



Gorgan University of Agricultural
Sciences and Natural Resources



Soil Science Society of Iran

(OPEN ACCESS)

Digital mapping of soil organic carbon using auxiliary data and machine learning models in the Badr watershed, Kurdistan province

Moslem Zarini Bahador^{*1}, Rahman Sharifi²

1. Corresponding Author, Ph.D. Graduate in Soil Science, Soil and Water Research Expert, Agricultural and Natural Resources Research and Training Center, Tehran Province, AREEO, Tehran, Iran. E-mail: moslem.zarini@gmail.com
2. Member of Scientific Board, Soil and Water Research Division, Tehran Agricultural and Natural Resources Research and Training Center, AREEO, Tehran, Iran. E-mail: r.sharifi@areeo.ac.ir

Article Info	ABSTRACT
Article type: Full Length Research Paper	Background and Objectives: Soil organic carbon (SOC) is a key factor in maintaining soil fertility and supporting soil ecosystem services, and it has been recognized by the United Nations Environment Programme as one of the major global environmental challenges. Studies in Iran have shown that, on average, each gram increase of organic carbon per kilogram of soil can enhance wheat grain yield by 286 kilograms per hectare. Furthermore, understanding changes in SOC is a central component in soil quality assessment. Therefore, this study was conducted with the aim of digitally mapping soil organic carbon using environmental auxiliary variables and predictive models in the Badr watershed, south of Qorveh city, in order to evaluate the effects of environmental characteristics and auxiliary variables on SOC prediction and to identify the most suitable model for estimating soil organic carbon reserves.
Article history: Received: 07.26.2024 Revised: 01.24.2025 Accepted: 01.25.2025	
Keywords: Auxiliary variables, Digital soil mapping, Linear multivariate regression, Satellite images	
	Materials and Methods: In this research, auxiliary data including Landsat 8 satellite imagery and a digital elevation model (DEM) with a spatial resolution of 10 meters were obtained from the National Cartographic Center. The geological map of Qorveh was acquired from the national geological database, from which the geological map of the Badr watershed was extracted and digitized in a Geographic Information System (GIS) environment, and the geomorphological map was prepared based on the Zinck geopedology method. The locations of the observation points were then determined, soil identification was carried out in the field, and soil samples were collected from different depths. These samples were analyzed for physical, chemical, and mineralogical properties, and the soils were classified accordingly. To implement the study, 125 soil profiles were selected using the Latin hypercube sampling method and excavated across the watershed; the samples were air-dried in the laboratory, ground, and passed through a 2 mm sieve, after which soil organic matter was measured using the Walkley–Black method. For modeling, artificial neural networks (ANN), decision tree analysis, and multivariate linear regression were used to predict soil properties, and to integrate the outputs of the models, the k-nearest neighbor (k-NN) method was applied for preparing and evaluating the digital maps of soil classes and characteristics.

Results: The prediction results were first evaluated using the 5-fold random validation method, and in the second step, decision tree analysis, artificial neural networks, k-nearest neighbor (KNN), and random forest models were applied for prediction, while the mixed multivariate linear regression method was used to integrate the outputs of the models. The accuracy of these predictions was then assessed using the 10-fold cross-validation method. The findings showed that among the applied models, multivariate linear regression (MLR) achieved the highest accuracy for predicting soil organic matter with a coefficient of determination (R^2) of 0.637 and a root mean square error (RMSE) of 0.232, whereas the lowest prediction accuracy was obtained from the KNN model.

Conclusion: Using the 5-fold random validation method, the comparison of models including Artificial Neural Network (ANN), Decision Tree Analysis (DTA), Multivariate Linear Regression (MLR), and K-Nearest Neighbor (KNN) revealed that the KNN model achieved the highest accuracy for predicting soil organic carbon, with a coefficient of determination (R^2) of 0.9906 and a root mean square error (RMSE) of 0.0361. However, given the spatial structure of the data, the 10-fold cross-validation method provides a more reliable evaluation, making it preferable to the 5-fold random validation method.

Cite this article: Zarini Bahador, Moslem, Sharifi, Rahman. 2025. Digital mapping of soil organic carbon using auxiliary data and machine learning models in the Badr watershed, Kurdistan province. *Journal of Soil Management and Sustainable Production*, 15 (2), 1-22.



© The Author(s).

DOI: 10.22069/ejsms.2025.22625.2162

Publisher: Gorgan University of Agricultural Sciences and Natural Resources



انجمن علوم خاک ایران

نشریه مدیریت خاک و تولید پایدار

شاپا چاپی: ۱۳۶۷-۲۳۲۲

شاپا آنالین: ۱۳۷۵-۲۳۲۲



دانشگاه گیلان و منابع طبیعی گیلان

نقشه‌برداری رقومی کربن آلی خاک با استفاده از داده‌های کمکی و مدل‌های یادگیری ماشین در حوزه آبخیز بدر، استان کردستان

مسلم زرینی بهادر^{۱*}، رحمان شریفی^۲

۱. نویسنده مسئول، فارغ‌التحصیل دکتری خاکشناسی، کارشناس تحقیقات خاک و آب مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران. رایانامه: moslem.zarini@gmail.com
۲. عضو هیأت علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران. رایانامه: r.sharifi@areeo.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله کامل علمی- پژوهشی تاریخ دریافت: ۱۴۰۳/۰۵/۰۵ تاریخ ویرایش: ۱۴۰۳/۱۱/۰۵ تاریخ پذیرش: ۱۴۰۳/۱۱/۰۶ واژه‌های کلیدی: تصاویر ماهواره‌ای، رگرسون خطی چندگانه، متغیرهای کمکی، نقشه‌برداری رقومی خاک	سابقه و هدف: کربن آلی خاک به عنوان عامل کلیدی در پایداری حاصلخیزی و باروری خاک و خدمات‌رسانی زیست‌بوم خاک محسوب می‌شود. کربن آلی خاک به‌عنوان یکی از موضوعات و چالش‌های مهم محیط زیستی در مقیاس جهانی در برنامه محیط زیست سازمان ملل متحد نیز گنجانده شده است. بررسی‌ها در ایران نشان می‌دهد به‌طور میانگین به ازای افزایش هر گرم کربن آلی در کیلوگرم خاک، عملکرد دانه گندم به طور میانگین ۲۸۶ کیلوگرم در هکتار افزایش می‌یابد. هم‌چنین آگاهی از تغییرات کربن آلی خاک یکی از مؤلفه‌های اصلی در ارزیابی کیفیت خاک است. اهداف این مطالعه نقشه‌برداری رقومی کربن آلی خاک و شناسایی اثرات ویژگی‌های محیطی بر پیش‌بینی‌های مقدار کربن آلی خاک، تجزیه و تحلیل اثر زیرمجموعه متغیرهای کمکی بر پیش‌بینی کربن آلی خاک و انتخاب بهترین مدل برای پیش‌بینی ذخایر کربن آلی خاک است. این مطالعه با هدف نقشه‌برداری رقومی کربن آلی خاک، با استفاده از متغیرهای کمکی محیطی و مدل‌های پیش‌بینی‌کننده و معرفی بهترین مدل‌ها در حوزه آبخیز بدر در جنوب شهرستان قروه انجام گرفت. مواد و روش‌ها: برای انجام این پژوهش در مرحله اول، داده‌های کمکی ازجمله تصاویر ماهواره‌ای لندست ۸ و مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر منطقه از سازمان نقشه‌برداری کشور تهیه گردید. نقشه زمین‌شناسی قروه از سایت زمین‌شناسی کشور تهیه و نقشه زمین‌شناسی حوزه آبخیز بدر از آن استخراج و در محیط سامانه اطلاعات جغرافیایی رقومی گردید. نقشه ژئومرفولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئودولوژی زینک در محیط سامانه اطلاعات جغرافیایی ترسیم شد. در مرحله دوم، محل نقاط مشاهداتی تعیین

گردید، شناسایی خاک‌ها در صحرا انجام پذیرفت، نمونه‌برداری از لایه‌های مختلف خاک انجام شد و اندازه‌گیری‌های فیزیکی و شیمیایی و کانی‌شناسی خاک‌ها انجام و خاک‌ها طبقه‌بندی شدند. در مرحله سوم، مدل‌سازی انجام، نقشه‌های رقومی کلاس‌ها و ویژگی‌های خاک تهیه شد و ارزیابی مدل‌ها صورت گرفت. برای انجام این مطالعه، بر اساس تکنیک ابر مکعب لاتین، محل ۱۲۵ خاک رخ در منطقه مطالعاتی انتخاب و حفر شد. نمونه‌های خاک، پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده و از الک ۲ میلی‌متری عبور داده شدند. سپس ماده آلی خاک به روش والکی بلاک اندازه‌گیری شد. جهت برآورد ویژگی‌های خاک، در مطالعه حاضر دو حالت مختلف مورد بررسی قرار گرفت. در حالت اول، مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم و رگرسیون خطی چندگانه برای پیش‌بینی مورد استفاده قرار گرفتند. هم‌چنین جهت ترکیب نتایج مدل‌ها، از مدل نزدیک‌ترین همسایه استفاده شد.

