



Digital mapping of soil properties (Calcium Carbonate and soil clay percentage) using landsat 8 and Prisma satellite images by the random forest algorithm

Sajjad Teimuri Bardyani ¹ | Fereydoon Sarmadian ²

1. Department of Soil Science and Engineering, Faculty of Agriculture, University of Tehran, Karaj, Iran. E-mail:

Sajjad.teimuri@ut.ac.ir

2. Corresponding Author, Department of Soil Science and Engineering, Faculty of Agriculture, University of Tehran, Karaj, Iran. E-mail: fsarmad@ut.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Aug. 22, 2023

Revised: Jan. 1, 2024

Accepted: Jan. 11, 2023

Published online: May. 2024

Keywords:

Calcium Carbonate,
Clay Percentage,
PRISMA Hyperspectral Satellite,
Random Forest.

ABSTRACT

Mapping soil properties using hyperspectral and multispectral satellite images, along with statistical approaches, and machine learning models such as Random Forests (RF), has shown great progress in accurately preparing agricultural maps. This study aimed to compare the performance of PRISMA and Landsat 8 images in modeling calcium carbonate and clay percentage using a Random Forest model. Firstly, Surface soil data was collected from Abik region of Qazvin province from October 2020 to October 2022. Furthermore, PRISMA and Landsat 8 spectral datasets were extracted from images downloaded from the websites of these two satellites, and soil reflectance data were obtained. The Random Forest regression model was then calibrated to estimate soil properties. The results of this study showed that the best accuracy in estimating soil characteristics using PRISMA data was obtained by using Auxiliary Variables such as principal components analysis, spectral indices, and indices extracted from the digital elevation model. The use of these three data sets provided the uppermost value for following statistical indices for estimating calcium carbonate and soil clay percentage: coefficient of determination (R^2), and Ratio of Performance to Inter Quartile range (RPIQ), Ratio Performance Deviation (RPD) and the lowest Root Mean Squared Error (RMSE) and Normalized Root Mean Squared Error (NRMSE). The best model for estimating clay percentage, using the Random Forest model and statistical indices, had an R^2 of 0.90, RMSE of 4.11, NRMSE of 0.18, RPIQ of 0.95, and RPD of 2.29. The best model for estimating calcium carbonate, using the Random Forest model and statistical indices, had an R^2 of 0.62, RMSE of 0.72, NRMSE of 0.20, RPIQ of 0.77, and RPD of 1.27. The results supported the expectation of the good ability of the PRISMA imager to estimate surface soil properties.

Cite this article: Teimuri Bardyani, S. & Sarmadian, F. (2024). Digital Mapping of Soil Properties Using Landsat 8 and PRISMA Satellite Images by the Random Forest Algorithm, *Iranian Journal of Soil and Water Research*, 55 (3), 381-399.

<https://doi.org/10.22059/ijswr.2024.363941.669558>

© The Author(s).

Publisher: The University of Tehran Press.



DOI: <https://doi.org/10.22059/ijswr.2024.363941.669558>

نقشه برداری رقومی کربنات کلسیم معادل و درصد رس خاک با استفاده از تصاویر ماهواره‌های لندست ۸ و پریسما توسط الگوریتم جنگل تصادفی

سجاد تیموری بردیانی^۱ | فریدون سرمیدیان^۲^۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران، رایانامه: Sajjad.timuri@ut.ac.ir^۲. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران. رایانامه: fsarmad@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۵/۳۱

تاریخ بازنگری: ۱۴۰۲/۱۰/۱۱

تاریخ پذیرش: ۱۴۰۲/۱۰/۲۱

تاریخ انتشار: خرداد ۱۴۰۳

واژه‌های کلیدی:

جنگل تصادفی،

درصد رس،

کربنات کلسیم معادل،

ماهواره ابرطیفی پریسما.

