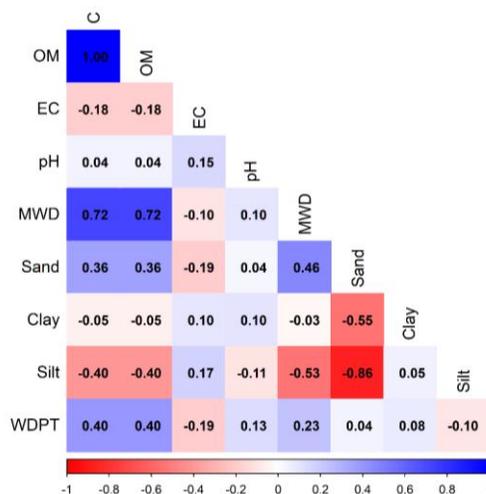
(Johnson, 2013; Liaw and Wiener, 2002). علاوه بر این برای بررسی روابط خطی میان متغیرهای عددی، ماتریس همبستگی بین خصوصیات فیزیکی و شیمیایی خاک ترسیم شد (شکل ۳).



شکل ۲- نمودار باکس پلات توزیع متغیرهای فیزیکی و شیمیایی خاک
Figure 2. Boxplot of the distribution of soil physicochemical variables

دارای همبستگی معنادار و قوی با یکدیگر هستند که می‌تواند نشان‌دهنده وابستگی‌های فیزیکوشیمیایی در خاک منطقه مطالعه باشد.

۳-۳- تحلیل همبستگی بین ویژگی‌های عددی خاک
برای بررسی روابط بین ویژگی‌های فیزیکی و شیمیایی خاک، از تحلیل همبستگی پیرسون استفاده شد. نتایج حاصل از ترسیم ماتریس همبستگی (شکل ۳) نشان می‌دهد که برخی از ویژگی‌ها



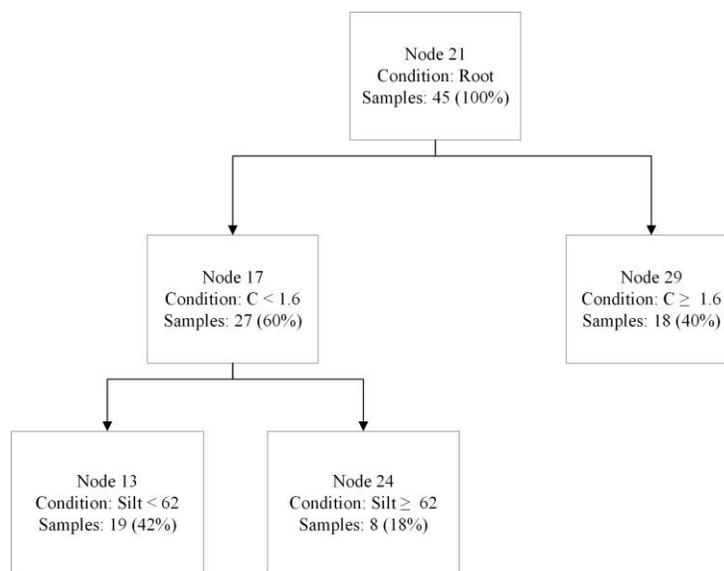
شکل ۳- ماتریس همبستگی بین ویژگی‌های فیزیکی و شیمیایی خاک
Figure 3- Correlation matrix of soil physicochemical properties

که این رابطه بیانگر تأثیر مثبت ماده آلی بر پایداری خاک‌دانه‌ها است. مطالعات انجام شده توسط Chenu et al. (2000) و Bouajila and Gallali (2010) نیز این موضوع را تأیید کرده‌اند

به‌طور مشخص، میانگین وزنی قطر خاک‌دانه‌ها دارای رابطه همبستگی مثبت و نسبتاً قوی با کربن آلی و ماده آلی است ($r=0.72$).

۳-۴- مدل درخت تصمیم برای پیش‌بینی شاخص WDPT
 برای بررسی تأثیر ویژگی‌های فیزیکی و شیمیایی خاک بر آب‌گریزی، یک مدل درخت تصمیم با الگوریتم CART و روش رگرسیون (method = "anova") بر داده‌های آموزشی اعمال شد (شکل ۴). نتایج نشان داد که کربن آلی (OC) مهم‌ترین متغیر در پیش‌بینی شاخص WDPT بوده و به‌عنوان گره ریشه انتخاب شد. آستانه ۱/۶ درصد برای متغیر کربن آلی معیار نخستین انشعاب در نظر گرفته شد که در مقادیر کم‌تر از این آستانه، متغیر درصد سیلت نقش تفکیک‌کننده ثانویه را ایفا می‌کند که با پژوهش‌های Lombardo et al., 2017، Emadi et al., 2020 و Wadoux et al., 2023 هم‌راستا است.

و بر نقش مواد آلی در افزایش پایداری خاک تأکید داشته‌اند. از سوی دیگر، بین درصد رس با درصد ماسه رابطه همبستگی منفی قوی دیده می‌شود ($r=-0.55$) که به‌دلیل ماهیت متقابل اجزای بافت خاک است (Raheem and Omar, 2021; Wang et al., 2009). نکته مهم، وجود رابطه مثبت بین زمان پایداری آب‌گریزی خاک (WDPT) با کربن آلی ($r=0.40$) است، که نشان می‌دهد افزایش مواد آلی ممکن است به تشدید آب‌گریزی منجر شود. این رابطه در مطالعات متعددی به‌ویژه در شرایط پس از آتش‌سوزی یا کاربرد کودهای آلی گزارش شده است (Hermansen et al., 2021; Raheem and Omar, 2021).

