

Digital Mapping of Soil Equivalent Calcium Carbonate Using Landsat 8 Satellite Images and Environmental Data by Machine Learning Models in Badr Watershed, Kurdistan Province

Moslem Zarinibahador^{1*} 

(Received: August 18-2024 ; Revised: October 25-2024 ; Accepted: December 23-2024)

1. Agricultural and Natural Resources Research and Training Center of Tehran Province, Agricultural Research, Education and Extension Organization, Tehran, Iran.

*: Corresponding author Email: moslem.zarini@ymail.com

Abstract

The calcium carbonate equivalent (CCE) in soil is one of the most important soil properties. Predicting the amount of calcium carbonate equivalent in soil is essential for sustainable soil fertility management. The present study aimed to digitally map calcium carbonate equivalent using auxiliary environmental variables, Landsat 8 satellite images, and predictive models and to present the best models in the Badr watershed in the south of Qorveh district. In the first phase, a geomorphologic map was created using a geologic map and based on the ZINC method in a geographic information system environment. In the second phase, the location of 125 survey profiles was determined using the Latin hypercube technique, and the calcium carbonate equivalent of the soil horizons was measured by acid titration. The auxiliary variables included derivatives of the digital elevation model, remote sensing indices from the Landsat 8 satellite, and a geopedological map. The principal component analysis (PCA) method was used to select suitable auxiliary variables. In the third phase, the modeling was carried out, digital maps of the soil classes and properties were created, and the models were evaluated. Two different cases were investigated in this study to estimate the calcium carbonate equivalent of the soil. In the first case, artificial neural network models, decision tree analysis, random forest, and the K-nearest neighbor model were used for prediction. The multiple linear regression model was also used to combine the results of the models. Among the models used to predict the equivalent amount of calcium carbonate using the 10-fold cross-validation method, the multiple linear regression (MLR) model had the highest prediction accuracy with a coefficient of determination of 0.796 and a mean square error of 6.514. In the 5-fold cross-validation method, the K-nearest neighbor (KNN) model had the highest predictive accuracy with a coefficient of determination of 0.9845 and a root mean square error of 2.1258. Due to the spatial nature of the 10-fold cross-validation method, the use of this method is preferable to the 5-fold cross-validation method. In addition, the most important auxiliary variables in order of importance to predict the calcium carbonate equivalent in soil were the carbonate index, slope direction, geomorphology, the base level of the catchment network, and the slope of the catchment.

Keywords: Artificial neural network, Predictive models, Carbonate index, Auxiliary variables

How to cite this article: Zarinibahador, M. 2025. Digital Mapping of Soil Equivalent Calcium Carbonate Using Landsat 8 Satellite Images and Environmental Data by Machine Learning Models in Badr Watershed, Kurdistan Province. *J. Water and Soil Sci.* 29(1):81-96. DOI: [10.47176/jwss.29.1.64611](https://doi.org/10.47176/jwss.29.1.64611)

Copyright © 2025 Zarinibahador. Published by Isfahan University of Technology.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

نقشه‌برداری رقومی کربنات کلسیم معادل خاک با استفاده از تصاویر ماهواره‌ای لندست ۸ و داده‌های کمکی توسط مدل‌های یادگیری ماشین در حوضه آبخیز بدر، استان کردستان

مسلم زرینی بهادر^{*۱} 

(تاریخ دریافت: ۱۴۰۳/۵/۲۸؛ تاریخ بازنگری: ۱۴۰۳/۸/۴؛ تاریخ پذیرش: ۱۴۰۳/۱۰/۳)

۱. مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان تهران، سازمان تحقیقات، آموزش و ترویج کشاورزی (تات)، تهران، ایران.

*: مسئول مکاتبات: پست الکترونیکی: moslem.zarini@ymail.com

چکیده

کربنات کلسیم معادل خاک (CCE) یکی از ویژگی‌های مهم خاک است. پیش‌بینی مقدار کربنات کلسیم معادل خاک برای مدیریت پایدار حاصلخیزی خاک ضروری است. مطالعه حاضر با هدف نقشه‌برداری رقومی کربنات کلسیم معادل با استفاده از متغیرهای کمکی محیطی و تصاویر ماهواره لندست ۸ و مدل‌های پیش‌بینی‌کننده و معرفی بهترین مدل‌ها، در حوضه آبخیز بدر در جنوب شهرستان قروه انجام گرفت. برای انجام این پژوهش در مرحله اول، نقشه ژئومورفولوژی با استفاده از نقشه زمین‌شناسی و بر اساس روش ژئوپدولوژی زینک در محیط سامانه اطلاعات جغرافیائی ترسیم شد. در مرحله دوم، محل ۱۲۵ خاکرخ مطالعاتی بر اساس تکنیک ابر مکعب لاتین تعیین شد و کربنات کلسیم معادل افق‌های خاک با روش تیتراسیون با اسید اندازه‌گیری شد. متغیرهای کمکی شامل مشتقات مدل رقومی ارتفاع، شاخص‌های سنجش‌ازدور دریافتی از ماهواره لندست ۸ و نقشه ژئوپدولوژی بودند که انتخاب متغیرهای کمکی مناسب با استفاده از روش تجزیه مؤلفه‌های اصلی (PCA) انجام شد. در مرحله سوم، مدل‌سازی انجام، نقشه‌های رقومی کلاس‌ها و ویژگی‌های خاک تهیه شد و ارزیابی مدل‌ها صورت گرفت. برای برآورد کربنات کلسیم معادل خاک، در مطالعه حاضر دو حالت مختلف بررسی شد. در حالت اول، مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم، جنگل تصادفی، مدل نزدیک‌ترین همسایه K برای پیش‌بینی استفاده شدند. همچنین به‌منظور ترکیب نتایج مدل‌ها، از مدل رگرسیون خطی چندگانه استفاده شد. در میان مدل‌های استفاده‌شده برای پیش‌بینی مقدار کربنات کلسیم معادل با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، مدل رگرسیون خطی چندگانه (MLR) با ضریب تعیین ۰/۷۹۶ و ریشه دوم متوسط مربعات خطا ۶/۵۱۴ از بیشترین دقت برای پیش‌بینی برخوردار بوده است. این در حالی است که با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی، مدل نزدیک‌ترین همسایه K (KNN) با ضریب تعیین ۰/۹۸۴۵ و ریشه دوم متوسط مربعات خطا ۲/۱۲۵۸ از بیشترین دقت برای پیش‌بینی برخوردار بوده است. به‌دلیل مکانی بودن روش اعتبارسنجی کافلد ۱۰ مکانی، استفاده از این روش بر روش اعتبارسنجی کافلد ۵ تصادفی ارجحیت دارد. همچنین متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل خاک به‌ترتیب اهمیت شامل شاخص کربنات، جهت شیب، ژئومورفولوژی، سطح مبنای شبکه آبراهه و شیب حوضه آبخیز بودند.

واژه‌های کلیدی: شبکه عصبی مصنوعی، مدل‌های پیش‌بینی‌کننده، شاخص کربنات، متغیرهای کمکی

حق انتشار این مستند، متعلق به دانشگاه صنعتی اصفهان است. ۱۴۰۴ ©.

این مقاله تحت گواهی زیر منتشر شده و هر نوع استفاده غیرتجاری از آن مشروط به استناد صحیح به مقاله و با رعایت

شرایط مندرج در آدرس زیر مجاز است: (<https://creativecommons.org/licenses/by-nc/4.0/>)



مقدمه

کربنات‌های پدوژنیک (PC) رسوبات کربنات ثانویه‌ای هستند که در داخل خاک، به‌ویژه در مناطق خشک و نیمه‌خشک تشکیل می‌شوند. آن‌ها از طریق فرایندهای پیچیده انحلال و رسوب مجدد کربنات‌های زمین‌شناسی، بیولوژیکی یا پدوژنیک موجود در خاک تشکیل می‌شوند (۳۰).

