



Publisher: Soil and Water Research Institute of Iran

Journal of Land Management

<https://lmj.areeo.ac.ir>

Digital mapping of soil pH using Landsat 8 satellite images and auxiliary data by machine learning models in Badr Watershed, Kurdistan Province

Zarinibahador, M. * 

PhD in Soil Science; Soil and Water Research Expert; Tehran Agricultural and Natural Resources Research and Education Center; Agricultural Research, Education and Extension Organization (AREEO); Varamin; Iran. Moslem.zarini@ymail.com.

Article Info

Article Type: Research Article

Received: November 23, 2024

Accepted: May 25, 2026

Keywords:

Random Forest, geomorphology, Multiple linear regression, Model evaluation

Corresponding author's email address:

Moslem.zarini@ymail.com

Extended Abstract

Objectives

Soil pH is a key chemical property that governs nutrient availability, microbial activity, and overall soil fertility, thereby exerting a significant influence on plant growth and productivity. Accurate spatial prediction of soil pH is essential for site-specific soil management and precision agriculture. The present study aims: 1) to prepare a digital map of the spatial distribution of soil pH in the 0–25 cm soil layer at Badr Watershed, southern Qorveh County, using environmental covariates derived from terrain attributes, remote sensing data, and geopedological information; 2) to evaluate the predictive performance of machine learning and statistical models, including Random Forest (RF), Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbor (KNN), and Multiple Linear Regression (MLR) used for soil pH estimation; 3) to identify the most influential environmental variables controlling the spatial variability of soil pH; and 4) to determine the modeling approach capable of producing the most accurate and reliable digital soil pH maps to support precision agriculture and sustainable soil management.

Material and Methods

This study was conducted in Badr Watershed located in the southern part of Qorveh County, Kurdistan Province, Iran, covering an area of approximately 6,700 ha. The objective was to develop a digital soil pH map of the 0–25 cm soil layer by integrating environmental covariates with machine learning and statistical modeling approaches. A geopedological map was first generated in a Geographic Information System (GIS) environment based on the Zink geopedological approach using geological and topographic information. Soil sampling locations were subsequently selected using the Latin Hypercube Sampling (LHS) technique to ensure representative coverage of the environmental variability across the study area. A total of 125 surface soil samples (0–25 cm) were collected and soil pH was determined in saturated soil paste using a calibrated pH meter following standard laboratory procedures.

A comprehensive set of environmental covariates was prepared to represent the soil-forming factors, including terrain attributes derived from a Digital Elevation Model (DEM), remote sensing indices extracted from Landsat 8 Imagery, and geopedological variables. These covariates were used as predictor variables for digital soil mapping.

Soil pH was modeled using five predictive approaches, including Random Forest (RF), Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbor (KNN), and Multiple Linear Regression (MLR), implemented in the R statistical software environment. Model performance was evaluated using both 10-fold cross-validation and 5-fold random validation procedures. Predictive accuracy was assessed based on the coefficient of determination (R^2), root mean square error (RMSE), and other relevant statistical performance indices. The resulting prediction maps were subsequently generated to characterize the spatial distribution of soil pH across the study area.

Results

The environmental covariates exhibited varying degrees of influence on the spatial prediction of soil pH. Variable importance analysis identified geomorphology as the most influential predictor, followed by watershed network base level, carbonate index, slope aspect, relative slope position, slope length factor (LS), and curvature index.

The the above-mentioned five statistical and machine learning models (namely, MLR, RF, ANN, DT, and KNN) were evaluated in terms of their predictive performance using both 10-fold cross-validation and 5-fold random validation procedures. The results revealed differences in prediction accuracy among the models evaluated.

Based on the 10-fold cross-validation results, the Multiple Linear Regression (MLR) model achieved the highest predictive performance, with a coefficient of determination (R^2) of 0.698 and a root mean square error (RMSE) of 0.190, indicating its strong capability of soil pH estimation across the study area. In contrast, the 5-fold random validation results identified the K-Nearest Neighbor (KNN) model as the most accurate and precise approach for soil pH prediction.

Comparison of the validation approaches further indicated that ensemble or combined prediction strategies would generally outperform individual models. In this regard, integrating the prediction outputs from multiple models improved the overall reliability and spatial accuracy of the digital soil pH maps generated, resulting in a more robust characterization of soil pH variability across the study watershed.

Conclusion

This study demonstrated the effectiveness of integrating environmental covariates with statistical and machine learning approaches for digital mapping of soil pH in Badr Watershed. The results confirmed that terrain- and geopedology-related variables played fundamental roles in explaining the spatial variability of soil pH, highlighting the importance of incorporating multiple environmental factors into digital soil mapping frameworks.

Model evaluation revealed that predictive performance varied depending on the validation strategy adopted. This is evidenced by the fact that Multiple Linear Regression (MLR) provided the highest predictive accuracy under the 10-fold cross-validation scheme, whereas the K-Nearest Neighbor (KNN) model performed best under the 5-fold random validation approach. These findings indicate that no single model is universally superior and that model performance depends on both the characteristics of the validation procedure and the spatial structure of the dataset.

Furthermore, the superior performance of combined prediction approaches suggests that, compared to individual models, ensemble modeling can improve the reliability and spatial accuracy of digital soil pH maps. The soil pH maps thus generated provide valuable spatial information that can support site-specific soil management, precision agriculture, and sustainable land-use planning in the study area.

Overall, the integration of geopedological information, terrain derivatives, remote sensing data, and advanced predictive models yields a robust framework for digital soil pH mapping. Future studies may be recommended to investigate hybrid and ensemble machine learning techniques, incorporate additional environmental covariates and higher-resolution remote sensing data, and evaluate the transferability of the modeling framework proposed herein to other regions with different soil and environmental conditions.

Cite this Article: Zarinibahador, M., 2026. Digital mapping of soil pH using Landsat 8 satellite images and auxiliary data by machine learning models in Badr Watershed, Kurdistan Province. *Journal of Land Management*, 14(1), PP.55-73.



DOI: <https://doi.org/10.22092/lmj.2026.368002.377>

نقشه‌برداری رقومی اسیدیته خاک با استفاده از تصاویر ماهواره‌ای لندست ۸ و داده‌های کمکی توسط مدل‌های یادگیری ماشین در حوزه آبخیز بدر، استان کردستان

مسلم زرینی بهادر* 

کارشناس تحقیقات خاک و آب مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران،

ایران. moslem.zarini@ymail.com

اطلاعات مقاله	چکیده
نوع مقاله: پژوهشی	
دریافت: ۱۴۰۳/۱۰/۳	
پذیرش: ۱۴۰۵/۳/۴	
واژه‌های کلیدی:	
جنگل تصادفی، ژئومورفولوژی، رگرسیون خطی چندگانه، ارزیابی مدل	
آدرس ایمیل نویسنده مسئول: moslem.zarini@ymail.com	
	pH خاک یکی از ویژگی‌های مهم خاک است که معیاری از اسیدی یا قلیایی بودن خاک می‌باشد و عمیقاً بر دسترسی به مواد مغذی و فعالیت میکروبی تأثیر می‌گذارد و مستقیماً بر رشد و سلامت گیاه تأثیر دارد. محصولات مختلف در محدوده‌های pH خاص رشد می‌کنند و حفظ سطح pH بهینه تضمین می‌کند که مواد مغذی ضروری به راحتی در دسترس گیاهان باشند. مطالعه حاضر با هدف نقشه‌برداری رقومی pH خاک با استفاده از متغیرهای کمکی محیطی و تصاویر ماهواره لندست ۸ و مدل‌های پیش‌بینی کننده و معرفی بهترین مدل‌ها، در حوزه آبخیز بدر در جنوب شهرستان قروه انجام گرفت. برای انجام این پژوهش در مرحله اول، نقشه‌ی ژئومورفولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئوپدولوژی زینک در محیط سامانه اطلاعات جغرافیائی ترسیم گردید. در مرحله دوم، محل ۱۲۵ خاکرخ مطالعاتی بر اساس تکنیک ابر مکعب لاتین تعیین گردید، و pH در گل اشباع توسط دستگاه پ‌هاش‌متر اندازه‌گیری شد. متغیرهای کمکی شامل مشتقات مدل رقومی ارتفاع، شاخص‌های سنجش از دور دریافتی از ماهواره لندست ۸ و نقشه ژئوپدولوژی بودند که انتخاب متغیرهای کمکی مناسب با استفاده از روش تجزیه مؤلفه‌های اصلی (PCA) انجام گرفت. در مرحله سوم، مدل‌سازی انجام، نقشه‌های رقومی کلاس‌ها و ویژگی‌های خاک تهیه گردید و ارزیابی مدل‌ها صورت گرفت. متغیرهای کمکی مهم در پیش‌بینی مقدار pH خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، سطح‌مبنای شبکه آبراهه‌ای، شاخص کرنات، جهت شیب، موقعیت نسبی شیب، عامل طول شیب و شاخص انحنا. پیش‌بینی pH توسط مدل‌های نزدیک‌ترین همسایه K، تحلیل درخت تصمیم، شبکه عصبی مصنوعی، جنگل تصادفی و رگرسیون خطی چندگانه صورت گرفت. در میان مدل‌های استفاده‌شده برای پیش‌بینی pH، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، مدل رگرسیون خطی چندگانه با ضریب تبیین ۰/۶۹۸ و ریشه دوم متوسط مربعات خطا ۰/۱۹۰ از بیشترین دقت برای پیش‌بینی برخوردار بود. بر اساس روش اعتبارسنجی کافلد ۵ تصادفی، دقت و صحت مدل نزدیک‌ترین همسایه بهترین عملکرد را در پیش‌بینی pH خاک دارا بوده است. به طور کلی، بر اساس روش‌های ارزیابی کافلد ۱۰ مکانی و کافلد ۵ تصادفی، می‌توان بیان داشت که روش‌های ترکیبی دارای قابلیت بیشتری برای پیش‌بینی pH می‌باشند. این بدین معنی است که ترکیب نتایج پیش‌بینی سایر مدل‌ها می‌تواند نقشه‌هایی با دقت بالاتر را تولید کند.

استناد: زرینی بهادر، مسلم، ۱۴۰۵. نقشه‌برداری رقومی اسیدیته خاک با استفاده از تصاویر ماهواره‌ای لندست ۸ و داده‌های کمکی توسط مدل‌های یادگیری

ماشین در حوزه آبخیز بدر، استان کردستان. نشریه مدیریت اراضی، ۱۴ (۱)، صص. ۵۵-۷۳.