یافته‌ها: نتایج این پیش‌بینی‌ها با استفاده از روش ارزیابی کافلد تصادفی ۵ مورد ارزیابی قرار گرفتند. در مرحله دوم، مدل‌های تحلیل درخت تصمیم، شبکه عصبی مصنوعی، نزدیک‌ترین همسایه و جنگل تصادفی برای پیش‌بینی به کار رفتند. هم‌چنین جهت ترکیب نتایج مدل‌ها در این حالت، از روش رگرسیون خطی چندگانه استفاده شد. با استفاده از روش ارزیابی کافلد مکانی ۱۰، نتایج پیش‌بینی مورد ارزیابی قرار گرفتند. نتایج نشان داد که در میان مدل‌های استفاده‌شده برای پیش‌بینی کربن آلی خاک، مدل رگرسیون خطی چندگانه با ضریب تعیین ۰/۶۳۷ و ریشه دوم متوسط مربعات خطا ۰/۲۳۲ از بیش‌ترین دقت برای پیش‌بینی برخوردار بوده است و کم‌ترین دقت پیش‌بینی به مدل نزدیک‌ترین همسایه K اختصاص یافته است.

نتیجه‌گیری: با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی، از بین مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم، رگرسیون خطی چندگانه و نزدیک‌ترین همسایه K؛ مدل نزدیک‌ترین همسایه K با ضریب تعیین ۰/۹۹۰۶ و ریشه دوم متوسط مربعات خطا ۰/۰۳۶۱ از بیش‌ترین دقت برای پیش‌بینی مقدار کربن آلی خاک برخوردار بوده است. ناگفته نماند که بخاطر مکانی بودن روش اعتبارسنجی کافلد ۱۰ مکانی، استفاده از این روش بر روش اعتبارسنجی کافلد ۵ تصادفی ارجحیت دارد.

استناد: زرینی بهادر، مسلم، شریفی، رحمان (۱۴۰۴). نقشه‌برداری رقومی کربن آلی خاک با استفاده از داده‌های کمکی و مدل‌های یادگیری ماشین در حوزه آبخیز بدر، استان کردستان. نشریه مدیریت خاک و تولید پایدار، ۱۵ (۲)، ۱-۲۲.

DOI: 10.22069/ejsms.2025.22625.2162



© نویسندگان.

ناشر: دانشگاه علوم کشاورزی و منابع طبیعی گرگان

مقدمه

توسعه فن‌آوری‌های نقشه‌برداری رقومی خاک^۱ (DSM) و کاربردهای آن (۱) نقشه‌برداری ذخایر کربن آلی خاک را از مقیاس محلی تا قاره‌ای، با استفاده از فاکتورهای اسکورپن (به عنوان مثال، خاک، آب و هوا، ارگانیک‌ها، سنگ مادر) امکان‌پذیر کرد (۲). نقشه‌برداری از ذخایر کربن آلی خاک در محیط‌های بزرگ و ناهمگن با استفاده از فاکتورهای اسکورپن به دلیل تنوع بالا و در دسترس بودن داده‌های نمونه یک چالش است، اما برای بهبود برآوردهای جهانی از ذخایر کربن آلی خاک و برجسته کردن محرک‌های اصلی برای حمایت از موجودی کربن و تصمیمات سیاسی بسیار مهم است. ذخایر کربن آلی خاک به تعادل بین ورودی‌ها و خروجی‌های کربن و شرایط محیطی (فاکتورهای اسکورپن) بستگی دارد که به‌طور مستقیم یا غیرمستقیم بر فرآیندهای کنترل‌کننده توزیع مکانی ذخایر کربن آلی خاک تأثیر می‌گذارد. در مقیاس جهانی و قاره‌ای، توزیع ذخایر کربن آلی خاک عمدتاً توسط شرایط آب و هوایی مانند دما و بارندگی کنترل می‌شود که با بارش بیش‌تر و دماهای پایین‌تر افزایش می‌یابد (۳و۲). بافت خاک هم‌چنین بر ذخایر کربن آلی خاک تأثیر می‌گذارد و مقادیر رس خاک برای محافظت از کربن آلی خاک به دلیل برهمکنش بین کربن آلی خاک و سطوح واکنش‌دهنده کانی‌های رسی کلیدی است (۴، ۵ و ۶).

پوشش گیاهی هم‌چنین نقش کلیدی در ذخایر کربن آلی خاک با توجه به زیست توده گیاهی در حال پوسیدگی، به‌ویژه در خاک سطحی (صفر تا ۳۰ سانتی‌متر) (۷) و شاخص‌های پوشش گیاهی، مانند شاخص نرمال‌شده تفاوت پوشش گیاهی، برای پیش‌بینی ذخیره کربن آلی خاک استفاده شده (۸) ایفا

می‌کند. در مناطقی با تغییرات توپوگرافی زیاد، کربن آلی خاک به طور منطقی با ویژگی‌های زمین، مانند ارتفاع، شیب و انحنای شیب پیش‌بینی می‌شود (۹، ۱۰ و ۱۱). ترکیب این عوامل محیطی مختلف شرایط محلی منحصر به فردی را ایجاد می‌کند که منجر به افزایش یا کاهش ذخایر کربن آلی خاک می‌شود و برای پیش‌بینی توزیع مکانی ذخایر کربن آلی خاک مفید است.

چندین مدل برای تخمین توزیع مکانی کربن آلی خاک از جمله الگوریتم‌های یادگیری ماشین، مانند جنگل‌های تصادفی (۳، ۱۲، ۱۳، ۱۴ و ۱۵)، کویست (۱۶، ۱۷ و ۱۸)، ماشین‌های بردار پشتیبان (۱۴، ۱۵، ۱۹ و ۲۰) و مدل‌های مبتنی بر کریجینگ (۲۱ و ۲۲) استفاده شده‌اند. کاربرد مدل‌های پیش‌بینی باید از اصل صرفه‌جویی پیروی کند که نشان می‌دهد بهترین مدل می‌تواند پدیده‌های مشابه را با استفاده از متغیرهای کم‌تری بدون از دست دادن عملکرد توضیح دهد (۲۳). استفاده از متغیرهای پیش‌بینی کم‌تر، درک بهتر محرک‌های اصلی و پردازش سریع‌تر رایانه را امکان‌پذیر می‌سازد (۲۴). با هدف ایجاد مدل‌های ساده‌تر، الگوریتم حذف ویژگی بازگشتی^۲ (RFE) به‌طور گسترده برای انتخاب بهینه زیرمجموعه‌های متغیرها به‌کار گرفته می‌شود، در حالی که عملکرد مدل را حفظ می‌کند (۲۴، ۲۵، ۲۶، ۲۷ و ۲۸).

نقشه‌برداری کربن خاک معمولاً در دو بعد (۲ بعدی) در یک فاصله عمقی خاک انجام می‌شود. با این حال، کربن آلی خاک با عمق خاک بسیار متغیر است و مطالعات توزیع عمودی کربن آلی خاک را با استفاده از توابع نمایی منفی و درونیابی مدل‌سازی کرده‌اند (۲۹ و ۳۰). خطوط درجه دوم مساحت مساوی نیز برای هماهنگ کردن داده‌های خاک بر اساس عمق استفاده می‌شوند (۳۱)، که امکان ایجاد مدل‌هایی را

برای ترسیم ویژگی‌های خاک به صورت سه‌بعدی فراهم می‌کند (۳۲ و ۳۳). مطالعه ذخایر کربن آلی در کل عمق خاک بسیار مهم است چرا که اطلاعات مربوط به منابع کربن؛ تغییرات آب و هوایی گذشته را نشان می‌دهد، زیرا کربن آلی خاک ذخیره شده در زیر سطح در مقایسه با لایه‌های بالایی از پایداری و حفاظت بالاتری برخوردار است (۳۴). توزیع ذخایر کربن آلی خاک در لایه‌های بالایی (صفر تا ۵۰ سانتی‌متر) عمدتاً توسط متغیرهای محیطی توضیح داده می‌شود، و یافتن متغیرهای کمکی که تنوع ذخایر کربن آلی خاک زیرسطحی را توضیح می‌دهند، چالش باقی می‌ماند. کشف چنین متغیرهای کمکی می‌تواند درک ما را از فرآیندهای منجر به فرونشست و انعطاف‌پذیری کربن در خاک‌های عمیق بهبود بخشد.

صرف‌نظر از مدل مورد استفاده برای نقشه‌برداری ذخایر کربن آلی خاک، فرآیند پیش‌بینی دارای عدم قطعیت است که می‌تواند توسط شبیه‌سازی مونت‌کارلو (۳۵)، بوت‌استرپ (۳۶) یا بر اساس فاصله پیش‌بینی^۱ (PI) بر اساس تفاوت بین باقی‌مانده بین خروجی‌های مدل‌سازی شده و داده‌های مشاهده شده مربوطه ارزیابی شود. در مواردی دیگر، با استفاده از رگرسیون کریجینگ روش فاصله پیش‌بینی را در نقشه‌برداری رقومی خاک برای ایجاد نقشه‌های پیش‌بینی و عدم قطعیت پیاده‌سازی کرد (۳۷، ۳۸ و ۳۹). پروژه نقشه خاک جهانی^۲ (GSM) عدم قطعیت را به عنوان فاصله پیش‌بینی ۹۰٪ تعریف می‌کند که محدوده مقادیری را که انتظار می‌رود مقدار واقعی در آن ۹ بار از ۱۰ بار رخ دهد، با احتمال ۱ از ۲۰ برای هر یک از دو دنباله توصیف می‌کند (۴۰).