نقشه برداری خصوصیات خاک با استفاده از تصاویر ماهواره‌ای ابرطیفی و چند طیفی در کنار رویکردهای آماری و با استفاده از مدل‌های یادگیری ماشین از جمله جنگل‌های تصادفی پیشرفت زیادی در دقت و صحت نقشه‌های تهیه شده داشته است. این تحقیق برای بررسی عملکرد تصاویر پریسما و لندست ۸ در مدل‌سازی کربنات کلسیم معادل و درصد رس با مدل جنگل تصادفی، در بخشی از اراضی شهرستان آبیگ استان قزوین از مهرماه سال ۱۳۹۹ تا مهرماه ۱۴۰۱ انجام شده است. در ابتدا، با استفاده از ۲۲۹ داده که از خاک سطحی جمع‌آوری شده در منطقه آبیگ استان قزوین به مساحت ۶۰ هزار هکتار اندازه‌گیری شد، در مرحله بعد مجموعه داده‌های طیفی دو ماهواره پریسما و لندست ۸، استخراج و داده‌های بازتاب خاک به دست آمدند. در این تحقیق از شاخص‌های طیفی، شاخص‌های مدل رقومی ارتفاع و تجزیه مولفه‌های اصلی به عنوان متغیر کمکی استفاده شد. در مرحله بعدی، مدل رگرسیون جنگل تصادفی جهت تخمین ویژگی‌های خاک با استفاده از ۸۰٪ از داده‌ها آموزش داده شد و از ۲۰٪ داده‌ها برای آزمون مدل استفاده شد. نتایج نشان داد که بهترین دقت در بازایی ویژگی‌های خاک سطحی توسط داده‌های پریسما، با استفاده از مجموعه داده‌های کمکی تجزیه مولفه‌های اصلی، شاخص‌های طیفی و شاخص‌های مستخرج از مدل رقومی ارتفاع به دست آمد. به طور دقیق‌تر، استفاده از این سه دسته داده، بیشترین ضریب تبیین، انحراف پیش‌بینی باقی‌مانده و نسبت عملکرد به فاصله بین چارکی و کمترین ریشه میانگین مربعات خطا و ریشه میانگین مربعات خطا نرمال شده را برای تخمین کربنات کلسیم معادل و درصد رس خاک نشان داد. بهترین مدل برای تخمین درصد رس، با مدل جنگل‌های تصادفی، شاخص‌های آماری (ضریب تبیین: ۰/۹۰؛ ریشه میانگین مربعات خطا: ۴/۹۱؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۳؛ نسبت دامنه بین چارکی: ۰/۸؛ نسبت انحراف عملکرد: ۲/۲۹) و بهترین مدل برای تخمین کربنات کلسیم معادل، با مدل جنگل‌های تصادفی، شاخص‌های آماری (ضریب تبیین: ۰/۶۱؛ ریشه میانگین مربعات خطا: ۰/۷۲؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۰؛ نسبت دامنه بین چارکی: ۰/۷۷؛ نسبت انحراف عملکرد: ۱/۲۷) به دست آمد.

استناد: تیموری بردیانی، سجاد و سرمیدیان، فریدون (۱۴۰۳). نقشه برداری رقومی کربنات کلسیم معادل و درصد رس خاک با استفاده از تصاویر ماهواره‌های لندست ۸ و

پریسما توسط الگوریتم جنگل تصادفی، مجله تحقیقات آب و خاک ایران، ۵۵ (۳)، ۳۸۱-۳۹۹.

<https://doi.org/10.22059/ijswr.2024.363941.669558>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2024.363941.669558>

مقدمه

کربنات کلسیم معادل یکی از اصلی ترین کانی های کربناته تشکیل دهنده خاک است و خصوصیات فیزیکی، شیمیایی و بیولوژیکی خاکها به توزیع و مقدار کربناتها در خاک وابسته هستند. کربناتها تأثیر مهمی بر روی سطح pH و ماده آلی خاک دارند و می توانند بر زیست فراهی فلزات سنگین تأثیرگذار باشند. (Wang et al., 2015; Martin., 2017) مواد معدنی خاک، به خصوص ذرات رس، خواص فیزیکی شیمیایی خاک را واپایش می کنند. رسها در ذخیره پتاسیم برای گیاهان و میکروارگانیسمهای خاک، جذب فلزات و آب، ظرفیت تبادل کاتیونی، جذب آلاینده های آلی و معدنی، کاهش تحرک آلاینده ها، حفاظت از خاک و سلامت محیط نقش دارند (Mousavifard et al., 2013).