DOI: <https://doi.org/10.22092/lmj.2026.368002.377>

مقدمه

pH بر خواص شیمیایی و بیولوژیکی خاک تأثیر می‌گذارد و ارتباط نزدیکی با سطح حاصلخیزی خاک، فعالیت‌های میکروبی و جانوری، نسبت C/N و تشکیل هوموس دارد (Meng et al., 2019; Liu et al., 2022; Zhao et al., 2022). pH خاک در ارتباط با عناصر غذایی (Zhang et al., 2017) و ریزمغذیها (Sillanpaa, 1982) از اهمیت ویژه‌ای برخوردار است و متغیر کلیدی مؤثر بر رشد گیاه می‌باشد (Gentili et al., 2018).

تخمین توزیع مکانی pH خاک اهمیت زیادی در پایش و مدیریت کیفیت خاک و اراضی دارد. با توجه به توزیع مکانی آن، به نظر می‌رسد pH خاک تحت تأثیر ارتفاع زمین (He et al., 2016) قرار می‌گیرد و همبستگی مکانی با دیگر عناصر خاک، مانند آهن را نشان می‌دهد (Tziachris et al., 2017). به‌طور کلی، تغییرپذیری مکانی ویژگی‌های خاک از یک نقطه به نقطه دیگر باعث گردیده‌است که تخمین‌های نقطه‌ای با استفاده از نقشه‌های سنتی خاک از دقت مناسبی برخوردار نباشند. علاوه بر آن، نقشه‌برداری سنتی خاک، نمی‌تواند اطلاعات کمی دقیقی در مورد توزیع ویژگی‌های خاک که به‌طور عمده در مطالعات محیطی به‌کار می‌روند را فراهم نماید (McBratney et al., 2003). مشکلات موجود در نقشه‌های سنتی خاک، نیاز شدید به اطلاعات دقیق از خاک با تفکیک مکانی بالا، رشد سریع فناوری اطلاعات همگام با اطلاعات حاصل از سنجش از دور و مدل‌های رقومی ارتفاع، موجب بروز و استفاده از روش‌های نوین نقشه‌برداری خاک شده‌است (McBratney et al., 2003). رویکردهای نقشه‌برداری رقومی خاک (DSM^۱) به‌طور گسترده برای تجزیه و تحلیل مکانی ویژگی‌ها/کلاس‌های خاک پذیرفته شده‌است، این رویکردها، سریع و مقرون به صرفه بوده و می‌توانند پیش‌بینی کمی همراه با عدم قطعیت ارائه دهند (Lu et al., 2023).

این روش‌ها، تشکیل خاک را تابعی از آب و هوا (C)، موجودات زنده (O)، توپوگرافی (R)، مواد مادری (P)، سن (A) و خاک و موقعیت مکانی توصیف می‌کنند. نقشه‌برداری رقومی خاک عبارت از ایجاد و جمع‌آوری سیستم‌های اطلاعات مکانی خاک با استفاده از روش‌های مشاهدات میدانی و آزمایشگاهی که با داده‌های محیطی از طریق ارتباطات کمی همراه شده‌اند. نقشه‌برداری رقومی خاک منجر به ایجاد خروجی به صورت نقشه رستری تخمین، همراه با عدم قطعیت پیش‌بینی می‌شود. افزایش دسترسی به داده‌های مکانی مانند مدل رقومی ارتفاع و تصاویر ماهواره‌ای، افزایش نیروی محاسباتی برای پردازش داده‌ها، توسعه ابزارهای داده‌کاوی و سامانه اطلاعات جغرافیایی و افزایش تقاضای جهانی به داده‌های مکانی دارای ارزیابی عدم قطعیت از عوامل تأثیرگذار در موفقیت نقشه‌برداری رقومی خاک بوده‌اند (pahlavanrad et al., 2017).

در حال حاضر، یادگیری ماشین، روش‌های آماری کلاسیک، روش‌های زمین‌آمار، روش‌های مبتنی بر سنجش‌ازدور و روش‌های هیبریدی به‌طور گسترده در نقشه‌برداری رقومی خاک استفاده می‌شود. در مقایسه با روش‌های زمین‌آمار، یادگیری ماشین می‌تواند به‌طور مؤثر رابطه غیرخطی بین ویژگی/کلاس خاک و متغیرهای محیطی را حل کند و نتایج خوبی در مطالعات قبلی به دست آمده‌است. از جمله مطالعات انجام شده در ایران، نقشه‌برداری رقومی اجزای بافت خاک با استفاده از رویکردهای یادگیری ماشین در دشت مرو دشت استان فارس بوده‌است که نتایج آن بیانگر عملکرد بهتر و دقت بیشتر مدل جنگل تصادفی، نسبت به مدل‌های SVR و Co-Kriging برای پیش‌بینی ذرات شن، رس و سیلت است (Khosravani et al., 2024). جنگل تصادفی (RF)، رگرسیون خطی، رگرسیون لجستیک، نزدیک‌ترین همسایه K (Leo., 2001) درخت تصمیم، و رگرسیون بردار

خاک را می‌توان با استفاده از الگوریتم‌های انتخاب ویژگی بهبود بخشید.

در کشاورزی مناطق خشک و نیمه‌خشک مانند حوضه آبخیز بدر، دستیابی به عملکرد پایدار و حفاظت از منابع خاک مستلزم شناخت دقیق و مکانی شده ویژگی‌های خاک است. در این میان، واکنش خاک (pH) به عنوان یک ویژگی کلیدی شیمیایی، نقشی اساسی در تعیین حاصلخیزی خاک، فراهمی عناصر غذایی برای گیاه، و فعالیت موجودات زنده خاک ایفا می‌کند. انتخاب و تهیه نقشه‌های رقومی pH خاک به دلایل متعددی برای مدیریت بهینه کشاورزی در این مناطق حیاتی است. اولاً، pH خاک به طور مستقیم قابلیت جذب عناصر غذایی پرمصرف و کم‌مصرف را کنترل می‌کند. در محدوده pH بهینه (عموماً ۶/۵ تا ۷/۵ برای بسیاری از گیاهان زراعی)، حلالیت و دسترسی عناصر اصلی مانند فسفر و ریزمغذی‌هایی نظیر آهن، روی و منگنز در حداکثر خود قرار دارد. در مناطق خشک و نیمه‌خشک به دلیل تبخیر بالا و مواد مادری غنی از کربنات‌ها، خاک‌ها اغلب دارای pH قلیایی هستند. در این شرایط، فراهمی عناصر ریزمغذی به شدت کاهش یافته و منجر به بروز کمبودهای تغذیه‌ای و افت عملکرد می‌شود. بنابراین، آگاهی از توزیع مکانی pH می‌تواند مدیریت کودهای شیمیایی را بهینه کرده و از مصرف بی‌رویه و ناکارآمد آنها جلوگیری نماید (Vasundhara et al., 2025).

ثانیاً، پایش تغییرات pH در طول زمان، که توسط مدل‌های رقومی میسر شده است، اطلاعات ارزشمندی درباره روندهای تخریب خاک مانند اسیدی شدن ناشی از مصرف بی‌رویه کودهای نیتروژنی یا آبیاری با کیفیت پایین فراهم می‌کند. این امر به ویژه در اکوسیستم‌های حساس و شکننده مناطق خشک که تعادل شیمیایی خاک به سرعت تحت تأثیر مدیریت ناصحیح قرار می‌گیرد، از اهمیت مضاعفی برخوردار است. به طور خلاصه، انتخاب پارامتر pH به عنوان یکی از لایه‌های اصلی اطلاعاتی در نقشه‌برداری رقومی خاک، نه تنها به درک بهتر از توزیع

پشتیبان (Smola, A.J. & Schölkopf., 2004) (SVR) از جمله الگوریتم‌های نماینده یادگیری ماشین هستند.

همچنین زراعت پیشه و همکاران (Zeraatpisheh et al., 2020) مقاله‌ای جامع درباره وضعیت گذشته، وضعیت فعلی و چشم‌انداز آینده نقشه‌برداری خاک در ایران ارائه می‌دهد که نشان می‌دهد DSM در ایران نسبتاً جدید است و با افزایش استفاده از داده‌های RS و GIS در حال توسعه است؛ همچنین بیشتر مطالعات در سطح استان‌ها انجام شده‌اند و هنوز نقشه‌های گسترده با وضوح بالا برای تمام کشور فراهم نیست.

تزی آکریس و همکاران (Tziachris et al., 2020) در مطالعه‌ای با عنوان مقایسه مدل‌های یادگیری ماشین و روش‌های زمین آماری ترکیبی با استفاده از متغیرهای محیطی و خاک برای پیش‌بینی pH خاک به این نتیجه رسیدند که مدل‌های یادگیری ماشین و به ویژه الگوریتم جنگل تصادفی نتایج بهتری را در پیش‌بینی مکانی pH خاک داشته‌اند. آتورک و همکاران (ÖZTÜRK et al., 2024) در پژوهشی با عنوان نقشه‌برداری رقومی pH و EC خاک و مقایسه تحلیلی نگرش‌های کریجینگ و یادگیری ماشین نشان دادند که مدل‌های یادگیری ماشین نسبت به روش‌های کریجینگ، نتایج بهتر و دقیق‌تری را در پیش‌بینی مکانی pH خاک ارائه دادند.