تشکیل آن‌ها تحت تأثیر ترکیبی از عوامل آب‌وهوایی، خواص شیمیایی و فیزیکی خاک، فعالیت بیولوژیکی و حرکت آب در خاک است. کربنات‌های پدوژنیک نقش کلیدی در ترسیب جهانی کربن بازی می‌کنند؛ زیرا می‌توانند کربن را در اشکال غیرآلی در بازه‌های زمانی طولانی به دام ببندازند (۱۰).

علاوه بر این، وجود کربنات‌ها می‌تواند به‌طور قابل توجهی بر ویژگی‌های فیزیکی خاک مانند ساختار خاک، تخلخل، هدایت هیدرولیکی و همچنین در دسترس بودن مواد مغذی و حاصلخیزی کلی خاک تأثیر بگذارد (۹). خاک‌های سطحی، فعالیت بیولوژیکی شدیدی از خود نشان می‌دهند که منجر به غلظت بیشتر کربن آلی خاک، pH کمتر و در نتیجه انحلال کامل ذرات کربنات می‌شود. با کاهش فعالیت بیولوژیکی با عمق، کاهش متناظر در محتوای کربن آلی خاک و در نتیجه افزایش مقادیر pH ایجاد می‌شود که نشان‌دهنده خثی شدن محیط اسیدی از طریق انحلال کربنات است. این تغییرات، شرایط مساعدی را برای رسوب کربنات‌های پدوژنیک در لایه‌های عمیق‌تر پروفیل خاک ایجاد می‌کند (۱۵).

کشاورزان، تصمیم‌گیرندگان کشاورزی و برنامه‌ریزان کاربری زمین برای مدیریت پایدار منابع طبیعی به ارزیابی‌های به‌روز، قابل اعتماد و دقیق از ویژگی‌های خاک نیاز دارند. خاک‌های شور و آهکی تهدیدی قابل توجه برای تولید محصولات زراعی هستند و می‌توانند به‌طور قابل توجهی بهره‌وری کشاورزی را به‌ویژه در مناطق خشک و نیمه‌خشک کاهش دهند. تجمع نمک در ناحیه ریشه فرایندهای خاک از جمله نیتروفیکاسیون، نترات‌زدایی و تجزیه باقی‌مانده را به دلیل کاهش فعالیت میکروارگانیسم‌ها و تنوع زیستی خاک محدود می‌کند. از سوی دیگر در خاک‌های آهکی به دلیل pH زیاد، دسترسی به برخی از عناصر غذایی مانند

فسفر، آهن و مس محدود است. با توجه به اینکه رویکردهای سنتی پایش و نقشه‌برداری پرهزینه و وقت‌گیر هستند، برآورد سریع و کارآمد خواص خاک توسط پژوهشگران با استفاده از داده‌های سنجش‌ازدور با اندازه‌گیری‌های میدانی مورد توجه قرار گرفته است (۱۸).

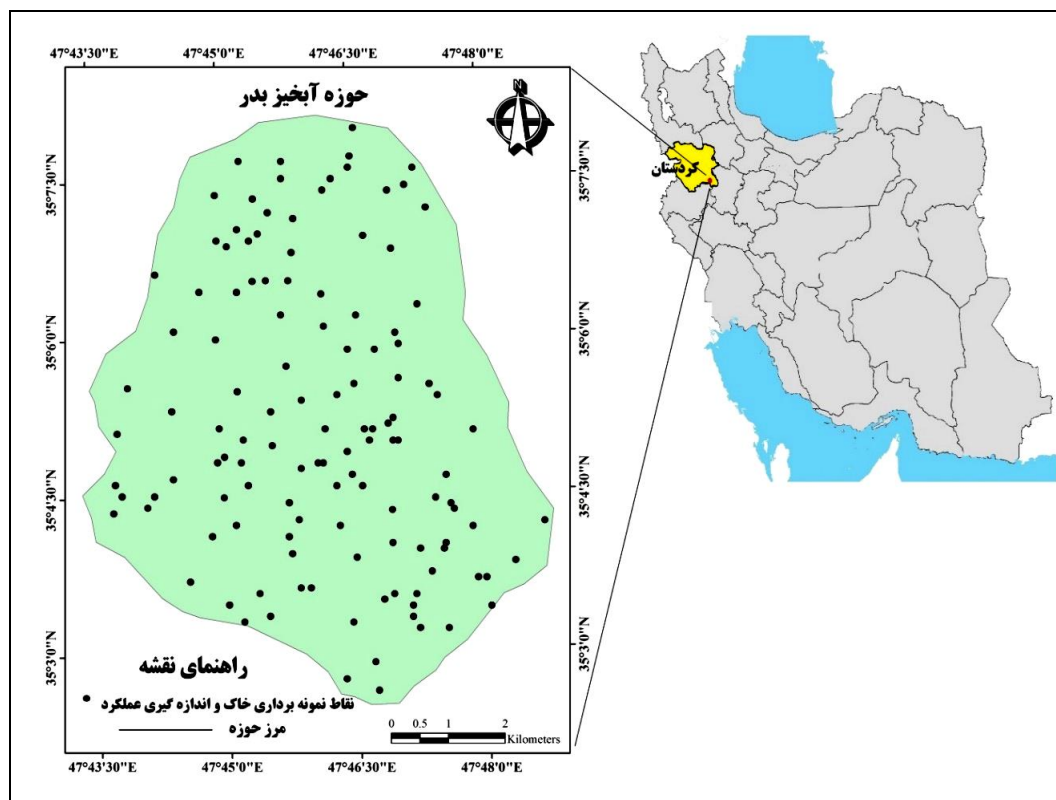
کربنات کلسیم معادل یکی از ویژگی‌های کلیدی خاک‌های مناطق خشک و نیمه‌خشک است که بررسی تغییرات سطحی و عمقی آن از اهمیت ویژه‌ای در بهره‌برداری پایدار از خاک‌های زراعی برخوردار است. از آنجایی که کربنات کلسیم معادل خاک (CCE) نقش مهمی در حاصلخیزی و مدیریت پایدار خاک در مناطق نیمه‌خشک ایفا می‌کند و توزیع مکانی این ویژگی را از بالا به پایین (تا عمق ۱ متر) با استفاده از داده‌های کمکی به‌دست می‌آورند، یعنی ویژگی‌های مرتبط با زمین (متغیرهای کمکی ثابت) و همچنین داده‌های سنجش‌ازدور (متغیرهای کمکی پویا) در توسعه کشاورزی پایدار مفید خواهد بود (۱۳).

در این پژوهش، از پنج روش یادگیری ماشین برای توسعه مدل‌های پیش‌بینی استفاده شد: شبکه عصبی مصنوعی، تحلیل درخت تصمیم، رگرسیون خطی چندگانه، نزدیک‌ترین همسایه K و رگرسیون خطی چندگانه. اطلاعات به‌دست‌آمده از مدل‌سازی CCE می‌تواند در ارزیابی تناسب اراضی برای کاربری‌های مختلف یا استراتژی‌های مدیریتی استفاده شود. از این‌رو، این پژوهش با هدف نقشه‌برداری رقومی کربنات کلسیم معادل با استفاده از داده‌های کمکی محیطی و تصاویر ماهواره لندست ۸ و مدل‌های پیش‌بینی‌کننده و معرفی بهترین مدل‌ها، در حوضه آبخیز بدر در جنوب شهرستان قروه انجام شد.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

اراضی مورد مطالعه با وسعتی حدود ۶۷۰۰ هکتار در جنوب شهرستان قروه و بین مختصات جغرافیائی ۴۷ درجه و ۲۳ 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).



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

لندست ۸ و مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر، از سازمان نقشه برداری کشور دریافت شد. نقشه ژئومورفولوژی با استفاده از نقشه زمین شناسی و بر اساس روش ژئوپدولوژی زینک در محیط سامانه اطلاعات جغرافیائی ترسیم شد. در مرحله دوم، محل ۱۲۵ خاکرخ مشاهداتی تعیین شد. شناسایی خاکها در صحرا انجام شد و از لایه های مختلف خاک نمونه برداشت شد. سپس اندازه گیری های فیزیکی و شیمیایی و کانی شناسی خاکها انجام و خاکها طبقه بندی شدند. در مرحله سوم، مدل سازی انجام، نقشه ویژگی های خاک تهیه و ارزیابی مدل ها صورت گرفت.