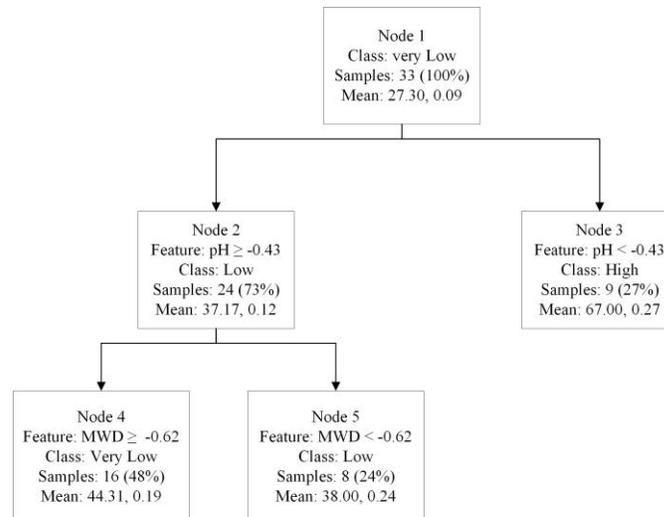


شکل ۴- درخت تصمیم حاصل از مدل CART برای پیش‌بینی زمان پایداری قطره آب (WDPT) بر اساس ویژگی‌های فیزیکی و شیمیایی خاک در خاک‌های لسی منطقه مورد مطالعه.

Figure 4. Decision Tree Generated by the CART Model for Predicting Water Drop Penetration Time (WDPT) Based on Soil Physicochemical Properties in Loess Soils of the Study Area

نشان می‌دهد که با وجود ساختار ساده و قابلیت تفسیر بالای این مدل، مدل از منظر آماری عملکرد ضعیفی داشته است ($RMSE=19/55$, $MAE=13/67$ و $R^2=0/02$) که بیانگر پیچیدگی پدیده آب‌گریزی خاک و یا ناکافی بودن حجم داده‌های آموزشی و تست باشد. در ادامه برای طبقه‌بندی WDPT، الگوریتم درخت تصمیم مجدداً اجرا شد (شکل ۵). در این مدل متغیرهای pH و میانگین وزنی قطر خاک‌دانه‌ها به‌ترتیب نقش کلیدی در گره‌های تصمیم‌گیری داشتند.

در این حالت، نمونه‌های با درصد سیلت کم‌تر از ۱۶ درصد، میانگین WDPT برابر با ۹/۹ ثانیه داشته‌اند (۲۴ درصد داده‌ها)، در حالی که این مقدار برای سیلت بیش‌تر یا مساوی ۱۶ درصد به ۱۷ ثانیه می‌رسید (۳۶ درصد داده‌ها). در سمت دیگر درخت، زمانی که کربن آلی بیش‌تر یا مساوی ۱/۶ درصد بود، میانگین WDPT برابر با ۳۰ ثانیه به‌دست آمد (۳۹ درصد داده‌ها). در مجموع، نتایج مدل درخت تصمیم نشان داد که کربن آلی نقش کلیدی در افزایش آب‌گریزی خاک دارد و در سطوح پایین‌تر آن درصد سیلت به‌عنوان متغیر مکمل اثرگذار است. تحلیل نتایج آماری این مدل (جدول ۴)



شکل ۵- نمودار درخت تصمیم برای طبقه‌بندی WDPT_class بر اساس متغیرهای خاکی (مقدارهای داخل گره‌ها نشان‌دهنده توزیع نسبی کلاس‌ها در هر گره است و رنگ‌ها نمایانگر کلاس غالب هستند)

Figure 5- Decision Tree Diagram for Classifying WDPT_class Based on Soil Variables (Values inside the nodes indicate the relative distribution of classes, and colors represent the dominant class in each node)

می‌شود. کاهش pH با افزایش آب‌گریزی رابطه معنی‌داری دارد. مطالعات پیشین مانند (García et al., Sephrnia et al. (2017) و (2022) و (Ellerbrock et al. (2005) رابطه غیرخطی میان کربن آلی و آب‌گریزی که به ماهیت قطبی و غیرقطبی ترکیبات آلی و نحوی اتصال آن‌ها با ذرات مرتبط است را تأیید می‌کنند. همچنین تأثیر متقابل کربن آلی و pH نیز خطی نبوده و بسته به سطح کربن آلی، تغییرات کیفی در رفتار آب‌گریزی مشاهده می‌شود.

گره ریش، کلاس غالب «خیلی کم» را نشان داد. اگر مقدار pH کم‌تر از -0.43 بود، داده‌ها عمدتاً در کلاس‌های «زیاد» و «خیلی زیاد» قرار گرفتند و در غیر این صورت، به سمت کلاس‌های «کم» یا «خیلی کم» تقسیم شدند. در مسیرهای دارای pH بالاتر، متغیر MWD تفکیک‌کننده بود به طوری که اگر مقدار آن بزرگ‌تر از -0.62 باشد، نمونه‌ها عمدتاً در کلاس «خیلی کم» قرار می‌گیرند و در غیر این صورت، توزیع داده‌ها بیش‌تر در کلاس «کم» مشاهده

جدول ۴- مقایسه شاخص‌های آماری ارزیابی عملکرد سه مدل یادگیری ماشین در پیش‌بینی شاخص WDPT