حوزه آبخیز بدر در استان کردستان یکی از مناطق مرتعی است که طی سال‌های اخیر با توجه به تغییر

کاربری‌های اتفاق افتاده تحت کشت گندم دیم قرار گرفته است، و از آنجایی که کربن آلی خاک در اکوسیستم‌های زراعی کارکردهای متفاوتی مانند افزایش حاصلخیزی خاک، کنترل فرسایش، افزایش نفوذپذیری آب در خاک و کاهش اثرات گازهای گلخانه‌ای دارد؛ از این‌رو یک شاخص کلیدی در تعیین سلامت خاک است که تمامی ویژگی‌های خاک را تحت‌تأثیر قرار می‌دهد. بنابراین هدف از این مطالعه نقشه‌برداری رقومی کربن آلی خاک و شناسایی اثرات ویژگی‌های محیطی بر مقدار پیش‌بینی‌های مقدار کربن آلی خاک، تجزیه و تحلیل اثر زیرمجموعه متغیرهای کمکی بر پیش‌بینی کربن آلی خاک و انتخاب بهترین مدل برای پیش‌بینی ذخایر کربن آلی خاک است.

مواد و روش‌ها

وضعیت عمومی 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، متوسط درجه حرارت سالیانه هوا ۱۲/۰۷ درجه سلسیوس و میانگین بارندگی سالیانه 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 با توجه به اطلاعات هواشناسی و داده‌های رطوبتی و حرارتی، رژیم رطوبتی خاک زریک (Xeric) و رژیم حرارتی آن مزیک (Mesic) تعیین شده است.

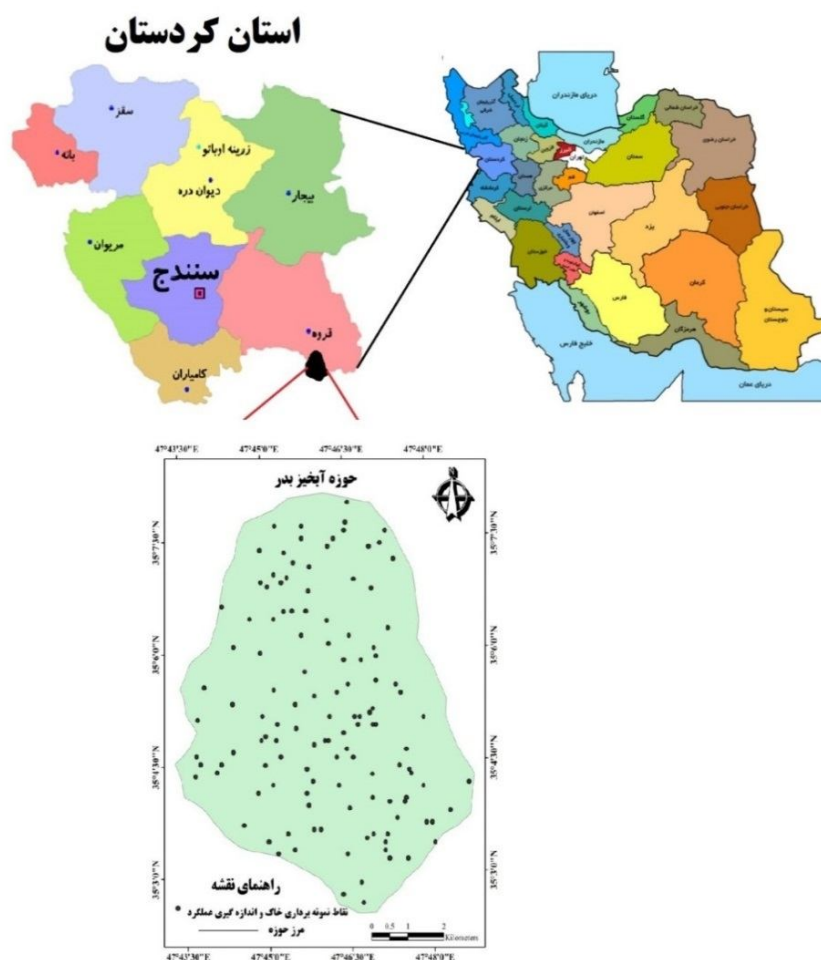
قدیمی‌ترین واحد چینه‌شناسی در منطقه، تریاس است. رسوبات تریاس از دولومیت کاملاً بلورین و

1- Prediction distance
2- Global Soil Map

دگرگونی سنگ‌های رسوبی - آتش فشانی (تریاس - ژوراسیک) گردیده و حاصل آن تشکیل مرمر، آمفیبولیت، شیست و فیلیت است (شکل ۲) (۴۲).

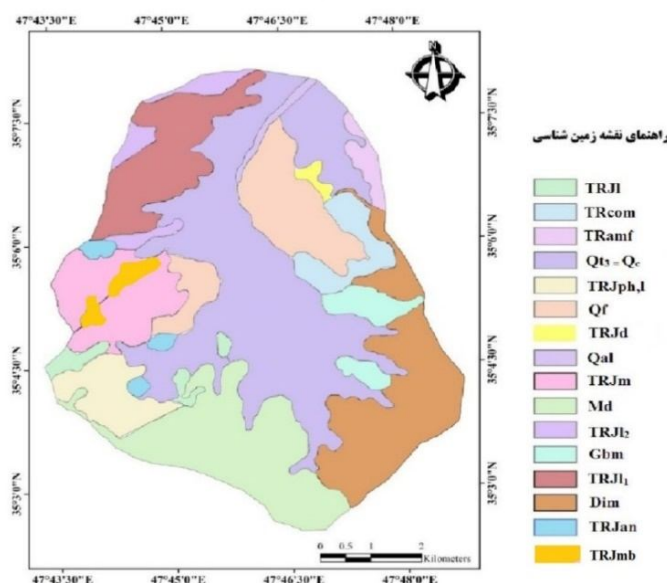
از لحاظ ژئومورفولوژی، حوزه آبخیز مورد مطالعه شامل کوه و اراضی پای کوهی (پیدمونت) است. مهم‌ترین محصولات زراعی منطقه شامل این موارد است؛ گندم دیم و آبی، جو دیم و آبی، نخود، لوبیا، یونجه و شبدر.

خردشده تشکیل شده است. روی این واحد چینه‌ای را شیل، شیست و ولکانیک‌های مشابه با سازند شمشک در البرز مرکزی، به سن ژوراسیک، می‌پوشانند. 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 (۴۱). اصلی‌ترین رخداد کوهزایی در این منطقه مربوط به کوهزایی کیمرین میانی است که سبب



شکل ۱- موقعیت منطقه مطالعاتی و پراکنش مکانی نقاط نمونه برداری.

Figure 1. Location of the study area and spatial distribution of sampling points.

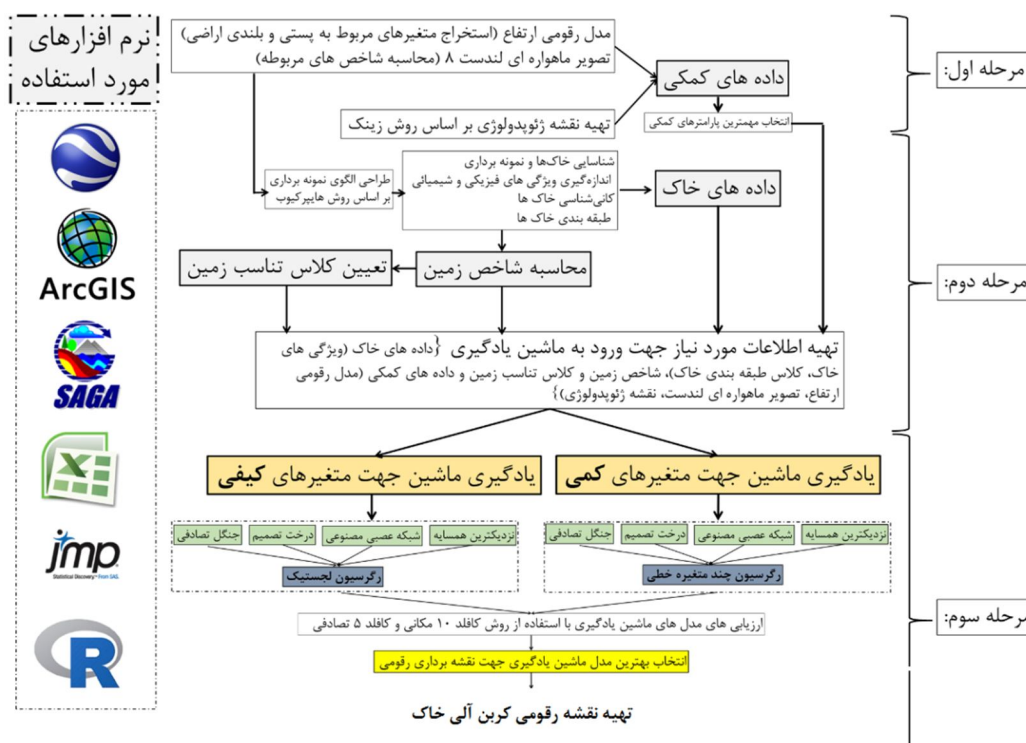


شکل ۲- زمین‌شناسی حوزه آبخیز بدر شهرستان قروه، استان کردستان.
Figure 2. Geology of Badr watershed, Qorveh city, Kurdistan province.

براساس روش USGS طبقه‌بندی شدند. در مرحله سوم، مدل‌سازی انجام، نقشه ویژگی‌های خاک تهیه و ارزیابی مدل‌ها انجام گرفت. جمع‌آوری داده‌های متغیرهای کمکی، پردازش و فرآیند انتخاب مهم‌ترین آن‌ها: در پژوهش حاضر، از سه منبع برای تهیه داده‌های کمکی بهره گرفته شد: تصاویر ماهواره‌ای، مدل رقومی ارتفاع و نقشه ژئومورفولوژی. مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر مبنای کار قرار گرفت. در این مرحله، با استفاده از این مدل و زیربرنامه‌های تعریف‌شده استاندارد موجود در نرم‌افزار Arc-GIS10 و ساگا^۱ مشتقات مختلف مرتبه‌های اول و دوم مدل تهیه شد. شاخص‌های سنجش‌از دور نیز با استفاده از داده‌های باندهای ۱ تا ۶ ماهواره لندست ۸ (سال ۲۰۱۳) محاسبه شدند.

