Research Article

Agricultural Engineering., 48(2) (2025) 205- 229 ISSN (E): 2588-526X

DOI: 10.22055/agen.2025.48843.1761 ISSN (P): 2588-5944

Digital mapping of soil physical and chemical properties using some machine learning algorithms and environmental variables in the Fereydan region, Isfahan Province

N.S. Emami[1], E. Chavoshi[2*], Sh. Ayoubi[3], N. Honarjoo[4] and M. Zeraatpisheh[5]

1. Ph.D. student, Department of Soil Science, College of Agriculture, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
2. Associate Professor, Department of Soil Science, College of Agriculture, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
3. Professor, Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, 841568311, Isfahan, Iran.
4. Associate Professor, Department of Soil Science, College of Agriculture, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
5. Associate Professor, Department of Soil Science, Gund Institute for Environment, University of Vermont, Burlington, USA.

Received: 22 January 2025 Accepted: 28 February 2025 *Corresponding Author: chavoshie@yahoo.com

Abstract

Introduction: Soil physical and chemical properties play critical roles in soil health, agricultural productivity, and natural resource management. These properties are key indicators for assessing soil sustainability and carbon storage. Understanding their spatial distribution is essential for effective land management and resource modeling, especially within the context of sustainable agriculture. The mean weight diameter of soil aggregates (MWD) is a crucial parameter that significantly influences the carbon cycle, soil structural stability, and agricultural productivity. Similarly, soil organic carbon (SOC) is vital for environmental sustainability. Digital soil mapping (DSM), which integrates data from environmental covariates and machine learning (ML) algorithms, is a powerful tool for predicting soil properties and enhancing sustainable land management practices. DSM techniques have proven highly effective in estimating soil parameters across diverse landscapes, reducing the need for extensive field surveys and enabling data-driven decision-making in precision agriculture. The primary objective of this study was to predict the spatial distribution of MWD and SOC using DSM. Additionally, this study aimed to rank the most influential environmental variables affecting these properties through ML-based modeling and to compare the performance of various ML methods.

Materials and Methods: This study was conducted in the Fereydan region (50º 6' E, 32º 56' N), Isfahan Province, Iran. We collected a total of 100 surface soil samples (0 to 20 cm depth) from various land uses, including irrigated croplands, dryland farms, and pastures, using a systematic sampling network. The samples were air-dried, sieved through a 2 mm mesh, and then analyzed in the laboratory. We employed six machine learning algorithms—artificial neural network (ANN),

© 2025 The Author(s). Published by Shahid Chamran University of Ahvaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

random forest (RF), cubist model, support vector machines (SVM), k-nearest neighbors (KNN), and boosted regression trees (BRT)—to relate environmental variables to the soil properties under investigation. For the modeling process, four datasets were used as predictor environmental covariates: remote sensing data, topography derived from a digital elevation model (DEM), soil properties, and classified land data. The models' accuracy was evaluated across four different scenarios, each utilizing one or more of these datasets. The most suitable environmental variables were selected through a multicollinearity test using the variance inflation factor (VIF). Finally, to model the two soil properties, the dataset was divided into calibration (70%) and validation (30%) subsets. The performance of the models was evaluated using various statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

Results and Discussion: The model validation results revealed that the KNN model achieved the highest prediction accuracy for MWD in the fourth scenario (R^2=0.59, RMSE=0.19, MAE=0.16). These metrics indicate a moderate yet significant ability of the KNN model to predict MWD. In contrast, the cubist model yielded the highest accuracy for SOC in the fourth scenario (R^2=0.78, RMSE=4.12, MAE=3.53). This suggests that the cubist model effectively captured the complex relationships between environmental variables and SOC, providing a robust framework for predicting soil carbon content. These findings underscore the strong potential of ML models to capture intricate interactions between soil properties and environmental factors, enabling precise spatial predictions of soil characteristics. The ability of these models to process large volumes of data and uncover hidden patterns is particularly valuable in soil science, where the relationship between soil properties and environmental variables can be complex and nonlinear. The results revealed that different environmental variables played varying roles in explaining the spatial variability of MWD and SOC. For MWD prediction, silt content emerged as the most influential variable, followed by VV and CHND, both of which significantly affected the spatial distribution of MWD. In contrast, for SOC prediction, the most important variables were the multi-resolution valley bottom flatness index (MRVBF) and clay content. These factors are critical for understanding the storage and cycling of organic carbon in soil, as both moisture retention and clay particles are known to influence the stabilization of soil organic matter. The relationship between these variables and SOC underscores the importance of considering soil texture and moisture dynamics when predicting soil carbon stocks.

Conclusion: The results of this study demonstrate that ML models, such as Cubist and KNN can effectively predict soil quality properties across different regions, particularly when combined with topographic and climatic data. These approaches provide an efficient and rapid method for generating digital soil maps and supporting sustainable land management. By leveraging multiple environmental datasets, ML-based digital soil mapping enhances modeling accuracy while reducing reliance on extensive soil sampling. These models are especially useful in areas with similar climates where full-scale soil sampling may not be feasible, offering a valuable tool for estimating soil quality attributes and improving agricultural productivity. Ultimately, the findings of this research can help policymakers and farmers identify critical areas and adopt optimal strategies to enhance soil quality and mitigate adverse environmental impacts. To further improve the accuracy and stability of predictions, future studies should incorporate high-resolution spatiotemporal data, advanced algorithms such as Cubist and KNN, and dynamic models for analyzing temporal variations. Additionally, investigating soil–water–plant interactions and assessing the impacts of climate change could play a crucial role in enhancing agricultural productivity and promoting the sustainable management of soil resources.

Keywords: Digital soil mapping. Machine learning. Environmental variables. Soil properties. Modeling accuracy. Feature selection.

مهندسی زراعی (مجله علمی کشاورزی) جلد ۴۸. شماره ۲، ۱۴۰۴

مقاله پژوهشی

نقشه‌برداری رقومی ویژگی‌های فیزیکی و شیمیایی خاک با استفاده از برخی الگوریتم‌های یادگیری ماشین و متغیرهای محیطی در منطقه فریدن، استان اصفهان

نازنین سادات امامی[۱]، الهام چاوشی[۲*]، شمس الله ایوبی[۳]، ناصر هنرجو[۴]، و مجتبی زراعت پیشه[۵]

۱- دانشجوی دکتری گروه علوم خاک، دانشکده کشاورزی، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران.

۲- دانشیار گروه علوم خاک، دانشکده کشاورزی، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران.

۳- استاد گروه علوم خاک، دانشکده کشاورزی، دانشگاه صنعتی اصفهان، اصفهان، ایران.

۴- دانشیار گروه علوم خاک، دانشکده کشاورزی، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران.

۵- دانشیار گروه علوم خاک، موسسه گاند برای محیط زیست، دانشگاه ورمانت، برلینگتون، آمریکا.

تاریخچه مقاله

دریافت: ۱۴۰۳/۱۱/۰۳

پذیرش نهایی: ۱۴۰۳/۱۲/۱۰

کلمات کلیدی:

نقشه‌برداری رقومی خاک،

یادگیری ماشین،

متغیرهای محیطی،

ویژگی‌های خاک،

دقت مدل‌سازی،

انتخاب ویژگی‌ها

چکیده

ویژگی‌های فیزیکوشیمیایی و مکانیکی خاک در سلامت خاک، بهره‌وری کشاورزی و ارزیابی پایداری آن نقش اساسی دارند. درک توزیع مکانی این ویژگی‌ها برای مدیریت مؤثر اراضی و کشاورزی پایدار ضروری است. نقشه‌برداری رقومی خاک (DSM) با استفاده از داده‌های محیطی و الگوریتم‌های یادگیری ماشین (ML)، ابزاری قدرتمند برای پیش‌بینی ویژگی‌های خاک است. هدف این پژوهش، پیش‌بینی توزیع مکانی میانگین وزنی قطر خاکدانه‌ها (MWD) و کربن آلی خاک (SOC) با استفاده از روش DSM و رتبه‌بندی متغیرهای محیطی مؤثر با مدل‌های ML و مقایسه این مدل‌ها در منطقه فریدن، استان اصفهان بود. شش الگوریتم ML شامل شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (RF)، مدل کوبیست (Cubist)، ماشین بردار پشتیبان (SVM)، الگوریتم K نزدیک‌ترین همسایه (KNN) و مدل رگرسیون درختی توسعه‌یافته (BRT) برای ارتباط دادن متغیرهای محیطی و ویژگی‌های خاک استفاده شدند. مدل‌سازی با چهار مجموعه‌داده شامل داده‌های سنجش از دور، توپوگرافی، ویژگی‌های خاک و داده‌های طبقه‌بندی شده انجام شد و عملکرد مدل‌ها با استفاده از ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) ارزیابی گردید. نتایج نشان داد که بهترین دقت پیش‌بینی برای MWD، در مدل KNN در سناریوی چهارم (R^2=۰/۵۹، RMSE=۰/۱۹ و MAE=۰/۱۶) و برای SOC، در مدل cubist در سناریوی چهارم (R^2=۰/۷۸، RMSE=٤/۱۲ و MAE=۳/۵۳)، به‌دست آمد. سیلت (Silt) مهم‌ترین متغیر برای MWD بود و VV و CHND در رتبه‌های بعدی قرار گرفتند. برای پیش‌بینی

* عهده دار مکاتبات

Email: chavoshie@yahoo.com

© 2025 The Author(s). Published by Shahid Chamran University of Ahvaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

SOC، MRVBF و Clay بیشترین تأثیر را داشتند. در نهایت، مدل‌های cubist و KNN با ترکیب داده‌های توپوگرافی و اقلیمی، به‌عنوان ابزاری مؤثر برای پیش‌بینی ویژگی‌های کیفی خاک و تولید نقشه‌های رقومی خاک انتخاب شدند. این روش‌ها به‌ویژه در مناطقی با اقلیم مشابه و شرایط نمونه‌برداری محدود، می‌توانند به بهبود بهره‌وری کشاورزی و ارتقای مدیریت پایدار اراضی کمک کرده و همچنین به سیاست‌گذاران در شناسایی مناطق بحرانی و کاهش اثرات منفی زیست‌محیطی یاری رسانند.

مقدمه

در سطح جهانی، تغییر کاربری اراضی و مدیریت ناپایدار کشاورزی به‌عنوان عوامل اصلی فرسایش آبی، از دست رفتن خاک و تخریب آن در مطالعات زیست‌محیطی گزارش شده‌اند و می‌توانند امنیت آب و غذا را تهدید کنند (۴۷). جنگل‌زدایی و تبدیل آن به اراضی زراعی از مهم‌ترین انواع تغییرات کاربری زمین هستند که می‌توانند تغییرات اقلیمی، فرسایش خاک و از دست رفتن خاک را تشدید کنند. با این‌حال، ساختار و کیفیت فیزیکی و شیمیایی خاک‌ها می‌توانند میزان آسیب‌پذیری و فرسایش‌پذیری آن‌ها در برابر فرسایش آبی و بادی را در منطقه تعیین کنند (۴۵، ۴۷). خاک به‌عنوان یکی از مهم‌ترین منابع طبیعی، نقشی اساسی در پایداری اکوسیستم، کشاورزی و امنیت غذایی ایفا می‌کند. ویژگی‌های خاک مانند ماده آلی خاک[1] (SOM) و میانگین وزنی قطر خاکدانه‌ها[2] (MWD) از جمله عوامل کلیدی هستند که تأثیر چشمگیری بر چرخه کربن، پایداری ساختاری خاک و بهره‌وری محصولات کشاورزی دارند (۴، ۴۵، ۴۷، ۴۹). به‌علاوه، کربن آلی نقشی حیاتی در پایداری محیط زیست دارد و از شاخص‌های مهم در ارزیابی کیفیت و سلامت خاک است (۴، ۳۲) میانگین وزنی قطر خاکدانه‌ها نیز شاخص مهمی برای ارزیابی کیفیت فیزیکی خاک است که بر جریان آب و هوا در خاک و در نهایت سلامت آن تأثیر می‌گذارد (۴۵). بنابراین، درک و ارزیابی ویژگی‌های خاک برای مدیریت پایدار اراضی و افزایش بهره‌وری کشاورزی ضروری است. با این‌حال، روش‌های سنتی نقشه‌برداری خاک به‌دلیل هزینه‌های بالا، زمان‌بر بودن و دقت محدود در مناطق پیچیده جغرافیایی با چالش‌هایی همراه هستند (۴۶، ۴۹). ازاین‌رو، استفاده بهینه از ویژگی‌های کلیدی خاک و توسعه روش‌های نوین پایش برای مقابله با تغییرات اقلیمی و تحقق اهداف امنیت غذایی امری حیاتی است.