لیو و همکاران (Lu et al, 2023) در مطالعه‌ای با هدف نقشه‌برداری رقومی کربنات‌ها و pH خاک در مقیاس اروپا از نه الگوریتم یادگیری ماشین استفاده کردند؛ که نتایج این مطالعه نشان داد مدل‌های یادگیری ماشین غیرخطی بهتر از مدل‌های خطی عمل کردند و مدل درختان بسیار تصادفی (ERT) بهترین دقت ارزیابی را برای پیش‌بینی pH خاک ارائه داد. ژائو و همکاران (Zhao et al., 2023) در مطالعه‌ای با عنوان نقشه‌برداری رقومی pH خاک بر اساس الگوریتم‌های یادگیری ماشین در شرق چین، برای پیش‌بینی pH خاک از داده‌های کمکی شامل شاخص‌های اقلیمی، توپوگرافی و پوشش گیاهی استفاده کردند و به این نتیجه رسیدند که دقت نقشه‌برداری رقومی

مواد و روشها

معرفی منطقه مورد مطالعه

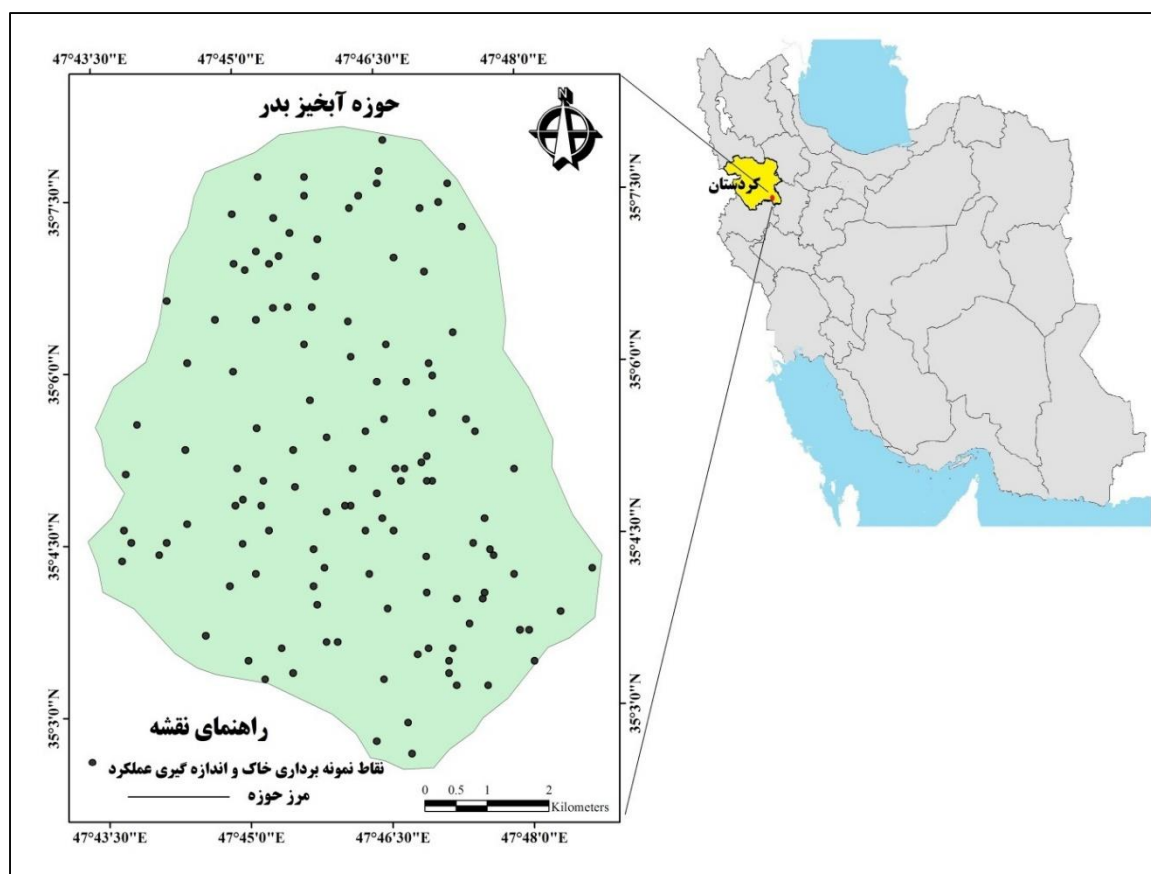
اراضی مورد مطالعه با وسعتی حدود ۶۷۰۰ هکتار در جنوب شهرستان قروه و بین مختصات جغرافیائی ۴۷ درجه و ۴۳ دقیقه و ۴۷ درجه و ۴۹ دقیقه طول شرقی و ۳۵ درجه و ۲ دقیقه و ۳۵ درجه و ۸ دقیقه عرض شمالی واقع شده است (شکل ۱). مطابق با آمار طولانی‌مدت هواشناسی سی‌ساله ایستگاه هواشناسی شهرستان قروه، متوسط درجه حرارت سالانه‌ی هوا ۱۲/۰۷ درجه سانتی‌گراد و میانگین بارندگی سالانه ۳۴۵/۸ میلی‌متر است. الگوی بارش، مدیترانه‌ای بوده و بخش اعظم بارندگی در فصول سرد سال اتفاق می‌افتد. بر اساس اقلیم نمای آمبرژه، اقلیم این ناحیه نیمه‌خشک و سرد است. در منطقه مورد مطالعه، رژیم رطوبتی خاک زریک (Xeric) و رژیم حرارتی آن مزیک (Mesic) تعیین شده است.

از نظر زمین‌شناسی نیز، آخرین توالی رسوبی به‌ظاهر پیوسته مزوزوئیک- ترشیاری، شیل‌ها و ماسه‌سنگ‌های دارای عدسی‌های آهکی به همراه سنگ‌های آتش‌فشانی هستند که به‌طور عمده همساز بر روی رسوبات کرتاسه بالائی (پالئوسن) قرار دارند (Hariri, 1995). اصلی‌ترین رخداد کوهزایی در این منطقه مربوط به کوهزایی کیمرین میانی است که سبب دگرگونی سنگ‌های رسوبی- آتش‌فشانی (تریاس- ژوراسیک) گردیده و حاصل آن تشکیل مرمَر، آمفیبولیت، شیست و فیلیت است (Hosseini, 1996). از لحاظ ژئومورفولوژی، حوضه آبخیز مورد مطالعه شامل کوه و اراضی دشت دامنه‌ای (پیدمونت) است (شکل ۲). کاربری‌های رایج اراضی کشاورزی، گندم دیم و آبی، جو دیم و آبی، نخود، لوبیا، یونجه، شبدر و... است.

مکانی حاصلخیزی خاک کمک می‌کند، بلکه پایه و اساس برنامه‌ریزی برای دستیابی به کشاورزی پایدار، افزایش بهره‌وری و حفظ محیط زیست در مناطق خشک و نیمه‌خشک را تشکیل می‌دهد (Filippi et al., 2018).

این پژوهش برای نخستین بار الگوی مکانی اسیدیته خاک را در مقیاس حوضه آبخیز بدر با استفاده از رویکرد نقشه‌برداری رقومی خاک ارائه می‌دهد و خلأ اطلاعاتی موجود در منطقه را برطرف می‌کند. همچنین از مدل‌های پیشرفته یادگیری ماشین برای پیش‌بینی اسیدیته خاک، نسبت به روش‌های سنتی زمین‌آمار و میان‌یابی کلاسیک استفاده شده که از نوآوری‌های این مقاله محسوب می‌شود.

مهم‌ترین خلأ پژوهش‌های نقشه‌برداری رقومی pH خاک، محدودیت متغیرهای کمکی مرتبط با فرآیندهای کنترل‌کننده pH خاک از جمله نادیده گرفتن کربنات کلسیم معادل، انتخاب ویژگی ناکافی و غیرسیستماتیک و نادیده گرفتن دانش خاکشناسی، کمبود داده‌های عمقی و زمانی، استفاده از Cross-validation ساده در اعتبارسنجی‌ها و فقدان نقشه‌های عدم قطعیت در مقیاس مدیریتی است؛ بنابراین این مطالعه در حوضه آبخیز بدر شهرستان قروه در استان کردستان با هدف نقشه‌برداری رقومی pH خاک با استفاده از متغیرهای کمکی محیطی و تصاویر ماهواره لندست ۸ و مدل‌های پیش‌بینی کننده همسایه K، تحلیل درخت تصمیم، شبکه عصبی مصنوعی، جنگل تصادفی و رگرسیون خطی چندگانه و معرفی بهترین مدل‌ها برای نقشه‌برداری رقومی pH خاک انجام گرفت.



شکل ۱ - موقعیت منطقه مطالعاتی و پراکنش مکانی نقاط نمونه برداری براساس روش cLHS
Figure 1 - Location of the study area and spatial distribution of sampling points based on the cLHS method

روش اجرا

نسخه SAGA GIS 2.1.0 مشتقات مختلف مرتبه‌های اول و دوم مدل رقومی ارتفاع تهیه شد. شاخص‌های سنجش از دور نیز با استفاده از داده‌های باندهای ۱ تا ۷ ماهواره‌ی لندست ۸ (سال ۲۰۱۳) محاسبه شدند (جدول ۱). در مرحله‌ی دوم، محل ۱۲۵ خاکرخ مشاهداتی تعیین شد، شناسایی خاک‌ها در صحرا انجام گردید، از لایه‌های مختلف خاک نمونه برداشت شد، سپس اندازه‌گیری اسیدیته خاک در گل اشباع انجام شد (Bohn et al., 2001) و در مرحله سوم، مدل‌سازی انجام، نقشه‌ی ویژگی‌های خاک تهیه و ارزیابی مدل‌ها صورت گرفت.

متغیرهای کمکی مورد استفاده در پیش‌بینی pH خاک

متغیرهای کمکی مورد استفاده در پیش‌بینی pH خاک که از مدل رقومی ارتفاع و تصاویر ماهواره‌ای استخراج گردیده در جدول ۱ نشان داده شده‌اند.