از آنجاکه دانش و درک ما از مشکلات و استفاده درست از خاک بدون شک توسط نقشه برداری و ارزیابی خاک معنی پیدا می کند نقشه های خاک ابزاری برای ارزیابی شرایط فعلی و آتی خاک خواهد بود (Carre et al., 2007). تصاویر سنجنش از دور، پایگاه داده مهمی را برای نقشه برداری خاک فراهم می کند، به موقع بودن و دقت شناسایی نوع خاک را تا حد زیادی بهبود می بخشد و به یک منبع مهم برای به دست آوردن اطلاعات نوع خاک در مقیاس بزرگ تبدیل شده است (Duan et al. 2022). نقشه رقومی خاک به طور گسترده ای برای ارزیابی توزیع مکانی ویژگی های خاک، از جمله کربنات کلسیم معادل و درصد نسبی ذرات خاک استفاده می شود (رحمانی و همکاران، ۱۳۹۹: Hengli et al, 2021; Zeraatpisheh, 2019).

حسگرهای تصویربرداری جزء اصلی هر سیستم سنجنش از راه دور هستند و دارای طیف گسترده ای از وضوح های مکانی، زمانی و طیفی هستند. (Toth and Jozkow., 2016) تصویربرداری چند طیفی با جمع آوری سیگنال های طیفی در چند باند مجزا عمل می کند که هر یک از باندها محدوده طیفی وسیعی از دهها تا صدها نانومتر را در بر می گیرد. تکنولوژی تصویربرداری فراطیفی، قادر است سیگنال های طیفی را در یک سری کانال پیوسته با پهنای باند طیفی باریک (مثلاً ۱۰ نانومتر) تشخیص دهد. به این ترتیب، این تکنولوژی قادر است ویژگی های طیفی در مقیاس دقیق را که ممکن است در تصویربرداری چندطیفی به ثبت نشوند را ضبط نماید. (Sahoo et al., 2015). از سال ۱۹۷۲، داده های لندست به عنوان منبع منحصربه فردی از اطلاعات سنجنش از دور برای محققین در حوزه های مختلف کاربرد دارند. با اضافه شدن لندست ۹، برنامه لندست در تصویربرداری از زمین به بیش از نیم قرن می رسد. سنجنده تصویرگر زمین عملیاتی ۱ در لندست ۸ دارای ۱۱ باند در بخش های مرئی مادون قرمز^۲، مادون قرمز نزدیک^۳ و مادون قرمز موج کوتاه^۴ است. همچنین دارای دو حسگر تصویربرداری زمین و حسگر مادون قرمز حرارتی است. تصاویر لندست ۸ دارای وضوح مکانی ۳۰ متری و پانکروماتیک ۱۵ متری در امتداد یک نوار ۱۸۵ کیلومتری هستند. (Vaughn Ihlen., 2019)

ماهواره پریسما، متعلق به آژانس فضایی ایتالیا^۵، دارای سنجنده الکترواپتیکی ابرطیفی است. این ماهواره توانایی ثبت تصاویر با وضوح مکانی ۳۰ متر در محدوده مرئی مادون قرمز و مادون قرمز موج کوتاه و ۵ متر در محدوده مرئی وضوح طیفی کمتر از ۱۲ نانومتر را دارد. پریسما از ۲۳۹ باند طیفی بین ۴۰۰ تا ۲۵۰۰ نانومتر، ۶۶ باند در مرئی مادون قرمز و ۱۷۳ باند در مادون قرمز موج کوتاه استفاده می کند. همچنین دارای دوربین پانکروماتیک با وضوح ۵ متر است. مطالعات (ASI., 2020; Karrayet al., 2023) بیانگر دقت خوب ماهواره ابرطیفی پریسمادر برآورد خصوصیات خاک است. (Gasmi et al., 2022)

پژوهش حاضر به منظور پهنه بندی و مقایسه خصوصیات خاک با استفاده از داده های ماهواره ابرطیفی پریسما و لندست ۸ انجام شده است. در این پژوهش، دو هدف اصلی مورد بررسی قرار گرفته اند. هدف اول، مقایسه دو ماهواره لندست ۸ و پریسما است که مشخص شود کدام یکی از دو ماهواره پریسما و لندست ۸ در پهنه بندی خصوصیات خاک عملکرد بهتری دارند. که با مقایسه دقت و صحت نتایج حاصل از این دو ماهواره، سعی شده است تا بهترین ماهواره برای این منظور شناسایی شود. هدف دوم تعیین بهترین ترکیب ورودی برای تهیه نقشه خصوصیات خاک است که به وسیله مقایسه ورودی های مختلف به مدل سعی بر پیدا کردن بهترین ترکیب ورودی برای مدلسازی خصوصیات شد.