نقشه‌برداری رقومی خاک[3] (DSM) روشی است که با استفاده از مدل‌سازی پیشرفته، روابط بین ویژگی‌های خاک و متغیرهای محیطی را تحلیل می‌کند. این روش، با پیش‌بینی ویژگی‌های خاک در مقیاس‌های جغرافیایی بزرگ، از سطح مزارع تا مناطق ملی و حوضه‌های آبخیز، و با کاهش هزینه‌های نمونه‌برداری، به بهبود مدیریت اراضی کمک می‌کند. مدل‌های DSM بر اساس رابطه کمی بین مشاهدات میدانی و پیش‌بینی‌کننده‌های محیطی طراحی می‌شوند تا الگوهای مکانی ویژگی‌های خاک شبیه‌سازی شوند. استفاده از الگوریتم‌های یادگیری ماشین[4] (ML)، به‌ویژه به‌دلیل دقت بالای آن در پیش‌بینی، به یکی از روش‌های رایج و مؤثر در برآورد خصوصیات خاک تبدیل شده است (۴۰، ۴۹). در دو دهه

1- Soil organic matter
2- Mean weight diameter
3- Digital soil mapping
4- Machine learning

گذشته، دانشمندان خاک از ML در طیف گسترده‌ای از سناریوها، از جمله نقشه‌برداری ویژگی‌های خاک یا طبقه‌بندی آن در مقیاس‌های مکانی از سطح محلی تا جهانی و در اعماق مختلف خاک استفاده کرده‌اند. گسترش استفاده از یادگیری ماشین (ML) در نقشه‌برداری رقومی خاک (DSM) به‌دلیل پیشرفت در دسترسی به داده‌های مکانی، افزایش قابلیت بهره‌برداری از اطلاعات محیطی، و توسعه نرم‌افزارها و ابزارهای محاسباتی پیشرفته برای تحلیل این داده‌ها تسهیل شده است (۲۷، ۳۲، ۴۰). مطالعات پیشین نشان داده‌اند که مدل‌هایی مانند جنگل تصادفی[1] (RF)، کوبیست[2]، ماشین بردار پشتیبان[3] (SVM)، الگوریتم K نزدیک‌ترین همسایه[4] (KNN)، رگرسیون درختی توسعه‌یافته[5] (BRT) و شبکه عصبی مصنوعی[6] (ANN) قادر به شبیه‌سازی ارتباطات غیرخطی و پیچیده بین متغیرهای محیطی و ویژگی‌های خاک هستند، که این امر موجب بهبود دقت نقشه‌برداری خاک می‌شود (۱۱، ۱۳، ۳۱، ۳۲، ۴۷، ۴۹). با این‌حال، بررسی مطالعات پیشین نشان می‌دهد که همچنان چالش‌ها و خلأهای تحقیقاتی در این حوزه وجود دارد، از جمله کمبود مدل‌هایی که بتوانند به طور همزمان دقت بالا، سرعت پردازش و امکان تفسیر نتایج را ارائه دهند (۴۰). علاوه بر این، تعیین مناسب‌ترین تکنیک مدل‌سازی متناسب با شرایط محیطی هر منطقه همچنان یکی از چالش‌های اساسی در نقشه‌برداری ویژگی‌های خاک محسوب می‌شود.

با پیشرفت‌های اخیر در نقشه‌برداری خاک، متغیرهای محیطی مانند تصاویر ماهواره‌ای، داده‌های زمین‌شناسی و اقلیمی که از منابع مختلف به‌دست آمده‌اند، با مشاهدات میدانی خاک ترکیب شده‌اند تا ویژگی‌های خاک به‌دقت پیش‌بینی شوند (۴۹). داده‌های سنجش از دور، به‌دلیل دسترسی رایگان، قدرت تفکیک مکانی بالا و قابلیت پایش مداوم، ابزارهای مؤثری برای DSM به شمار می‌روند (۳). در این میان، سری ماهواره‌های سنتینل، به‌ویژه سنتینل ۱ و ۲، با ارائه داده‌های رایگان، محصولات نزدیک به زمان واقعی، و فرکانس بازدید بالا، نقش چشمگیری در پایش ایمنی محیطی و شبیه‌سازی روابط خاک- منظر اراضی ایفا کرده‌اند (۲۵، ۴۹). این داده‌ها، همراه با مدل‌های پیش‌بینی‌کننده و داده‌های ژئومورفولوژیکی، اقلیمی و مدل‌های رقومی ارتفاع (DEM)، تحلیل و پیش‌بینی دقیق توزیع مکانی ویژگی‌های خاک را ممکن ساخته‌اند (۱۱، ۴۵، ۴۶، ۴۷).

مطالعات اخیر بر نقش ML در نقشه‌برداری و مدل‌سازی داده‌های خاک تأکید داشته‌اند (۲۷). در پژوهشی، به مقایسه عملکرد مدل‌های ML پرداخته و نشان داده شد که مدل‌های RF و ANN در پیش‌بینی ویژگی‌های خاک بهترین عملکرد را دارند (۱۳). همچنین، تحقیق دیگری به چالش‌های مدل‌سازی داده‌های محیطی و تفسیر روابط بین متغیرهای محیطی و خاک پرداخته و مشخص کرد که ادغام دانش خاک‌شناسی با یادگیری ماشین می‌تواند دقت نقشه‌ها را بهبود دهد (۴۰). در مطالعه‌ای بر اساس معیارهای دقت، مدل‌های cubist و RF دقیق‌ترین پیش‌بینی‌ها را برای ویژگی‌های خاک ارائه دادند. همچنین، متغیرهای سنجش از دور مانند شاخص پوشش گیاهی و باند ۴ تصاویر ماهواره‌ای سنتینل-۲، به‌همراه ویژگی‌های توپوگرافی، مهم‌ترین عوامل تأثیرگذار بر پیش‌بینی مدل‌ها بودند (۴۶). امیرعابدی و همکاران[7] (۱۳۹۵) گزارش کردند که ANN در مقایسه با مدل‌های رگرسیونی، عملکرد و دقت بالاتر و خطای کمتری در برآورد MWD خاک داشته است. همچنین، هر دو روش بهترین عملکرد را با متغیرهای ورودی شامل

1- Random forest
2- Cubist
3- Support vector machines
4- K-nearest neighbors
5- Boosted regression trees
6- Artificial neural network

7- Amir Abedi *et al.*

جرم ویژه ظاهری، کربن آلی و درصد رس نشان دادند (۱).

منطقه فریدن در استان اصفهان، به‌عنوان یکی از قطب‌های صنعتی و کشاورزی ایران، با اقلیمی کوهستانی، به داده‌های دقیق در خصوص ویژگی‌های خاک نیاز دارد تا مدیریت بهتری برای منابع آب و خاک صورت گیرد. با وجود پیشرفت‌های اخیر در DSM، چالش‌هایی مانند انتخاب الگوریتم بهینه، تفسیر روابط پیچیده بین متغیرها و محدودیت داده‌های محیطی در مناطق خاص همچنان وجود دارد. بنابراین، اطلاعات مربوط به پیش‌بینی ویژگی‌های خاک و عواملی که بر تغییرات آن‌ها تأثیر می‌گذارند، هنوز ناقص و نیازمند تحقیقات بیشتری است. هدف اصلی این پژوهش، رفع شکاف‌های موجود در مدل‌سازی پیش‌بینی ویژگی‌های خاک و ارزیابی شش الگوریتم ML شامل ANN، RF، cubist، SVM، KNN، و BRT برای پیش‌بینی ویژگی‌های خاک (MWD و SOC) در منطقه فریدن است. همچنین، شناسایی و تحلیل متغیرهای محیطی مؤثر بر تغییرپذیری ویژگی‌های خاک، یکی دیگر از اهداف این مطالعه است. این پژوهش به‌عنوان اولین تلاش برای استفاده از این الگوریتم‌ها در پیش‌بینی ویژگی‌های خاک در منطقه فریدن، می‌تواند به درک بهتر روابط بین ویژگی‌های خاک و متغیرهای محیطی کمک کند. همچنین، ارزیابی ترکیب داده‌های مبتنی بر سناریوهای مختلف، از جمله ادغام سنجش از دور، ویژگی‌های توپوگرافی، طبقه‌بندی زمین و ویژگی‌های خاک، بینش‌های قابل‌توجهی در مدل‌سازی کیفیت خاک فراهم می‌آورد. تهیه نقشه‌های دقیق از ویژگی‌های خاک می‌تواند به بهبود مدیریت پایدار اراضی در مناطق حساس زیست‌محیطی کمک کند. تحلیل جامع عملکرد مدل‌ها بر اساس معیارهای دقت مختلف، بهترین رویکردها را برای DSM شناسایی کرده و به درک عمیق‌تری از روابط پیچیده بین خاک و متغیرهای محیطی منجر می‌شود.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

این پژوهش در منطقه فریدن استان اصفهان، ایران، در طول جغرافیایی ۵۰ درجه و ۶ دقیقه و عرض جغرافیایی ۳۲ درجه و ۵۶ دقیقه شمالی، با میانگین ارتفاع ۲۲۹۰ متر از سطح دریا انجام شد. مساحت محدوده مطالعاتی حدود ۱۲۳۹ کیلومتر مربع است. این منطقه دارای میانگین بارندگی سالیانه ۳۲۰ میلی‌متر (حداقل ۲۵۰ و حداکثر ۴۰۰ میلی‌متر) و میانگین دمای سالانه ۱۱/۵ درجه سلسیوس است. میانگین حداقل دمای سالانه در منطقه ۱۷/۶- درجه سلسیوس و میانگین حداکثر دمای سالانه ۲۳ درجه سلسیوس است.

ویژگی‌های کوهستانی و ارتفاع زیاد منطقه موجب بارش‌های فراوان، به‌ویژه به صورت برف، شده است که در سرسبزی و رشد گیاهان متنوع نقش مؤثری دارد و مراتع سرسبزی را ایجاد می‌کند. پوشش گیاهی این منطقه شامل گونه‌های مختلف گون، گز، کرفس کوهی و درختان پراکنده بلوط است. اقلیم فریدن دارای تابستان‌های معتدل و زمستان‌های بسیار سرد و سخت است و بیش از هشت ماه از سال پوشیده از برف می‌باشد (۶). این شهرستان دارای تپه ماهورهای فراوانی است و اغلب کشاورزان برای امرار معاش، اراضی شیب‌دار را با تغییر کاربری از مرتع به دیم‌زارهای غلات تبدیل کرده‌اند. این اقدام به کاهش کیفیت خاک در منطقه منجر شده است (۳۰).

در شهرستان فریدن، آهک‌های مارنی فسیل‌دار به‌صورت دگرشیب بر روی سنگ‌های دگرگونی قرار گرفته‌اند. میان این لایه‌ها، شیل ماسه‌ای و میکروکنگلومرای خاکستری یا سبزرنگ تیره مشاهده می‌شود. در شمال این منطقه، ضخامت کنگلومرای قاعده پرمین به ۳۵۰ تا ۴۰۰ متر می‌رسد (۶، ۳۰).

خاک‌های منطقه مورد مطالعه بر اساس سامانه رده‌بندی آمریکایی خاک[1] (۲۰۱۴) در رده‌های Entisols، Inceptisols و Alfisols قرار می‌گیرند و به‌ترتیب در زیرگروه‌های Typic Xerorthents، Typic Calcixerepts و Calcic Haploxeralfs دسته‌بندی می‌شوند (۳۶).

ژئومورفولوژی منطقه

در این پژوهش، واحدهای ژئومورفیک منطقه مورد مطالعه با استفاده از تصاویر ماهواره‌ای گوگل ارث و بررسی‌های میدانی شناسایی و طبقه‌بندی شدند (شکل ۱). برای انجام این کار، معیارهای مختلفی مانند شیب، جهت شیب، ویژگی‌های زمین‌شناسی، نقشه‌های سفره‌های زیرزمینی، مناظر و اشکال زمین مورد تجزیه و تحلیل قرار گرفت. با توجه به پیچیدگی‌های متنوع موجود در مناظر، از یک سیستم سلسله مراتبی چهار سطحی برای طبقه‌بندی واحدهای ژئومورفیک استفاده شد (۲۴). این سلسله مراتب شامل سطوح زیر است:

۱- زمین‌نما: این سطح به عوارض بزرگ زمین‌ساختی مانند کوه‌ها، دشت‌ها و فلات‌ها اشاره دارد.

۲- لندفرم: در این سطح، اشکال زمین در مقیاس کوچک‌تر مانند دره‌ها، تپه‌ها و دشت‌های آبرفتی مورد بررسی قرار می‌گیرند.

۳- سنگ‌شناسی: این سطح بر نوع سنگ‌ها و رسوبات موجود در منطقه تمرکز دارد.