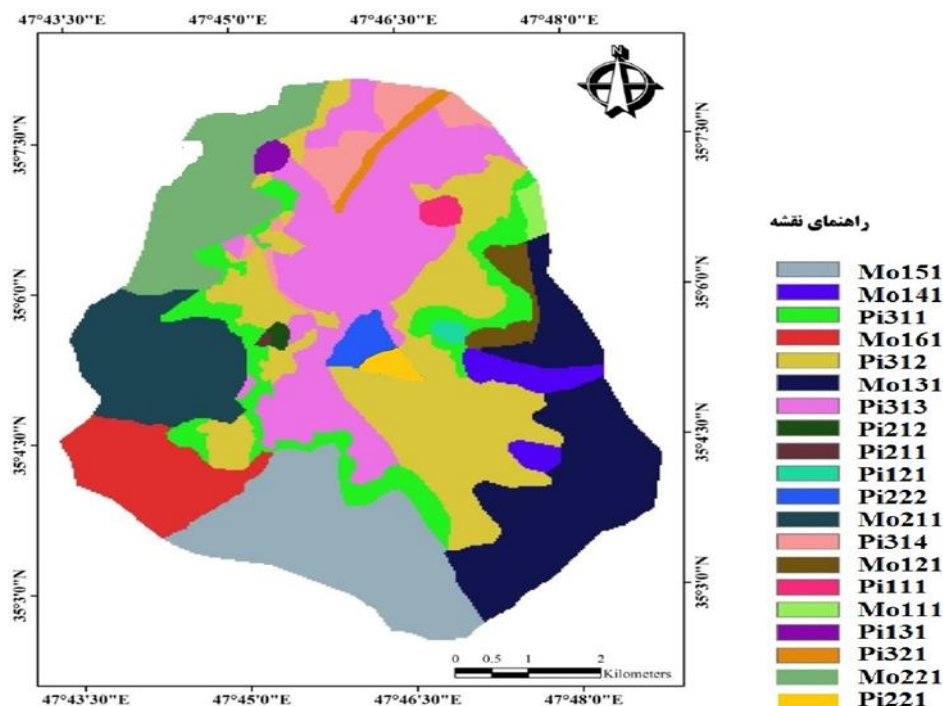
در پژوهش حاضر، از سه منبع برای تهیه داده‌های کمکی بهره گرفته شد: تصاویر ماهواره‌ای، مدل رقومی ارتفاع و نقشه ژئوپدولوژی. برای اجرای این پژوهش مراحل زیر، طبق شکل ۳ انجام گرفت: در مرحله‌ی اول، داده‌های کمکی از جمله: تصاویر ماهواره‌ای لندست ۸ سال ۲۰۱۳ و مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر، از سامانه سازمان نقشه‌برداری کشور دریافت گردید و تغییر اندازه پیکسل‌ها به منظور یکسان نمودن اندازه آن‌ها با دستور Resampling و تابع نزدیک‌ترین همسایه در نرم افزار Arc GIS انجام شد. نقشه‌ی ژئوپدولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئوپدولوژی زینک در محیط سامانه اطلاعات جغرافیائی ترسیم گردید. در این مرحله، با استفاده از این مدل و زیربرنامه‌های تعریف شده‌ای استاندارد موجود در نرم افزار Arc-GIS10 و ساگا

جدول ۱- متغیرهای کمکی مورد استفاده جهت پیش‌بینی اسیدیته خاک حوزه آبخیز بدره
Table 1- Auxiliary variables used to predict soil acidity in the Badr watershed

فاکتور SCORPAN SCORPAN factor	نماد پارامتر Parameter symbol	تعریف Description	مشتقات اولیه و ثانویه پستی و بلندی، مستخرج از مدل رقومی ارتفاع/ عرض‌های باندی مختلف، برگرفته از تصاویر ماهواره‌ای Primary and secondary derivatives of the digital elevation model
Relief	Elev	ارتفاع از سطح دریای آزاد Elevation above sea level	ارتفاع Elevation
Relief	BL Channel network base level	ارتفاع سطح پایه شبکه آبراهه Channel network base level elevation	سطح پایه شبکه آبراهه Channel network base level
Relief	Modified catchment area	تجمعات جریان آب و پارامترهای مربوطه Water flow gatherings and related parameters	مساحت حوضه‌های آبخیز اصلاح شده Modified catchment area
Relief	Valley depth	شاخص نسبت پهنای دره به عمق دره Valley width to valley depth index	عمق دره Valley depth
Relief	MRVBF	میزان همواری کف دره The flatness of the valley floor	شاخص همواری دره با درجه تفکیک بالا Multi-Resolution Valley Bottom Flatness Index
Relief	MRRTF	میزان همواری نوک قله The flatness of the ridge top	شاخص همواری قله برآمدگی با درجه تفکیک بالا Multi-Resolution Ridge Top Flatness Index
Climate	WI	معیاری از تجمع رطوبت An index of moisture accumulation	شاخص خیسگی Wetness Index
Relief	Slope (S)	زاویه بین سطح تماس و سطح افقی در یک نقطه درسته زمین The angle between the contact and the horizontal surfaces at a real point of the earth	شیب Slope
Relief	Aspect (Asp)	بیشترین سرعت تغییر در ارتفاع در هر سلول از مدل ارتفاعی The highest rate of change in altitude in each cell from the altitude model	جهت Aspect
Relief	Relative slope position	شیب بالائی یا میانی یا پائینی Upper, middle or lower slope	موقعیت نسبی شیب Relative slope position
Relief	Catchment slope	میزان گرادیان مسیر حرکت Gradient of the movement path	شیب حوضه Catchment slope

ادامه جدول ۱-			
Relief	Catchment aspect	جهت شیب مسیر حرکت Slope aspect of the movement path	جهت شیب حوضه Catchment aspect
Relief	Catchment height	میزان ارتفاع مسیر حرکت height of the movement path	ارتفاع حوضه Catchment height
Relief	Slope curvature	انحنای در صفحه عمودی خط جریان Curve on the vertical plane of the flow line	انحنای شیب Slope curvature
Relief	Slope surface curvature	اندازه همگرایی و واگرایی جریان Degree of convergence and divergence of the current	انحنای سطحی شیب Slope surface curvature
Relief	Vertical distance of channel network	فاصله‌ی عمودی تا شبکه آبراهه را نشان می‌دهد Indicates the vertical distance to the waterway network	فاصله‌ی عمودی تا شبکه آبراهه Vertical distance of channel network
Relief	LS Factor	طول شیب را نشان می‌دهد Indicates the length of the slope	طول شیب Slope length
Relief	Convergence index	مقدار پستی‌وبلندی را نشان می‌دهد Shows the value of the post and elevation	شاخص انحنا Convergence index
Relief	Cross-sectional curvature	مقدار انحنا در جهت عمود بر شیب اصلی دامنه The amount of curvature in the direction	انحنای عمود بر شیب اصلی (انحنای عرضی) Cross-sectional curvature
Relief	Longitudinal curvature	مقدار انحنا در جهت شیب اصلی دامنه Curvature in the direction of the overall slope	انحنای طولی Longitudinal curvature
Organisms	NDVI	$(NIR^* - RED^{**}) / (NIR + RED)$	شاخص گیاهی تفاضلی نرمال شده Normalized differential vegetation index
Organisms	SAVI	$[(NIR - RED) / (NIR + RED + L^{***})] \times [1 + L^v]$	شاخص گیاهی تعدیل‌کننده اثر خاک Soil-adjusted vegetation index
Remote Sensing	Clay Index	B5 / B7	شاخص رس Clay Index
Remote Sensing	Carbonate Index	B3 / B2	شاخص کربنات Carbonate Index

^۱- اثر خاک در شاخص SAVI با پارامتری به اسم L مشخص می‌شود. برای پوشش گیاهی ضعیف این مقدار صفر است که در این صورت این شاخص با شاخص NDVI برابر است. برای پوشش گیاهی غنی مقدار این شاخص برابر با عدد یک و به صورت پیشفرض معادل با عدد ۰/۵ در نظر گرفته می‌شود.



شکل ۲- ژئوپدولوژی حوزه آبخیز بدر شهرستان قروه، استان کردستان
Figure 2- Geopedology of Badr watershed, Qorve city, Kurdistan province

جدول ۲- تحلیل ژئوپدولوژی در حوزه آبخیز بدر

Table 2- Geopedological analysis in the Badr watershed

Landscape ^۱	Relief ^۲ /Molding ^۳	Lithology ^۴	Landform ^۵
Mountain (Mo)	High Mountain (Mo1)	TRamf (Mo11)	Slope facet complex (Mo111)
		TRcom (Mo12)	Slope facet complex (Mo121)
		Dim (Mo13)	Slope facet complex (Mo131)
		Gbm (Mo14)	Slope facet complex (Mo141)
		Md (Mo15)	Slope facet complex (Mo151)
		RJph,l (Mo16)	Slope facet complex (Mo161)
	Low Mountain (Mo2)	TRJm (Mo21)	Slope facet complex (Mo211)
		TRJl1 (Mo22)	Slope facet complex (Mo221)
Piedmont (Pi)	Hill (Pi1)	TRJd (Pi11)	Slope facet complex (Pi111)
		TRcom (Pi12)	Slope facet complex (Pi121)
		Q _{u1} (Pi13)	Slope facet complex (Pi131)
	Alluvial fan (Pi2)		Apical (Pi211)
		Qf1 (Pi21)	Distal (Pi212)
			Apical (Pi221)
		Qf2 (Pi22)	Distal (Pi222)
	Glacis (Pi3)		Slope > 20% (Pi311)
			20% > Slope > 10% (Pi312)
		Qt3 (Pi31)	10% > Slope > 5% (Pi313)
			5% > Slope > 0% (Pi314)
		Qal (Pi32)	Slope facet complex (Pi321)

^۱- آخرین، همگونی‌ترین و کوچکترین واحد ژئوپدولوژیکی است که با شکل، دینامیک و سن مشخص می‌شود (زینک، ۱۹۸۹)

^۲- به جنس سنگ‌های سخت و رخساره تشکیلات پوششی نرم اشاره دارد (زینک، ۱۹۸۹)

^۳- شکلی از زمین است که تعیین کننده آن مجموعه توپوگرافی و زمین شناسی است (زینک، ۱۹۸۹)

^۴- شکلی از زمین است که تعیین کننده آن فرآیندهای مورفوتنیکی یا شرایط مورفواقلیمی است (زینک، ۱۹۸۹)

^۵- قسمت بزرگی از زمین که در آن یا ریلیف‌های مشابه تکرار شده اند یا مجموعه ای از ریلیف‌های غیر مشابه دیده می‌شوند (زینک، ۱۹۸۹)