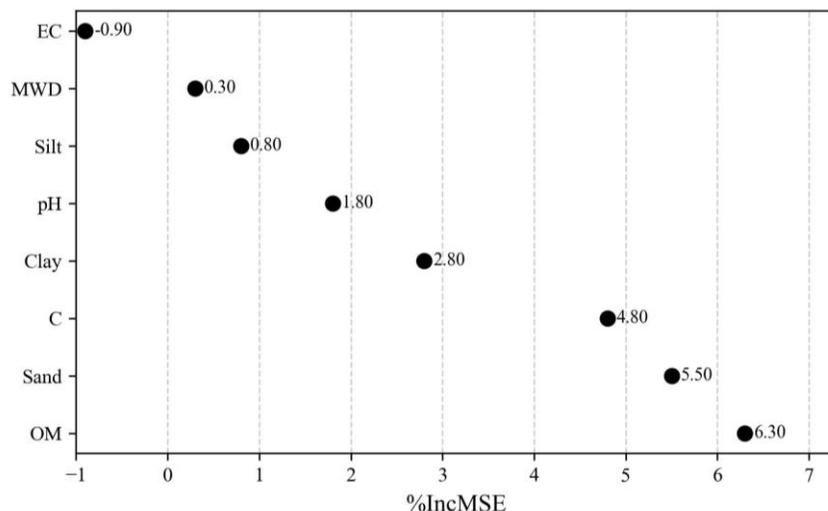
Table 4. Comparison of Statistical Performance Metrics of Three Machine Learning Models in Predicting WDPT

Model	RMSE	MAE	R ²
Decision Tree	55/19	67/13	0/02
Random Forest	17/81	11/31	0/02
Optimized Random Forest	15/00	11/93	0/42
XGBoost	19/88	16/02	-0/39
Optimized XGBoost	14/70	11/56	0/42

افزایش تعداد تکرارها از ۳ به ۵ دقت مدل افزایش یافت و مقدار RMSE و R² به ترتیب به ۱۵ و ۰/۴۲ بهبود یافتند (جدول ۴). تحلیل اهمیت متغیرها (شکل ۶) نشان داد که کربن آلی، درصد ماسه و ماده آلی مهم‌ترین عوامل مؤثر در پیش‌بینی شاخص WDPT هستند. در مقابل متغیرهایی مثل EC و pH نقش کم‌تری داشتند. این یافته‌ها یا مطالعات قبلی (Doerr and Thomas, 2004; Mataix-Solera and Doerr, 2000) و یافته‌های اخیر (García et al., 2022; Wang et al., 2024; Wang et al., 2025) هم‌خوانی دارد.

۳-۵- مدل جنگل تصادفی (Random Forest)

مدل جنگل تصادفی به‌عنوان یک الگوریتم تجمیعی مبتنی بر درخت تصمیم، برای پیش‌بینی شاخص آب‌گریزی خاک (WDPT) به‌کار گرفته شد. در اجرای اولیه با تنظیمات پیش‌فرض، عملکرد مدل ضعیف بود (RMSE=۱۷/۸۱، MAE=۱۱/۳۱ و R²=-۰/۰۲) و تنها ۲ درصد از واریانس داده‌ها را توضیح می‌داد (جدول ۵). برای بهبود عملکرد مدل، بهینه‌سازی پارامترها در دو مرحله صورت گرفت: ابتدا مقدار mtry با استفاده از اعتبارسنجی متقاطع تکرار شونده (Repeated 10-fold CV) بهینه‌سازی شد (mtry=2). سپس با

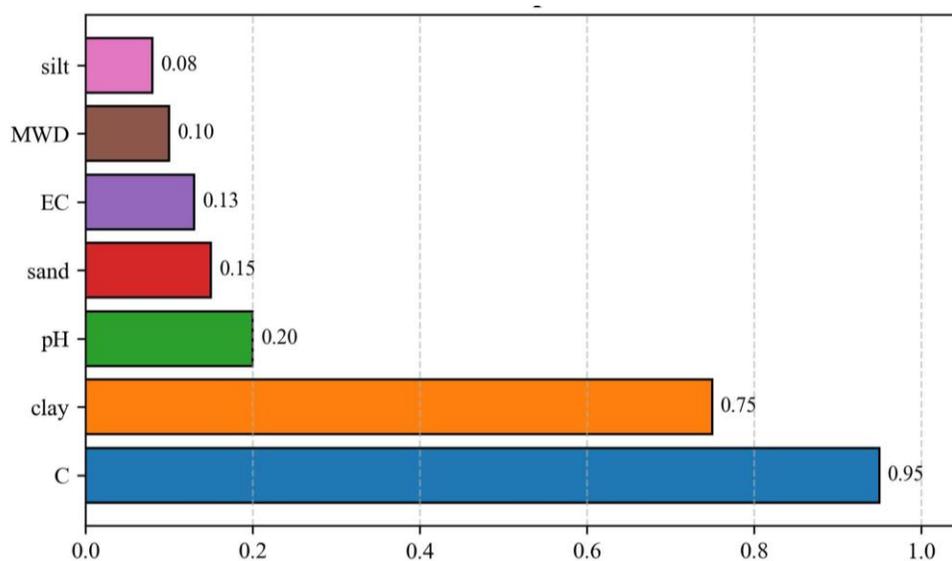


شکل ۶- اهمیت نسبی ویژگی‌های خاک در مدل جنگل تصادفی بر اساس درصد افزایش MSE در صورت حذف متغیر
Figure 6. Relative Importance of Soil Properties in the Random Forest Model Based on Percent Increase in MSE Upon Variable Removal

Guestrin (2016) با بهینه‌سازی پارامترهای کلیدی مثل نرخ یادگیری، عمق درخت، وزن گره‌ها و درصد نمونه‌برداری، عملکرد مدل به‌طور چشم‌گیری بهبود پیدا کرد (RMSE=۱۴/۷). بررسی اهمیت ویژگی‌ها (شکل ۷) نشان داد که کربن آلی مهم‌ترین عامل در مدل XGB بوده و متغیرهای درصد رس، pH و ماسه در رتبه‌های بعدی قرار دارند. این یافته‌ها با نتایج Baghbani et al. (2025) هم‌راستا هستند.