جمع آوری داده های متغیرهای کمکی، پردازش و فرایند

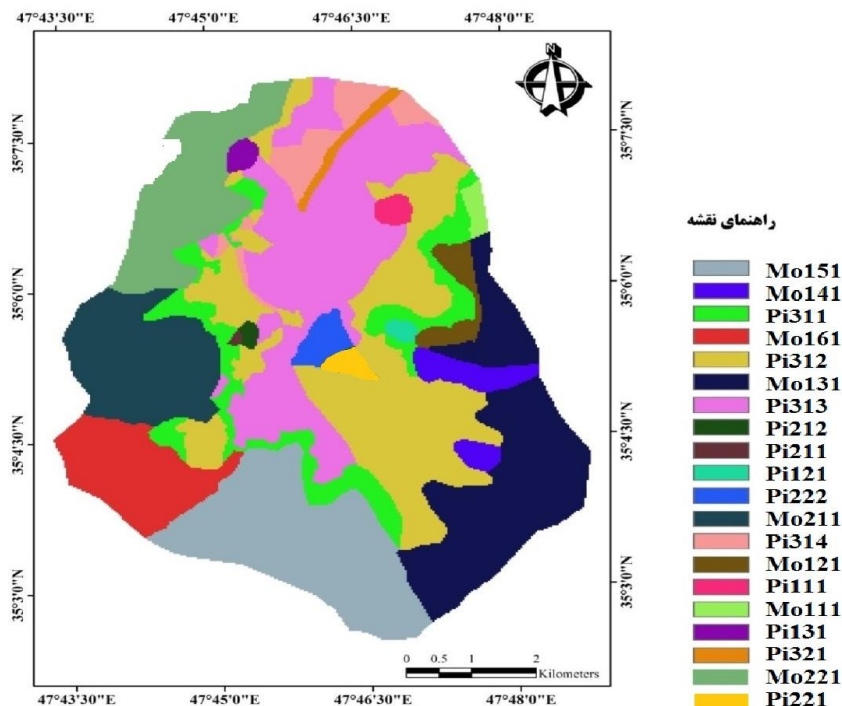
انتخاب مهم ترین آنها

در پژوهش حاضر، از سه منبع برای تهیه داده های کمکی بهره گرفته شد: تصاویر ماهواره ای، مدل رقومی ارتفاع و نقشه ژئومورفولوژی.

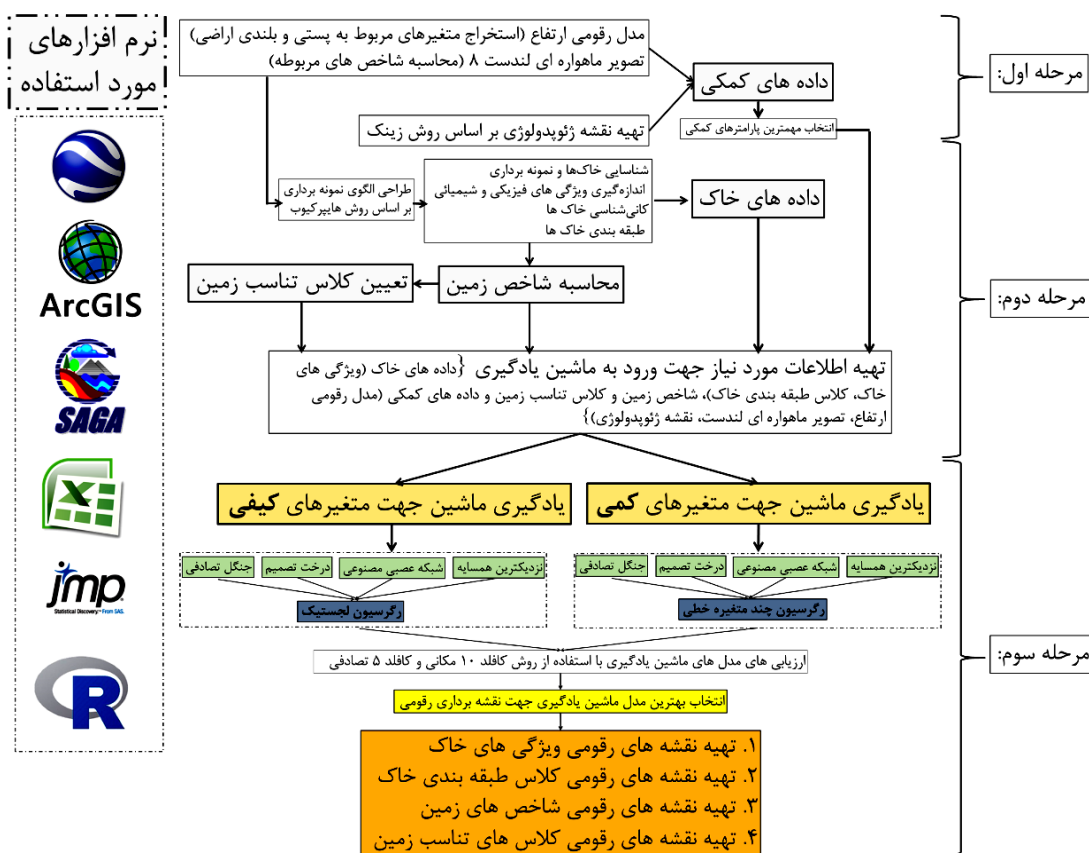
مطابق با آمار طولانی مدت هواشناسی، متوسط درجه حرارت سالیانه هوا ۱۲/۰۷ درجه سانتی گراد و میانگین بارندگی سالیانه ۳۴۵/۸ میلی متر است. الگوی بارش، مدیترانه ای بوده و بخش اعظم بارندگی در فصول سرد سال اتفاق می افتد. بر اساس اقلیم نمای آمبرژه، اقلیم این ناحیه نیمه خشک و سرد است. در منطقه مورد مطالعه، رژیم رطوبتی خاک زریک (Xeric) و رژیم حرارتی آن مزیک (Mesic) تعیین شده است. از لحاظ ژئومورفولوژی، حوزه آبخیز مورد مطالعه شامل کوه و اراضی پای کوهی (پیدمونت) است (شکل ۲). کاربری های رایج اراضی کشاورزی، گندم دیم و آبی، جو دیم و آبی، نخود، لوبیا، یونجه، شبدر و... است.

روش اجرا

برای اجرای این پژوهش مراحل زیر، طبق شکل ۳ انجام گرفت: در مرحله اول، داده های کمکی از جمله: تصاویر ماهواره ای



شکل ۲. ژئوپدولوژی حوضه آبخیز بدر شهرستان قروه، استان کردستان



شکل ۳. مراحل اجرای پژوهش

نمونه برداری، تجزیه و تحلیل آزمایشگاهی و پردازش داده‌ها

در این پژوهش از نرم افزار R برای مشخص کردن نقاط حفر خاکرخ‌ها استفاده و داده‌های استفاده شده برای وارد کردن به نرم افزار مشتمل بر DEM، RS و نقشه ژئومورفولوژی بودند. موقعیت تمامی خاکرخ‌ها توسط سیستم موقعیت یاب جهانی (GPS) ثبت شد. خاکرخ‌ها بر اساس راهنمای تشریح و نمونه برداری خاک‌ها در صحرا (۲۴) تشریح و از تمامی افق‌های آن نمونه برداری صورت گرفت. نمونه‌های خاک پس از هوا خشک شدن در محیط آزمایشگاه، کوبیده و از الک ۲ میلی متری عبور داده شدند. سپس کربنات کلسیم معادل با روش تیتراسیون با اسید اندازه گیری شد.

متوسط وزنی عمق ۱۰۰ سانتی متری کربنات کلسیم معادل، برای نقشه برداری رقومی ویژگی‌های خاک بکار رفت. به منظور محاسبه متوسط وزنی، مقدار هر ویژگی در افق‌های خاک، ضخامت هر افق و ضریب وزنی عمق (۱۱) استفاده شد.