نمونه برداری، تجزیه و تحلیل آزمایشگاهی

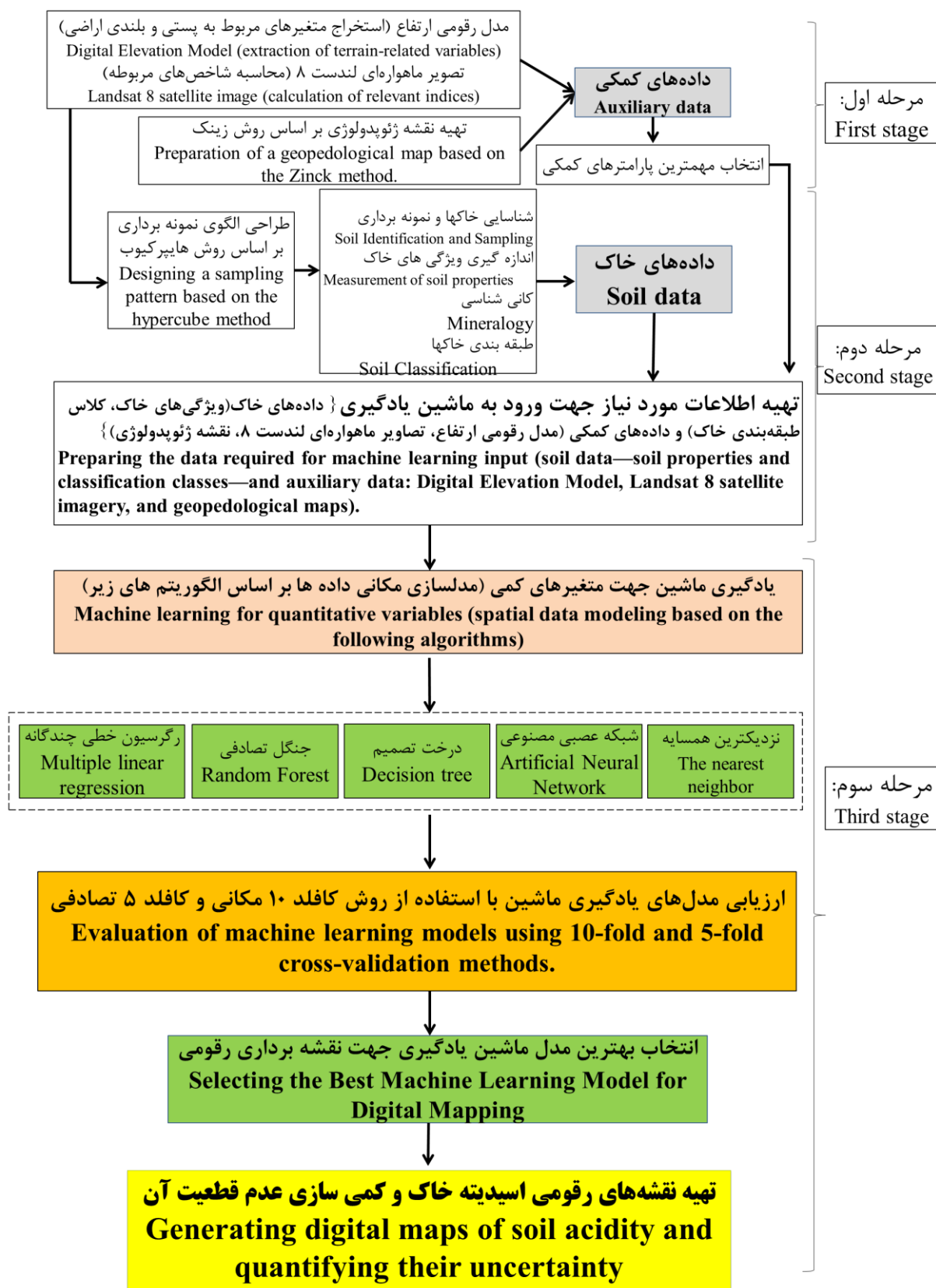
در این تحقیق از نرم افزار R برای مشخص کردن نقاط حفر خاکرخ‌ها بهره گرفته شد و داده‌های مورد نیاز برای ورود به نرم‌افزار شامل مدل رقومی ارتفاع و نقشه ژئودولوژی و شاخص‌های برگرفته از تصاویر ماهواره لندست ۸ بودند. برای انجام این کار، ابتدا بسته نرم‌افزاری clhs در نرم‌افزار R فراخوانی شد، سپس داده‌های متغیرهای کمکی حوزه آبخیز بدر از جمله DEM، تصاویر ماهواره‌ای لندست ۸ و نقشه ژئودولوژی وارد محیط نرم‌افزار شدند و ۱۲۵ نقطه نمونه‌برداری به روش ابر مکعب لاتین انتخاب گردید (شکل ۱).

موقعیت تمامی خاکرخ‌ها نیز توسط سیستم موقعیت‌یاب جهانی (GPS) ثبت شد و خاکرخ‌ها بر

اساس راهنمای تشریح و نمونه‌برداری خاک‌ها در صحرا (Schoeneberger et al., 2012) تشریح و از تمامی افق‌های آن نمونه‌برداری صورت گرفت. نمونه‌های خاک، پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده و از الک ۲ میلی‌متری عبور داده شدند. سپس pH در گل اشباع توسط دستگاه pH متر اندازه‌گیری گردید.

پردازش اطلاعات

متوسط وزنی عمق ۲۵ سانتیمتری pH، برای نقشه‌برداری رقومی خاک بکار رفت. به منظور محاسبه متوسط وزنی، مقدار هر ویژگی در افق‌های خاک، ضخامت هر افق و ضریب وزنی عمق (Sys et al., 1991) مورد استفاده قرار گرفت.



شکل ۳- مراحل اجرای پژوهش

Figure 3- Research implementation steps

یادگیری ماشین

الف- مدل سازی مکانی

در این تحقیق از رویکردهای شبکه عصبی مصنوعی، تحلیل درخت تصمیم، جنگل تصادفی، نزدیک ترین همسایه K و رگرسیون خطی چندگانه برای مدل سازی مکانی pH خاک در نرم افزار R استفاده شد.

رگرسیون خطی چندگانه

با استفاده از این روش می توان همزمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. در صورت وجود متغیرهای مستقل $X_1, X_2, \dots, X_n$ اگر بخواهیم ارتباط خطی بین آنها و متغیر Y که وابسته به آنهاست ایجاد کنیم رابطه ۱ بین آنها برقرار است:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (1)$$

که در این رابطه، از مقادیر $a_1, a_2, \dots, a_n$ به عنوان ضرایب رگرسیون یاد می شود. این ضرایب، ضرایبی نامشخص اند که در حقیقت مسئول برآورد پارامتر پاسخ هستند (Were et al., 2015).

شبکه های عصبی مصنوعی

در روش شبکه های عصبی مصنوعی، سه لایه ی ورودی، پنهان و خروجی وجود دارد و این لایه ها با یکدیگر در ارتباط هستند. در این روش، از طریق وزن دهی داده های ورودی توسط فرد و در حقیقت آموزش سیستم، وزن نهایی به دست می آید (Salehi & Khademi, 2008).

مدل میانگین گیری نزدیک ترین همسایه K

میانگین گیری نزدیک ترین همسایه K، ارزش پاسخ برای یک مشاهده داده شده را با استفاده از پاسخ های مشاهداتی در همسایه های محلی پیش بینی می کند. می توان از این مدل با یک پاسخ قطعی برای طبقه بندی و یک پاسخ پیوسته برای پیش بینی استفاده کرد (Nemes et al., 2006; Zolfaghari et al., 2013).

مدل تحلیل درخت تصمیم

مدل درختی طبقه بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر است متغیرهای کمی یا متغیرهای طبقه بندی شده را بر اساس مجموعه ای از متغیرهای پیش بینی کننده کمی و کیفی برآورد کند، در این روش مجموعه ای از شرط های منطقی به صورت یک الگوریتم با ساختار درختی برای طبقه بندی یا پیش بینی کمی بک متغیر به کار می رود. ایجاد درخت تصمیم دو مرحله دارد. مرحله اول، مرحله ایجاد و رشد درخت است. این مرحله شامل پیوند و انشعاب است. مرحله دوم، مرحله توقف و هرس است. هدف این مرحله به حداقل رساندن خطای پیش بینی است (Scull et al., 2005; Moore et al., 1991).

مدل جنگل تصادفی

جنگل های تصادفی در حقیقت مجموعه هایی از درخت های پیش بینی کننده با احتمال و پراکندگی یکسان هستند. اساس این روش، انتخاب خطای داده های بزرگ تر به عنوان خطای اصلی و ایجاد همبستگی بین خطاهای دیگر داده ها به ترتیب است (Breiman., 2001).

ارزیابی مدل های یادگیری ماشین

در این پژوهش، جهت ارزیابی مدل های یادگیری ماشین از دو روش بهره گرفته شد: ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی. در این نوع اعتبارسنجی ها، داده ها به K زیرمجموعه دسته بندی شدند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و $K-1$ تای دیگر برای آموزش بکار رفتند. این روال، K بار تکرار شد و همه داده ها دقیقاً یک بار برای آموزش و یک بار برای اعتبارسنجی بکار رفتند. در نهایت، میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده شد. لازم به ذکر می باشد که تفاوت ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی، در انتخاب مجموعه داده های خاک جهت آموزش و ارزیابی می باشد. در روش ارزیابی متقابل تصادفی، داده های به طور تصادفی به دو دسته تقسیم می شوند:

توان دوم خطاها زمانی که از متغیرهای مستقل (x) ها استفاده نشود.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

نتایج و بحث

خلاصه آماری داده‌ها

جدول ۳، آمار توصیفی خصوصیات فیزیکی و شیمیایی خاک‌ها در منطقه مورد مطالعه را نشان می‌دهد. ویژگی‌های خاک که ضریب تغییرات آن‌ها بیش از ۳۵، بین ۱۵ و ۳۵ و کمتر از ۱۵ است، به ترتیب دارای تغییرات زیاد، متوسط و کم می‌باشند (Wilding & Dress., 1983). بر این اساس؛ آهک، کربن آلی و سنگریزه به ترتیب با ضریب تغییرات ۱۷۴/۴، ۶۲/۴ و ۶۱/۳ درصد، از تغییرات زیاد؛ شوری، ظرفیت تبادل کاتیونی، رس، سیلت و شن؛ به ترتیب با ضریب تغییرات ۲۵/۴، ۱۶/۹، ۳۰/۴، ۲۳/۵ و ۲۰/۳ درصد؛ دارای تغییرات متوسط و pH، با ضریب تغییرات ۳/۶ درصد؛ کمترین میزان تغییرات را در خاک دارد.

متغیرهای کمکی مهم در پیش‌بینی pH خاک

متغیرهای کمکی مهم در پیش‌بینی pH خاک، در منطقه مورد مطالعه به ترتیب اهمیت نسبی در جدول ۴ ارائه شده‌اند. متغیرهای کمکی مهم در پیش‌بینی مقدار pH خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، سطح‌مبنای شبکه آبراه‌ای، شاخص کربنات، جهت شیب، موقعیت نسبی شیب، عامل طول شیب و شاخص انحنا. درجه‌بندی اهمیت با به‌کارگیری الگوریتم ReliefAttributeEval نرم‌افزار WEKA مشخص گردید (جدول ۴).