۴- سطوح ژئومورفیک: در این سطح، جزئی‌ترین اشکال زمین مانند گودال‌ها، شیارها و رسوبات رودخانه‌ای بررسی می‌شوند.

در نهایت نقشه ژئومورفولوژی منطقه که شامل ۳۰ واحد ژئومورفیک است، تهیه شد (شکل ۱).

نمونه‌برداری و تجزیه‌های آزمایشگاهی

برای جمع‌آوری نمونه‌های خاک، از روش نمونه‌برداری شبکه‌ای سیستماتیک با ابعاد ۱۶۰×۱۶۰ متر استفاده شد. مختصات تمام نقاط نمونه‌برداری به‌صورت تصادفی نظام‌دار در هر سلول شبکه شناسایی گردید. پس از انجام مطالعات میدانی، تعداد ۱۰۰ نمونه از خاک سطحی (عمق ۰ تا ۲۰ سانتی‌متر) از کاربری‌های مختلف شامل مزارع آبی، دیم و مراتع جمع‌آوری شد (شکل ۲). موقعیت مکانی هر نمونه با استفاده از دستگاه GPS ثبت شد (۱۰). نمونه‌ها پس از جمع‌آوری و انتقال به آزمایشگاه، هوا خشک شده، از الک با منافذ ۲ میلی‌متری عبور داده شدند و سپس آنالیزهای آزمایشگاهی روی آن‌ها انجام گرفت.

ویژگی‌های خاک

۱- ویژگی فیزیکی: میانگین وزنی قطر خاکدانه‌ها (MWD، میلی‌متر) با استفاده از روش الک تر (۱۲) و از طریق معادله (۱) محاسبه شد.

$$MWD=\sum W_i X_i \qquad (1)$$

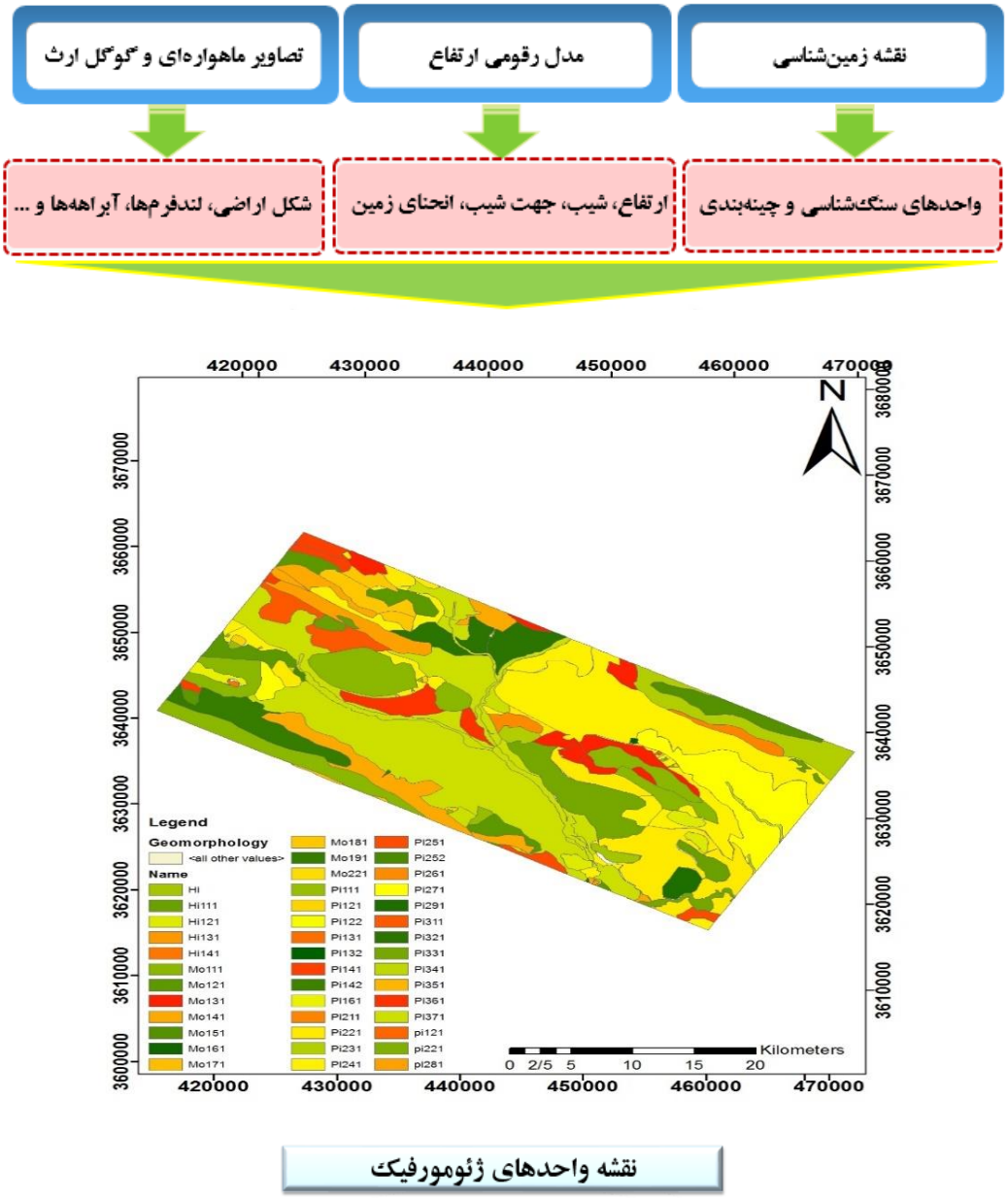
که در این معادله، Wi میانگین قطر منافذ دو الک متوالی (میلی‌متر) و Xi نسبت جرم خاکدانه‌های باقی‌مانده روی هر الک به جرم کل خاک پس از تصحیح شن است.

۲- ویژگی‌های شیمیایی: کربن آلی خاک (SOC، درصد) با استفاده از روش والکلی- بلک اندازه‌گیری شد (۴۱).

مراحل پژوهش

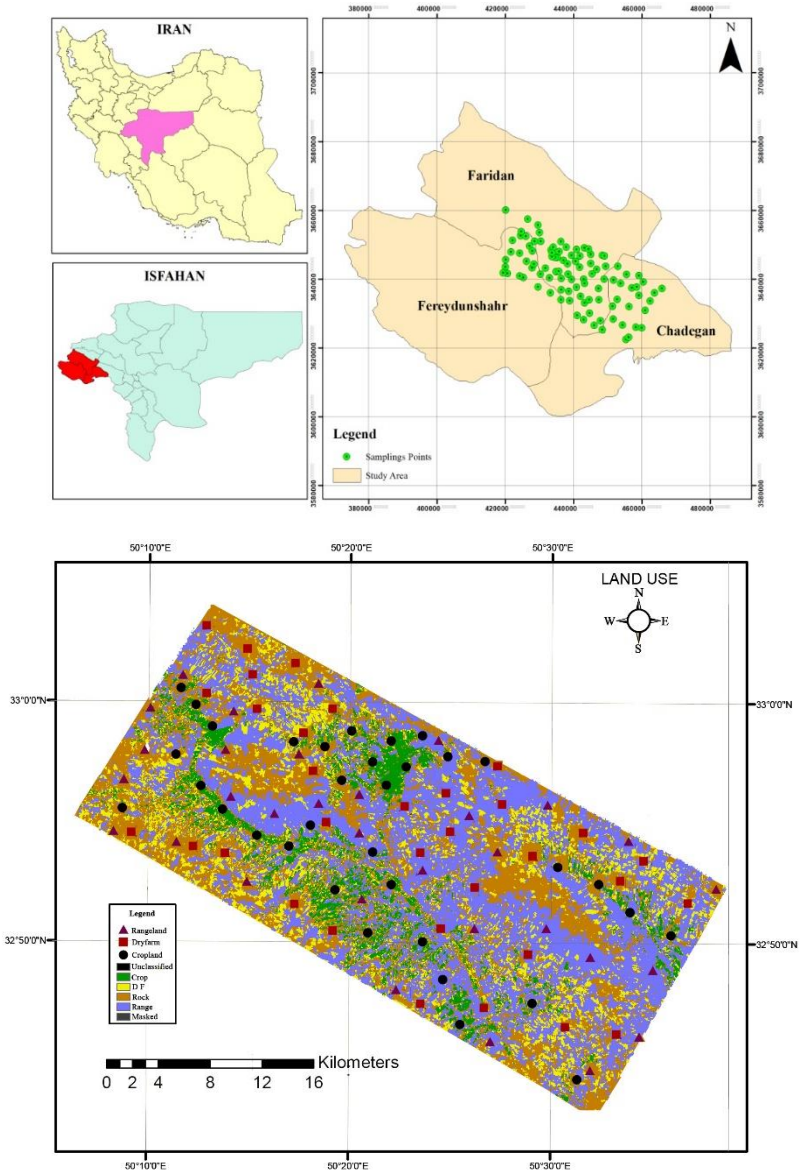
روند تهیه نقشه چهار ویژگی خاک در منطقه مطالعاتی در پنج مرحله اصلی سازمان‌دهی شده است که عبارتند از: ۱- نمونه‌برداری و جمع‌آوری اطلاعات خاک و متغیرهای کمکی محیطی از منطقه مطالعاتی؛ ۲- مدل‌سازی دو ویژگی خاک با استفاده از متغیرهای کمکی محیطی و روش‌های یادگیری ماشین؛ ۳- اعتبارسنجی کارایی مدل‌ها در منطقه مطالعاتی؛ ۴- تعیین مهم‌ترین فاکتورهای کمکی محیطی در پیش‌بینی دو ویژگی خاک؛ ۴- تولید نقشه پیش‌بینی دو ویژگی خاک بر اساس بهترین مدل مکان‌محور.

1- USDA Soil Taxonomy



شکل (۱) فلوچارت سیستم سلسله مراتب چهار سطحی واحدهای ژئومورفیک

Figure (1) Flowchart of the four-level hierarchical system of geomorphic units



شکل (۲) موقعیت منطقه مورد مطالعه (شهرستان فریدن، استان اصفهان، ایران)، همراه با تنوع کاربری‌های اراضی و توزیع نقاط نمونه‌برداری

Figure (2) Location of the studied area (Fereydan district, Isfahan province, Iran), including land use diversity and sampling point distribution

متغیرهای کمکی محیطی برای نقشه‌برداری رقومی

برای نقشه‌برداری رقومی ویژگی‌های خاک، چهار مجموعه‌داده به‌عنوان متغیرهای کمکی پیش‌بینی‌کننده برای فرایندهای مدل‌سازی استفاده شد. این مجموعه‌ها شامل داده‌های سنجش از دور، متغیرهای کمکی توپوگرافی که از مدل رقومی ارتفاع (DEM) به‌دست آمده، داده‌های ویژگی‌های خاک و داده‌های طبقه‌بندی شده زمین بودند.

۱- متغیرهای کمکی سنجش از دور: برای ایجاد مجموعه‌داده‌ای شامل شاخص‌های پوشش گیاهی، از تصاویر سنتینل ۲[1] و نرم‌افزار SAGA GIS[2] برای استخراج متغیرهای محیطی زیر استفاده شد: شاخص گیاهی تفاضلی نرمال‌شده[3]

1- Sentinel-2
2- System for automated geoscientific analysis
3- Normalized difference vegetative index

(NDVI) (۶)، شاخص پوشش گیاهی تنظیم‌شده با خاک[1] (SAVI) (۲۹)، شاخص پوشش گیاهی افزایش‌یافته[2] (EVI) و نوارهای تصاویر سنتینل ۲ (B2-B12). محاسبه شاخص‌های پوشش گیاهی با استفاده از معادلات جدول ۱ انجام شد (۲۵).

۲- متغیرهای کمکی ویژگی‌های توپوگرافی: توپوگرافی به‌عنوان یکی از عوامل اصلی تشکیل خاک در مناطق خشک و نیمه‌خشک در نظر گرفته می‌شود (۴۶). برای دست‌یابی به ویژگی‌های توپوگرافی، از پایگاه داده Alaska Satellite Facility از ماهواره ALOS استفاده شد. ویژگی‌های زمین که از نقشه رقومی ارتفاع به‌دست آمد عبارت‌اند از: ارتفاع، شاخص رطوبت توپوگرافی[3] (TWI)، جهت شیب، انحنای مقطعی[4]، سطح پایه شبکه کانال[5] (CNBL)، حوضه آبریز، تجمع جریان، ضریب LS، موقعیت شیب نسبی[6] (RSP)، شاخص هموار کف دره با وضوح چندگانه[7] (MRVBF)، شاخص هموار بالای پشته[8] (MRRTF)، شیب، عمق دره (Val.Depth)، و شاخص همگرایی (جدول ۱) (۳۷، ۵۰).