متوسط وزنی درصد کربنات کلسیم معادل تا عمق یک متری با استفاده از ضرایب وزنی جدول ۱ و معادله (۱) محاسبه شد (۲۶).

$$X_i = \frac{\sum D_i X_i C_i}{\sum D_i} \quad (1)$$

در این معادله:

X_i = مقدار هر یک از پارامترهای خاکی برای افق i ام

D_i = ضخامت افق i ام

C_i = ضریب وزنی عمق افق i ام

ارزیابی مدل‌های یادگیری ماشین

در این پژوهش، به منظور برآورد کربنات کلسیم معادل از مدل‌های شبکه عصبی مصنوعی، نزدیک‌ترین همسایه K ، درخت تصمیم و جنگل تصادفی استفاده شد. علاوه بر این، از مدل رگرسیون خطی چندگانه برای ترکیب مدل‌های یاد شده نیز بهره گرفته شد. به منظور ارزیابی مدل‌های یادگیری ماشین از دو روش بهره گرفته شد: ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی. در این نوع اعتبارسنجی‌ها، داده‌ها به K

مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر مبنای کار قرار گرفت. در این مرحله، با استفاده از این مدل و زیر برنامه‌های تعریف شده استاندارد موجود در نرم افزار Arc-GIS10 و ساگا (System for Automated Geoscientific Analyses, SAGA) مشتقات مختلف مرتبه‌های اول و دوم مدل تهیه شد. شاخص‌های سنجش از دور نیز با استفاده از داده‌های باندهای ۱ تا ۶ ماهواره لندست ۸ (سال ۲۰۱۳) محاسبه شدند. تفکیک واحدهای ژئومرفیک موجود در منطقه مطالعاتی با توجه به تفاوت رنگ، الگوی آبراهه‌ای، 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 (۳۲) 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، خاک (های) مربوط به هر شکل زمین نیز معرفی شد. در پایان، نقشه ژئومورفولوژی در سامانه اطلاعات جغرافیایی (GIS) رقومی شد.

برای انتخاب پارامترهای مؤثر در مدل سازی، از روش «انتخاب ویژگی به ترتیب اهمیت» استفاده شد. به این معنی که تمام اطلاعات کمکی پس از استخراج از نرم افزار ساگا، وارد نرم افزار وکا (Weka) شد و با استفاده از الگوریتم ReliefAttributeEval، اطلاعات کمکی مهم و مؤثر در پیش بینی ویژگی‌ها و کلاس‌های طبقه بندی خاک و رده بندی کیفی تناسب اراضی به ترتیب اهمیت، انتخاب و در مدل سازی استفاده شد. این الگوریتم، ارزش یک زیر مجموعه از ویژگی‌ها، 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، زیر مجموعه یاد شده که با ویژگی‌ها و کلاس‌های مورد پیش بینی همبستگی زیادی دارند، در حالی که همبستگی درونی آن‌ها کم است، ترجیح داده می شوند.

جدول ۱. ضرایب وزنی عمق‌های متوسط‌گیری، مورد استفاده در محاسبه‌ی متوسط وزنی میزان کربنات کلسیم معادل خاک

عمق متوسط‌گیری (سانتی‌متر)	تعداد قسمت‌های مساوی عمق	ضرایب وزنی
۱۰۰ >	۴	۰/۰-۲۵/۱-۷۵/۱-۲۵/۷۵
۱۰۰-۷۵	۴	۰/۰-۲۵/۱-۷۵/۱-۲۵/۷۵
۷۵-۵۰	۳	۰/۰-۶/۱-۹/۵
۵۰-۲۵	۲	۰/۱-۸/۲
۲۵ <	۱	ضریب وزنی نیاز نیست

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 کربنات کلسیم معادل خاک در نرم‌افزار R استفاده شد.

۱- رگرسیون خطی چندگانه

با استفاده از این روش می‌توان همزمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. برای به‌دست‌آوردن نتایج مطلوب‌تر از طریق رگرسیون خطی چندگانه، نمونه‌ها باید دارای فراوانی زیاد و همچنین از دقت قابل توجهی برخوردار باشند؛ زیرا این روش در مقابل اطلاعات نادرست، حساسیت زیادی دارد و ورود چنین داده‌هایی ممکن است منجر به خطاهای بزرگی در نتایج به‌دست‌آمده 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 مد نظر 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، معادله (۲) بین آن‌ها برقرار است:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (2)$$

که در این معادله، 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 (۵ و ۱۶).

زیرمجموعه دسته‌بندی شدند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار رفتند. این روال، K بار تکرار شد و همه داده‌ها دقیقاً یک‌بار برای آموزش و یک‌بار برای اعتبارسنجی بکار رفتند. درنهایت، میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده شد. گفتنی است که تفاوت ارزیابی متقابل تصادفی و ارزیابی تصادفی مکانی، در انتخاب مجموعه داده‌های خاک به منظور آموزش و ارزیابی است. در روش ارزیابی متقابل تصادفی، داده‌ها به طور تصادفی به دو دسته تقسیم می‌شوند: داده‌هایی که برای آموزش به کار می‌روند و داده‌هایی که برای اعتبارسنجی یا آزمون استفاده می‌شوند. 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 برای 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 (۲۸).

همچنین، برای بیان کمی میزان خطا از معیارهای معمول در یادگیری ماشین استفاده شد. بدین منظور، برای بررسی دقت و صحت مقادیر پیش‌بینی‌شده داده‌های کمی مانند کربنات کلسیم معادل خاک توسط مدل‌های یادگیری ماشین، شاخص‌هایی از قبیل جذر میانگین مربع خطا (Root Mean Square Error: RMSE)، جذر میانگین مربع خطای نسبی (Relative RMSE: RMSE%)، ضریب تعیین (R^2) و ضریب همبستگی (r) محاسبه شدند.

۲- شبکه‌های عصبی مصنوعی

رایانه‌ها، دارای سرعت قابل توجهی نسبت به انسان هستند؛ 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 شد. مبنای شبکه‌های عصبی مصنوعی روابط پیچیده ریاضی منطبق بر سامانه 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 (۲۳). در پژوهشی برآورد واحدهای خاک را با استفاده از شبکه‌های عصبی مصنوعی و با به کارگیری داده‌های رقومی ارتفاع انجام داده و بیان کردند که ترکیب تکنیک شبکه‌های عصبی مصنوعی و اطلاعات به دست آمده از مدل رقومی ارتفاع، علاوه بر صرفه‌جویی در زمان و هزینه‌ها، نتایج قابل قبولی ارائه می‌دهد (۶).

در مطالعه‌ای الگوی توزیع خاک‌ها در منطقه بروجن استان چهارمحال و بختیاری را با استفاده از شبکه عصبی مصنوعی بررسی کردند. نتایج نشان داد که استفاده از ویژگی‌های مدل رقومی ارتفاع، به تنهایی نمی‌تواند منجر به تهیه نقشه خاک با دقت و صحت زیاد شود و اطلاعات دیگری مانند واحدهای ژئومرفیک نیز لازم است (۴). در این پژوهش، با ورود ویژگی ژئوفرم، دقت و صحت برآورد واحد خاک هم در درون‌یابی و هم در برون‌یابی افزایش یافت. نتایج این مطالعه نشان داد که با ارزیابی دقیق، نظر افراد خبره و ورود اطلاعات محیطی کافی به مدل، این تکنیک می‌تواند با کاهش هزینه و زمان، دقت و صحت نقشه را نیز افزایش دهد. در پژوهشی دیگر عملکرد محصول گندم دیم را با استفاده از مدل‌های شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره تعیین کردند و بیان کردند که مدل شبکه‌های عصبی مصنوعی، قدرت زیادی در پیش‌بینی عملکرد محصول دارد (۱۷).

۳- 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 (۲۲). 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.