۳-۶- مدل XGBoost

مدل XGB به‌عنوان یکی از الگوریتم‌های قدرتمند مبتنی بر گرادینان تقویتی، برای پیش‌بینی شاخص WDPT به‌کار گرفته شد. نسخه اولیۀ مدل عملکرد ضعیفی داشت (RMSE=۱۹/۸۸، MAE=۱۶/۰۲ و $R^2=-۰/۳۹$) که نشان‌دهندۀ عدم توانایی در بازتولید واریانس داده‌های واقعی و وجود انحراف قابل‌توجه میان مقادیر پیش‌بینی‌شده و واقعی بود (جدول ۴). این نتایج با توجه به ساختار غیرخطی داده‌ها، قابل انتظار و مشابه یافته‌های Chen and



شکل ۷- اهمیت نسبی ویژگی‌های خاک در مدل XGB
Figure 7. Relative Importance of Soil Properties in the XGBoost Model

مقابل، درصد رس الگوی متفاوتی داشت و بیش‌تر در بخش‌های مرکزی و شمالی متمرکز بود که احتمالاً ناشی از شرایط زمین‌شناسی و فرآیندهای ژئومورفولوژیکی است. نقشه پراکنش شاخص WDPT نیز نشان داد که نواحی با آب‌گریزی بالا عمدتاً با مناطق دارای کربن آلی زیاد هم‌پوشانی دارند که بر نقش کلیدی کربن آلی در افزایش آب‌گریزی تأکید دارد (Shakesby et al., 2000). اگر چه درصد رس در برخی نقاط با تغییرات WDPT هم‌راستا بود، اما همبستگی آن ضعیف‌تر و گاه متناقض با رفتار آب‌گریزی بود (Doerr and Shakesby, 2011). در مجموع، نتایج تحلیل مکانی بیانگر آن است که کربن آلی و درصد رس الگوهای فضایی مشخصی دارند و با پراکنش WDPT مرتبط هستند. این یافته‌ها می‌تواند منبای توسعه مدل‌های پیش‌بینی مکانی دقیق‌تر برای آب‌گریزی خاک، به‌ویژه در مناطق دارای شرایط مشابه، قرار گیرد.

۳-۹- تحلیل حساسیت مدل‌ها

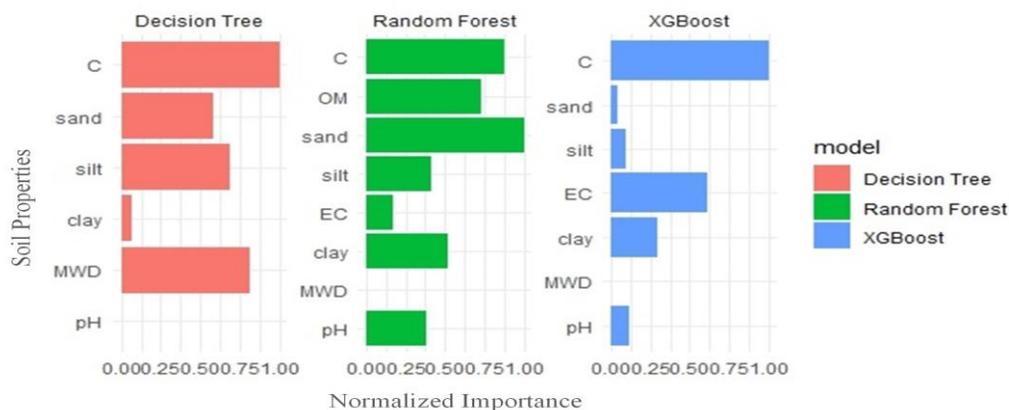
برای ارزیابی نقش نسبی هر ویژگی خاک در پیش‌بینی WDPT، تحلیل حساسیت مبتنی بر اهمیت متغیرها در سه مدل یادگیری ماشین (درخت تصمیم، جنگل تصادفی و XGB) انجام شد (شکل ۸). کربن آلی در هر سه مدل مهم‌ترین متغیر بود که با یافته‌های پیشین مثل (Brungard et al., 2015) و (Ng et al., 2019) مطابقت دارد. در مدل جنگل تصادفی درصد ماسه و ماده آلی در رتبه‌های بعدی قرار گرفتند، در حالی که در مدل XGB و درخت تصمیم، درصد رس اهمیت بالایی داشتند. این اختلاف در رتبه‌بندی متغیرها به تفاوت در مکانیسم‌های یادگیری و ارزیابی اهمیت در هر مدل باز می‌گردد (Friedman, 2001; Lipton, 2016).

۳-۷- مقایسه عملکرد مدل‌های یادگیری ماشین در پیش‌بینی آب‌گریزی خاک

در این پژوهش، سه الگوریتم یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی و XGB برای پیش‌بینی شاخص WDPT مورد ارزیابی قرار گرفتند. نسخه اولیه این مدل‌ها عملکرد ضعیفی داشتند، اما پس از بهینه‌سازی پارامترها در مرحله اولیه، عملکرد دو مدل جنگل تصادفی و XGB به‌طور قابل‌توجهی افزایش پیدا کرد. بر اساس نتایج نهایی (جدول ۴) مدل جنگل تصادفی بهترین عملکرد را داشت و مدل XGB در رتبه بعدی قرار گرفت و درخت تصمیم نیز ضعیف‌ترین عملکرد را از خود نشان داد. نتایج نشان دادند که تنظیم مؤثر پارامترها نقشی کلیدی در ارتقا دقت مدل‌ها دارد و در هر دو مدل جنگل تصادفی و XGB کاهش بیش از ۱۵ درصد در RMSE حاصل شد. این یافته‌ها با مطالعات پیشین مانند (Wang et al., 2024) و (Baghbani et al., 2025) هم‌راستا است که در آن‌ها نیز استراتژی‌های بهینه‌سازی موجب بهبود معنادار مدل‌سازی شدند.