1.Operational Land Imager
2.The visible and near-infrared, VNIR
3.Near-infrared, NIR
4.Short-Wave Infrared, SWIR
5.Italian Space Agency

پیشینه پژوهش

تحقیقات انجام شده از مجموعه‌ای از روش‌های تحلیلی، از جمله یادگیری ماشین مانند جنگل تصادفی برای نقشه‌برداری ذرات رس با استفاده از تصاویر ابرطیفی و چند طیفی نشان داد که این مدل عملکرد خوبی در مدل‌سازی مقادیر رس خاک دارد (Akpa et al., 2014); همچنین عملکرد مدل جنگل تصادفی و ماهواره لندست در مطالعاتی که در مورد مقادیر کربنات کلسیم معادل خاک انجام گرفته نشان می‌دهند که مدل جنگل تصادفی و اطلاعات ماهواره لندست به خوبی برای مدل‌سازی مقادیر کربنات کلسیم معادل خاک کارایی دارند (رحمانی و همکاران، ۱۳۹۸، موسوی و همکاران، ۱۴۰۰).

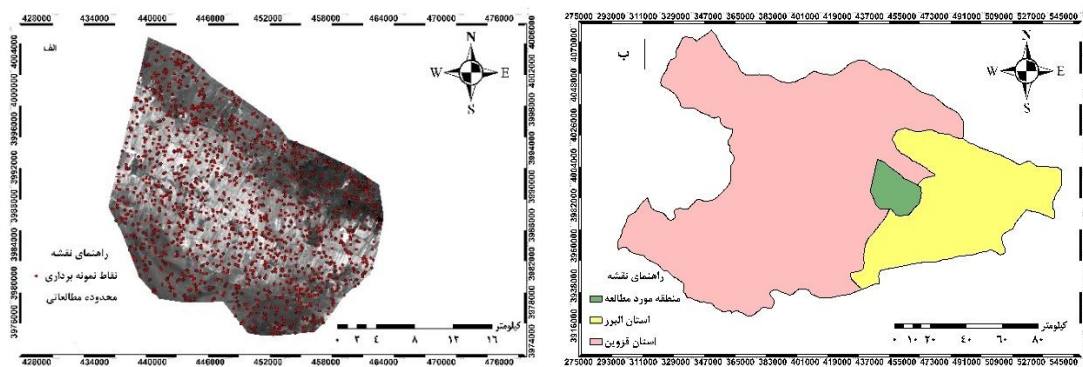
(Casa et al., 2023) در تحقیقی که انجام دادند، به این نتیجه رسیدند که تصاویر ماهواره پریسما توانایی قابل قبولی برای نقشه‌برداری خصوصیات خاک دارند. نتایج تحقیق (Mazid et al., 2022) نشان داد که تصاویر ابرطیفی پریسما قابلیت تخمین رطوبت خاک را داشته و می‌توانند به عنوان یک ابزار مؤثر در مطالعات رطوبت خاک استفاده شوند.