روش نزدیک‌ترین همسایگی K دارای یکی از ساده‌ترین و کارآمدترین الگوریتم‌های آموزش برای پیش‌بینی داده‌ها است (۱). 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 (۲۲ و ۳۳). به این ترتیب که در ابتدا فواصل اقلیدسی بین خاک مورد آزمون و تمامی خاک‌های مرجع محاسبه شده، سپس تعداد K عدد از نزدیک‌ترین خاک‌ها بر اساس فاصله به دست آمده وزن‌دهی می‌شوند. در انتها با توجه به وزن هر یک از خاک‌ها در مجموعه‌ای با K عدد خاک (نزدیک‌ترین همسایه)، برآوردی از داده‌های هدف که مورد آزمون واقع شده‌اند، صورت می‌پذیرد.

همچنین در مطالعه‌ای با عنوان مقایسه روش‌های K نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی برای پهنه‌بندی رقومی شوری خاک در منطقه‌ی چاه افضل اردکان، به این نتیجه دست یافتند که مدل K نزدیک‌ترین همسایگی دارای دقت بیشتری نسبت به شبکه عصبی مصنوعی به منظور پیش‌بینی شوری خاک

مرحله است. مرحله اول ایجاد و رشد درخت است که شامل پیوند و انشعاب است و مرحله دوم توقف و هرس است که هدف آن کمینه کردن خطای پیش‌بینی است (۱۴).

نتایج و بحث

تشریح خاک‌های منطقه مطالعاتی

رده‌بندی خاک‌رخ‌های حفر شده بر مبنای سامانه رده‌بندی خاک آمریکایی (۲۵) تا سطح فامیل انجام شد. بر اساس نتایج حاصل از این رده‌بندی، می‌توان بیان کرد که در منطقه مورد مطالعه؛ سه رده Entisols و Mollisols، Inceptisols؛ سه زیررده Xerepts، Xerolls و Orthents؛ چهار گروه بزرگ Calcixerepts، Haploxerepts، Haploxerolls و Xerorthents وجود دارند. گروه بزرگ Calcixerepts از دو زیرگروه Typic Calcixerepts و Vertic Calcixerepts تشکیل شده است.

زیرگروه‌های Haploxerepts عبارت‌اند از: Typic Haploxerepts، Fluventic Haploxerepts، Calcic Haploxerepts و Vertic Haploxerepts و Lithic Haploxerepts.

گروه بزرگ Haploxerolls شامل زیرگروه‌های Vertic Haploxerolls و Typic Haploxerolls و گروه بزرگ Xerorthents مشتمل بر زیرگروه‌های Lithic Xerorthents و Typic Xerorthents است. در مجموع این یازده زیرگروه به سی و دو فامیل طبقه‌بندی می‌شوند.

همان‌طور که در شکل ۴ دیده می‌شود، ۴/۸، ۶/۴ و ۴۴/۸ درصد خاک‌ها به ترتیب در رده‌های اینسپتی‌سولز، مالی‌سولز و انتی‌سولز قرار دارند.

خلاصه آماری داده‌ها

جدول ۲، آمار توصیفی خصوصیات فیزیکی و شیمیایی خاک‌ها و عملکرد گندم 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

است؛ به‌طوری که این مدل به خوبی توانسته ارتباط قوی بین داده‌های شوری خاک و اطلاعات محیطی برقرار کند (۳).

۴- مدل تحلیل درخت تصمیم

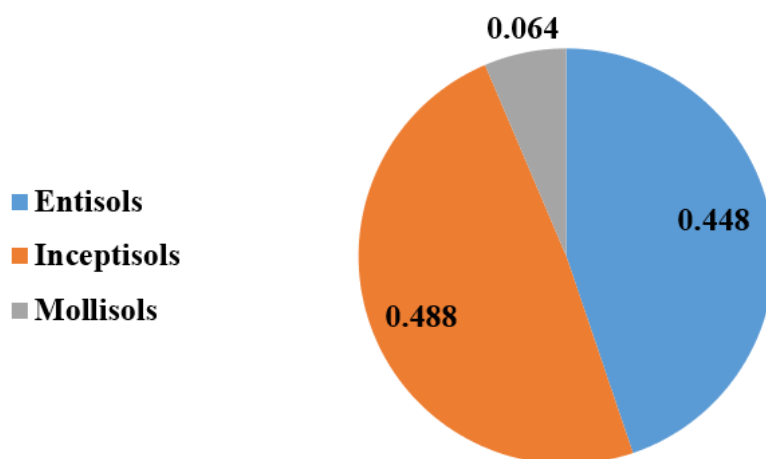
مدل درختی طبقه‌بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه‌بندی شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است. در این روش، مجموعه‌ای از شرط‌های منطقی به صورت یک الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود. ایجاد درخت تصمیم دو مرحله دارد. مرحله اول، مرحله ایجاد و رشد درخت است. این مرحله شامل پیوند و انشعاب است. مرحله دوم، مرحله توقف و هرس است. هدف این مرحله کمینه کردن خطای پیش‌بینی است (۲۱).

۵- مدل جنگل تصادفی

مدل جنگل تصادفی یک روش ناپارامتری است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه‌بندی شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است. این روش اولین بار توسط بریمن (Breiman) گسترش یافت و یکی از جدیدترین روش‌های داده‌کاوی است. مدل جنگل تصادفی (RF) در حقیقت مجموعه‌ای از درختان پیش‌بینی کننده با احتمال و پراکندگی یکسان هستند. اساس این روش بر مبنای انتخاب خطای داده‌های بزرگ‌تر به عنوان خطای اصلی و ایجاد همبستگی بین خطاهای دیگر داده‌ها به ترتیب است (۷).

مدل جنگل تصادفی (RF) مدل توسعه یافته از مدل طبقه‌بندی و رگرسیون درختی (Classification and Regression Tree) است. روش CART روشی است که داده‌ها را به صورت تکراری برای ایجاد ارتباط بین متغیر پاسخ و متغیرهای مستقل و انجام پیش‌بینی مکانی جداسازی می‌کند (۸).

تکنیک RF مجموعه‌ای از شرط‌های منطقی است که به صورت یک الگوریتم برای پیش‌بینی کمی یک متغیر با ساختار درختی به کار می‌رود. ایجاد درخت در این روش شامل دو



شکل ۴. رده‌بندی خاک‌های منطقه مطالعاتی بر اساس سامانه رده‌بندی خاک آمریکایی

جدول ۲. آمار توصیفی برخی ویژگی‌های فیزیکی و شیمیایی خاک و عملکرد اندازه‌گیری شده

ویژگی‌های فیزیکی و شیمیایی خاک و عملکرد اندازه‌گیری شده	واحد	کمینه	بیشینه	میانگین	انحراف معیار	ضریب تغییرات
pH	—	۶/۸	۸	۷/۵	۰/۳	۳/۶
هدایت الکتریکی	dS.m ⁻¹	۰/۱	۰/۴	۰/۲	۰/۱	۲۵/۵
کربن آلی	درصد	۰/۱	۲/۳	۰/۶	۰/۴	۶۲/۴
کربنات کلسیم معادل	درصد	۰	۶۵	۹/۸	۱۷	۱۷۴/۴
ظرفیت تبادل کاتیونی	cmol(+)kg ⁻¹	۸/۹	۲۶/۵	۱۴/۳	۲/۴	۱۶/۹
رس	درصد	۹	۴۳	۲۲/۵	۶/۸	۳۰/۴
سیلت	درصد	۱۵/۹	۴۸/۶	۲۹/۱	۶/۸	۲۳/۵
شن	درصد	۲۷/۱	۷۰	۴۸/۳	۹/۸	۲۰/۳
سنگریزه	درصد حجمی	۱/۵	۴۸	۱۷/۱	۱۰/۴	۶۱/۳
عملکرد مشاهداتی	kg ha ⁻¹	۰	۲۱۶۰	۹۴۹/۴	۸۳۲/۲	۸۷/۶

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 بکار می‌رود. این شاخص تعیین‌کننده کلاس تناسب زمین برای گندم دیم است. پارامترهای محیطی که از

کم هستند (۲۹). 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

۵ تصادفی (جدول ۶)، از بین مدل‌های شبکه عصبی مصنوعی (Artificial Neural Network, ANN)، تحلیل درخت تصمیم (Decision Tree Analysis, DTA)، رگرسیون خطی چندگانه (Multi Linear Regression, MLR) و نزدیک‌ترین همسایه K (K Nearest Neighbor)، مدل نزدیک‌ترین همسایه K (KNN) با ضریب تعیین ۰/۹۸۴۵ و ریشه دوم متوسط مربعات خطا ۲/۱۲۵۸ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

آنالیز مکانی کربنات کلسیم معادل خاک

مقادیر کمینه و بیشینه کربنات کلسیم معادل در حوضه آبخیز بدر، ۰ و ۶۵ درصد است. در نواحی شمال غربی حوضه، یعنی جایی که کوه‌ها از سنگ مادر مرمر تشکیل شده‌اند، بیشترین مقدار کربنات کلسیم معادل وجود دارد (شکل ۵). در غرب، جنوب و شرق مقدار این ماده به صفر می‌رسد؛ زیرا در این مناطق، کوه‌ها از سنگ‌های آذرین و دگرگونی غیر آهکی تشکیل شده‌اند. پایین‌ترین دشت‌های آبرفتی که تحت تأثیر سنگ مادر آهکی هستند، دارای خاک آهکی هستند. در شکل ۵، 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.

متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل خاک به ترتیب اهمیت عبارت‌اند از: شاخص کربنات، جهت شیب، ژئومورفولوژی، سطح‌مبنای شبکه آبراهه و شیب حوضه آبخیز (جدول ۴). مهم‌ترین عامل نشان‌دهنده مقدار کربنات کلسیم معادل، شاخص کربنات است که از نسبت باند ۳ به باند ۲ ماهواره لندست ۸ به‌دست می‌آید. اثر جهت شیب بر پیش‌بینی مقدار آهک، در درجه دوم اهمیت قرار دارد. شیب‌های به‌طرف شمال سردتر و مرطوب‌ترند؛ در نتیجه تراکم پوشش گیاهی بیشتر می‌شود. ترشح اسید از ریشه‌های گیاه، یونیزه شدن کربنات کلسیم و کاهش مقدار آن در خاک را باعث می‌شود. ژئومورفولوژی و شیب حوضه آبخیز، از نظر تأثیرگذاری بر میزان کربنات به ترتیب در درجات سوم و پنجم اهمیت قرار دارند.

مدل رقومی ارتفاع و تصاویر ماهواره لندست ۸ به‌دست آمدند، ویژگی‌های خاک یادشده را تحت تأثیر قرار می‌دهند.

متغیرهای کمکی مورد استفاده در پیش‌بینی ویژگی‌ها و کلاس‌های خاک

متغیرهای کمکی استفاده شده در پیش‌بینی ویژگی‌ها و کلاس‌های خاک که از نرم‌افزار ساگا و تصاویر ماهواره‌ای استخراج شده در جدول ۳ نشان داده شده‌اند.

متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل

متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل در خاک‌های منطقه مورد مطالعه به ترتیب اهمیت در جدول ۴ ارائه شده است. در این جدول از چپ به راست، از اهمیت متغیرهای کمکی کاسته می‌شود. درجه‌بندی اهمیت متغیرهای کمکی با به کارگیری الگوریتم ReliefAttributeEval نرم‌افزار WEKA مشخص شد. همچنین متغیرهای کمکی مهم با استفاده از روش تحلیل مؤلفه اصلی (PCA) انتخاب شدند.

مقایسه دقت مدل‌های مورد استفاده

نتایج ارزیابی صحت پیش‌بینی مقدار کربنات کلسیم معادل خاک‌های مورد مطالعه، با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی، در جدول ۵ ارائه شده است. این پیش‌بینی توسط مدل‌های نزدیک‌ترین همسایه K (KNN)، تحلیل درخت تصمیم (DTA)، شبکه عصبی مصنوعی (ANN)، جنگل تصادفی (R.F.) و رگرسیون خطی چندگانه (MLR) صورت گرفته است. در میان مدل‌های استفاده شده برای پیش‌بینی مقدار کربنات کلسیم معادل، مدل رگرسیون خطی چندگانه (MLR) با ضریب تعیین ۰/۷۹۶ و ریشه دوم متوسط مربعات خطا ۶/۵۱۴ از بیشترین دقت برای پیش‌بینی برخوردار بوده است.

در جدول ۵ دقت پیش‌بینی مقدار کربنات کلسیم معادل، از بالا به طرف پایین افزایش می‌یابد. کمترین دقت پیش‌بینی به مدل نزدیک‌ترین همسایه K (KNN) اختصاص یافته است. این در حالی است که با استفاده از روش اعتبارسنجی کافلد

جدول ۳. متغیرهای کمکی استفاده شده برای پیش‌بینی ویژگی‌ها و کلاس‌های خاک (۲)

تعریف	مشتقات اولیه و ثانویه پستی و بلندی، مستخرج از مدل رقومی ارتفاع/ عرض‌های باندی مختلف، برگرفته از تصاویر ماهواره‌ای
ارتفاع از سطح دریای آزاد	ارتفاع
ارتفاع سطح پایه شبکه آبراهه	سطح پایه شبکه آبراهه
تجمعات جریان آب و پارامترهای مربوطه	مساحت حوضه‌های آبخیز اصلاح‌شده
شاخص نسبت پهنای دره به عمق دره	عمق دره
میزان همواری کف دره	شاخص همواری دره با درجه تفکیک زیاد
میزان همواری نوک قله	شاخص همواری قله برآمدگی با درجه تفکیک زیاد
معیاری از تجمع رطوبت	شاخص خیسی
زاویه بین سطح تماس و سطح افقی در یک نقطه درسته زمین	شیب
بیشترین سرعت تغییر در ارتفاع در هر سلول از مدل ارتفاعی	جهت
شیب بالایی یا میانی یا پائینی	موقعیت نسبی شیب
میزان گرا دیان مسیر حرکت	شیب حوضه
جهت شیب مسیر حرکت	جهت شیب حوضه
میزان ارتفاع مسیر حرکت	ارتفاع حوضه
انحنای در صفحه عمودی خط جریان	انحنای شیب
اندازه همگرایی و واگرایی جریان	انحنای سطحی شیب
فاصله عمودی تا شبکه آبراهه را نشان می‌دهد	فاصله عمودی تا شبکه آبراهه
طول شیب را نشان می‌دهد	طول شیب
مقدار پستی و بلندی را نشان می‌دهد	شاخص انحنای
مقدار انحنای در جهت عمود بر شیب اصلی دامنه	انحنای عمود بر شیب اصلی (انحنای عرضی)
مقدار انحنای در جهت شیب اصلی دامنه	انحنای طولی
$(NIR^* - RED^{**}) / (NIR + RED)$	شاخص گیاهی تفاضلی نرمال شده (NDVI)
$[(NIR - RED) / (NIR + RED + L^{***})] \times [1 + L]$	شاخص گیاهی تعدیل‌کننده اثر خاک (SAVI)
$B5 / B7$	شاخص رس
$B3 / B2$	شاخص کربنات
نمایش عمق آب و ردیابی ذرات گردوغبار و دود و دارای طول موج μm	باند ۱ ماهواره لندست ۸
$0.433 - 0.453$	

جدول ۴. متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل خاک به ترتیب اهمیت