۳-۸- تحلیل مکانی پراکنش ویژگی‌های کلیدی خاک

باتوجه به اهمیت نقش تحلیل الگوی فضایی ویژگی‌های خاک در درک روابط محیطی و شناسایی عوامل مؤثر، سه متغیر کلیدی شامل کربن آلی، درصد رس و WDPT با استفاده از مختصات نمونه‌برداری ترسیم و تحلیل شدند. نتایج نشان داد که مقادیر بالای کربن آلی عمدتاً در نواحی جنوب شرقی و غربی منطقه مشاهده می‌شوند که می‌تواند تحت تأثیر عواملی مثل تراکم پوشش گیاهی، کاربری اراضی و مدیریت حفاظتی باشد (Zornoza et al., 2008). در

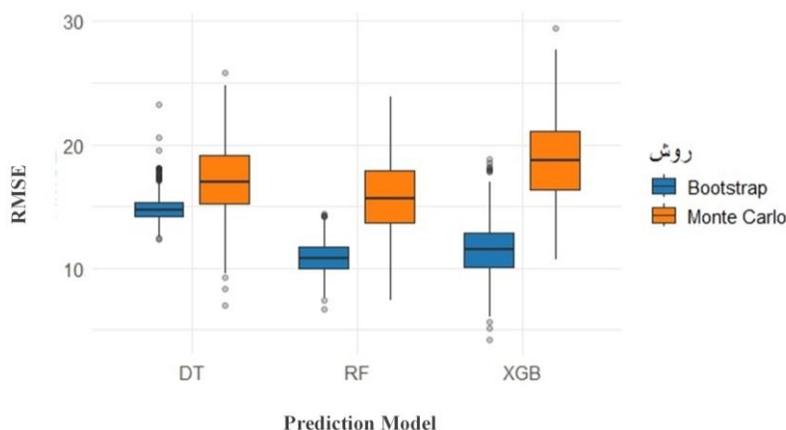


شکل ۸- مقایسه اهمیت نسبی ویژگی‌های خاک در پیش‌بینی WDPT توسط سه مدل یادگیری ماشین (درخت تصمیم، جنگل تصادفی و XGB) بر اساس مقدار نرمال‌شده اهمیت متغیرها.

Figure 8. Comparison of the Normalized Relative Importance of Soil Properties in Predicting WDPT Using Three Machine Learning Models (Decision Tree, Random Forest, and XGBoost)

جنگل تصادفی مجدداً بهترین عملکرد را با RMSE برابر با ۱۵/۷۵۸ و انحراف معیار ۲/۷۹۷ نشان داد. در مقابل مدل XGB بیش‌ترین حساسیت را به داده‌های ناپایدار از خود نشان داد (RMSE=۱۸/۷۱۶) که نشان از کاهش دقت و ثبات در محیط‌های نویزی دارد. این نتایج با یافته‌های پژوهش Padarian et al. (2019) هم‌سو هستند.

مقایسه دو رویکرد تحلیل عدم قطعیت مورد بررسی در این پژوهش نشان داد که مدل جنگل تصادفی نه تنها دارای کم‌ترین میانگین خطا بلکه دارای پایداری بالاتر در شرایط عدم قطعیت است (شکل ۹). در حالی که مدل XGB به رغم قدرت بالا در شناسایی روابط پیچیده، در محیط‌های واقعی با داده‌های نویزی، ثبات کم‌تری دارد.



شکل ۹- نمودار مقایسه‌ای RMSE حاصل از دو رویکرد Bootstrap و Monte Carlo برای مدل‌های DT، RF و XGB به‌منظور

تحلیل عدم قطعیت پیش‌بینی‌ها

Figure 9. Comparative RMSE Chart from Bootstrap and Monte Carlo Approaches for DT, RF, and XGB Models in Predictive Uncertainty Analysis

بیش‌تری در برابر نویز داده‌ها و سناریوهای ناپایدار محیطی دارد. این یافته‌ها می‌توانند مبنایی برای شناسایی نواحی بحرانی از نظر آب‌گریزی و طراحی برنامه‌های مدیریتی مؤثرتر در مناطق دارای خاک‌های لسی فراهم کنند. پیشنهاد می‌شود که در مطالعات آتی از مدل‌های ترکیبی (مانند RF-XGB) یا ساختارهای یادگیری عمیق برای بهبود دقت پیش‌بینی استفاده شود. هم‌چنین، بهره‌گیری از داده‌های چندزمانی (پیش و پس از آتش‌سوزی، بارندگی یا تغییر کاربری) می‌تواند به درک عمیق‌تری از پویایی آب‌گریزی خاک کمک کند.

سپاس‌گزاری

بدین‌وسیله از حمایت‌های دانشگاه علوم کشاورزی و منابع طبیعی گرگان برای انجام این طرح پژوهشی سپاسگزاری به‌عمل می‌آید.