داده‌هایی که برای آموزش به‌کار می‌روند و داده‌هایی که برای اعتبارسنجی یا آزمون مورد استفاده قرار می‌گیرند. در چنین شرایطی ممکن است داده‌های آموزش و آزمون از لحاظ مکانی در کنار یکدیگر در فاصله بسیار نزدیک قرار گیرند. اما در روش ارزیابی متقابل مکانی، در ابتدا منطقه مطالعاتی به تعدادی قطعه، تقسیم‌بندی شده و سپس داده‌های قرار گرفته در یک قطعه جهت ارزیابی مدل استفاده شده و بقیه قطعه‌ها برای آموزش مدل به‌کار می‌روند. در چنین شرایطی، می‌توان مطمئن شد که مدل آموزش دیده، در کل منطقه مورد ارزیابی قرار گرفته است (Valavi et al., 2018).

در این پژوهش، برای بیان کمی میزان خطا از معیارهای معمول در یادگیری ماشین استفاده گردید. بدین منظور، جهت بررسی دقت و صحت مقادیر پیش‌بینی شده داده‌های کمی خاک مانند pH توسط مدل‌های یادگیری ماشین شاخص‌هایی از قبیل جذر میانگین مربع خطا (Root Mean Square Error: RMSE) و جذر میانگین مربع خطای نسبی (Relative RMSE: RMSE%)، ضریب تعیین (R^2) و ضریب همبستگی (r) محاسبه گردیدند.

$$RMSE = \sqrt{\frac{\sum (x'_i - x_i)^2}{n}} \quad (2)$$

در این رابطه، x'_i مقدار برآورد شده در نقطه x_i مقدار مشاهده در نقطه x_i و n تعداد نقاط است.

$$RMSE\% = \frac{RMSE}{\bar{x}} * 100 \quad (3)$$

در این رابطه، $\bar{x}$ میانگین مقادیر واقعی مشاهدات است.

$$R^2 = SSR/SST \quad (4)$$

پارامتر SSR را مجموع توان دوم رگرسیون نامید و کاهش در مجموع توان دوم خطاها به خاطر استفاده از متغیرهای مستقل (x) را نشان می‌دهد. و SST مجموع

جدول ۳- آمار توصیفی برخی ویژگی‌های فیزیکی و شیمیایی خاک و عملکرد اندازه‌گیری شده

Table 3- Descriptive statistics of some physical and chemical characteristics of the soil and measured performance

ویژگی‌های فیزیکی و شیمیایی خاک Soil physical and chemical properties	واحد Unit	حداقل Min	حداکثر Max	میانگین Mean	انحراف معیار SD	ضریب تغییرات CV
اسیدیته pH	-	6.8	8	7.5	0.3	3.6
هدایت الکتریکی EC	dS.m ⁻¹	0.1	0.4	0.2	0.1	25.5
کربن آلی OC	%	0.1	2.3	0.6	0.4	62.4
کربنات کلسیم معادل CCE	%	0	65	9.8	17	174.4
ظرفیت تبادل کاتیونی CEC	cmol(+).kg ⁻¹	8.9	26.5	14.3	2.4	16.9
رس Clay	%	9	43	22.5	6.8	30.4
سیلت Silt	%	15.9	48.6	29.1	6.8	23.5
شن Sand	%	27.1	70	48.3	9.8	20.3
سنگریزه Gravel	درصد حجمی Volume %	1.5	48	17.1	10.4	61.3

جدول ۴- متغیرهای کمکی مهم در پیش‌بینی pH خاک

Table 4- Important covariates in predicting soil pH

ویژگی (Feature)	متغیرهای کمکی مهم به ترتیب اهمیت (Important covariates in order of importance)
pH	Geomorphology, Channel network base level, Carbonate index, Aspect, Relative slope position, LS Factor, Convergence index

نتایج ارزیابی دقت مدل‌های یادگیری ماشین

نتایج ارزیابی صحت پیش‌بینی pH خاک، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، در جدول ۵ ارائه شده است. این پیش‌بینی توسط مدل‌های نزدیک‌ترین همسایه K، تحلیل درخت تصمیم، شبکه عصبی مصنوعی، جنگل تصادفی و رگرسیون خطی چندگانه صورت گرفته است. در میان مدل‌های استفاده‌شده برای پیش‌بینی pH، مدل رگرسیون خطی چندگانه با ضریب تعیین ۰/۶۹۸ و ریشه دوم متوسط مربعات خطا ۰/۱۹۰ از بیشترین دقت برای پیش‌بینی برخوردار بوده است و مدل جنگل تصادفی

برای پیش‌بینی pH خاک، جایگاه دوم را از نظر قدرت پیش‌بینی به خود اختصاص داده است که با نتایج واندانا و همکارانش (Vandana et al., 2024) هم‌راستا می‌باشد. واندانا و همکاران (Vandana et al., 2024) بیان کردند که در بین تکنیک‌های نقشه‌برداری رقومی خاک، مدل جنگل تصادفی (RF) با مقادیر RMSE کمتر، ME و R² بالاتر برای پیش‌بینی مکانی pH (RMSE=0.014، ME=0.28، R²=0.81) و EC خاک (ME=0.022، R²=0.73) از بهترین عملکرد برخوردار بوده است.

جدول ۵- نتایج ارزیابی صحت پیش‌بینی pH، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی

Table 5- The results of evaluating the accuracy of pH prediction, using the 10-fold cross validation method

مدل Model	RMSE	MAE	R ²
KNN	0.258	0.218	0.239
DTA	0.194	0.156	0.562
ANN	0.193	0.154	0.607
RF	0.191	0.154	0.626
MLR	0.190	0.144	0.698

درخت تصمیم، رگرسیون خطی چندگانه و نزدیک‌ترین همسایه K؛ مدل نزدیک‌ترین همسایه K با ضریب تعیین ۰/۹۴۵۴ و ریشه دوم متوسط مربعات خطا ۰/۰۶۴۷ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

در جدول ۵، دقت پیش‌بینی pH از بالا به طرف پائین افزایش می‌یابد. کمترین دقت پیش‌بینی به مدل نزدیک‌ترین همسایه K اختصاص یافته است. این در حالی است که با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی (جدول ۶)، از بین مدل‌های شبکه عصبی مصنوعی، تحلیل

جدول ۶- نتایج ارزیابی صحت پیش‌بینی pH، با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی

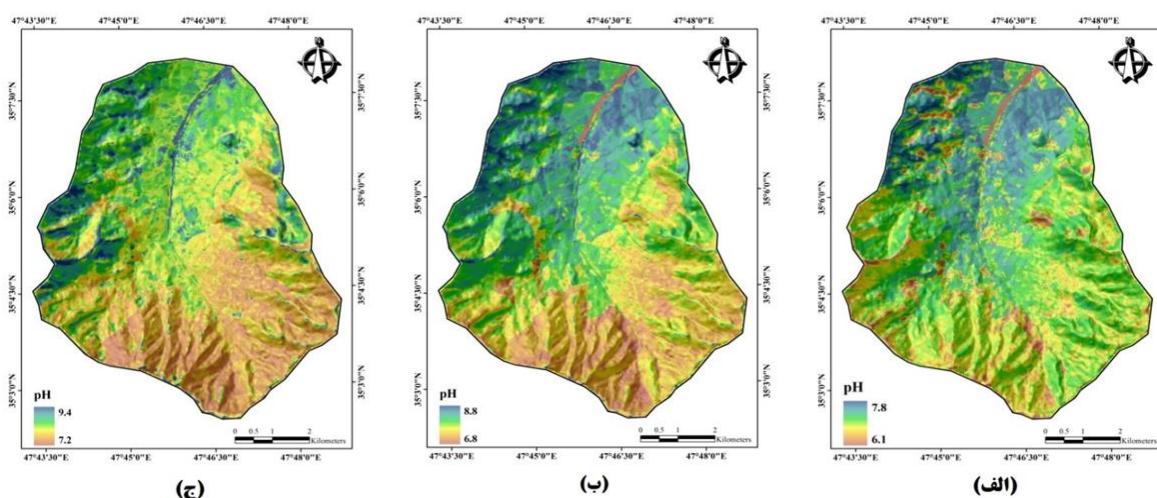
Table 6- The results of evaluating the accuracy of pH prediction, using the 5-fold random validation method

مدل Model	RMSE	^N RMSE=RMSE/Mean real	R ²	ME
ANN	0.1042	0.01384	0.7785	-0.00598
DTA	0.11508	0.01528	0.821	0
MLR	0.6088	0.08086	0.7413	0
KNN	0.0647	0.00859	0.9454	0.00792

۴ نشان داده شده‌اند. محدوده تغییرات pH خاک در این حوزه آبخیز بین ۶/۸۵ و ۸ است.

نقشه پیش‌بینی اسیدیته خاک

نقشه‌های پیش‌بینی pH با استفاده از مدل رگرسیون خطی چندگانه به عنوان مدل با دقت بالا در شکل



شکل ۴- نقشه‌ی پیش‌بینی pH خاک با استفاده از مدل رگرسیون خطی چندگانه. با روش اعتبارسنجی کافلد ۱۰ مکانی

Figure 4- Soil pH prediction map using multi linear regression model with 10-fold cross validation methods

میزان pH خاک وجود دارد به این صورت که با افزایش شاخص کربنات میزان pH هم بیشتر شده است.

از نظر ترتیب اهمیت، جهت شیب چهارمین عامل تأثیرگذار بر مقدار pH خاک به شمار می‌رود. در شمال، شمال غرب و غرب حوضه که خاک‌ها آهکی هستند، شیب‌های جنوبی از شیب‌های شرقی pH بیشتری دارند. این پدیده را این گونه می‌توان توجیه نمود که شیب‌های جنوبی از شیب‌های شرقی خشک‌ترند. این خشکی بیشتر باعث می‌گردد تا آهک موجود در خاک کمتر از خاک شسته شود و در نتیجه، pH خاک بالاتر باقی بماند. چن و همکاران (Chen et al., 1997) در مطالعه‌ای با عنوان روابط ویژگی‌های خاک با توپوگرافی و پوشش گیاهی در یک جنگل بارانی نیمه گرمسیری در جنوب تایوان به این نتیجه رسیدند که جهت شیب عامل اصلی تأثیرگذار بر pH خاک است.