۳- متغیرهای کمکی خاک: متغیرهای کمکی خاک مانند مواد آلی، چگالی ظاهری و توزیع اندازه ذرات، ابتدا از طریق نمونه‌برداری و آزمایش‌های آزمایشگاهی اندازه‌گیری شده و سپس با استفاده از روش درون‌یابی کریجینگ در نرم‌افزار ArcGIS 8.10 برای برآورد مقادیر در نقاط مجهول استفاده شده‌اند.

۴- داده‌های دسته‌بندی انتخاب شده: در این پژوهش، واحدهای نقشه خاک، کاربری اراضی و زمین‌شناسی به‌عنوان متغیرهای کمکی برای طبقه‌بندی انتخاب شدند. سپس، با توجه به این نکته که نقشه کاربری اراضی برای کل کشور ایران موجود بود، نقشه کاربری اراضی منطقه مورد مطالعه با استفاده از کدهای مربوطه در محیط گوگل ارث انجین تهیه شد.

تجزیه و تحلیل پیش‌پردازش

هنگامی که مدلی وجود دارد که تعداد زیادی پیش‌بینی کننده دارد، بهبود قابلیت و عملکرد کلی آن نیاز به جستجوی جامع از همه زیرمجموعه‌های ممکن برای انتخاب بهترین پیش بینی‌کننده‌ها دارد (۱۹، ۳۴). علاوه بر این، همه پیش‌بینی‌کننده‌ها در بسته "caret" ضروری نیستند و استفاده از انتخاب ویژگی برای این مدل‌ها کارآمدتر خواهد بود. در این پژوهش، ابتدا برای بررسی چندخطی بودن بین پیش‌بینی‌کننده‌های در نظر گرفته‌شده، همبستگی‌های زوجی بین متغیرهای کمکی محاسبه گردید؛ از آن‌جا که هیچ متغیری با ضریب همبستگی بالای ۰/۸ مشاهده نشد، تحلیل همبستگی به‌صورت جداگانه ارائه نشده است. سپس، عامل تورم واریانس[9] (VIF) برای هر متغیر کمکی محاسبه شد و متغیرهای کمکی با VIF>10 از مجموعه‌داده‌ها حذف شدند (۸). در مجموع، از بین متغیرهای ورودی انتخاب‌شده، ۹ ویژگی توپوگرافی، ۵ شاخص سنجش از دور و ۵ داده کمکی دیگر (داده‌های برخی ویژگی‌های خاک، کاربری اراضی و زمین‌شناسی) در مدل‌سازی ویژگی‌های خاک وارد شدند (جدول ۲).

چهار سناریو برای تعیین قابلیت داده‌های سنجش از دور ترکیبی، ویژگی‌های توپوگرافی، داده‌های برخی ویژگی‌های خاک و داده‌های طبقه‌بندی انتخابی برای نقشه‌برداری رقومی ویژگی‌های خاک طراحی شد:

سناریو ۱: استفاده از متغیرهای مشتق‌شده از داده‌های سنجش از دور سنتینل ۱ و ۲

سناریو ۲: استفاده از متغیرهایی که از داده‌های زمین مشتق‌شده از DEM به‌دست آمده و تمام پارامترهای سناریو ۱

سناریو ۳- استفاده از تمام پارامترهای سناریو ۲ و داده‌های طبقه‌بندی‌شده

سناریو ۴- استفاده از تمام پارامترهای سناریو ۳ و برخی از ویژگی‌های خاک

1- Soil adjusted vegetation index
2- Enhanced vegetation index
3- Topographic wetness index
4- Cross-sectional curvature
5- Channel network base level
6- Relative slope position
7- Multi-resolution valley bottom flatness index
8- Multi-resolution ridge top flatness index
9- Variance inflation factor

جدول (۱) پارامترهای کمکی برای پیش‌بینی مکانی ویژگی‌های خاک

Table (1) Auxiliary parameters for spatial prediction of soil properties

داده‌های محیطی Environmental data	نام متغیر Variable name	توضیحات Description
داده‌های سنجش از دور Remote sensing data	باندهای تصاویر سنتینل Sentinel Bands (B2-B7)	-
	شاخص نرمال شده پوشش گیاهی Normalized Difference Vegetation Index (NDVI)	$(B5 - B4)/(B5 + B4)$
	شاخص بهبود یافته گیاهی Enhanced Vegetation Index (EVI)	$2/5 \times \frac{(B5 - B4)}{(B5 + 6 \times B4) - (7/5 \times E}$
	شاخص تعدیل شده پوشش گیاهی خاک Soil Adjusted Vegetation Index (SAVI)	$\frac{B5 - B4}{B5 + B4 + L} \times (1 + L)$
مدل رقومی ارتفاع Digital elevation model (DEM)	ارتفاع (Elevation)	m
	سطح پایه شبکه کانال Channel Network Base Level	CNBL
	حوضه آبخیز (Catchment Area)	Catchment. Area
	شاخص توپوگرافی (LS Factor)	LS. Factor
	شاخص همواری بالای پشته با درجه تفکیک بالا Multi-resolution of Ridge Top Flatness Index	MRRTF
	شاخص همواری کف دره با درجه تفکیک بالا Multi-resolution Valley Bottom Flatness Index	MRVBF
	انحنای کلی (Total Curvature)	Total. Curv
	شاخص همگرایی (Convexity)	Convexity
	شاخص شیب نسبی Slope Position Relative	RSP
	شاخص رطوبت توپوگرافی Topographic Wetness Index	TWI
	عمق دره (Valley Depth)	Val.Depth
داده‌های طبقه‌بندی شده Classified data	واحدهای نقشه خاک (Soil Units Map)	-
	نقشه کاربری اراضی (Land Use Map)	انواع کاربری‌ها و استفاده از اراضی Land use types and patterns
	نقشه زمین‌شناسی (Geology Map)	سازند سنگ‌شناسی Lithological formations
	نقشه ژئومورفولوژی (Geomorphology Map)	سطوح ژئومورفولوژی Geomorphology surfaces
ویژگی‌های خاک Soil properties	رس (Clay), سیلت (Silt), شن (Sand), وزن مخصوص ظاهری (Bulk density, BD), کربن آلی خاک (SOC)	-

نقشه‌برداری رقومی ویژگی‌های خاک

در این پژوهش، از الگوریتم‌های RF، cubist، SVM، KNN، BRT و ANN برای مدل سازی مکانی ویژگی‌های خاک استفاده شد (۱۳، ۴۶، ۴۷، ۴۹). ویژگی‌های خاک ارتباط نزدیکی با ویژگی‌های توپوگرافی و توزیع پوشش گیاهی دارند (۱۵). همچنین، از نرم‌افزار R4.2.1 برای پیش‌بینی ویژگی‌های خاک و تهیه نقشه رقومی این ویژگی‌ها استفاده شد. مدل‌سازی و تمامی انتخاب‌ها با استفاده از بسته‌های caret و caretEnsemble انجام شد.

ارزیابی اعتبار مدل

برای مدل‌سازی دو ویژگی خاک، داده‌ها به دو گروه داده‌های واسنجی (۷۰ درصد داده‌ها) و داده‌های اعتبارسنجی (۳۰ درصد داده‌ها) تقسیم شدند. در نهایت، به‌منظور ارزیابی دقت مدل‌ها، از آماره‌های مختلفی مانند ضریب تعیین[1] (R^2)، ریشه میانگین مربعات خطا[2] (RMSE) و میانگین خطای مطلق[3] (MAE) استفاده شد. معادلات این روش‌ها در روابط ۲، ۳ و ۴ ارایه شده است.

$$R^2=\frac{\left(n\sum_{i=1}^{n}X_{Pi}X_{mi}-\sum_{i=1}^{n}X_{Pi}\sum_{i=1}^{n}X_{mi}\right)^2}{\left[n\sum_{i=1}^{n}X_{pi}^{2}-\left(\sum_{i=1}^{n}X_{Pi}\right)^2\right]\left[n\sum_{i=1}^{n}X_{mi}^{2}-\left(\sum_{i=1}^{n}X_{mi}\right)^2\right]} \quad (۲)$$

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}\left(X_{pi}-X_{mi}\right)^2}{n}} \quad (۳)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|X_{pi}-X_{mi}\right| \quad (۴)$$

در این معادلات، X_{mi} مقادیر اندازه‌گیری یا مشاهده شده ویژگی‌های خاک، X_{pi} مقادیر پیش‌بینی شده ویژگی‌های خاک و n تعداد نمونه‌ها هستند. ترکیبی از پارامترهای آماری ذکر شده برای ارزیابی مدل کافی است.

نتایج و بحث

میانگین وزنی قطر خاکدانه‌ها (MWD) یکی از شاخص‌های مهم ساختاری خاک است که بر پایداری و سلامت آن تأثیر می‌گذارد. در سناریوی اول، مدل‌های RF، SVM و KNN بالاترین مقدار R^2 (۰/۵۶) را ثبت کردند. با این‌حال، مدل SVM به‌دلیل مقدار RMSE (۰/۱۸) و MAE (۰/۱۶)، عملکرد مشابهی با RF داشت، اما مدل cubist با R^2 برابر با ۰/۵ عملکرد ضعیف‌تری نشان داد. مدل BRT نیز با R^2 برابر با ۰/۵ و RMSE برابر با ۰/۱۹ کمترین دقت را داشت (جدول ۲).

در سناریوی دوم، مدل ANN با R^2 برابر با ۰/۵۷ بهترین عملکرد را داشت. مقادیر RMSE و MAE در این مدل به‌ترتیب برابر با ۰/۱۸ و ۰/۱۵ بودند که نشان‌دهنده دقت مناسب مدل در پیش‌بینی ویژگی MWD است. مدل cubist نیز با R^2 برابر با ۰/۵۵ و MAE برابر با ۰/۱۵ عملکرد خوبی ارائه کرد، اما مدل SVM با کاهش قابل‌توجه در مقدار R^2 (۰/۴۴) و افزایش مقادیر RMSE (۰/۲۴) و MAE (۰/۲۱) ضعیف‌ترین عملکرد را نشان داد.

در سناریوی سوم، مدل‌های ANN و cubist با R^2 برابر با ۰/۵۷ بهترین عملکرد را داشتند. مدل ANN با RMSE برابر با ۰/۱۷ نسبت به cubist (۰/۱۹) دقیق‌تر بود و MAE در این مدل‌ها به‌ترتیب ۰/۱۵ و ۰/۱۶ بود. در مقابل، مدل‌های SVM و BRT با مقادیر R^2 به‌ترتیب برابر با ۰/۴۸ و ۰/۴۹ عملکرد ضعیف‌تری نسبت به سایر مدل‌ها داشتند (جدول ۲).

در سناریوی چهارم، مدل KNN با R^2 برابر با ۰/۵۹ بالاترین دقت را ارائه داد. RMSE این مدل برابر با ۰/۱۹ و MAE آن ۰/۱۶ بود. مدل‌های cubist و ANN به‌ترتیب با R^2 برابر با ۰/۵۷ و ۰/۵۵ در جایگاه‌های دوم و سوم قرار گرفتند. مدل BRT با R^2 برابر با ۰/۴۶ ضعیف‌ترین عملکرد را نشان داد (جدول ۲). بنابراین، در پیش‌بینی ویژگی MWD، مدل KNN در سناریوی چهارم با مقدار R^2 برابر با ۰/۵۹ بهترین عملکرد را ارائه داد و به‌عنوان قوی‌ترین مدل در میان شش مدل و چهار سناریو شناخته شد (جدول ۲). این نتیجه بر اهمیت ترکیب داده‌های متنوع در سناریوی چهارم تأکید دارد، داده‌هایی که شامل متغیرهای مشتق‌شده از داده‌های سنجش از دور سنتینل ۱ و ۲، داده‌های زمینی حاصل از DEM، داده‌های طبقه‌بندی‌شده

1- Coefficient of determination
2- Root mean square error
3- Mean absolute error

کاربری اراضی، زمین‌شناسی، ژئومورفولوژی، و ویژگی‌های خاک (مانند ماده آلی، چگالی ظاهری و ذرات خاک) است. این ترکیب داده‌ها دقت مدل را به طور قابل‌توجهی افزایش داده است. مدل ANN در سناریوهای دوم و سوم و مدل‌های RF و cubist در برخی سناریوها عملکردی قابل‌قبول و نسبتاً مشابه داشتند، اما دقت آن‌ها در سناریوی چهارم به سطح KNN نرسید. در مقابل، مدل SVM به‌دلیل مقادیر پایین R^2 و مقادیر بالای RMSE و MAE، به‌ویژه در سناریوهای دوم و سوم، به‌عنوان یکی از مدل‌های ضعیف‌تر در این تحلیل شناخته شد. این نتایج با یافته‌های مطالعات پیشین هم‌راستا است که نشان می‌دهند مدل‌های پیچیده‌تر مانند RF به‌دلیل قابلیت‌های بالاتر در تعمیم و پردازش روابط پیچیده بین متغیرها، عملکرد بهتری نسبت به مدل‌های ساده‌تر مانند SVM دارند. به‌طور خاص، مدل SVM به‌دلیل استفاده از مقادیر گسسته در گره‌های پایانی و عدم توانایی در مدل‌سازی روابط پیچیده، دقت کمتری در پیش‌بینی ویژگی‌های خاک دارد (۵۱).