۳-۱- تحلیل عدم قطعیت با رویکرد Bootstrap و Monte Carlo

با استفاده از رویکرد Bootstrap و محاسبه RMSE در تکرارهای مختلف، مشخص شد که مدل جنگل تصادفی عملکرد دقیق‌تر و پایدارتری (RMSE=۱۰/۸۳۷) در مقایسه با مدل XGB (RMSE=۱۱/۵۱۱) و درخت تصمیم (RMSE=۱۴/۸۴۱) دارد. این یافته‌ها با نتایج مطالعه متاآنالیز شده‌ای از Cutler et al. (2007) و Hastie et al. (2009) هم‌راستا هستند، که دلیل برتری مدل جنگل تصادفی را مکانیسم ترکیبی و تصادفی آن در کاهش بیش‌برازش و افزایش تعمیم‌پذیری می‌دانند. نتایج تحلیل عدم قطعیت با رویکرد Monte Carlo نیز نشان داد که در مواجهه با نویز نرمال‌شده، مدل

۴- نتیجه‌گیری

این مطالعه نشان داد که الگوریتم‌های پیشرفته یادگیری ماشین، به‌ویژه مدل جنگل تصادفی (Random Forest)، قابلیت بالایی در مدل‌سازی شدت آب‌گریزی خاک (WDPT) در خاک‌های لسی شمال ایران دارند. در بین سه مدل به‌کاررفته، مدل RF با RMSE برابر با ۱۵ و R^2 برابر با ۰/۴۲ دقیق‌ترین پیش‌بینی را ارائه داد. مدل XGBoost نیز عملکردی قابل‌مقایسه داشت، اما نسبت به تغییرات داده‌ای پایداری کم‌تری نشان داد. در مقابل، مدل درخت تصمیم به‌دلیل ساختار ساده خود، دقت پایین‌تری در مدل‌سازی روابط پیچیده بین متغیرها داشت. تحلیل اهمیت متغیرها نشان داد که کربن آلی، کلید اصلی در پیش‌بینی شدت آب‌گریزی است و پس از آن، درصد ماسه، رس، و هدایت الکتریکی نقش‌های مؤثری ایفا کردند. نتایج تحلیل عدم قطعیت نیز نشان داد که مدل RF، پایداری

تضاد منافع نویسندگان

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه تضاد منافی در خصوص نگارش و انتشار مطالب و نتایج این پژوهش ندارند.

دسترسی به داده‌ها

داده‌ها و نتایج استفاده‌شده در این پژوهش از طریق مکاتبه با نویسنده مسئول در اختیار قرار خواهد گرفت.

مشارکت نویسندگان

علی محمدیان بهبهانی: ایده‌پردازی، تأمین منابع، روش‌شناسی، تحلیل و بررسی؛ کهزاد حیدری: کار میدانی و آزمایشگاهی، نگارش نسخه نهایی مقاله، تحلیل نتایج، ترسیم و مصورسازی؛ محسن حسینعلی‌زاده: روش‌شناسی، ویرایش مقاله، تحلیل و بررسی.

- Chenu, C., Bissonnais, Y. Le, & Arrouays, D. (2000). Organic Matter Influence on Clay Wettability and Soil Aggregate Stability. *Soil Science Society of America Journal*, 64(4), 1479–1486. doi: 10.2136/SSSAJ2000.6441479X
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792. doi: 10.1890/07-0539.
- Danielsen, A. C. S., Hermansen, C., Weber, P. L., Mikstas, D., Pesch, C., de Carvalho Gomes, L., Gutierrez, S., Nielsen, P. H., Greve, M. H., Møldrup, P., Normand, S., & de Jonge, L. W. (2025). Soil Water Repellency in Natural and Semi-Natural Habitats: A Nexus Between Abiotic Factors and Prokaryotic Communities. *European Journal of Soil Science*, 76(2), e70063. doi: 10.1111/EJSS.70063
- de Blas, E., Almendros, G., & Sanz, J. (2013). Molecular characterization of lipid fractions from extremely water-repellent pine and eucalyptus forest soils. *Geoderma*, 206, 75–84. doi: 10.1016/j.geoderma.2013.04.027
- Doerr, S. H., & Shakesby, R. A. (2011). Handbook of Soil Sciences Properties and Processes, second edition. In Handbook of Soil Sciences Properties and Processes, second edition (pp. 515–525). CRC press, Taylor and Francis group. doi: 10.1016/B978-0-444-51269-7.50023-0
- Doerr, S. H., & Thomas, A. D. (2000). The role of soil moisture in controlling water repellency: new evidence from forest soils in Portugal. *Journal of Hydrology*, 231–232, 134–147. doi: 10.1016/S0022-1694(00)00190-6

منابع

- امامی، مریم، خرمالی، فرهاد، پهلوان راد، محمدرضا و ابراهیمی، سهیلا (۱۴۰۳). تهیه نقشه‌های سه‌بعدی اجزای بافت خاک با تلفیق الگوریتم جنگل رگرسیونی چندکی و تابع عمق اسپیلاین در استان گلستان. *تحقیقات آب و خاک ایران*، ۱۵۵(۱)، ۵۱–۶۸. doi: 10.22059/ijswr.2023.366978.669594
- پهلوان راد، محمدرضا، تومانیان، نورایر و خرمالی، فرهاد (۱۳۹۵). معرفی نقشه‌برداری رقومی خاک. *مدیریت اراضی* ۴(۲)، ۹۷–۱۱۴. doi: 10.22092/lmj.2017.109482
- حیدری، کهزاد، نجفی نژاد، علی، محمدیان بهبهانی، علی و اونق، مجید (۱۳۹۷). بررسی شدت آب‌گریزی خاک و تغییرات زمانی آن پس از آتش سوزی تجویزی در مناطق جنگلی آبخیز توشن استان گلستان. *پژوهش‌های حفاظت آب و خاک* ۲۵(۴)، ۴۷–۲۷. doi: 10.22069/jwsc.2018.14663.2960