در جنوب، جنوب شرق و شرق که خاک‌ها غیرآهکی می‌باشند، شیب‌های شمالی نسبت به شیب‌های غربی، از pH کمتری برخوردار می‌باشند. در این قسمت از حوضه نیز شیب‌های شمالی از شیب‌های غربی مرطوب‌ترند و این رطوبت بیشتر، شسته شدن بیشتر کاتیون‌های بازی و کاهش بیشتر pH را باعث می‌گردد، چیتیری و همکاران (Chytrý et al., 2007) نیز در مطالعه‌ای دریافتند که pH خاک با افزایش بارندگی و افزایش رطوبت یک روند نزولی نشان می‌دهد که تأیید کننده نتایج به دست آمده در این پژوهش می‌باشد. ژانگ و همکاران (Zhang et al., 2019) نشان دادند که اهمیت نسبی متغیرهای مورد مطالعه نسبت به pH خاک با افق متفاوت است. محدوده دمای سالانه (ATR) و شاخص رطوبت زمین (TWI) عوامل مهمی بودند که تغییرپذیری pH خاک را در افق A کنترل کردند. مواد مادری، بارش، ATR و TWI متغیرهای مهم کنترل کننده pH در افق B بودند. مواد اولیه، ATR و بارش عوامل کلیدی کنترل کننده pH در افق C بودند که این نتایج نیز با نتایج مطالعات ما

در شکل ۴؛ نقشه الف، پیش‌بینی با دامنه پراکندگی پنج درصد کمتر از میانگین، نقشه ب، پیش‌بینی مقدار میانگین، و نقشه ج، پیش‌بینی با دامنه پراکندگی پنج درصد بیشتر از میانگین را نشان می‌دهند. مهم‌ترین عاملی که تعیین‌کننده مقدار pH خاک‌هاست، سنگ مادر است که با مطالعات ژائو و همکارانش (Zhao et al., 2023) مطابقت دارد. خاک‌هایی که از سنگ‌های حاوی کربنات کلسیم منشأ گرفته‌اند، دارای pH بالاتر و آن‌هایی که از سنگ مادر آذرین (بدون کربنات کلسیم) به وجود آمده‌اند، از کمترین pH در منطقه برخوردار می‌باشند. یک سری دیگر از محققان گزارش دادند که تغییرات pH خاک تحت تأثیر مواد مادری و سنگ بستر قرار دارد. به عنوان مثال، خاک‌هایی که از ماسه‌سنگ‌های تریاس و ماسه‌های کواترنر ایجاد شده‌اند، مقادیر pH بسیار متفاوتی دارند (Gruba and Socha, 2016).

رویتر و همکاران (Reuter et al., 2008) گزارش دادند که توزیع مکانی pH خاک به شدت به ماهیت ماده مادری و سنگ بستر بستگی دارد. فابیان و همکاران (Fabian et al., 2014) مقادیر pH پایین خاک را بر روی سنگ‌های کریستالی و مقادیر pH بالا را در مناطقی که زیر آن سنگ آهک وجود دارد، گزارش نمودند که با مطالعات ما در این پژوهش کاملاً همخوانی دارد.

متغیرهای کمکی مهم در پیش‌بینی مقدار pH خاک به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، سطح‌مبنای شبکه آبراهه‌ای، شاخص کربنات، جهت شیب، موقعیت نسبی شیب، طول شیب و شاخص انحنا. بعد از سنگ مادر، ژئومورفولوژی مهم‌ترین متغیر کمکی در پیش‌بینی pH خاک به شمار می‌رود. در شمال، شمال غرب و غرب حوضه که خاک‌ها آهکی هستند، از اراضی پای کوهی به طرف کوه، pH افزایش نشان داد و در جنوب، جنوب شرق و شرق که خاک‌ها غیرآهکی می‌باشند، از اراضی پای کوهی به طرف کوه، pH کاهش می‌یابد. شاخص کربنات در پیش‌بینی pH خاک در درجه سوم اهمیت قرار دارد و رابطه مستقیمی بین شاخص کربنات و

مبنی بر تأثیر مستقیم مواد مادری بر pH خاک همخوانی دارد و نتایج در یک راستا می‌باشند.

توجه به متغیرهای انسانی و مدیریتی در ارزیابی عدم قطعیت نقشه‌های پیش‌بینی شده pH خاک در حوضه‌های آبخیز با کاربری و مدیریت اراضی متنوع، از اهمیت ویژه‌ای برخوردار است. در یک حوضه ۶۷۰۰ هکتاری با واحدهای ژئومورفولوژی متفاوت و کاربری‌های ترکیبی شامل مرتع، کشاورزی دیم و آبی، متغیرهای مدیریتی مانند عمق و نوع شخم، میزان و نوع کودهای مصرفی، و سیستم‌های آبیاری، منابع مهمی از تغییرپذیری مکانی و در نتیجه عدم قطعیت در پیش‌بینی‌های مکانی pH محسوب می‌شوند. این متغیرها که اغلب در چارچوب سستی SCORPAN به‌طور صریح در نظر گرفته نمی‌شوند، می‌توانند با تأثیرگذاری بر فرایندهای شیمیایی خاک در مقیاس محلی، موجب انحراف از روابط پیش‌بینی‌شده بر اساس عوامل طبیعی (مانند مواد مادری و توپوگرافی) گردند. وایس و همکاران (Vaysse et al., 2017) با تأکید بر اهمیت تلفیق عدم قطعیت در نقشه‌های حاصلخیزی خاک برای مدیریت اراضی در مقیاس محلی، نشان دادند که عدم قطعیت ناشی از متغیرهای مدیریتی و کاربری اراضی می‌تواند تأثیر قابل‌توجهی بر قضاوت‌های تصمیم‌گیری در سطح واحدهای مدیریتی مانند آبخیزها داشته باشد. این موضوع در مناطق خشک و نیمه‌خشک که مدیریت نادرست شخم و کوددهی می‌تواند به سرعت منجر

به تخریب کیفیت خاک و تغییرات شدید pH گردد، از اهمیت مضاعفی برخوردار است و نادیده گرفتن آن در فرآیند مدل‌سازی می‌تواند به برآورد نادرست از قابلیت اراضی و ارائه توصیه‌های مدیریتی ناکارآمد منجر شود.

نتیجه‌گیری

به ترتیب کاهش اهمیت، ژئومورفولوژی، سطح شبکه آبراهه، شاخص کربنات، جهت شیب، موقعیت نسبی شیب، LS Factor و Convergence index مهم‌ترین عوامل مؤثر در پیش‌بینی pH خاک در منطقه مورد مطالعه می‌باشند. بر اساس روش اعتبارسنجی کافلد ۱۰ مکانی، بهترین مدل برای پیش‌بینی pH خاک، رگرسیون خطی چندگانه است. برای پیش‌بینی pH خاک، مدل جنگل تصادفی، از نظر قدرت پیش‌بینی، رتبه دوم را به خود اختصاص می‌دهد. به‌طور کلی، نزدیک‌ترین همسایه K، پائین‌ترین عملکرد را در پیش‌بینی pH خاک دارا بوده است. بر اساس روش اعتبارسنجی کافلد ۵ تصادفی، دقت و صحت مدل نزدیک‌ترین همسایه بهترین عملکرد را در پیش‌بینی pH خاک دارا بوده است.

تعارض منافع

در این مقاله تعارض منافعی وجود ندارد و این مسئله مورد تأیید نویسندگان مقاله است.

References

1. Angiulli, F., 2005, August. Fast condensed nearest neighbor rule. In *Proceedings of the 22nd international conference on Machine learning*, pp. 25-32. <https://doi.org/10.1145/1102351.1102355>
2. Bohn, H. L., & McNeal, B. L. O Connor, GA (2001). Soil Chemistry.
3. Breiman, L., 2001. Random forests. *Machine learning*, 45. pp.5-32. <https://doi.org/10.1023/A:1010933404324>
4. Chen, Z.S., Hsieh, C.F., Jiang, F.Y., Hsieh, T.H. and Sun, I.F., 1997. Relations of soil properties to topography and vegetation in a subtropical rain forest in southern Taiwan. *Plant Ecology*, 132, pp.229-241. <https://doi.org/10.1023/A:1009762704553>
5. Cheng-Jim, J.I., Yuan-He, Y.A.N.G., Wen-Xuan, H.A.N., Yan-Fang, H.E., Smith, J. and Smith, P., 2014. Climatic and edaphic controls on soil pH in alpine grasslands on the Tibetan Plateau, China: a quantitative analysis. *Pedosphere*, 24(1), pp.39-44. [https://doi.org/10.1016/S1002-0160\(13\)60078-8](https://doi.org/10.1016/S1002-0160(13)60078-8)