به طور کلی، نتایج تحلیل این پژوهش نشان داد که استفاده از مدل KNN در سناریوی چهارم و مدل ANN در سناریوهای دوم و سوم برای پیش‌بینی دقیق‌تر ویژگی MWD توصیه می‌شود. همچنین، مقادیر پایین RMSE و MAE برای مدل KNN در سناریوی چهارم تأیید می‌کند که داده‌های ترکیبی نقشی کلیدی در افزایش دقت پیش‌بینی ویژگی‌های مرتبط با ساختار خاک دارند.

کربن آلی خاک (SOC) نقش اساسی در ارتقای کیفیت خاک ایفا می‌کند و از عوامل مؤثر بر شاخص‌های پایداری خاکدانه‌ها محسوب می‌شود (۴۷). در پژوهش حاضر، در سناریوی اول، مدل ANN با R^2=۰/۶۳ بهترین دقت را ارائه کرد، اما مقادیر RMSE (۹/۶) و MAE (۷/۹۷) نشان‌دهنده خطای نسبی بالا بود. مدل cubist با R^2 برابر با ۰/۵۴ و RMSE برابر با ۵/۵۶ عملکرد دقیق‌تری نسبت به ANN داشت. مدل‌های RF (R^2=۰/۴۷) و BRT (R^2=۰/۵۲) در این سناریو دقت کمتری داشتند (جدول ۲).

در سناریوی دوم، مدل cubist با R^2 برابر با ۰/۶۸ و مقادیر RMSE و MAE به‌ترتیب ۵/۵۲ و ۴/۶۴ بهترین عملکرد را ارائه داد. مدل SVM (R^2=۰/۶۴) نیز دقت بالایی داشت، اما خطای آن بیشتر از cubist بود. مدل ANN (R^2=۰/۳۹) پایین‌ترین دقت را در این سناریو داشت (جدول ۲).

در سناریوی سوم، مدل cubist با R^2 برابر با ۰/۵۳ و مقادیر RMSE و MAE به‌ترتیب ۵/۷۳ و ۴/۹۲، عملکرد بهتری نسبت به سایر مدل‌ها داشت. مدل‌های KNN (R^2=۰/۴۷) و RF (R^2=۰/۴۱) دقت کمتری نشان دادند (جدول ۲). در سناریوی چهارم، مدل cubist با R^2 برابر ۰/۷۸ و کمترین مقادیر RMSE (R^2=۴/۱۲) و MAE (R^2=۳/۵۳) بهترین عملکرد را ارائه داد. مدل RF (R^2=۰/۷۲) نیز دقت بالایی داشت، اما مقادیر RMSE و MAE (۴/۴۴ و ۳/۷۰) اندکی بیشتر بود. مدل‌های KNN (R^2=۰/۵۶) و BRT (R^2=۰/۶۳) عملکرد ضعیف‌تری نسبت به cubist داشتند (جدول ۲).

در تمامی سناریوها، به‌ویژه سناریوی چهارم، مدل cubist بهترین عملکرد را در پیش‌بینی SOC ارائه داد و بالاترین مقدار R^2 و کمترین مقادیر RMSE و MAE را به ثبت رساند. این نتایج نشان‌دهنده توانایی بالای این مدل در مدیریت داده‌های پیچیده و ترکیب متغیرهای ورودی است. مدل RF نیز در برخی سناریوها، به‌ویژه در سناریوی چهارم، عملکرد مناسبی داشت و در رتبه دوم قرار گرفت. سایر مدل‌ها مانند ANN، SVM، KNN و BRT در اکثر سناریوها دقت کمتری داشتند و برای پیش‌بینی SOC مناسب نبودند. برای مثال، مدل ANN در سناریوی اول با R^2=۰/۶۳ عملکرد نسبتاً خوبی داشت اما در سایر سناریوها دقت آن کاهش یافت. مدل cubist به‌دلیل دقت بالا و خطای کم، بهترین گزینه برای پیش‌بینی SOC است. پیشنهاد می‌شود در مطالعات آینده از این مدل در کنار سایر الگوریتم‌های پیشرفته و استفاده از داده‌های متنوع زمینی و ماهواره‌ای برای بهبود دقت پیش‌بینی بهره گرفته شود. مطالعه ژو و همکاران[1] (۲۰۲۰) نشان داد مدل‌های RF و cubist در

1- Zhou *et al.*

پیش‌بینی ویژگی‌های خاک با استفاده از داده‌های ماهواره‌ای سنتینل و شاخص‌های توپوگرافی عملکرد بسیار مطلوبی دارند. این یافته‌ها با تأکید پژوهش ما بر اهمیت استفاده از داده‌های کمکی هم‌خوانی دارد (۵۰). در هم‌راستایی با پژوهش ما، مطالعه ملیهو و همکاران[1] (۲۰۲۳)، مدل cubist (R^2=۰/۷۷ و RMSE=۱/۲) را به‌عنوان ابزاری مؤثر برای پیش‌بینی SOC معرفی کرد، به‌ویژه زمانی که از متغیرهای محیطی استفاده شده بود (۲۳). همچنین، مطالعه جان و همکاران[2] (۲۰۲۰) مدل‌های RF (R^2=۰/۶۸) و cubist (R^2=۰/۵۱) را در پیش‌بینی SOC و مواد مغذی خاک مؤثر دانسته‌اند، به‌ویژه زمانی که شاخص‌های پوشش گیاهی و توپوگرافی به‌کار گرفته شده بودند (۱۱). در مقابل، مطالعه‌ای دیگر مدل‌های پیشرفته مانند تقویت گرادیانی شدید[3] (XGBoost) را برای پیش‌بینی SOC به‌عنوان مدل برتر معرفی کرده است (۴۴).

انتخاب ویژگی و اهمیت متغیرهای محیطی در پیش‌بینی ویژگی‌های خاک

در مدل‌سازی با استفاده از متغیرهای کمکی، برخی از پارامترها همواره بیشترین تأثیر را در پیش‌بینی دارند که می‌توان آن‌ها را از طریق آنالیز انتخاب ویژگی شناسایی کرد. این فرآیند، اهمیت متغیرهای محیطی را در مدل‌های ML برجسته می‌کند و بینش‌های ارزشمندی در مورد ماهیت مجموعه داده‌ها ارائه می‌دهد. علاوه بر این، آنالیز تحلیل انتخاب ویژگی باعث افزایش کارایی پیش‌بینی‌های مدل می‌شود. بررسی این ویژگی‌ها امکان ارزیابی و تخصیص امتیاز به متغیرهای مختلف ورودی را فراهم کرده و شناسایی مهم‌ترین عوامل تأثیرگذار در مدل‌سازی را تسهیل می‌کند (۴۶، ۴۷). مطالعات پیشین نشان می‌دهند که مقادیر زیادی از متغیرهای کمکی محیطی به‌طور گسترده برای بررسی ویژگی‌های خاک استفاده شده‌اند. با این‌حال، هنوز یک راهنمای مشخص و استاندارد برای انتخاب مجموعه‌ای مشخص از این متغیرها وجود ندارد (۴۶، ۴۷).

اهمیت نسبی متغیرهای محیطی در پیش‌بینی دو ویژگی خاک MWD و SOC با استفاده از مدل‌های بهینه KNN و cubist بر اساس بهترین سناریو در شکل ۳ نشان داده شده است. در مدل KNN و سناریوی ۴ برای پیش‌بینی MWD، نتایج نشان داد که سیلت به‌عنوان مهم‌ترین متغیر تأثیرگذار، بیشترین اهمیت را در مدل دارد. این امر نشان‌دهنده نقش اساسی درصد سیلت خاک در پایداری خاکدانه‌ها و خواص فیزیکی خاک است. سیلت، با دانه‌های ریز خود، می‌تواند چسبندگی ذرات خاک به یکدیگر را بهبود بخشیده و موجب پایداری بیشتر خاکدانه‌ها و در نتیجه افزایش MWD شود.

برای پیش‌بینی SOC، مدل cubist و سناریوی ۴، نتایج نشان داد که شاخص MRVBF مهم‌ترین متغیر تأثیرگذار است (شکل ۳). در تأیید این نتایج، دهروماراجان و هگده[4] (۲۰۲۲) نیز MRVBF را به‌عنوان مهم‌ترین متغیر در پیش‌بینی توزیع مکانی ذرات خاک معرفی کرده‌اند (۴). این یافته‌ها نشان می‌دهند که شکل زمین، به‌ویژه مناطق هموار دره‌ها، تأثیر چشمگیری بر کیفیت خاک دارد. مناطق هموار دره‌ها معمولاً دارای خاک‌هایی با ویژگی‌های رسوبی و آبرفتی هستند که اغلب از نظر مواد آلی غنی‌ترند. به عبارت دیگر، این مناطق به‌دلیل ویژگی‌های خاص خود محل تجمع و ذخیره مواد آلی خاک هستند و این امر موجب بهبود کیفیت خاک می‌شود.

پس از MRVBF، میزان رس (Clay) به‌عنوان متغیر مهم بعدی شناخته شد (شکل ۳). این موضوع نشان می‌دهد که درصد رس در خاک تأثیر عمده‌ای بر محتوای مواد آلی دارد، زیرا خاک‌های رسی ظرفیت بالایی برای نگهداری آب و مواد مغذی دارند. در مقابل، زمین‌شناسی (Geology) به‌عنوان کم‌اثرترین متغیر در پیش‌بینی SOC شناخته شد. این نتایج نشان می‌دهند که ویژگی‌های توپوگرافی و ساختار خاک نقش کلیدی در پیش‌بینی SOC دارند. مطالعات دیگر نیز تأیید کرده‌اند که متغیرهای توپوگرافی تأثیر قابل توجهی بر تغییرات مکانی ویژگی‌های خاک دارند (۱۶، ۲۲).

1- Meliho *et al.*
2- John *et al.*
3- Extreme gradient boosting machine
4- Dharumarajan and Hegde

جدول (۲) نتایج اعتبارسنجی مدل‌های یادگیری ماشین در پیش‌بینی ویژگی‌های خاک

Table (2) Validation results of machine learning models for predicting soil properties

ویژگی‌های خاک Soil Properties	مدل Model	سناریو (۱) Scenario (1)			سناریو (۲) Scenario (2)			سناریو (۳) Scenario (3)			سناریو (۴) Scenario (4)		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
میانگین وزنی قطر خاکدانه‌ها Mean weight diameter of soil aggregates (MWD, mm)	ANN	0.54	0.18	0.16	0.57	0.18	0.15	0.57	0.17	0.15	0.55	0.18	0.15
	RF	0.56	0.19	0.17	0.54	0.18	0.16	0.50	0.46	0.16	0.52	0.17	0.15
	Cubist	0.50	0.17	0.15	0.55	0.18	0.15	0.57	0.19	0.16	0.57	0.18	0.15
	SVM	0.56	0.18	0.16	0.44	0.24	0.21	0.48	0.32	0.24	0.53	0.23	0.20
	KNN	0.56	0.19	0.16	0.45	0.19	0.16	0.51	0.18	0.16	0.59	0.19	0.16
	BRT	0.50	0.19	0.16	0.53	0.20	0.17	0.49	0.18	0.15	0.46	0.19	0.16
ماده آلی خاک Soil organic carbon (SOC, %)	ANN	0.63	9.60	7.97	0.39	9.62	8.03	0.55	9.58	8.05	0.59	9.74	8.16
	RF	0.47	6.31	5.58	0.44	5.70	5.0	0.41	5.74	5.08	0.72	4.44	3.70
	Cubist	0.54	5.56	4.65	0.68	5.52	4.64	0.53	5.73	4.92	0.78	4.12	3.53
	SVM	0.51	5.73	4.80	0.64	7.50	6.25	0.59	7.96	6.36	0.59	7.01	5.62
	KNN	0.57	5.95	5.18	0.56	5.45	4.77	0.47	5.78	5.01	0.56	5.92	5.15
	BRT	0.52	6.05	5.28	0.63	5.91	5.09	0.58	6.01	4.98	0.63	4.69	4.04

ANN: شبکه عصبی مصنوعی، RF: جنگل تصادفی، Cubist: کوبیست، SVM: ماشین بردار پشتیبان، KNN: الگوریتم K نزدیک‌ترین همسایه، BRT: رگرسیون درختی توسعه‌یافته، R^2: ضریب تعیین، RMSE: ریشه میانگین مربعات خطا و MAE: میانگین خطای مطلق.