References

- Baghbani, A., Kiany, K., Abuel-Naga, H., & Lu, Y. (2025). Predicting the Compression Index of Clayey Soils Using a Hybrid Genetic Programming and XGBoost Model. *Applied Sciences*, Vol. 15, Page 1926, 15(4), 1926. doi: 10.3390/APP15041926
- Blaesbjerg, N. H., Weber, P. L., de Jonge, L. W., Moldrup, P., Greve, M. H., Arthur, E., Knadel, M., & Hermansen, C. (2022). Water repellency prediction in high-organic Greenlandic soils: Comparing vis-NIRS to pedotransfer functions. *Soil Science Society of America Journal*, 86(3), 643–657. doi: 10.1002/SAJ2.20407
- Bouajila, A., & Gallali, T. (2010). Land use effect on soil and particulate organic carbon, and aggregate stability in some soils in Tunisia. *African Journal of Agricultural Research*, 5(8), 764–774. doi: 10.5897/AJAR10.183
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239–240, 68–83. doi: 10.1016/J.GEODERMA.2014.09.019
- Chen, J., McGuire, K. J., & Stewart, R. D. (2020). Effect of soil water-repellent layer depth on post-wildfire hydrological processes. *Hydrological Processes*, 34(2), 270–283. doi:org/10.1002/HYP.13583
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794. doi: 10.1145/2939672.2939785

- Ellerbrock, R. H., Gerke, H. H., Bachmann, J., & Goebel, M.-O. (2005). Composition of Organic Matter Fractions for Explaining Wettability of Three Forest Soils. *Soil Science Society of America Journal*, 69(1), 57. doi: 10.2136/SSSAJ2005.0057
- Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020). Predicting and Mapping of Soil Organic Carbon Using Machine Learning Algorithms in Northern Iran. *Remote Sensing*, 12(14). doi: 10.3390/rs12142234
- Emami, M., Khormali, F., Pahlavan Rad, M. reza, & Ebrahimi, S. (2024). Preparation of three-dimensional maps of soil particle size fractions by combining quantile regression forest algorithm and spline depth function in Golestan Province. *Iranian Journal of Soil and Water Research*, 55(1), 51–68. doi: 10.22059/ijswr.2023.366978.669594 [In Persian]
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. 29(5), 1189–1232. doi: 10.1214/AOS/1013203451
- García, E. M., Alberti, M. G., & Arcos Álvarez, A. A. (2022). Measurement-While-Drilling Based Estimation of Dynamic Penetrometer Values Using Decision Trees and Random Forests. *Applied Sciences* 2022, 12(9), 4565. doi: 10.3390/APP12094565
- Hallett, P. D., Bachmann, J., Czachor, H., Urbanek, E., & Zhang, B. (2011). Hydrophobicity of Soil. *Encyclopedia of Earth Sciences Series*, Part 4, 378–384. doi: 10.1007/978-90-481-3585-1_195
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. doi: 10.1007/978-0-387-84858-7
- Heidary, K., Najafi Nejad, A., Dekker, L. W., Ownegh, M., & Mohammadian Behbahani, A. (2018). Impact of Soil Water Repellency on Hydrological and Erosion Processes; A Review. *ECOPERSEA*, 6(4), 269–284. doi: 10.1001.1.23222700.2018.6.4.6.7
- Heidary, K., Najafinejad, A., Mohammadian Behbahani, A., & Ownegh, M. (2018a). Assessment of Soil Water Repellency Intensity and Its Temporal Variability after Prescribed Fire in Forest Areas of Toshen Watershed, Golestan Province. *Journal of Water and Soil Conservation*, 25(4), 27–47. doi: 10.22069/jwsc.2018.14663.2960 [In Persian]
- Heidary, K., Najafinejad, A., Mohammadian Behbahani, A., & Ownegh, M. (2018b). Assessment of Soil Water Repellency Intensity and Its Temporal Variability after Prescribed Fire in Forest Areas of Toshen Watershed, Golestan Province. *Journal of Water and Soil Conservation*, 25(4), 27–47. doi: 10.22069/jwsc.2018.14663.2960
- Hermansen, C., Norgaard, T., de Jonge, L. W., Weber, P. L., Moldrup, P., Greve, M. H., Tuller, M., & Arthur, E. (2021). Linking water vapor sorption to water repellency in soils with high organic carbon contents. *Soil Science Society of America Journal*, 85(4), 1037–1049. doi: 10.1002/saj2.20248
- Kariminejad, N., Hosseinalizadeh, M., & Pourghasemi, H. R. (2022). Digital soil mapping of soil bulk density in loess derived-soils with complex topography. *Computers in Earth and Environmental Sciences: Artificial Intelligence and Advanced Technologies in Hazards and Risk Management*, 593–599. doi: 10.1016/B978-0-323-89861-4.00018-X
- Kavian, A., Azmoodeh, A., & Solaimani, K. (2014). Deforestation effects on soil properties, runoff and erosion in northern Iran. *Arabian Journal of Geosciences*, 7(5), 1941–1950. doi: 10.1007/S12517-013-0853-1
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. doi: 10.18637/jss.v028.i05
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. *Applied Predictive Modeling*, 1–600. doi: 10.1007/978-1-4614-6849-3/COVER
- Li, Y., & Liu, D. (2024). Effects of under-forest economic activities on soil water repellency, soil hydraulic properties and preferential flow in karst forests. *Ecohydrology and Hydrobiology*, 24(1), 128–153. doi: 10.1016/j.ecohyd.2023.12.006
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Lipton, Z. C. (2016). The Mythos of Model Interpretability. *Communications of the ACM*, 61(10), 35–43. doi: 10.1145/3233231
- Lombardo, L., Saia, S., Schillaci, C., Mai, P. M., & Huser, R. (2017). Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks. *Geoderma*, 318, 148–159. doi: 10.1016/j.geoderma.2017.12.011
- Maleki, S., Khormali, F., Chen, S., Pourghasemi, H. R., & Hosseinalizadeh, M. (2022). Digital soil mapping of organic carbon at two depths in loess hilly region of Northern Iran. *Computers in Earth and Environmental Sciences: Artificial Intelligence and Advanced Technologies in Hazards and Risk Management*, 467–475. doi: 10.1016/B978-0-323-89861-4.00033-6
- Mao, J., Li, Y., Zhang, J., Zhang, K., Ma, X., Wang, G., & Fan, L. (2022a). Organic carbon and silt determining subcritical water repellency and field capacity of soils in arid and semi-arid region.