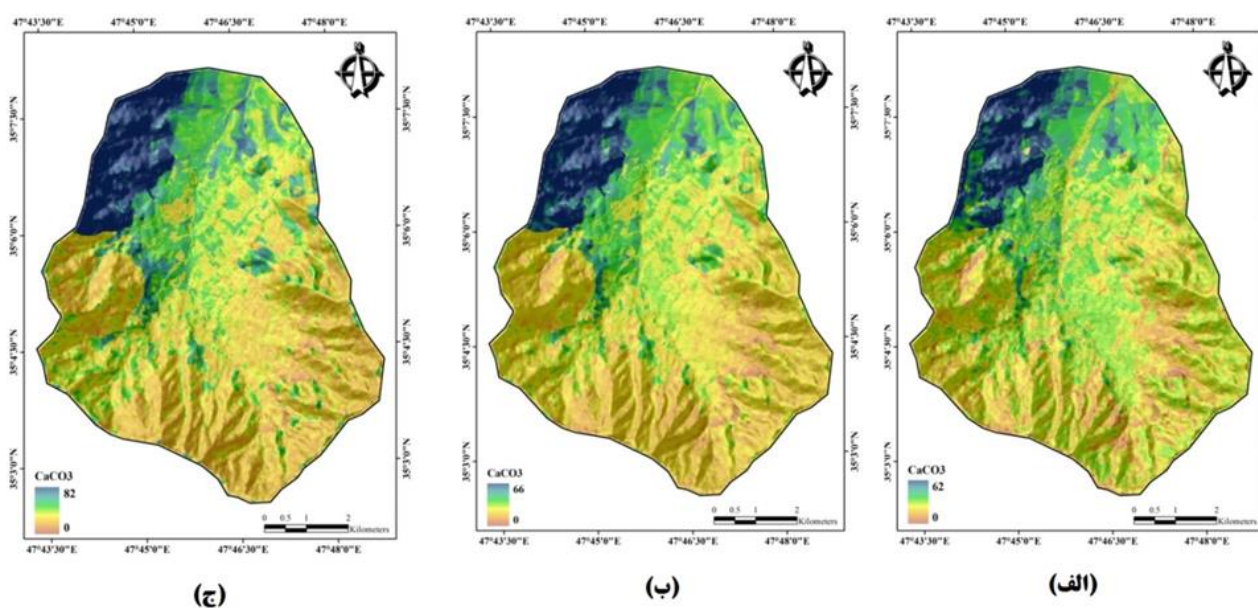
ویژگی	متغیرهای کمکی مهم در پیش‌بینی کربنات کلسیم معادل خاک به ترتیب اهمیت
CaCO3	B1, Aspect, Geomorphology, Channel network base level, Catchment slope

جدول ۵. نتایج ارزیابی صحت پیش‌بینی آهک با استفاده از روش اعتبارسنجی کافلد ۱۰ مکانی

مدل	RMSE	MAE	R ²
KNN	۱۳/۹۳۶	۳۲۹/۹	۰/۳۴۷
DTA	۸/۱۱۶	۵/۳۲۵	۰/۷۲۳
ANN	۷/۸۱۷	۵/۳۲۵	۰/۷۷۷
RF	۷/۸۱۵	۴/۵۶۸	۰/۷۸۰
MLR	۶/۵۱۴	۴/۳۸۰	۰/۷۹۶

جدول ۶. نتایج ارزیابی صحت پیش‌بینی مقدار کربنات کلسیم معادل با استفاده از روش اعتبارسنجی کافلد ۵ تصادفی

مدل	RMSE	^N RMSE=RMSE/Mean real	r ^C	R ²	ME= mean predict- mean real
ANN	۲/۲۴۶۶	۰/۲۳۰۲۶	۰/۹۹۰۶	۰/۹۸۱۹	۰/۱۰
DTA	۶/۳۸۷۷	۰/۶۵۴۷	۰/۹۲۳۶	۰/۸۵۸	۰
MLR	۷/۲۹۹۸	۰/۷۴۸۱	۰/۹۳۵۲	۰/۸۷۸۳	۰
KNN	۲/۱۲۵۸	۰/۲۱۷۸۸	۰/۹۹۲۱	۰/۹۸۴۵	۰/۲۷



شکل ۵. نقشه پیش‌بینی کربنات کلسیم معادل خاک با استفاده از مدل رگرسیون خطی چندگانه

موسوی و همکاران در پژوهشی (۲۰) مدل‌سازی تغییرات عمقی کربنات کلسیم معادل خاک را با استفاده از سه الگوریتم یادگیری ماشین، جنگل تصادفی، رگرسیون درخت تصمیم و

در شمال غرب حوضه مورد مطالعه، از کوه به طرف اراضی پای کوهی که شیب حوضه کاهش می‌یابد، از مقدار کربنات کلسیم معادل نیز کاسته می‌شود.

پیچیده بین مواد مادری و توپوگرافی را با استفاده از مدل رقومی ارتفاع و نقشه‌های سنتی ارزیابی خاک بررسی کردند که مدل جنگل تصادفی را به‌عنوان یک رویکرد یادگیری ماشین و داده‌کاوی اثربخش اعلام کردند که با نتایج ما در این بررسی در یک راستا و همسو است.

نتیجه‌گیری کلی

به‌منظور برآورد کربنات کلسیم معادل خاک، در مطالعه حاضر دو حالت مختلف بررسی شد. در حالت اول، مدل‌های شبکه عصبی مصنوعی، تحلیل درخت تصمیم و رگرسیون خطی چندگانه برای پیش‌بینی استفاده شدند. همچنین برای ترکیب نتایج مدل‌ها، از مدل نزدیک‌ترین همسایه استفاده شد. نتایج این پیش‌بینی‌ها با استفاده از روش ارزیابی کافلد تصادفی ۵ ارزیابی شدند. در مرحله دوم، مدل‌های تحلیل درخت تصمیم، شبکه عصبی مصنوعی، نزدیک‌ترین همسایه و جنگل تصادفی برای پیش‌بینی به‌کار رفتند. همچنین به‌منظور ترکیب نتایج مدل‌ها در این حالت، از روش رگرسیون خطی چندگانه استفاده شد. با استفاده از روش ارزیابی کافلد مکانی ۱۰، نتایج پیش‌بینی ارزیابی شدند. نتایج به‌طور کلی نشان داد که:

۱) بر اساس روش اعتبارسنجی کافلد ۱۰ مکانی، بهترین مدل برای پیش‌بینی تمام ویژگی‌های خاک، رگرسیون خطی چندگانه است. برای همه ویژگی‌های خاک، مدل جنگل تصادفی، از نظر قدرت پیش‌بینی، رتبه دوم را به خود اختصاص می‌دهد. به‌طورکلی، نزدیک‌ترین همسایه K، کمترین عملکرد را در پیش‌بینی کربنات کلسیم معادل خاک دارا بوده است.

۲) بر اساس روش اعتبارسنجی کافلد ۵ تصادفی، دقت و صحت مدل نزدیک‌ترین همسایه بهترین عملکرد را در پیش‌بینی کربنات کلسیم معادل خاک دارا بوده است.

۳) به‌طور کلی، بر اساس روش‌های ارزیابی کافلد ۱۰ مکانی و کافلد ۵ تصادفی، می‌توان بیان کرد که روش‌های ترکیبی دارای قابلیت بیشتری هستند. این بدین معنی است که ترکیب نتایج پیش‌بینی سایر مدل‌ها می‌تواند نقشه‌هایی با دقت بیشتر را تولید کند.

نزدیک‌ترین همسایه K در دشت قزوین بررسی کردند که نتایج آن‌ها نشان داد الگوریتم جنگل تصادفی در مقایسه با دو الگوریتم دیگر بیشترین مقدار دقت و کمترین خطا را ارائه کرد که با نتایج این مطالعه هم‌راستا و هم‌جهت است. در میان مدل‌های استفاده‌شده برای پیش‌بینی مقدار کربنات کلسیم معادل در این مطالعه، مدل رگرسیون خطی چندگانه (MLR) با ضریب تعیین ۰/۷۹۶ و ریشه دوم متوسط مربعات خطا ۶/۵۱۴ از بیشترین دقت و مدل جنگل تصادفی با ضریب تعیین ۰/۷۸۰ و ریشه دوم متوسط مربعات خطا ۷/۸۱۵ از نظر قدرت و دقت پیش‌بینی در جایگاه دوم قرار گرفته است.

در یک مطالعه مقایسه‌ای با عنوان نقشه‌برداری رقومی ویژگی‌های خاک با استفاده از یادگیری ماشین در مناطق نیمه‌خشک مرکز ایران (۳۱)، به این نتیجه رسیدند که برای برآورد مقدار کربنات کلسیم معادل خاک پارامترهای سنجش‌ازدور مهم‌ترین متغیر کمکی بودند و مدل‌های جنگل تصادفی و کوبیست بیشترین دقت را در پیش‌بینی مقدار کربنات کلسیم معادل خاک داشتند.