ANN: Artificial neural networks, RF: Random forest, SVM: Support vector machine, KNN: K-nearest neighbors, BRT: Boosted regression trees, R^2: Coefficient of determination, RMSE: Root-mean-square error, and MAE: Mean absolute error.

در مطالعه‌ای که توسط خورده‌بین و همکاران[1] (۱۳۹۹) انجام شد، برای تخمین ذخیره SOC در بخشی از اراضی بهبهان، از متغیرهای کمکی شامل اجزای سرزمین، داده‌های سنجش از دور و نقشه کاربری اراضی استفاده شد. یافته‌ها نشان داد که مدل RF با R^2 بالا (۰/۹۶) و مقدار RMSE پایین، بهترین عملکرد را در این زمینه داشت (۱۴). در پژوهشی دیگر در شهرستان کامیاران استان کردستان، ۱۱۰ نمونه خاک از کاربری‌های مختلف جمع‌آوری و برای پیش‌بینی SOC از ۱۰۱ متغیر کمکی شامل DEM، تصاویر ماهواره‌ای و متغیرهای اقلیمی استفاده شد. نتایج نشان داد که مدل ANN با R^2 برابر با ۰/۹۷ عملکرد بهتری نسبت به رگرسیون خطی چند متغیره با R^2=۰/۵۹ داشت و پراکنش SOC بیشتر تحت تأثیر عوامل توپوگرافی و اقلیمی بود (۲۱). نتایج مطالعه رحمتی و همکاران[2] (۱۴۰۳) نشان داد که الگوریتم RF در برآورد میزان SOC عملکرد بهتری نسبت به روش رگرسیون خطی چندمتغیره داشته است، به‌طوری‌که مقادیر RMSE و R^2 برای جنگل تصادفی برابر با ۰/۱۲ و ۰/۷۹ و برای رگرسیون خطی چندمتغیره به ترتیب ۰/۱۹ و ۰/۵۷ به دست آمد. همچنین، یافته‌های آن‌ها نشان داد که متغیرهای محیطی تأثیرگذار بر توزیع SOC در مدل‌های مورد بررسی یکسان نبودند؛ به‌گونه‌ای که در مدل RF، شاخص‌های مرتبط با پوشش گیاهی و در مدل رگرسیون خطی چندمتغیره، شاخص‌های توپوگرافی نقش برجسته‌تری در پیش‌بینی SOC ایفا کرده‌اند (۳۲).

نتایج مطالعات مختلف نشان داد که مشتقات DEM و پیش‌بینی‌کننده‌های استخراج‌شده از سنتینل ۱ و سنتینل ۲، به‌ویژه ترکیب آن‌ها، برای مدل‌سازی مؤثر ویژگی‌های خاک اهمیت دارد و موجب بهبود عملکرد پیش‌بینی می‌شود (۳۹، ۴۸). ترکیب این متغیرهای کمکی، بالاترین دقت مدل‌سازی را به دست آورد، به‌طوری‌که ۴۴٪ از تغییرپذیری SOC را توضیح داد (۴۸). پژوهشگران دیگر نیز به نتایجی مشابه دست یافتند و تأکید کردند که متغیرهای توپوگرافی مانند ضریب LS، شیب و ارتفاع، مهم‌ترین پیش‌بینی‌کننده‌ها برای نقشه‌برداری ویژگی‌های مختلف خاک هستند (۹، ۴۸). علاوه بر این، در پژوهشی دیگر گزارش شد که شاخص‌های پوشش گیاهی و انعکاس سنجش از دور از عوامل کلیدی برای پیش بینی ویژگی‌های خاک به‌شمار می‌روند (۴۳). همچنین در گزارشی در شمال ایران، ویژگی‌های توپوگرافی و شاخص‌های پوشش گیاهی به‌عنوان متغیرهای کمکی در مدل‌سازی SOC شناسایی شدند که می‌توانند برای ارزیابی سریع و مقرون‌به‌صرفه غلظت SOC به‌منظور انجام اقدامات مدیریتی در خاک‌های جنگلی استفاده شوند (۳۷).

نتایج تجزیه و تحلیل اهمیت متغیرها در مطالعه حاضر نشان می‌دهند که ترکیبی از ویژگی‌های توپوگرافی، زمین‌شناسی و خصوصیات فیزیکی خاک تأثیر چشمگیری در پیش‌بینی ویژگی‌های خاک دارد. این تحلیل‌ها علاوه بر تأکید بر اهمیت متغیرهای محیطی برای پیش‌بینی تغییرات مکانی خاک، مسیرهایی برای بهینه‌سازی مدل‌های ML در مدیریت خاک و کشاورزی فراهم می‌آورند. تأثیر متغیرهای کمکی در پیش‌بینی کیفیت خاک بسته به ویژگی خاص خاک و مدل به کار رفته متفاوت است، و انتخاب سناریوی مناسب می‌تواند دقت پیش‌بینی‌ها را افزایش دهد. استفاده از متغیرهای ترکیبی مانند ویژگی‌های توپوگرافی و خصوصیات فیزیکی خاک می‌تواند دقت مدل‌ها را ارتقاء داده و در طراحی پایدار مدیریت خاک و ارتقای بهره‌وری آن در شرایط مختلف اکوسیستمی نقش مؤثری ایفا کند.

توزیع مکانی ویژگی‌های خاک

در شکل ۴، توزیع مکانی ویژگی‌های خاک شامل MWD و SOC بر اساس بهترین مدل‌ها و سناریوها ارائه شده است. بر اساس مدل KNN در سناریوی ۴، مقادیر MWD بین ۰/۴۱ تا ۰/۷۴ میلی‌متر متغیر بود. مقادیر

1- Khordehbin *et al.*
2- Rahmati *et al.*

پایین‌تر ممکن است ناشی از فرسایش خاک، کاهش مواد آلی یا فعالیت‌های کشاورزی نامناسب باشد، در حالی‌که مقادیر بالاتر نشان‌دهنده پایداری بهتر خاکدانه‌ها و کیفیت مناسب‌تر خاک است.

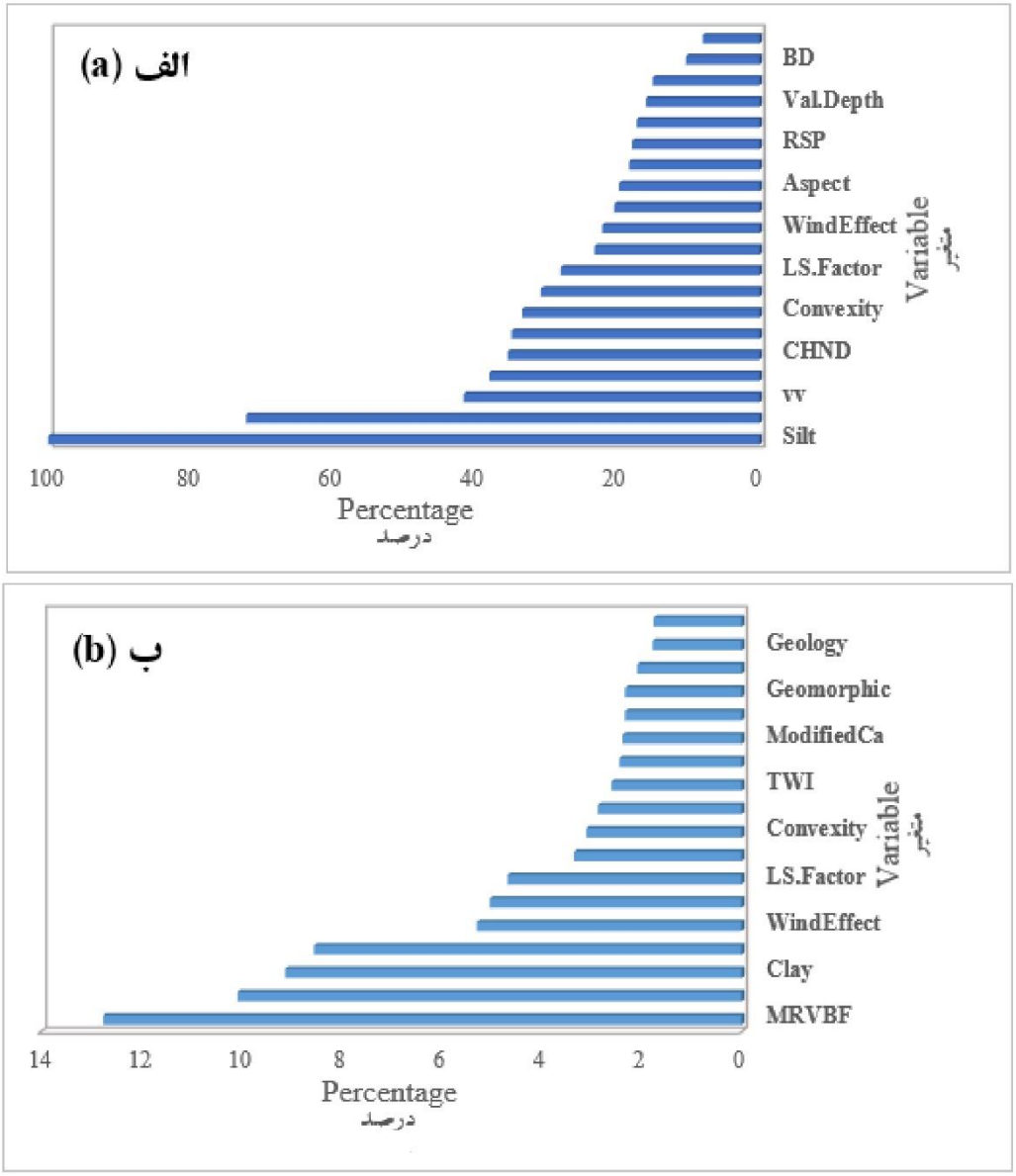
مقادیر SOC بر اساس مدل cubist در سناریوی ۴، بین ۲/۸۰ تا ۱۹/۷۳ درصد متغیر بود و با تغییرات رنگ از آبی (مقادیر کم) به قرمز (مقادیر بالا) نشان داده شد. مقادیر بالای SOC معمولاً در مناطقی با شرایط مساعد، مانند پوشش گیاهی انبوه، تجمع مواد آلی و رطوبت کافی، مشاهده می‌شود، درحالی‌که مقادیر پایین آن ناشی از فرسایش خاک، کاهش پوشش گیاهی در منطقه برای تثبیت کربن آلی خاک، یا بهره‌برداری نامناسب است. الگوی پراکندگی نشان می‌دهد که مقدار SOC به‌طور قابل‌توجهی در مناطق مختلف متفاوت است و این تفاوت به عواملی مانند توپوگرافی، ویژگی‌های زمین‌شناسی، و نوع پوشش گیاهی مرتبط است. شناسایی مناطق با مقادیر بالای SOC می‌تواند در برنامه‌ریزی کشاورزی و حفاظت از خاک نقش مؤثری داشته باشد.

در پژوهشی گزارش شد که ارتفاع یک متغیر مؤثر برای پیش‌بینی ویژگی‌های خاک است و ذخیره SOC با افزایش ارتفاع افزایش می‌یابد. تفاوت‌ها در شیب سطح ممکن است به طور غیرمستقیم بر ورودی و از دست رفتن کربن از طریق ارتباط نزدیک آن با پوشش گیاهی، دما و بارش تأثیر بگذارد (۳۸). در نواحی کوهستانی منطقه مطالعه ما، بارش نسبتاً زیاد و دماهای پایین برای تجمع ماده آلی و کربن در خاک مساعد است. زمین‌های با ارتفاع پایین تر که بارش نسبتاً کمی دارند و دماهای بالاتری در منطقه تجربه می کنند، محتوای SOC کمتری داشتند و عمدتاً پوشیده از اراضی کشاورزی بودند. این یافته مشابه نتایج مطالعه دیگری است که نشان داد محتوای بالاتر SOC در مناطق جنگلی و محتوای کمتر SOC در مناطق کشاورزی وجود دارد (۴۲) که می‌توان آن را با افزایش تجزیه مواد آلی و از دست رفتن از طریق فرسایش و کشاورزی توضیح داد.

مناطق با مقادیر پایین ویژگی‌های کیفیت خاک نیازمند مدیریت پایدار خاک، نظیر افزودن مواد آلی، کاهش فرسایش و استفاده از شیوه‌های کشاورزی حفاظتی هستند. اطلاعات مکانی ارائه‌شده می‌تواند به سیاست‌گذاران و کشاورزان در برنامه‌ریزی‌های محلی برای بهبود کیفیت خاک کمک کند.