روش اجرا: برای اجرای این پژوهش مراحل زیر، طبق شکل ۳ انجام گرفت: در مرحله اول، داده‌های کمکی ازجمله: تصاویر ماهواره‌ای لندست ۸ از سایت ارث اکسپلورر (earthexplorer) دانلود گردید، و مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر منطقه از سازمان نقشه‌برداری کشور تهیه گردید و بر روی این تصاویر نمونه‌گیری مجدد انجام شد. نقشه زمین‌شناسی قروه از سایت زمین‌شناسی کشور تهیه و نقشه زمین‌شناسی حوزه آبخیز بدر از آن استخراج و در محیط سامانه اطلاعات جغرافیائی رقومی گردید. نقشه ژئومورفولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئودولوژی زینک (۴۳) در محیط سامانه اطلاعات جغرافیائی ترسیم گردید. در مرحله دوم، محل نقاط مشاهداتی بر اساس تکنیک ابرمکعب لاتین (۲۹) تعیین، شناسایی خاک‌ها در صحرا، نمونه‌برداری از لایه‌های مختلف خاک، اندازه‌گیری‌های فیزیکی و شیمیایی و کانی‌شناسی خاک‌ها (به منظور تخمین نگهداشت مقدار کربن آلی خاک) انجام و خاک‌ها

1- System for Automated Geoscientific Analyses, SAGA



شکل ۳- مراحل اجرای پژوهش.

Figure 3. Steps of research implementation.

نرم افزار ساگا، وارد نرم افزار وکا (Weka) شد و با استفاده از الگوریتم انتخاب ویژگی ها، اطلاعات کمی مهم و مؤثر در پیش بینی ویژگی ها و کلاس های طبقه بندی خاک و رده بندی کیفی تناسب اراضی به ترتیب اهمیت، انتخاب و در مدل سازی مورد استفاده قرار گرفت. درجه بندی اهمیت متغیرهای کمی با به کارگیری الگوریتم ReliefAttributeEval نرم افزار WEKA مشخص گردید. این الگوریتم، ارزش یک زیرمجموعه از ویژگی ها، با در نظر گرفتن توانایی پیش بینی فردی هر یک از آنها و همچنین میزان فراوانی بین آنها را ارزیابی می کند. در این الگوریتم، زیرمجموعه فوق الذکر که با ویژگی ها و کلاس های مورد پیش بینی همبستگی بالایی دارند، درحالی که همبستگی درونی آنها کم است، ترجیح داده می شوند.

تفکیک واحدهای ژئومرفیک موجود در منطقه مطالعاتی با توجه به تفاوت رنگ، الگوی آبراهه ای، شیب، شکل شیب و ارتفاع که در تصاویر سه بعدی گوگل ارث قابل مشاهده هستند، بر اساس نقشه زمین شناسی و سامانه سلسله مراتبی زینک (۴۳) انجام شد. این سامانه از بالاترین تا پایین ترین سطح طبقه بندی شامل زمین نما (Landscape)، پستی و بلندی (Relief) و یا قالب (Molding)، سنگ و مواد مادری (Lithology) و شکل زمین (Landform) است. به منظور نشان دادن رابطه بین زمین ریخت (Geoform) و خاک (های) مربوط به هر شکل زمین نیز معرفی گردید. در پایان، نقشه ژئومرفولوژی در سامانه اطلاعات جغرافیائی رقومی گردید.

برای انتخاب پارامترهای مؤثر در مدل سازی، از روش انتخاب ویژگی به ترتیب اهمیت استفاده شد. به این معنی که تمام اطلاعات کمی پس از استخراج از

جمع‌آوری داده‌های خاک و عملکرد محصول: در نقشه‌برداری رقومی خاک، شبکه نمونه‌برداری (مکان‌های نمونه‌برداری) خاک برای ایجاد روابط کمی بین خاک و متغیرهای محیطی تأثیر بنیادین دارد. با توجه به اهمیت این موضوع، در پژوهش حاضر، برای افزایش کارایی مدل‌های نقشه‌برداری رقومی، نسبت به استفاده از روش نمونه‌برداری ابرمکعب لاتین (۲۹) اقدام شد. این تکنیک، یک روش نمونه‌برداری مونت‌کارلوی محدودشده و یک طرح تصادفی طبقه‌بندی‌شده است که به کمک توزیع چندمتغیره، نمونه‌برداری قابل‌قبولی را باعث می‌شود. بر اساس تکنیک ابر مکعب لاتین، محل ۱۲۵ خاک‌رخ در منطقه مطالعاتی انتخاب شد. در این پژوهش از نرم‌افزار R نسخه ۳,۵,۳ برای مشخص کردن نقاط حفر خاک‌رخ‌ها استفاده و داده‌های استفاده شده برای وارد کردن به نرم‌افزار مشتعل بر مدل رقومی ارتفاع، داده‌های سنجش از دور و نقشه ژئومورفولوژی بودند. موقعیت تمامی خاک‌رخ‌ها توسط سیستم موقعیت‌یاب جهانی^۱ (GPS) ثبت شد.

مطالعات خاکشناسی برای تهیه نقشه‌های رقومی کلاس‌های خاک و ویژگی‌های آن انجام شد و فاصله خاک‌رخ‌ها از یکدیگر بر اساس روش ابرمکعب لاتین تعیین شد. خاک‌رخ‌ها بر اساس راهنمای تشریح و نمونه‌برداری خاک‌ها در صحرا (۴۴) تشریح و از تمامی افق‌های آن‌ها نمونه‌برداری انجام گرفت. سطوح مختلف ژئومورفولوژی نیز برای هر پروفیل یادداشت‌برداری شد.

جمع‌آوری داده‌های خاک و عملکرد محصول: در نقشه‌برداری رقومی خاک، شبکه نمونه‌برداری (مکان‌های نمونه‌برداری) خاک برای ایجاد روابط کمی بین خاک و متغیرهای محیطی تأثیر بنیادین دارد. با توجه به اهمیت این موضوع، در پژوهش حاضر،

برای افزایش کارایی مدل‌های نقشه‌برداری رقومی، نسبت به استفاده از روش نمونه‌برداری ابرمکعب لاتین (۲۹) اقدام شد. این تکنیک، یک روش نمونه‌برداری مونت‌کارلوی محدودشده و یک طرح تصادفی طبقه‌بندی‌شده است که به کمک توزیع چندمتغیره، نمونه‌برداری قابل‌قبولی را باعث می‌شود. بر اساس تکنیک ابرمکعب لاتین، محل ۱۲۵ خاک‌رخ در منطقه مطالعاتی انتخاب شد. در این پژوهش از نرم‌افزار R نسخه ۳,۵,۳ برای مشخص کردن نقاط حفر خاک‌رخ‌ها استفاده و داده‌های استفاده شده برای وارد کردن به نرم‌افزار مشتعل بر مدل رقومی ارتفاع، داده‌های سنجش از دور و نقشه ژئومورفولوژی بودند. موقعیت تمامی خاک‌رخ‌ها توسط سیستم موقعیت‌یاب جهانی (GPS) ثبت شد.

مطالعات خاکشناسی برای تهیه نقشه‌های رقومی کلاس‌های خاک و ویژگی‌های آن انجام شد و فاصله خاک‌رخ‌ها از یکدیگر بر اساس روش ابرمکعب لاتین تعیین شد. خاک‌رخ‌ها بر اساس راهنمای تشریح و نمونه‌برداری خاک‌ها در صحرا (۴۴) تشریح و از تمامی افق‌های آن‌ها نمونه‌برداری انجام گرفت. سطوح مختلف ژئومورفولوژی نیز برای هر پروفیل یادداشت‌برداری شد.

اندازه‌گیری مشخصات فیزیکی و شیمیایی خاک: نمونه‌های خاک، پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده و از الک ۲ میلی‌متری عبور داده شدند و میزان کربن آلی خاک به روش والکی بلاک (۱۹۳۴) اندازه‌گیری گردید (۴۵).

پردازش اطلاعات: متوسط وزنی عمق ۲۵ سانتی‌متری درصد کربن آلی برای نقشه‌برداری رقومی ویژگی‌های خاک به‌کار رفت. به‌منظور محاسبه متوسط وزنی، مقدار هر ویژگی در افق‌های خاک، ضخامت هر افق و ضریب وزنی عمق (۴۶) مورد استفاده قرار گرفت.

یادگیری ماشین

الف- مدل‌های یادگیری ماشین، به کار رفته در این تحقیق: در این پژوهش، جهت برآورد داده‌های کمی خاک مانند کربن آلی خاک از مدل‌های شبکه عصبی مصنوعی، نزدیک‌ترین همسایه، تحلیل درخت تصمیم و جنگل تصادفی استفاده گردید. علاوه بر این، از مدل رگرسیون خطی چندگانه جهت ترکیب مدل‌های مذکور نیز بهره گرفته شد.