در مطالعه دیگری که به‌منظور پیش‌بینی مکانی برخی ویژگی‌های خاک سطحی با استفاده از مدل‌های درونیابی و یادگیری ماشین در استان کردستان انجام شد که نتایج آن حاکی از برتری مدل جنگل تصادفی در مقایسه با مدل‌های درونیابی کریجینگ معمولی و وزن‌دهی معکوس فاصله در پیش‌بینی CCE بود (۱۹).

همچنین در مطالعه تیموری و همکاران (۲۷)، بهترین مدل برای تخمین کربنات کلسیم معادل، مدل جنگل تصادفی است که ضریب تبیین ۰/۶۱، ریشه میانگین مربعات خطا ۰/۷۲، ریشه میانگین مربعات خطا نرمال‌شده ۰/۲۰، نسبت دامنه بین چارکی ۰/۷۷ و نسبت انحراف عملکرد ۱/۲۷ را دارد که با نتایج این مطالعه مبنی بر برتری مدل جنگل تصادفی نسبت به مدل‌ها و الگوریتم‌های دیگر در پیش‌بینی مقدار کربنات کلسیم معادل خاک در یک راستا است. در مطالعه دیگری هیونگ و همکاران (۱۲) با استفاده از مدل جنگل تصادفی روابط سلسله‌مراتبی

منابع مورد استفاده

References

1. Angiulli, F. 2005. Fast condensed nearest neighbor rule. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany.
2. Arrouays, D., N. McKenzie, J. Hempel, A. Richer, A. de Forges and B. McBratney. 2013. GlobalSoilMap: Basis of the global spatial soil information system, CRC Press, London.
3. Ayoubi, S., R. Taghizadeh, Z. Namazi, A. Zulfaqari and F. Rustae Sadrabadi. 2016. 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* 20 (76): 59-71 (in Farsi).
4. Bagheri Badaghabadi, M., M. H. Salehi, A. Esfandiarpour Borojni, J. Mohammadi, A. Karimi Karubih and N. Tomanian. 2011. Evaluation and investigation of the generalizability of soil-land inferential model (SoLIM) in digital soil mapping using the digital height model and its derivatives. *Journal of Agricultural Sciences and Techniques and Natural Resources, Water and Soil Sciences*. 61:155-165 (in Farsi).
5. Balan, B., S. Mohaghegh and S. Ameri. 1995. State- of- Art- in permeability determination from well log data: Part1- A comparative study, model development. *Society of Petroleum Engineers* 30978: 17-25.
6. Behrens, T., H. Forster, T. Scholten, U. Steninrucken, E. Spies and M. Goldschmitt. 2005. Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science* 168: 21-33.
7. Breiman, L. 2001. Random forests. *Machine learning*, 45: 5-32.
8. Camera, C. Z., J. S. Zomeni, A. M. Noller, I. C. Zissimos, B. Christoforou and A. Bruggeman. 2017. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma* 285: 35-49.
9. Castellini, M., A. M. Stellacci, M. Tomaiuolo and E. Barca. 2019. Spatial variability of soil physical and hydraulic properties in a durum wheat field: An assessment by the BEST-Procedure. *Water* 11(7):1434.
10. Egli M., M. Bösiger, K. Lamorski, C. Sławiński, M. Plötze, G. L. Wiesenber, D. Tikhomirov, A. Musso, S. Y. Hsu and S. Raimondi. 2021. Pedogenesis and carbon sequestration in transformed agricultural soils of Sicily. *Geoderma* 15(402): 115355.
11. Givi, J. 1997. Qualitative assessment of land suitability for agricultural and garden plants, Soil and Water Research Institute. *Technical Magazine* (1015): 100 (in Farsi).
12. Heung, B., C. E. Bulmer and M. G. Schmidt. 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. *Geoderma* 214:141-154.
13. Lotfollahi, L., M. A. Delavar, A. Biswas, M. Jamshidi and R. Taghizadeh-Mehrjardi. 2023. Modeling the spatial variation of calcium carbonate equivalent to depth using machine learning techniques. *Environmental Monitoring and Assessment* 195(5): 607.
14. Maghsodi, Z., M. Rostaminia, M. Faramarzi, A. Keshavarzi, A. Rahmani and S. R. Mousavi. 2020. Digital mapping of soil family class using the machine learning approach (a case study: semi-arid lands in the west of Iran). *Journal of Water and Soil Science* 24(2): 153-163 (in Farsi).
15. Matan, K., A. Bensa, D. Domínguez-Villar and K. Krklec, 2024. Morphology and formation processes of pedogenic carbonates in Mediterranean soils: a case study from the North Dalmatian Plain, Croatia. In: *Abstract Book, 7th Regional Scientific Meeting on Quaternary Geology* 69-70.
16. Matinfar, H. R., Z. Mghsodi, S. R. Mossavi and M. Jalali. 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 (in Farsi).
17. Mehnatkesh, A., S. Ayoubi, A. Jalalian and K. Sahrawat. 2013. Relationships between Soil depth and terrain attributes in a semi arid hilly region in western Iran. *Journal of Mountain Science* 10: 163-172.
18. Mirzaei, S., N. Rasooli and S. Pignatti. 2024. Evaluating the spatial variations of soil salinity and calcium carbonate in marginal lands of Sirjan playa (Iran), using PRISMA hyperspectral imagery. *EGU General Assembly Conference Abstracts* 20310.
19. Mousavi S.R., F. Parsayi, A. Rahmani, M. H. Sedri and M. Kohsar Bostani. 2020. Spatial Prediction Some of the Surface Soil Properties Using Interpolation and Machine Learning Models. *Journal of Soil Management and Sustainable Production* 10(3): 27-49 (in Farsi).
20. Mousavi, S. R., F. Sarmadian, M. Omid and P. Bogaert. 2021. Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Water and Soil* 35(5): 719-734 (in Farsi).
21. Nabiollahi, K., A. Haidari and R. Taghizade Mehrjardi. 2015. Digital mapping of soil texture using regression tree and artificial neural network in Bijar, Kurdistan. *Water and Soil* 28(5): 1025-1036.
22. Nemes, A., W.J. Rawls and Y.A. Pachepsky. 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal* 70: 327-336.

23. Salehi, M. H. and H. Khademi. 2008. Basics of soil mapping. Isfahan University Jihad Publications, Isfahan (in Farsi).
24. Schoeneberger, P. J., D. A. Wysocki and E. C. Benham. 2012. Field book for describing and sampling soils. 3rd Version. National Soil Survey Center, Lincoln.
25. Soil Survey Staff. 2014. Soil Taxonomy: A basic systems of soil classification for making and interpreting soil surveys. 12th Edition. NRCS. USDA.
26. Sys, C. E. Vanranst and J. Debaveye. 1991. Land evaluation. Part I. Principles in land evaluation and crop production calculations. General Administration for Development Cooperation, Brussels.
27. Teimouri Bardyani, S. and F. Sarmadian. 2024. Digital mapping of soil properties (Calcium Carbonate and soil clay percentage) using landsat 8 and Prisma satellite images by the random forest algorithm. *Iranian Journal of Soil and Water Research*, 55(3): 381-399 (in Farsi).
28. Valavi, R., J. Elith, J. José, M. Lahoz and G. Gurutzeta. 2018. Block CV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv* 357798.
29. Wilding, L.P. and L.R. Dress. 1983. Spatial variability and pedology. In: Wilding, L.P. (ed.), *Pedogenesis and Soil Taxonomy: Concepts and Interactions*. Elsevier scientific publishing, New York.
30. Zamanian, K., K. Pustovoytov and Y. Kuzyakov. 2016. Pedogenic carbonates: Forms and formation processes. *Earth-Science Reviews* 157: 1-17.
31. Zeraatpisheh, M., S. Ayoubi, A. Jafari, S. Tajik and P. Finke. 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338: 445-452.
32. Zinck, J.K. 1989. Physiography and soil. Lecture notes for K 6 course. Soil Division. ITC. Enschede, The Netherlands.
33. Zolfaghari, A., S. M. Tirgar, T. Soltani, B. T. Afshari and F. Sarmadian. 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 (in Farsi).