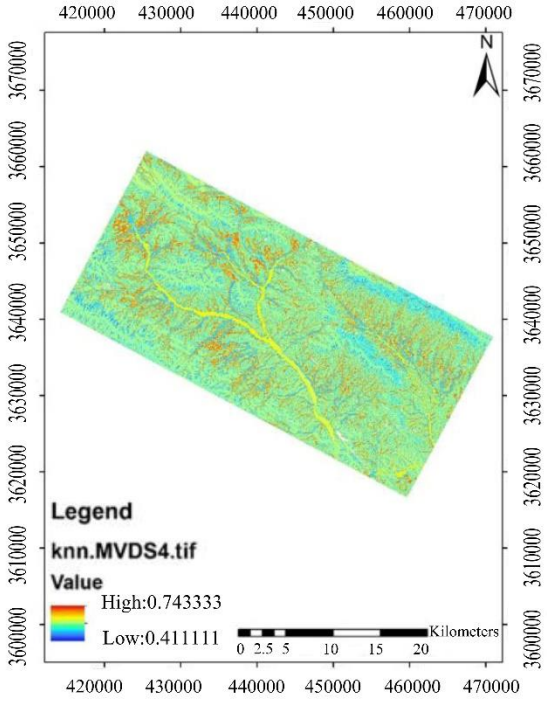
آگاهی از توزیع مکانی ویژگی‌های خاک برای مدیریت خاک، تولید آب و پایداری در مقیاس حوزه آبریز بسیار اهمیت زیادی دارد. با این‌حال، کمبود داده‌های خاک مانع از بهره‌برداری کامل از این ابزارها می‌شود، بنابراین نیاز به برآورد ویژگی‌های خاک و انجام توزیع مکانی آن احساس می‌شود (۷).

در مطالعه‌ای شاخص‌های پایداری تجزیه خاک، از جمله MWD، میانگین هندسی قطر خاکدانه‌ها (GMD)، تجمعات مقاوم به آب (WSA) و SOC در بخش‌های مختلف تجمعی خاک با استفاده از DSM و مدل‌های ML پیش‌بینی مکانی شدند. مدل‌های RF، KNN، SVM وANN همراه با ترکیب چهار مدل مستقل، با استفاده از روش اعتبارسنجی متقابل ۱۰ برابری تکراری آموزش داده شدند. نتایج نشان داد که مدل RF عملکرد بهتری در پیش‌بینی MWD (R^2=۰/۷۴)، GMD (R^2=۰/۷۵) و WSA (R^2=۰/۵۸) داشت، در حالی‌که مدل‌های kNN و SVM پیش‌بینی بهتری برای SOC ارائه دادند. با این‌حال، مدل ترکیبی دقت پیش‌بینی‌ها را برای تمام اهداف خاک افزایش داد. علاوه بر این، تحلیل اهمیت متغیرها نشان داد که ویژگی‌های خاک مانند SOM و داده‌های سنجش از دور عمدتاً تغییرات شاخص‌های پایداری تجزیه خاک و SOC در انواع بخش‌های تجمعی خاک را توضیح می‌دهند (۴۷).

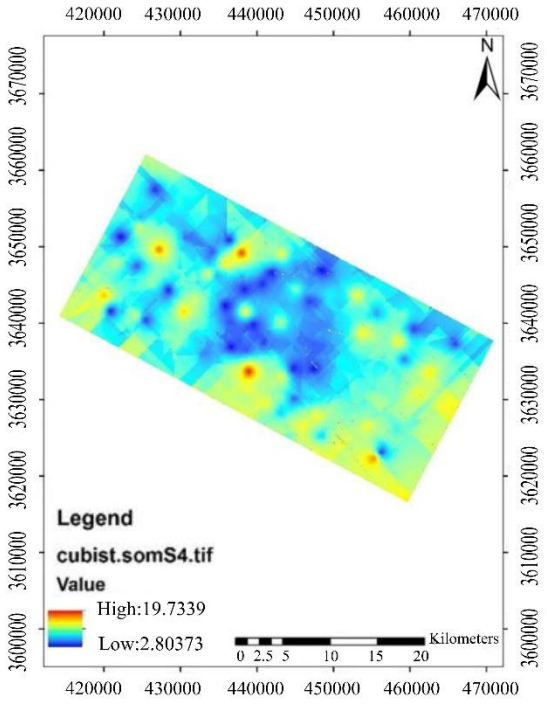


شکل (۳) اهمیت نسبی متغیرهای کمکی در پیش‌بینی ویژگی‌های کیفیت خاک با استفاده از مدل پیش‌بینی بهینه. (الف) میانگین وزنی قطر خاکدانه‌ها (MWD، میلی‌متر)، الگوریتم K نزدیک‌ترین همسایه (KNN)، سناریوی ٤ و (ب) کربن آلی خاک (SOC، درصد)، مدل کوبیست، سناریوی ٤

Figure (3) The relative importance of auxiliary variables in predicting soil quality attributes using the optimal prediction model, (a) Mean weight diameter of soil aggregates (MWD, mm)using the K-nearest neighbors (KNN) algorithm, Scenario 4; (b) Soil organic carbon (SOC, %) using the Cubist model, Scenario 4.



الف(a)



ب(b)

شکل (٤) توزیع مکانی ویژگی‌های پیش‌بینی شده کیفیت خاک بر اساس بهترین مدل و سناریو. (الف) میانگین وزنی قطر خاکدانه‌ها (MWD، میلی‌متر)، الگوریتم K نزدیک‌ترین همسایه (KNN)، سناریوی ٤ و (ب) کربن آلی خاک (SOC، درصد)، مدل کوبیست، سناریوی ٤

Figure (4) Spatial distribution of predicted soil quality properties based on the best model and scenario. (a) Mean weight diameter of soil aggregates (MWD, mm) using the K-nearest neighbors (KNN) algorithm, Scenario 4, and (b) soil organic carbon (SOC, %) using the Cubist model, Scenario 4.

در تحقیقی دقت مدل‌های رگرسیون خطی چندگانه (MLR)، ANN و برنامه‌ریزی بیان ژن (GEP) در برآورد MWD خاکدانه‌های تر مقایسه شد و داده‌ها از ۸۰ نمونه خاک سطحی از اراضی جنگلی، مرتعی و زراعی استان اردبیل جمع‌آوری و ویژگی‌های فیزیکی و شیمیایی آن‌ها اندازه‌گیری شد. نتایج نشان داد که مدل ANN با داشتن R^2 بالا، RMSE پایین و ضریب نش-ساتکلیف (NS) مناسب، بهترین عملکرد را در پیش‌بینی MWD نسبت به مدل‌های MLR و GEP داشت. در حالی‌که مدل MLR دقت کمتری و خطای بیشتری در برآورد MWD نشان داد. متغیرهای مهم در مدل‌ها شامل شن، کربن آلی و GWD خاکدانه‌های خشک بودند (۲).

مطالعات پیشین نشان داده‌اند که مدل RF و SVM از روش‌های کارآمد در DSM برای پیش‌بینی اجزای بافت خاک هستند. صحرایی و همکاران[1] (۱۴۰۱) در پژوهشی بر روی اراضی استان خوزستان گزارش کردند که مدل RF با ارائه مقادیر بالاتر R^2 و خطای کمتر نسبت به SVM، عملکرد دقیق‌تری در پیش‌بینی ذرات شن، سیلت و رس داشته است. همچنین، از میان متغیرهای کمکی مشتق‌شده از DEM و شاخص‌های طیفی و گیاهی سنجش از دور، شاخص‌هایی مانند شوری، روشنایی، شبکه آبراهه‌ها و باند ۶ ماهواره لندست ۸ به‌عنوان مهم‌ترین متغیرهای محیطی در پیش‌بینی بافت خاک شناسایی شدند (۳۳).

در پژوهشی، مدل RF عملکرد بهتری نسبت به مدل SVM در نقشه‌برداری SOC در سوئد نشان داده است (۳۵). در مقابل، در مطالعه‌ای دیگر، مدل SVM در مقایسه با RF در نقشه‌برداری SOC در کنیا عملکرد برتری داشته است (۴۲). علاوه بر این، در تحقیقاتی دیگر، نتایج معکوسی در مقایسه مدل‌های BRT و RF در پیش‌بینی ویژگی‌های خاک گزارش شده است (۲۸). همچنین در تحقیقی در زمینه الگوریتم‌ها و متغیرهای کمکی DSM، الگوریتم رگرسیون بردار پشتیبان (SVR) پیش‌بینی‌های SOC را با دقت کمتری نسبت به مدل‌های RF و BRT ارائه داده است (۱۸). این نتایج نشان می‌دهد که هیچ مدل پیش‌بینی واحدی نمی‌تواند در تمامی شرایط بهترین عملکرد را داشته باشد. عدم تطابق در نتایج مطالعات مختلف ممکن است ناشی از تفاوت در مساحت مناطق مطالعه، ماهیت مجموعه داده‌ها، کیفیت و کمیت متغیرهای محیطی و تراکم نمونه‌برداری باشد. افزون بر این، ارزیابی مدل‌ها با استفاده از مجموعه داده‌های مستقل می‌تواند تفاوت‌های قابل توجهی در عملکرد این الگوریتم‌ها ایجاد کند (۱۷). بر اساس این نتایج، پیشنهاد می‌شود برای کالیبراسیون مدل‌های پیش‌بینی رقابتی، از مجموعه داده‌های آزمایشی خاص و مناسب استفاده شود تا ارزیابی عملکرد مدل‌ها با دقت بیشتری انجام گیرد (۱۸).

نتیجه‌گیری

این پژوهش به ارزیابی کارایی مدل‌های ML در پیش‌بینی ویژگی‌های خاک بر اساس داده‌های چندمنظوره، از جمله داده‌های سنجش از دور، ویژگی‌های توپوگرافی، داده‌های طبقه‌بندی‌شده و خصوصیات خاک در منطقه فریدن استان اصفهان پرداخته است. مدل‌های cubist و KNN، به‌ویژه در سناریوی چهارم، توانایی بالایی در پیش‌بینی دقیق‌تر و پایدارتر ویژگی‌های کلیدی خاک مانند MWD و SOC نشان دادند. ترکیب متغیرهای محیطی شامل شاخص‌های زمین‌شناختی، توپوگرافی و پوشش گیاهی، دقت پیش‌بینی‌ها را به‌طور قابل توجهی افزایش داده و تحلیل مکانی این ویژگی‌ها را تسهیل کرده است. نتایج این مطالعه نشان داد که برای پیش‌بینی هر یک از ویژگی‌های کلیدی خاک، برخی متغیرها از اهمیت بیشتری برخوردار بودند. برای MWD، Silt به‌عنوان مهم‌ترین متغیر تأثیرگذار شناسایی شد و پس از آن VV و CHND در رتبه‌های بعدی قرار گرفتند. در مورد SOC، متغیرهای MRVBF و Clay بیشترین تأثیر را داشتند. این

1- Sahraei *et al.*

پژوهش تأکید دارد که ترکیب ویژگی‌های فیزیکی، شیمیایی و توپوگرافی خاک در کنار مدل‌های پیشرفته ML، ابزارهای قدرتمندی برای تحلیل دقیق کیفیت خاک و مدیریت پایدار منابع طبیعی فراهم می‌کند. نقشه‌های رقومی تولیدشده می‌توانند سیاست‌گذاران و کشاورزان را در شناسایی مناطق بحرانی و طراحی برنامه‌های بهبود کیفیت خاک یاری دهند. به طور کلی، یافته‌های این پژوهش نشان داد که بهره‌گیری از نقشه‌برداری رقومی می‌تواند فرآیند تهیه نقشه‌های خاک را بهبود بخشیده و سرعت و کارایی دسترسی به اطلاعات و داده‌های مورد نیاز کارشناسان خاک و آب را افزایش دهد. همچنین، این تحقیق بستری برای مطالعات آینده در زمینه مدل‌سازی پیشرفته کیفیت خاک و ارتقای بهره‌وری زراعی در شرایط اقلیمی و توپوگرافی مشابه فراهم می‌آورد. برای ارتقای دقت و پایداری پیش‌بینی‌ها، پیشنهاد می‌شود در مطالعات آینده از داده‌های ترکیبی با وضوح مکانی و زمانی بالا، الگوریتم‌های پیشرفته‌ای نظیر cubist و KNN، و مدل‌های دینامیکی برای تحلیل تغییرات زمانی استفاده شود. همچنین، بررسی تعاملات میان خاک، آب و گیاه و تحلیل اثرات تغییرات اقلیمی می‌تواند گامی مؤثر در ارتقای بهره‌وری زراعی و مدیریت پایدار منابع خاک باشد.