ب- ارزیابی مدل‌های یادگیری ماشین: در این پژوهش، جهت ارزیابی مدل‌های یادگیری ماشین از دو روش بهره گرفته شد: ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی. در این نوع اعتبارسنجی‌ها، داده‌ها به K زیرمجموعه دسته‌بندی شدند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و $K-1$ تای دیگر برای آموزش به کار رفتند. این روال، K بار تکرار شد و همه داده‌ها دقیقاً یکبار برای آموزش و یکبار برای اعتبارسنجی به کار رفتند. در نهایت، میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده شد. لازم به ذکر است که تفاوت ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی، در انتخاب مجموعه داده‌های خاک جهت آموزش و ارزیابی است. در روش ارزیابی متقابل تصادفی، داده‌های به طور تصادفی به دو دسته تقسیم می‌شوند: داده‌هایی که برای آموزش به کار 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 به کار می‌روند. در چنین

شرایطی، می‌توان مطمئن شد که مدل آموزش دیده، در کل منطقه مورد ارزیابی قرار گرفته است (۴۷). 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 کربن آلی توسط مدل‌های یادگیری ماشین شاخص‌هایی مانند جذر میانگین مربع خطا^۱ و جذر میانگین مربع خطای نسبی^۲، ضریب تعیین (R^2) و ضریب همبستگی (r) محاسبه شدند.

$$RMSE = \sqrt{\frac{\sum(x' - x)^2}{n}} \quad (1)$$

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

$$RMSE\% = \frac{RMSE}{\bar{x}} * 100 \quad (2)$$

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

مدل‌سازی مکانی: در این پژوهش از رویکردهای شبکه عصبی مصنوعی، تحلیل درخت تصمیم، جنگل تصادفی، نزدیک‌ترین همسایه K و رگرسیون خطی چندگانه برای مدل‌سازی مکانی کربن آلی خاک استفاده شد.

رگرسیون خطی چندگانه: با استفاده از این روش می‌توان هم‌زمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. برای به دست آوردن نتایج مطلوب‌تر از طریق رگرسیون خطی چندگانه، نمونه‌ها باید دارای فراوانی بالا و هم‌چنین از دقت قابل توجهی برخوردار باشند. زیرا این روش در مقابل اطلاعات نادرست، حساسیت بالایی دارد و ورود چنین داده‌هایی ممکن

1- Root Mean Square Error: RMSE

2- Relative RMSE: RMSE%

است منجر به خطاهای بزرگی در نتایج به دست آمده شود. علاوه بر این، برای استفاده از این روش متغیرها باید توزیع نرمال داشته باشند و تغییر آن‌ها از یک رابطه خطی پیروی کند. رگرسیون خطی چندگانه در حقیقت ارتباط بین یکسری از متغیرهای پیش‌بینی‌کننده را با متغیر پاسخ موردنظر بیان می‌کند. 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 زیر بین آن‌ها برقرار است:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

که در این رابطه، از مقادیر $a_1, a_2, \dots, a_n$ با عنوان ضرایب رگرسیون یاد می‌شود. این ضرایب، ضرایبی نامشخص‌اند که در حقیقت مسئول برآورد پارامتر پاسخ هستند (۴۸ و ۴۹).

شبکه‌های عصبی مصنوعی: رایانه‌ها، دارای سرعت قابل‌توجهی نسبت به انسان می‌باشند؛ ولی قدرت تفکر ندارند و در هنگام کمبود اطلاعات و یا وجود داده‌های پرت، عملکرد آن‌ها به شدت کاهش می‌یابد و رضایت‌بخش نیست. از آن‌جاکه داده‌هایی که در دسترس پژوهشگران هستند، معمولاً محدود و گاه پرت هستند؛ ایده شبکه‌های عصبی مصنوعی که یک سامانه محاسباتی هوشمند و دارای قدرت یادگیری است، مطرح گردید. مبنای شبکه‌های عصبی مصنوعی روابط پیچیده ریاضی منطبق بر سامانه‌ی عصبی انسان است. در روش شبکه‌های عصبی مصنوعی، سه لایه ورودی، پنهان و خروجی وجود دارد و این لایه‌ها با یکدیگر در ارتباط هستند. در این روش، از طریق وزن‌دهی داده‌های ورودی توسط فرد و در حقیقت آموزش سیستم، وزن نهایی به دست می‌آید (۵۰ و ۵۱) برآورد واحدهای خاک را با استفاده از شبکه‌های عصبی مصنوعی و با به کارگیری داده‌های رقومی ارتفاع انجام داده و بیان نمودند که ترکیب تکنیک

شبکه‌های عصبی مصنوعی و اطلاعات به دست آمده از مدل رقومی ارتفاع، علاوه بر صرفه‌جویی در زمان و هزینه‌ها، نتایج قابل‌قبولی ارائه می‌دهد.

مدل میانگین‌گیری نزدیک‌ترین همسایه K : میانگین‌گیری نزدیک‌ترین همسایه K ، ارزش پاسخ برای یک مشاهده داده‌شده را با استفاده از پاسخ‌های مشاهداتی در همسایه‌های محلی پیش‌بینی می‌کند. می‌توان از این مدل با یک پاسخ قطعی برای طبقه‌بندی و یک پاسخ پیوسته برای پیش‌بینی استفاده کرد (۵۸). این روش، یک روش ناپارامتری است که بر اساس فاصله تا مشاهدات همسایه، پایه‌گذاری شده است. به این دلیل، میانگین‌گیری نزدیک‌ترین همسایه K قادر به طبقه‌بندی مشاهدات با استفاده از مرزهای ارزش پیش‌بینی نامنظم است. با این حال، الگوریتم، حساس به پیش‌بینی‌کننده‌های بی‌ربط است؛ بنابراین انتخاب پیش‌بینی‌کننده‌ها، قبل از اجرای مدل می‌تواند مفید باشد.

آنگوایی (۵۲) معتقد است، روش نزدیک‌ترین همسایگی K دارای یکی از ساده‌ترین و کارآمدترین الگوریتم‌های آموزش جهت پیش‌بینی داده‌ها است. بنا به نظر نمس و همکاران (۵۳) و ذوالفقاری و همکاران (۵۴)، در این تکنیک، تعداد بهینه‌ای از نمونه‌های موجود در یک مجموعه که واجد شبیه‌ترین ویژگی‌ها به نمونه هدف باشند در نظر گرفته می‌شود و سپس با توجه به فواصل اقلیدسی نمونه‌های هدف و هر یک از نمونه‌های بانک مرجع، وزن‌دهی عناصر آموزش‌دهنده انجام می‌پذیرد. در یک بانک داده مرجع خاک، K عدد از نزدیک‌ترین (مشابه‌ترین) خاک‌ها به خاک مورد آزمون انتخاب می‌شوند و این خاک‌ها با توجه به میزان مشابهت به خاک مورد آزمون وزن‌دهی می‌شوند. به این ترتیب که در ابتدا فواصل اقلیدسی بین خاک مورد آزمون و تمامی خاک‌های مرجع

الف- وقتی داده‌ها کمی و کیفی هستند، تفسیر کردن آن‌ها راحت‌تر است. ب- بسیار راحت‌تر با داده‌های گمشده یا داده‌های پرت برخورد می‌کند. ج- توانایی درک و تحلیل روابط غیرخطی خصوصیات خاک را دارد د- هیچ فرضیه‌ای برای توزیع داده‌ها نیاز ندارد. ه- بسیار راحت به روز می‌شود.

مدل جنگل تصادفی: جنگل‌های تصادفی در حقیقت مجموعه‌هایی از درخت‌های پیش‌بینی‌کننده با احتمال و پراکندگی یکسان هستند. اساس این روش، انتخاب خطای داده‌های بزرگ‌تر به عنوان خطای اصلی و ایجاد همبستگی بین خطاهای دیگر داده‌ها به ترتیب است. انتخاب خطای عمومی بر اساس بزرگ‌ترین داده‌ها، پیش‌بینی دقیق و قابل استنادی ارائه می‌دهد. زمانی این پیش‌بینی دقیق‌تر می‌شود که داده‌های ورودی به مدل بدون اعمال تغییرات باشد (۵۸).

هنگل و همکاران (۳) در بخشی از آفریقا اقدام به تهیه نقشه رقومی خصوصیات خاک با دو مدل رگرسیون چندگانه خطی و جنگل‌های تصادفی کردند که نتایج اعتبارسنجی مدل‌ها نشان داد مدل جنگل‌های تصادفی توانسته به‌طور معنی‌داری دقت بیشتری نسبت به مدل رگرسیون خطی چندگانه از خود نشان دهد. هیونگ و همکاران (۵۹) اقدام به تهیه نقشه مواد مادری خاک با استفاده از متغیرهای کمکی و مدل جنگل‌های تصادفی کردند که نتایج آن‌ها نشان داد زمانی که از متغیرهای کمکی مستخرج از تصاویر ماهواره‌ای و متغیرهای کمکی مدل رقومی ارتفاع به‌عنوان داده‌های ورودی به مدل استفاده شد، ضریب کاپای اعتبار سنجی مدل ۸۹/۶ درصد استخراج گردید.

نتایج و بحث

تشریح خاک‌های منطقه مطالعاتی: رده‌بندی خاک‌رخ‌های حفرشده بر مبنای سامانه رده‌بندی خاک آمریکایی (۲۰۱۴) تا سطح فامیل انجام شد. بر اساس نتایج

محاسبه شده، سپس تعداد K عدد از نزدیک‌ترین خاک‌ها بر اساس فاصله به‌دست آمده، وزن‌دهی می‌شوند. در انتها با توجه وزن هر یک از خاک‌ها در مجموعه‌ای با K عدد خاک (نزدیک‌ترین همسایه)، بر آوردی از داده‌های هدف که مورد آزمون واقع شده‌اند، صورت می‌پذیرد.