6. Chytrý M., Danihelka J., Ermakov N., Hájek M., Hájková P., Kočí M. Plant species richness in continental southern Siberia: effects of pH and climate in the context of the species pool hypothesis [J]. *Global Ecology and Biogeography*, 2007, 16(5): 668–678.
<https://doi.org/10.1111/j.1466-8238.2007.00320.x>
7. Fabian, C., Reimann, C., Fabian, K., Birke, M., Baritz, R., Haslinger, E. and GEMAS Project Team, 2014. GEMAS: Spatial distribution of the pH of European agricultural and grazing land soil. *Applied Geochemistry*, 48, pp.207-216.
<https://doi.org/10.1016/j.apgeochem.2014.07.017>
8. Filippi, P., Cattle, S. R., Bishop, T. F., Odeh, I. O., & Pringle, M. J. (2018). Digital soil monitoring of top-and sub-soil pH with bivariate linear mixed models. *Geoderma*, 322, 149-162. **<https://doi.org/10.1016/j.geoderma.2018.02.033>**
9. Gentili, R., Ambrosini, R., Montagnani, C., Caronni, S. and Citterio, S., 2018. Effect of soil pH on the growth, reproductive investment and pollen allergenicity of *Ambrosia artemisiifolia* L. *Frontiers in plant science*, 9, p.1335.
<https://doi.org/10.3389/fpls.2018.01335>
10. Gruba, P. and Socha, J., 2016. Effect of parent material on soil acidity and carbon content in soils under silver fir (*Abies alba* Mill.) stands in Poland. *Catena*, 140, pp.90-95.
<https://doi.org/10.1016/j.catena.2016.01.020>
11. Grunwald, S. 2006. Environmental soil-landscape modeling: Geographic information technologies and pedometrics. CRC Press, New York, 300p.
12. Hariri, A. 1995. An attitude on the origin of a group of different rocks in the Qorve area. Master's thesis, *Shahid Beheshti University*, Tehran. (In Persian).
13. He, X., Hou, E., Liu, Y. and Wen, D., 2016. Altitudinal patterns and controls of plant and soil nutrient concentrations and stoichiometry in subtropical China. *Scientific reports*, 6(1), p.24261. **<https://doi.org/10.1038/srep24261>**
14. Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E. and Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, pp.62-77.
<https://doi.org/10.1016/j.geoderma.2015.11.014>
15. Hosseini, M. 1996. Description of Geological Map 1:100000 Quarter Corners (Map Attachment), *Geological and Mineral Exploration Organization of the country*. (In Persian).
16. Khosravani, P., Baghernejad, M., Moosavi, A.A. and FallahShamsi, S.R., 2024. Digital Mapping of Soil Texture Particles with Machine Learning Models and Environmental Covariates. *Water & Soil*, 37(6). (In Persian).
<https://doi.org/10.22067/jsw.2023.84413.1331>
17. Leo, B., 2001. Random forests. *Machine learning*, 45, pp.5-23.
18. Liu, K., Liu, Z., Zhou, N., Shi, X., Lock, T.R., Kallenbach, R.L. and Yuan, Z., 2022. Diversity-stability relationships in temperate grasslands as a function of soil pH. *Land Degradation & Development*, 33(10), pp.1704-1717. **<https://doi.org/10.1002/ldr.4259>**
19. Lu, Q., Tian, S. and Wei, L., 2023. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. *Science of the Total Environment*, 856, p.159171. **<https://doi.org/10.1016/j.scitotenv.2022.159171>**
20. McBratney, A.B., Santos, M.M. and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117(1-2), pp.3-52. **[https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)**
21. Meng, C., Tian, D., Zeng, H., Li, Z., Yi, C. and Niu, S., 2019. Global soil acidification impacts on belowground processes. *Environmental Research Letters*, 14(7), p.074003. **<https://doi.org/10.1088/1748-9326/ab239c>**
22. Moore, I.D., Grayson, R.B. and Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), pp.3-30. **<https://doi.org/10.1002/hyp.3360050103>**
23. Moore, I.D., Gessler, P.E., Nielsen, G.A.E. and Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil science society of america journal*, 57(2), pp.443-452. **<https://doi.org/10.2136/sssaj1993.03615995005700020026x>**

24. Nemes, A., Rawls, W.J. and Pachepsky, Y.A., 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal*, 70(2), pp.327-336. <https://doi.org/10.2136/sssaj2005.0128>
25. ÖZTÜRK, M., KILIÇ, M. and GÜNAL, H., 2024. Digital Mapping of Soil pH and Electrical Conductivity: A Comparative Analysis of Kriging and Machine Learning Approaches. *MAS Journal of Applied Sciences*, 9(4), pp.1168-1185. <https://doi.org/10.5281/zenodo.14542860>
26. pahlavanrad, M. , Toomanian, N. and Khormali, F. (2017). Digital soil mapping. *Land Management Journal*, 4(2), 97-114. <https://doi.org/10.22092/lmj.2017.109482>
27. Reuter, H.I., Lado, L.R., Hengl, T. and Montanarella, L., 2008. Continental-scale digital soil mapping using European soil profile data: soil pH. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, 19(1), pp.91-102.
28. Salehi, M. H., and Khademi, H. 2008. Basics of soil mapping. Isfahan University Jihad Publications. 210 pages. (In Persian).
29. Schoeneberger, P.J., Wysocki, D.A. and Benham, E.C. eds., 2012. *Field book for describing and sampling soils*. Government Printing Office.
30. Scull, P., Franklin, J. and Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological modelling*, 181(1), pp.1-15. <https://doi.org/10.1016/j.ecolmodel.2004.06.036>
31. Sillanpää, M., 1982. Micronutrients and the nutrient status of soils: a global study (Vol. 48). Food & Agriculture Org..
32. Simonson, R.W., 1995. Factors of soil formation. A system of quantitative pedology. *Geoderma*, 68(4), pp.334-335. [https://ui.adsabs.harvard.edu/link_gateway/1995Geode..68..334S/doi:10.1016/0016-7061\(95\)90014-4](https://ui.adsabs.harvard.edu/link_gateway/1995Geode..68..334S/doi:10.1016/0016-7061(95)90014-4)
33. Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and computing*, 14, pp.199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
34. Soil Survey Staff. 2014. Soil Taxonomy: A basic systems of soil classification for making and interpreting soil surveys. *12th Edition*. NRCS. USDA.
35. Sreenivas, K., Dadhwal, V.K., Kumar, S., Harsha, G.S., Mitran, T., Sujatha, G., Suresh, G.J.R., Fyze, M.A. and Ravisankar, T., 2016. Digital mapping of soil organic and inorganic carbon status in India. *Geoderma*, 269, pp.160-173. <https://doi.org/10.1016/j.geoderma.2016.02.002>
36. SYS, C., Van Ranst, E. and DEBAVEYE, J., Land Evaluation. Part I: principles in land evaluation and crop production calculations. Agricultural Publications nr. 7, GADC, Brussels, Belgium, 1991.
37. Tziachris, P., Metaxa, E., Papadopoulos, F. and Papadopoulou, M., 2017. Spatial modelling and prediction assessment of soil iron using kriging interpolation with pH as auxiliary information. *ISPRS International Journal of Geo-Information*, 6(9), p.283. <https://doi.org/10.3390/ijgi6090283>
38. Tziachris, P., Aschonitis, V., Chatzistathis, T., Papadopoulou, M. and Doukas, I.J.D., 2020. Comparing machine learning models and hybrid geostatistical methods using environmental and soil covariates for soil pH prediction. *ISPRS International Journal of Geo-Information*, 9(4), p.276. <https://doi.org/10.3390/ijgi9040276>
39. Valavi, R., Elith, J., Lahoz-Monfort, J.J. and Guillera-Arroita, G., 2018. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv*, p.357798. [doi: https://doi.org/10.1101/357798](https://doi.org/10.1101/357798)
40. Vandana, N., Suresh, G., Mitran, T. and Mahadevappa, S.G., 2024. Digital Mapping of Soil pH and Electrical Conductivity Using Geostatistics and Machine Learning. *International Journal of Environment and Climate Change*, 14(2), pp.273-286. DOI: 10.9734/IJECC/2024/v14i23944
41. Vasu, D., Singh, S.K., Tiwary, P., Chandran, P., Ray, S.K. and Duraisami, V.P., 2016. Pedogenic processes and soil–landform relationships for identification of yield-limiting soil properties. *Soil Research*, 55(3), pp.273-284. <https://doi.org/10.1071/SR16111>

42. Vasu, D., Singh, S.K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V.P., Ramamurthy, V., Lalitha, M. and Kalaiselvi, B., 2017. Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management. *Soil and Tillage Research*, 169, pp.25-34. <https://doi.org/10.1016/j.still.2017.01.006>
43. Vasundhara, R., Dharumarajan, S., Hegde, R., Moharana, P. C., Kalaiselvi, B., Prakash, H., & Patil, N. (2025). Modeling Soil pH at regional scale using environmental covariates and machine learning algorithm. *Environmental Monitoring and Assessment*, 197(7), 799. <https://doi.org/10.1007/s10661-025-14254-5>
44. Vaysse, K., Heuvelink, G. B., & Lagacherie, P. (2017). Spatial aggregation of soil property predictions in support of local land management. *Soil Use and Management*, 33(2), 299-310. <https://doi.org/10.1111/sum.12350>
45. Were, K., Bui, D.T., Dick, Ø.B. and Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, 52, pp.394-403. <https://doi.org/10.1016/j.ecolind.2014.12.028>
46. Wilding, L. and Drees, L.R., 1983. Spatial variability and pedology. In *Developments in Soil Science* (Vol. 11, pp. 83-116). Elsevier. [https://doi.org/10.1016/S0166-2481\(08\)70599-3](https://doi.org/10.1016/S0166-2481(08)70599-3)
47. Zeraatpisheh, M., Jafari, A., Bodaghabadi, M. B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., & Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *Catena*, 188, 104424. <https://doi.org/10.1016/j.catena.2019.104424>
48. Zhang, Y., Biswas, A. and Adamchuk, V.I., 2017. Implementation of a sigmoid depth function to describe change of soil pH with depth. *Geoderma*, 289, pp.1-10. <https://doi.org/10.1016/j.geoderma.2016.11.022>
49. Zhang, Y.Y., Wu, W. and Liu, H., 2019. Factors affecting variations of soil pH in different horizons in hilly regions. *Plos one*, 14(6), p.e0218563. <https://doi.org/10.1371/journal.pone.0218563>
50. Zhao, X., He, C., Liu, W.S., Liu, W.X., Liu, Q.Y., Bai, W., Li, L.J., Lal, R. and Zhang, H.L., 2022. Responses of soil pH to no-till and the factors affecting it: A global meta-analysis. *Global Change Biology*, 28(1), pp.154-166. <https://doi.org/10.1111/gcb.15930>
51. Zhao, Z.D., Zhao, M.S., Lu, H.L., Wang, S.H. and Lu, Y.Y., 2023. Digital mapping of soil pH based on machine learning combined with feature selection methods in east China. *Sustainability*, 15(17), p.12874. <https://doi.org/10.3390/su151712874>
52. Zinck, J.K. 1989. Physiography and soil. Lecture notes for K 6 course. *Soil Division. ITC. Enschede, The Netherlands*. 156 p.
53. Zolfaghari, A. A. Tirgar Soltani, M. T. Dyck, M. and Weldeyohannes, A. 2013. Comparison of K-nearest neighbor and artificial neural network methods for predicting cation exchange capacity of soil. 50th anniversary Alberta soil science workshop. *Book of Abstracts*. 77-94. (In Persian). <https://dor.isc.ac/dor/20.1001.1.23221267.1392.3.1.5.2>