References

1. Amir Abedi, H., Asghari, S.A., Mesri Ghandomshin, T., and Balandeh, N. 2016. Prediction of mean weight diameter of aggregates using artificial neural network and regression models. Applied Soil Research, 4(1): 39-53. (In Persian with English Abstract)

2. Asghari, S., Heidari, K., Hasanpour Kashani, M., and Shahab Arkhazloo, H. 2025. Evaluation of regression and intelligent models for estimating mean weight diameter of wet aggregates. Water and Soil, 38(5): DOI: 10.22067/jsw.2025.91071.1454. (In Persian with English Abstract)

3. Cho, K.H., Beon, M.S., and Jeong, J.C. 2018. Dynamics of soil salinity and vegetation in a reclaimed area in Saemangeum, Republic of Korea. Geoderma. 321: 42–51. DOI: 10.1016/j.geoderma.2018.01.031.

4. Dai, L., Wang, Z., Zhuo, Z., Ma, Y., Shi, Z., and Chen, S. 2025. Prediction of soil organic carbon fractions in tropical cropland using a regional visible and near-infrared spectral library and machine learning. Soil and Tillage Research, 245: 106297. DOI: 10.1016/j.still.2024.106297.

5. Dharumarajan, S., and Hegde, R. 2022. Digital mapping of soil texture classes using Random Forest classification algorithm. Soil Use and Management, 38(1): 135-149. DOI: 10.1111/sum.12668.

6. Emami, N.S., Chavoshi, E., Ayoubi, S., Honarjoo, N., and Zeraatpisheh, M. 2024. Comprehensive assessment of soil quality in various land uses: a comparative analysis of soil quality index models. Environmental Earth Sciences, 83(17): 498.

7. Ghavami, M.S., Ayoubi, S., Mosaddeghi, M.R., and Naimi, S. 2023. Digital mapping of soil physical and mechanical properties using machine learning at the watershed scale. Journal of Mountain Science, 20(10): 2975-2992. DOI: 10.1007/s11629-023-8056-z.

8. Guo, L., Fu, P., Shi, T., Chen, Y., Zhang, H., Meng, R., and Wang, S. 2020. Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images. Soil and Tillage Research, 196: 104477. DOI: 10.1016/j.still.2019.104477.

9. Hinge, G., Surampalli, R.Y., and Goyal, M.K. 2018. Prediction of soil organic carbon stock using digital mapping approach in humid India. Environmental Earth Sciences, 77: 1-10. DOI: 10.1007/s12665-018-7374-x.

10. Jiang, L. 2021. A fast and accurate circle detection algorithm based on random sampling. Future Generation Computer Systems, 123: 245-256.

11. John, K., Abraham Isong, I., Michael Kebonye, N., Okon Ayito, E., Chapman Agyeman, P., and Marcus Afu, S. 2020. Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. Land, 9(12): 487. DOI: 10.3390/land9120487.

12. Kemper, W., and Rosenau, R. 1986. Aggregate stability and size distribution. In: Klute A (Ed) Methods of soil analysis: Part 1 Physical and mineralogical methods, Wiley, 5: 425-442. DOI: 10.2136/sssabookser5.1.2ed.c17.

13. Khaledian, Y., and Miller, B.A. 2020. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling, 81: 401-418. DOI: 10.1016/j.apm.2019.12.016.

14. Khordehbin, S., Hojatia, S.A.E.I.D., Landi, A., and Ahmadianfar, I. 2020. Comparison of different data mining methods in predicting soil organic carbon storage in some lands of Behbahan city. Iranian Journal of Soil and Water Research, 51(4): 1041-1054. DOI: 10.22059/ijswr.2020.288526.668308. (In Persian with English Abstract)

15. Khormali, F., Ajami, M., Ayoubi, S., Srinivasarao, C., and Wani, S.P. 2009. Role of deforestation and hillslope position on soil quality attributes of loess-derived soils in Golestan province, Iran. Agriculture, Ecosystems and Environment, 134(3-4): 178-189. DOI: 10.1016/j.agee.2009.06.017.

16. Khosravani, P., Baghernejad, M., Moosavi, A.A., and Rezaei, M. 2023. Digital mapping and spatial modeling of some soil physical and mechanical properties in a semi-arid region of Iran. Environmental Monitoring and Assessment, 195(11): 1367. DOI: 10.1007/s10661-023-11980-6.

17. Khosravi, V., Gholizadeh, A., Kodešová, R., Agyeman, P.C., Saberioon, M., and Borůvka, L. 2025. Visible, near-infrared, and shortwave-infrared spectra as an input variable for digital mapping of soil organic carbon. International Soil and Water Conservation Research, 13(1): 203-214. DOI: 10.1016/j.iswcr.2024.10.002.

18. Lamichhane, S., Kumar, L., and Wilson, B. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma, 352: 395-413. DOI: 10.1016/j.geoderma.2019.05.031.

19. Ließ, M., Schmidt, J., and Glaser, B. 2016. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. PLOS One, 11(4): e0153673. DOI: 10.1371/journal.pone.0153673.

20. Loiseau, T., Chen, S., Mulder, V.L., Dobarco, M.R., Richer-de-Forges, A.C., Lehmann, S., ... and Arrouays, D. 2019. Satellite data integration for soil clay content modelling at a national scale. International Journal of Applied Earth Observation and Geoinformation, 82: 101905. DOI: 10.1016/j.jag.2019.101905.

21. Mahmoudzadeh, H., Matinfar, H.R., and Taghizadeh, M.R. 2021. Digitization of soil organic carbon with artificial neural network and multivariate linear regression in Kurdistan province. Journal of Soil Management and Sustainable Production, 10(4): 77-98. (In Persian with English Abstract)

22. Maleki, S., Karimi, A., Mousavi, A., Kerry, R., and Taghizadeh-Mehrjardi, R. 2023. Delineation of soil management zone maps at the regional scale using machine learning. Agronomy, 13(2): 445. DOI: 10.3390/agronomy13020445.

23. Meliho, M., Boulmane, M., Khattabi, A., Dansou, C.E., Orlando, C.A., Mhammdi, N., and Noumonvi, K.D. 2023. Spatial prediction of soil organic carbon stock in the Moroccan High Atlas using machine learning. Remote Sensing, 15(10): 2494. DOI: 10.3390/rs15102494.

24. Moameni, A. and Zinck, J.A., 1997. Application of statistical quality control charts and geostatistics to soil quality assessment in a semi-arid environment of south-central Iran. International Institute for Aerospace Survey and Earth Sciences (ITC), 3(4): 1-26.

25. Naimi, S., Ayoubi, S., Zeraatpisheh, M., and Dematte, J.A.M. 2021. Ground observations and environmental covariates integration for mapping of soil salinity: a machine learning-based approach. Remote Sensing, 13(23): 4825. DOI: 10.3390/rs13234825.

26. Nelson, R. 1983. Carbonate and gypsum. In: Page AL (ed) Methods of soil analysis: Part 2 Chemical and microbiological properties, Wiley, 9: 181-197. DOI: 10.2134/agronmonogr9.2.2ed.c11.

27. Nozari, S., Pahlavan-Rad, M.R., Brungard, C., Heung, B., and Boruvka, L. 2024. Digital soil mapping using machine learning-based methods to predict soil organic carbon in two different districts in the Czech Republic. Soil and Water Research, 19(1): 32-49. DOI: 10.17221/119/2023-SWR.

28. Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., and Papritz, A. 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil, 4(1): 1-22. DOI: 10.5194/soil-2017-14.

29. Qiu, Y.L., Chen, C., Han, J., Wang, X.T., Wei, S.Y., and Zhang, Z.T. 2019. Satellite remote sensing estimation model of soil salinity in Jiefangzha irrigation under vegetation coverage. Water Saving Irrigation, 0(10): 108-112

30. Rahimi Ashjerdi, M., and Ayoubi, S. 2013. Impacts of land use change and slope positions on some soil properties and magnetic susceptibility in Ferydunshahr district, Isfahan province. Water and Soil, 27(5): 882-895. DOI: 10.22067/jsw.v0i0.31182. (In Persian with English Abstract)

31. Rahimi mashkale, M., Delavar, M. A., jamshidi, M., and sharififar, A. 2023. Improving the classification of soil imbalanced data using machine learning algorithms in some part of Zanjan provice land. Agricultural Engineering, 46(1): 61-82. DOI: 10.22055/agen.2023.43838.1667. (In Persian with English Abstract)

32. Rahmati, F., Rngzan, K., and Landi, A. 2024. Modelling and digitally mapping of surface soil organic carbon in Semirom county employing several machine learning algorithms. Agricultural Engineering, 47(4): 585-605. DOI: 10.22055/agen.2024.47815.1740. (In Persian with English Abstract)

33. Sahraei, N., Landi, A., and Hojati, S. 2022. Digital mapping of soil texture components in part of Khuzestan plain lands using machine learning models. Iranian Journal of Soil and Water Research, 53(10): 2261-2276. DOI: 10.22059/ijswr.2022.348442.669360. (In Persian with English Abstract)

34. 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): 1-15. DOI: 10.1016/j.ecolmodel.2004.06.036.

35. Siewert, M.B. 2018. High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. Biogeosciences, 15(6): 1663-1682. DOI: 10.5194/bg-15-1663-2018.

36. Soil Survey Staff. 2014. Keys to Soil Taxonomy, 12nd Ed. USDA-Natural Resources Conservation Service, Washington, DC.

37. Tajik, S., Ayoubi, S., and Zeraatpisheh, M. 2020. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. Geoderma Regional, 20: e00256. DOI: 10.1016/j.geodrs.2020.e00256.

38. Tsui, C.C., Tsai, C.C., and Chen, Z.S. 2013. Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. Geoderma, 209: 119-127. DOI: 10.1016/j.geoderma.2013.06.013.

39. Vaudour, E., Gomez, C., Fouad, Y., and Lagacherie, P. 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sensing of Environment, 223: 21-33. DOI: 10.1016/j.rse.2019.01.006.

40. Wadoux, A.M.C., Minasny, B., and McBratney, A.B. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews, 210: 103359. DOI: 10.1016/j.earscirev.2020.103359.

41. Walkley A., and Black, I.A. 1934. An examination of the Degtjareff method for determining organic carbon in soils: effect of variations in digestion conditions and of inorganic soil constituents. Soil Science, 63: 251-263. DOI: 10.1097/00010694-194704000-00001.

42. Were, K., Bui, D.T., Dick, O.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: 394-403. DOI: 10.1016/j.ecolind.2014.12.028.

43. Yang, R.M., Zhang, G.L., Liu, F., Lu, Y.Y., Yang, F., Yang, F., Yang, M., Zhao, Y.G., and Li, D.C. 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic

carbon concentration in an alpine ecosystem. Ecological Indicators, 60: 870-878. DOI: 10.1016/j.ecolind.2015.08.036.

44. Ye, Z., Sheng, Z., Liu, X., Ma, Y., Wang, R., Ding, S., Liu, M., Li, Z., and Wang, Q. 2021. Using machine learning algorithms based on GF-6 and Google Earth engine to predict and map the spatial distribution of soil organic matter content. Sustainability, 13(24): 14055. DOI: 10.3390/su132414055.

45. Ye, L., Tan, W., Fang, L., and Ji, L. 2019. Spatial analysis of soil aggregate stability in a small catchment of the Loess Plateau, China: II. Spatial prediction. Soil and Tillage Research, 192: 1-11. DOI: 10.1016/j.still.2019.03.009.

46. Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., and Finke, P. 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma, 338915: 445-452. DOI: 10.1016/j.geoderma.2018.09.006.

47. Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M.R., and Xu, M. 2021. Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. Geoderma Regional, 27: e00440. DOI: 10.1016/j.geodrs.2021.e00440.

48. Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., and Lausch, A. 2020. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Science of The Total Environment, 729: 138244. DOI: 10.1016/j.scitotenv.2020.138244.

49. Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., ... and Lausch, A. 2021. Prediction of soil organic carbon and the C: N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Science of The Total Environment, 755: 142661. DOI: 10.1016/j.scitotenv.2020.142661.

50. Zhou, Y., Ma, H., Xie, Y., Jia, X., Su, T., Li, J., and Shen, Y. 2020. Assessment of soil quality indexes for different land use types in typical steppe in the loess hilly area, China. Ecological Indicators, 118: 106743. DOI: 10.1016/j.ecolind.2020.106743.

51. Zolfaghari, A.A., Yazdani, M.R., Khosravi. M., and Mahmoudi S.M. 2019. Comparison of different data mining methods for digital mapping of soil particle-size fractions in lands of Semnan plain. Iranian Journal of Soil and Water Research, 51(2): 375-385. DOI: 10.22059/ijswr.2019.285195.668259. (In Persian with English Abstract)