ایوبی و همکاران (۵۵) در مطالعه‌ای با عنوان مقایسه روش‌های K نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی برای پهنه‌بندی رقومی شوری خاک در منطقه چاه افضل اردکان، به این نتیجه دست یافتند که مدل K نزدیک‌ترین همسایگی دارای دقت بیشتری نسبت به شبکه عصبی مصنوعی به‌منظور پیش‌بینی شوری خاک است، به‌طوری‌که این مدل به خوبی توانسته ارتباط قوی بین داده‌های شوری خاک و اطلاعات محیطی برقرار کند.

مدل تحلیل درخت تصمیم^۱: مدل درختی طبقه‌بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر است متغیرهای کمی یا متغیرهای طبقه‌بندی‌شده را بر اساس مجموعه‌ای از متغیرهای پیش‌بینی‌کننده کمی و کیفی برآورد کند. در این روش، مجموعه‌ای از شرط‌های منطقی به‌صورت یک الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به‌کار می‌رود. ایجاد درخت تصمیم دو مرحله دارد. مرحله اول، مرحله ایجاد و رشد درخت است. این مرحله شامل پیوند^۲ و انشعاب^۳ است. مرحله دوم، مرحله توقف و هرس^۴ است. هدف این مرحله به حداقل رساندن خطای پیش‌بینی است (۵۶).

بعضی از مزایای استفاده از درخت تصمیم که مور و همکاران (۵۷) در مقایسه با سایر روش‌های طبقه‌بندی بیان کردند عبارت‌اند از:

- 1- Decision Tree Analysis (DTA)
- 2- Merging
- 3- Splitting
- 4- Pruning

حاصل از این رده‌بندی، می‌توان بیان نمود که در منطقه مورد مطالعه؛ سه رده Mollisols, Inceptisols و Entisols؛ سه زیررده Xerolls, Xerepts و Orthents؛ چهار گروه بزرگ Calcixerepts, Haploxerepts و Haploxerolls وجود دارند.

گروه بزرگ Calcixerepts از دو زیرگروه Vertic Calcixerepts و Typic Calcixerepts تشکیل شده است، گروه بزرگ Haploxerepts از پنج زیرگروه Fluventic Haploxerepts, Calcic Haploxerepts, Typic Haploxerepts, Lithic Haploxerepts و Vertic Haploxerepts تشکیل شده است، گروه بزرگ Haploxerolls از دو زیرگروه Vertic Haploxerolls و Typic Haploxerolls تشکیل شده است و گروه بزرگ Xerorthents از دو زیرگروه Typic Xerorthents و Lithic xerorthents تشکیل شده است.

خلاصه آماری داده‌ها: جدول ۱، آمار توصیفی خصوصیات فیزیکی و شیمیایی خاک‌ها و عملکرد

گندم دیم 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 (۶۰). 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. کلاس اندازه ذرات اغلب خاک‌ها، لومی ریز است. عملکرد اندازه‌گیری شده نیز با ضریب تغییرات ۸۷/۷ درصد، تغییرات زیادی را در خاک‌های منطقه نشان داد. درجه تناسب ویژگی‌های خاک تأثیرگذار بر عملکرد، در محاسبه شاخص زمین به کار می‌رود. این شاخص تعیین‌کننده کلاس تناسب زمین برای گندم دیم است. پارامترهای محیطی که از مدل رقومی ارتفاع و تصاویر ماهواره لندست ۸ به دست آمدند، ویژگی‌های خاک فوق‌الذکر را تحت تأثیر قرار می‌دهند.

جدول ۱- آمار توصیفی برخی ویژگی‌های فیزیکی و شیمیایی خاک و عملکرد اندازه‌گیری شده.

Table 1. Descriptive statistics of some physical and chemical characteristics of soil and measured performance.

ضریب تغییرات Coefficient of Variation	انحراف معیار Standard Deviation	میانگین Average	حداکثر Maximum	حداقل Minimum	واحد Unit	عملکرد اندازه‌گیری شده physical and chemical characteristics of soil and measured performance
3.6	0.3	7.5	8	6.8	-	pH
25.5	0.1	0.2	0.4	0.1	m ⁻¹ .dS	EC
62.4	0.4	0.6	2.3	0.1	%	O.C
174.4	17	9.8	65	0	%	CaCO ₃
16.9	2.4	14.3	26.5	8.9	cmol(+).kg ⁻¹	CEC
30.4	6.8	22.5	43	9	%	Clay
23.5	6.8	29.1	48.6	15.9	%	Silt
20.3	9.8	48.3	70	27.1	%	Sand
61.3	10.4	17.1	48	1.5	درصد حجمی Volume percentage	Gravel
87.6	832.2	949.4	2160	0	kg ha ⁻¹	عملکرد مشاهداتی Observational performance

ضريب تعيين $0/637$ و ريشه دوم متوسط مربعات خطا $0/232$ از بيش‌ترين دقت برای پيش‌بینی برخوردار بوده است. در روش استفاده از رگرسیون خطی چندگانه، پيش‌بینی توسط ترکیبی از چهار مدل نزديک‌ترین همسایه، تحلیل درخت تصميم، شبکه عصبی مصنوعی و جنگل تصادفی انجام گرفت (شکل ۳). در جدول ۲، دقت پيش‌بینی کربن آلی خاک، از بالا به طرف پائين افزایش می‌یابد.

نتایج ارزیابی صحت پيش‌بینی کربن آلی خاک (SOC)، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، در جدول ۲ ارائه شده است. این پيش‌بینی توسط مدل‌های نزديک‌ترین همسایه K (KNN)، تحلیل درخت تصميم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (R.F.) و رگرسیون خطی چندگانه (MLR) انجام گرفته است. در میان مدل‌های استفاده شده برای پيش‌بینی کربن آلی خاک، مدل رگرسیون خطی چندگانه (MLR) با

جدول ۲- نتایج ارزیابی صحت پيش‌بینی کربن آلی خاک، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی.

Table 2. The results of evaluating the accuracy of soil organic carbon prediction, using the 10- fold cross validation method.

مدل Model	RMSE	MAE	R^2
KNN	0.356	0.281	0.220
DTA	0.264	0.207	0.544
ANN	0.256	0.200	0.551
RF	0.243	0.194	0.569
MLR	0.232	0.191	0.637

با ضريب تعيين $0/9906$ و ريشه دوم متوسط مربعات خطا $0/0361$ از بيش‌ترين دقت برای پيش‌بینی برخوردار بوده است. ناگفته نماند که به‌خاطر مکانی بودن روش اعتبارسنجی کافلد ۱۰ مکانی، استفاده از این روش بر روش اعتبارسنجی کافلد ۵ تصادفی ارجحیت دارد.

کم‌ترین دقت پيش‌بینی به مدل نزديک‌ترین همسایه K، اختصاص یافته است. این در حالی است که با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی (جدول ۳)، از بین مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصميم، رگرسیون خطی چندگانه و نزديک‌ترین همسایه K، مدل نزديک‌ترین همسایه K

جدول ۳- نتایج ارزیابی صحت پيش‌بینی کربن آلی خاک، با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی.

Table 3. The results of evaluating the accuracy of predicting soil organic carbon, using the 5fold random validation method.

مدل Model	RMSE	N RMSE=RMSE/Mean real	r^C	R^2	ME
ANN	0.1065	0.1781	0.9415	0.8983	-0.00375
DTA	0.163	0.2726	0.8937	0.808	0
MLR	0.2774	0.4640	0.7767	0.6349	0
KNN	0.0361	0.0603	0.9953	0.9906	-0.00103

شاخص رس، شاخص همواری قله با درجه تفکیک بالا و شاخص همواری کف دره با درجه تفکیک بالا. درجه‌بندی اهمیت با به‌کارگیری الگوریتم ReliefAttributeEval نرم‌افزار WEKA مشخص گردید.

متغیرهای کمکی مهم در پیش‌بینی کربن آلی خاک به ترتیب اهمیت در جدول ۴ ملاحظه می‌گردد. این متغیرها به ترتیب اهمیت عبارت‌اند از: ژئومورفولوژی، انحنای مقطع عرضی، موقعیت نسبی شیب، جهت شیب، شاخص همگرایی، شیب، سطح‌مبنای شبکه آبراهه‌ای، شاخص خیسی، شاخص کربنات،

جدول ۴- متغیرهای کمکی مهم در پیش‌بینی کربن آلی خاک، به‌ترتیب اهمیت. از چپ به راست، از اهمیت متغیر کمکی کاسته می‌شود.

Table 4. Important auxiliary variables in predicting soil organic carbon, in order of importance. from left to right, the importance of the covariate decreases.