- Frontiers in Environmental Science*, 10, 1031237. doi: 10.3389/FENVS.2022.1031237/BIBTEX
- Mao, J., Li, Y., Zhang, J., Zhang, K., Ma, X., Wang, G., & Fan, L. (2022b). Organic carbon and silt determining subcritical water repellency and field capacity of soils in arid and semi-arid region. *Frontiers in Environmental Science*, 10, 1031237. doi: 10.3389/FENVS.2022.1031237/BIBTEX
- Mataix-Solera, J., & Doerr, S. (2004). Hydrophobicity and aggregate stability in calcareous topsoils from fire-affected pine forests in southeastern Spain. *Geoderma*.
- Mehta, V., Hasanvand, S., Sepahvand, A., Sihag, P., Beiranvand, N., & Singh, B. (2024). A benchmark comparison of AI-based modeling of soil infiltration rates. *Journal of Hydroinformatics*, 26(12), 3060–3079. doi: 10.2166/hydro.2024.086
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, 352, 251–267. doi: 10.1016/j.geoderma.2019.06.016
- Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, 16. doi: 10.1016/J.GEODRS.2018.E00198
- Pahlavanrad, M., Toomanian, N., & Khormali, F. (2017). Digital soil mapping. *Journal of land Management*, 4(2), 97–114. doi: 10.22092/lmj.2017.109482 [In Persian]
- Raheem, A. M., & Omar, N. Q. (2021). Investigation of distinctive physico-chemical soil correlations for Kirkuk city using spatial analysis technique incorporated with statistical modeling. *International Journal of Geo-Engineering*, 12(1), 1–21. doi: 10.1186/S40703-021-00147-2/FIGURES/9
- Sepahvand, A., Golkarian, A., Billa, L., Wang, K., Rezaie, F., Panahi, S., Samadianfard, S., & Khosravi, K. (2022). Evaluation of deep machine learning-based models of soil cumulative infiltration. *Earth Science Informatics* 2022 15:3, 15(3), 1861–1877. doi: 10.1007/S12145-022-00830-7
- Sepehri, N., Hajabbasi, M. A., Afyuni, M., & Lichner, L. (2017). Soil water repellency changes with depth and relationship to physical properties within wettable and repellent soil profiles. *Journal of Hydrology and Hydromechanics*, 65(1), 99–104. doi: 10.1515/JOHH-2016-0055
- Shakesby, R. A., Doerr, S. H., & Walsh, R. P. D. (2000). The erosional impact of soil hydrophobicity: current problems and future research directions. *Journal of Hydrology*, 231–232, 178–191. doi: 10.1016/S0022-1694(00)00193-1
- Tarek, Z., Elshewey, A. M., Shohieb, S. M., Elhady, A. M., El-Attar, N. E., Elseuofi, S., & Shams, M. Y. (2023). Soil Erosion Status Prediction Using a Novel Random Forest Model Optimized by Random Search Method. *Sustainability* 2023, Vol. 15, Page 7114, 15(9), 7114. doi: 10.3390/SU15097114
- Wadoux, A. M. J. C., Saby, N. P. A., & Martin, M. P. (2023). Shapley values reveal the drivers of soil organic carbon stock prediction. *SOIL*, 9(1), 21–38. doi: 10.5194/SOIL-9-21-2023
- Wang, D., Regentova, E., Muthukumar, V., Berli, M., & Harris, F. C. (2024). A machine learning framework to measure Water Drop Penetration Time (WDPT) for soil water repellency analysis. *Machine Learning with Applications*, 18, 100595. doi: 10.1016/J.MLWA.2024.100595
- Wang, J., Wang, W., Ren, X., Wu, Q., Chai, X., Qu, Y., Xu, X., & Du, F. (2025). Aliphatic carbon regulates soil water repellency in a chronosequence of grassland enclosure in the Loess Hilly Region. *Soil and Tillage Research*, 246, 106356. doi: 10.1016/J.STILL.2024.106356
- Wang, T., Wedin, D., & Zlotnik, V. A. (2009). Field evidence of a negative correlation between saturated hydraulic conductivity and soil carbon in a sandy soil. *Water Resources Research*, 45(7). doi: 10.1029/2008wr006865
- Weber, P. L., Hermansen, C., Norgaard, T., Pesch, C., Moldrup, P., Greve, M. H., Müller, K., Arthur, E., & de Jonge, L. W. (2021). Moisture-dependent Water Repellency of Greenlandic Cultivated Soils. *Geoderma*, 402. doi: 10.1016/J.GEODERMA.2021.115189
- Yang, Y., & Mei, G. (2022). A Deep Learning-Based Approach for a Numerical Investigation of Soil–Water Vertical Infiltration with Physics-Informed Neural Networks. *Mathematics* 2022, Vol. 10, Page 2945, 10(16), 2945. doi: 10.3390/MATH10162945
- Zhang, Q., Yang, J., & Kong, Y. (2025). Effects of tung oil and its induced hydrophobicity on the cracking behavior of purple soil exposed to wetting–drying cycle conditions. *CATENA*, 256, 109112. doi: 10.1016/J.CATENA.2025.109112
- Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V., & Mataix-Beneyto, J. (2008). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology and Biochemistry*, 40(7), 1923–1930. doi: 10.1016/J.SOILBIO.2008.04.003