ویژگی‌ها Characteristics	متغیرهای کمکی مهم به ترتیب اهمیت Important covariates in order of importance
کربن آلی O.C.	Geomorphology, Cross sectional Curvature, Relative slope position, Aspect, Convergence index, Slope, Channel network base level, Topographic wetness index, Carbonate index, Clay index, MRRTF, MRVBF

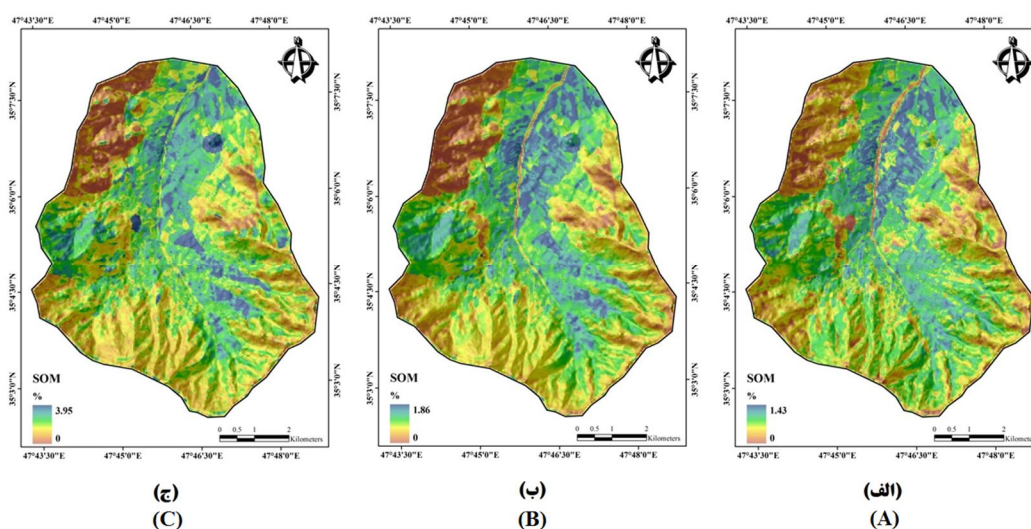
در شکل ۴ ملاحظه می‌گردد، آن قسمت از شیب‌های شمالی و شرقی که دارای سنگ‌آهک و یا خاک آهکی نیستند، از کربن آلی بیشتری برخوردارند. شاخص خیسی از دیگر متغیرهای کمکی است که در جدول ۴ از عوامل مؤثر بر مقدار کربن آلی خاک محسوب شده است. نقاط با شاخص خیسی بالا و نقاطی که به‌صورت موضعی رطوبت بالاتری دارند، معمولاً از کربن آلی بالاتری برخوردارند. ماندابی در خاک، باعث کاهش اکسیژن و در نتیجه تأخیر در تجزیه کربن آلی و تجمع آن در خاک می‌گردد. در حوزه آبخیز بدر، اراضی پست و خیس کف دره‌ها، دارای کربن آلی بیشتری می‌باشند. خاک‌های رسی نیز نسبت به خاک‌های درشت‌بافت، آب بیشتری در خود نگه می‌دارند، در نتیجه اکسیژن آن‌ها کم‌تر شده و کربن آلی در آن‌ها کم‌تر تجزیه و بیش‌تر تجمع می‌یابد. به‌این‌علت، شاخص رس نیز از متغیرهای کمکی تأثیرگذار بر افزایش کربن آلی به شمار آمده است که این مطلب با مطالعات میرزایی و همکاران (۲۰۱۶) با عنوان عنوان تغییرپذیری مکانی کربن آلی خاک با

تأثیرگذارترین متغیر کمکی در مقدار کربن آلی خاک، ژئومورفولوژی است. نقشه‌های پیش‌بینی مقدار کربن آلی در شکل ۴ نشان داده شده است. مقادیر کربن آلی در زمین‌نمای کوه، به‌ویژه کوه‌های آهکی کم‌تر و در زمین‌نمای پای کوه، به‌خصوص در پای کوه‌های غیر آهکی، افزایش می‌یابد که با مطالعات یو و همکاران در پهنه‌بندی کربن آلی خاک هم‌راستا است (۶۱). شاخص آهک نیز جزو متغیرهای کمکی تأثیرگذار بر میزان کربن آلی خاک در جدول ۴ ملاحظه می‌گردد. از قله کوه به‌طرف دره (موقعیت نسبی شیب در جدول ۴) نیز مقدار کربن آلی سیر صعودی دارد (۶۲). جهت شیب نیز، از عوامل تعیین‌کننده میزان کربن آلی گزارش شده است. با توجه به این‌که جهات شمالی و شرقی شیب کوه تابش کم‌تری از خورشید دریافت می‌کنند، تبخیر در آن‌ها کم‌تر و در نتیجه مرطوب‌ترند. رطوبت بیش‌تر باعث تراکم بیش‌تر پوشش گیاهی و در نتیجه افزایش کربن آلی در این جهات می‌گردد که این مورد با مطالعات بانگرو و همکاران (۶۲) مطابقت دارد. همان‌گونه که

پراکندگی ۵ درصد بیشتر از میانگین را نشان می‌دهند. در این شکل و در سایر نقشه‌های پیش‌بینی ویژگی‌های خاک، فقط از مدل رگرسیون خطی چندگانه که طبق روش اعتبارسنجی کافلد ۱۰ مکانی، به‌عنوان بهترین مدل جهت پیش‌بینی این ویژگی‌ها شناخته شد، استفاده گردید.

کمک گرفتن از داده‌های سنجش از دور در یک راستا می‌باشد (۶۳).

در شکل ۴، سه نقشه برای پیش‌بینی مقدار کربن آلی خاک ارائه شده است. نقشه ب، پیش‌بینی مقدار میانگین، نقشه الف، پیش‌بینی با دامنه پراکندگی ۵ درصد کمتر از میانگین و نقشه ج، پیش‌بینی با دامنه



شکل ۴- نقشه‌های پیش‌بینی کربن آلی خاک با استفاده از رگرسیون خطی چندگانه. نقشه ب، پیش‌بینی مقدار میانگین، نقشه الف، پیش‌بینی با دامنه پراکندگی ۵ درصد کمتر از میانگین و نقشه ج، پیش‌بینی با دامنه پراکندگی ۵ درصد بیشتر از میانگین را نشان می‌دهند.

Figure 4. Prediction maps of soil organic carbon using linear multivariate regression. Map B shows the prediction of the average value, Map A shows the prediction with a dispersion range 5% lower than the average, and Map C shows the prediction with a dispersion range 5% higher than the average.

مکانی، بهترین مدل برای پیش‌بینی کربن آلی خاک، رگرسیون خطی چندگانه است. برای کربن آلی خاک، مدل جنگل تصادفی، از نظر قدرت پیش‌بینی، رتبه دوم را به خود اختصاص می‌دهد و مدل نزدیک‌ترین همسایه K، پائین‌ترین عملکرد را در پیش‌بینی کربن آلی خاک دارا بوده است. هم‌چنین نتایج این مطالعه نشان داد؛ ژئومرفولوژی، مشتقات مستخرج از مدل رقومی ارتفاع و داده‌های دریافتی از تصاویر ماهواره‌ای به ترتیب مهم‌ترین ویژگی‌های مؤثر در پیش‌بینی وضعیت کربن آلی خاک بودند.

نتیجه‌گیری کلی

در این پژوهش از رویکردهای شبکه عصبی مصنوعی، تحلیل درخت تصمیم، جنگل تصادفی، نزدیک‌ترین همسایه K و رگرسیون خطی چندگانه برای مدل‌سازی مکانی کربن آلی خاک استفاده شد. به‌طورکلی، روش‌های ارزیابی کافلد ۱۰ مکانی و کافلد ۵ تصادفی، نشان دادند که روش‌های ترکیبی دارای قابلیت بیشتری برای پیش‌بینی کربن آلی خاک می‌باشند. این بدین معنی است که ترکیب نتایج پیش‌بینی مدل‌ها می‌تواند نقشه‌هایی با دقت بالاتر را تولید کند. بر اساس روش اعتبارسنجی کافلد ۱۰

منابع

1. McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52. **doi.org/10.1016/S0016-7061(03)00223-4.**
2. Minasny, B., McBratney, A. B., Malone, B. P., & Wheeler, I. (2013). Digital mapping of soil carbon. *Advances in agronomy*, 118, 1-47. **doi.org/10.1016/B978-0-12-405942-9.00001-3.**
3. Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., ... & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one*, 10(6), e0125814. **doi.org/10.1371/journal.pone.0125814.**
4. Powers, J. S., Corre, M. D., Twine, T. E., & Veldkamp, E. (2011). Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation. *Proceedings of the National Academy of Sciences*, 108(15), 6318-6322. **doi.org/10.1073/pnas.1016774108.**
5. Grüneberg, E., Schöning, I., Hessenmöller, D., Schulze, E. D., & Weisser, W. W. (2013). Organic layer and clay content control soil organic carbon stocks in density fractions of differently managed German beech forests. *Forest Ecology and Management*, 303, 1-10. **DOI:10.1016/J.FORECO.2013.03.014.**
6. Mayer, L. M. (1994). Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chemical Geology*, 114(3-4), 347-363. **doi.org/10.1016/0009-2541(94)90063-9.**
7. Bui, E., Henderson, B., & Viergever, K. (2009). Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Global biogeochemical cycles*, 23(4). **doi.org/10.1029/2009GB003506.**
8. Yang, R. M., Zhang, G. L., Yang, F., Zhi, J. J., Yang, F., Liu, F., ... & Li, D. C. (2016). Precise estimation of soil organic carbon stocks in the northeast Tibetan Plateau. *Scientific reports*, 6(1), 21842. **DOI: 10.1038/srep21842.**
9. Fissore, C., Dalzell, B. J., Berhe, A. A., Voegtli, M., Evans, M., & Wu, A. (2017). Influence of topography on soil organic carbon dynamics in a Southern California grassland. *Catena*, 149, 140-149. **DOI:10.1016/j.catena.2016.09.016.**
10. Nyssen, J., Temesgen, H., Lemenih, M., Zenebe, A., Haregeweyn, N., & Haile, M. (2008). Spatial and temporal variation of soil organic carbon stocks in a lake retreat area of the Ethiopian Rift Valley. *Geoderma*, 146(1-2), 261-268. **DOI:10.1016/j.geoderma.2008.06.007.**
11. Oueslati, I., Allamano, P., Bonifacio, E., & Claps, P. (2013). Vegetation and topographic control on spatial variability of soil organic carbon. *Pedosphere*, 23(1), 48-58. **DOI:10.1016/S1002-0160(12)60079-4.**
12. Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., ... & 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.**
13. Hounkpatin, O. K., de Hipt, F. O., Bossa, A. Y., Welp, G., & Amelung, W. (2018). Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). *Catena*, 166, 298-309. **DOI:10.1016/j.catena.2018.04.013.**
14. Ließ, M., Schmidt, J., & 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.org/10.1371/journal.pone.0153673.**

15. Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., & Li Liu, D. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*, 630, 367-378. **doi.org/10.1016/j.scitotenv.2018.02.204.**
16. Adhikari, K., Hartemink, A. E., Minasny, B., Bou Kheir, R., Greve, M. B., & Greve, M. H. (2014). Digital mapping of soil organic carbon contents and stocks in Denmark. *PloS one*, 9(8), e105519. **doi.org/10.1371/journal.pone.0105519.**
17. Gray, J. M., Bishop, T. F., & Yang, X. (2015). Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia. *Soil Research*, 53(1), 24-42. **DOI:10.1071/SR13306.**
18. Rossel, R. V., Brus, D. J., Lobsey, C., Shi, Z., & McLachlan, G. (2016). Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. *Geoderma*, 265, 152-163. **DOI:10.1016/j.geoderma.2015.11.016.**
19. Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., & Van Orshoven, J. (2017). Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological indicators*, 77, 139-150. **DOI:10.1016/j.ecolind.2017.02.010.**
20. Minasny, B., Setiawan, B. I., Saptomo, S. K., & McBratney, A. B. (2018). Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma*, 313, 25-40. **DOI:10.1016/j.geoderma.2017.10.018.**
21. Bonfatti, B. R., Hartemink, A. E., Giasson, E., Tornquist, C. G., & Adhikari, K. (2016). Digital mapping of soil carbon in a viticultural region of Southern Brazil. *Geoderma*, 261, 204-221. **DOI:10.1016/j.geoderma.2015.07.016.**
22. Gamble, J. D., Feyereisen, G. W., Papiernik, S. K., Wentz, C., & Baker, J. (2017). Regression-Kriged Soil Organic Carbon Stock Changes in Manured Corn Silage–Alfalfa Production Systems. *Soil Science Society of America Journal*, 81(6), 1557-1566. **DOI:10.2136/sssaj.2017.04.0138.**
23. Batty, M., & Torrens, P. M. (2001). Modelling complexity: the limits to prediction. *Cybergeo: European Journal of Geography*. **https://doi.org/10.4000/cybergeo.1035.**
24. Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards Jr, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239, 68-83. **DOI:10.1016/j.geoderma.2014.09.019.**
25. Ballabio, C. (2009). Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*, 151(3-4), 338-350. **doi.org/10.1016/j.geoderma.2009.04.022.**
26. Hounkpatin, O. K., de Hipt, F. O., Bossa, A. Y., Welp, G., & Amelung, W. (2018). Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). *Catena*, 166, 298-309. **DOI:10.1016/j.catena.2018.04.013.**
27. Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PloS one*, 8(6), e66409. **doi.org/10.1371/journal.pone.0066409.**
28. Vašát, R., Kodešová, R., Borůvka, L., Jakšík, O., Klement, A., & Brodský, L. (2017). Combining reflectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation. *Geoderma*, 303, 133-142. **doi.org/10.1016/j.geoderma.2017.05.018.**
29. Minasny, B., McBratney, A. B., Mendonça-Santos, M. D. L., Odeh, I. O. A., & Guyon, B. (2006). Prediction and digital mapping of soil carbon

- storage in the Lower Namoi Valley. *Soil Research*, 44(3), 233-244. DOI:10.1071/SR05136.
30. Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D., & Van Meirvenne, M. (2009). Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Science Society of America Journal*, 73(2), 614-621. doi.org/10.2136/sssaj2007.0410.
31. Bishop, T. F. A., McBratney, A. B., & Laslett, G. M. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma*, 91(1-2), 27-45. DOI:10.1016/S0016-7061(99)00003-8.
32. Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., ... & Greve, M. H. (2013). High-resolution 3-D mapping of soil texture in Denmark. *Soil Science Society of America Journal*, 77(3), 860-876. DOI:10.2136/sssaj2012.0275.
33. Mulder, V. L., Lacoste, M., Richer-de-Forges, A. C., Martin, M. P., & Arrouays, D. (2016). National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, 263, 16-34. DOI: 10.1016/j.geoderma.2015.08.035.
34. Fontaine, S., Barot, S., Barré, P., Bdioui, N., Mary, B., & Rumpel, C. (2007). Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature*, 450(7167), 277-280. DOI:10.1038/nature06275.
35. Meersmans, J., van Wesemael, B., De Ridder, F. A., & Van Molle, M. (2009). Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, 152(1-2), 43-52. DOI:10.1016/j.geoderma.2009.05.015.
36. Pan, L., & Politis, D. N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177, 1-27. doi.org/10.1016/j.jspi.2014.10.003.
37. Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural networks*, 19(2), 225-235. DOI:10.1016/j.neunet.2006.01.012.
38. Solomatine, D. P., & Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, 45(12). doi.org/10.1029/2008WR006839.
39. Malone, B. P., McBratney, A. B., & Minasny, B. (2011). Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma*, 160(3-4), 614-626. DOI:10.1016/j.geoderma.2010.11.013.
40. Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., ... & Zhang, G. L. (2014). GlobalSoilMap: Toward a fine-resolution global grid of soil properties. *Advances in agronomy*, 125, 93-134. DOI: 10.1016/B978-0-12-800137-0.00003-0.
41. 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]
42. Hosseini, M. (1996). Description of Geological Map 1:100000 Quarter Corners (Map Attachment), *Geological and Mineral Exploration Organization of the country*. [In Persian]
43. Zinck, J. A. (1989). Physiography and soils. Lecture notes for soil students. *Soil Science Division. Soil survey courses subject matter: K6 ITC, Enschede, The Netherlands*.
44. Schoeneberger, P. J., Wysocki, D. A., & Benham, E. C. (Eds.). (2012). *Field book for describing and sampling soils*. Government Printing Office.
45. Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil science*, 37(1), 29-38. doi.org/10.1097/00010694-193401000-00003.

46. Givi, J. (1997). Qualitative assessment of land suitability for agricultural and garden plants, Soil and Water Research Institute. *Technical magazine* 1015, 100 p. [In Persian]
47. Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillerá-Arroita, G. (2018). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv*, 357798. DOI:10.1111/2041-210X.13107.
48. Balan, B., Mohaghegh, S., & Ameri, S. (1995, September). State-of-the-art in permeability determination from well log data: Part 1-A comparative study, model development. In *SPE Eastern Regional Meeting* (pp. SPE-30978). SPE. DOI:10.2118/30979-MS.
49. Matinfar, H. R., Mghsodi, Z., Mossavi, S. R., & Jalali, M. (2021). Evaluation of Machine Learning Methods in Digital Mapping of Soil Organic Carbon (part of Khorramabad Plain). *Journal of Water and Soil Science*, 24 (4), 327-342. DOI:10.47176/jwss.24.4.8411. (In Farsi)
50. Salehi, M. H., & Khademi, H. (2017). Basics of soil mapping. Isfahan University Jihad Publications. 210 p. [In Persian]
51. Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E. D., & Goldschmitt, M. (2005). Digital soil mapping using artificial neural networks. *Journal of plant nutrition and soil science*, 168(1), 21-33. <https://doi.org/10.1002/jpln.200421414>.
52. Angiulli, F. (2005, August). Fast condensed nearest neighbor rule. In *Proceedings of the 22nd international conference on Machine learning* (pp. 25-32).
53. Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2006). Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal*, 70(2), 327-336. DOI:10.2136/sssaj.2005.0128.
54. Zolfaghari, A. A., Tirgar Soltani, M. T., Afshart, T., & Sarmadian, F. (2013). Comparison of K-nearest neighbor and artificial neural network methods for predicting cation exchange capacity of soil. *Journal of Soil Management and Sustainable Production*, 3(1), 77-94. <https://dor.isc.ac/dor/20.1001.1.23221267.1392.3.1.5.2>. [In Persian]
55. Ayoubi, S., Taghizadeh, R., Namazi, Z., Zulfqari, A., & Rustae Sadrabadi, F., (2015). Comparison of k-nearest neighbor and artificial neural network methods for digital zoning of soil salinity in Afzal Ardakan well area. *Journal of Water and Soil Science*. <https://civilica.com/doc/1201298>. <http://dx.doi.org/10.18869/acadpub.jstnar.20.76.59>. [In Persian]
56. Scull, P., Franklin, J., & 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.
57. Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), 3-30. <https://doi.org/10.1002/hyp.3360050103>.
58. Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
59. Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma*, 214, 141-154. DOI:10.1016/j.geoderma.2013.09.016.
60. Wilding, L., & 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).
61. Yoo, K., Amundson, R., Heimsath, A. M., & Dietrich, W. E. (2006). Spatial patterns of soil organic carbon on

- hillslopes: Integrating geomorphic processes and the biological C cycle. *Geoderma*, 130(1-2), 47-65. [https:// doi.org/10.1016/j.geoderma.2005.01.008](https://doi.org/10.1016/j.geoderma.2005.01.008).
62. Bangroo, S. A., Najar, G. R., & Rasool, A. (2017). Effect of altitude and aspect on soil organic carbon and nitrogen stocks in the Himalayan Mawer Forest Range. *Catena*, 158, 63-68. **DOI:10.1016/j.catena.2017.06.017.**
63. Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H., & Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, 145, 118-127. **DOI:10.1016/j.catena.2016.05.023.**