جنگل تصادفی یک الگوریتم یادگیری ماشینی است که با استفاده از چندین درخت تصادفی به عنوان یک مجموعه کار می‌کند. هر درخت به صورت مستقل از دیگر درخت‌ها ساخته شده و با استفاده از روش جداسازی و مرتب کردن، داده‌های ورودی را به دسته‌های مختلف تقسیم می‌کند (Cutler et al., 2007). در نهایت، با ترکیب خروجی‌های تمام درخت‌ها، الگوریتم پاسخ نهایی را ارائه می‌دهد. الگوریتم جنگل تصادفی بر اساس مجموعه‌ای از درخت‌های تصمیم‌گیری ساخته می‌شود که با استفاده از یک نمونه بوت‌استرپ^۱ تصادفی از داده‌های آموزشی، برازش می‌شوند. برای افزایش تصادفی بودن مدل، یک زیرمجموعه تصادفی از متغیرهای پیش‌بینی کننده برای ایجاد هر قانون تقسیم‌بندی در درخت‌های تصمیم‌گیری جداگانه انتخاب می‌شود. الگوریتم جنگل تصادفی یک مدل پیش‌بینی کننده قوی است که برای کاهش اثرات بیش‌ازحد برازش مدل در نظر گرفته شده است (Zolfaghari Nia., 2022).

روش‌شناسی پژوهش

منطقه مورد مطالعه

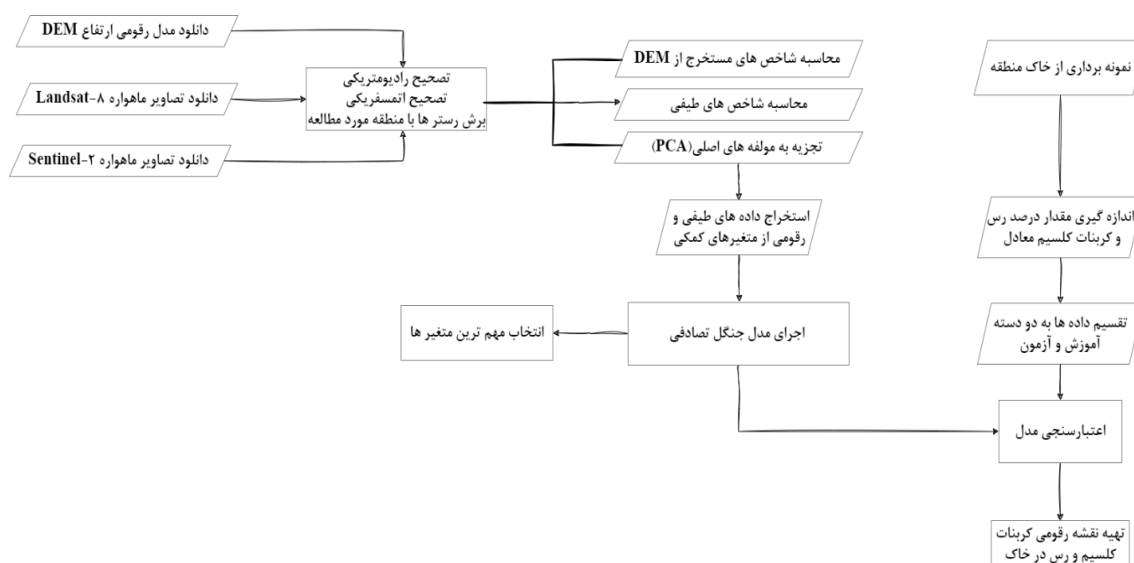
منطقه مورد مطالعه، بخشی از اراضی شهرستان آبیک است که در مابین استان البرز و استان قزوین واقع شده و در زون 39 با عرض جغرافیایی ۳۹۷۳۰۰۰ تا ۴۰۰۷۲۰۰ متر و طول جغرافیایی ۴۲۸۵۰۰ تا ۴۶۵۳۰۰ متر قرار دارد. این منطقه تقریباً حدود ۶۰۰۰۰ هکتار را پوشش می‌دهد (شکل ۱). میانگین بارندگی سالیانه در این منطقه ۲۹۶ میلی‌متر و متوسط دمای سالیانه ۱۴٫۸ درجه سانتیگراد است. رژیم‌های رطوبتی و حرارتی خاک منطقه به ترتیب زیریک و تریمیک طبقه‌بندی شده‌اند (رحمانی و همکاران، ۱۴۰۱). همچنین، این منطقه از لحاظ پستی و بلندی، بیشتر مسطح است و بلندترین ارتفاع آن ۱۳۸۰ متر و کمترین ارتفاع آن ۱۱۵۰ متر است.



شکل ۱. الف: منطقه مورد مطالعه در استان‌های قزوین و البرز ب: موقعیت نمونه‌ها در منطقه مورد مطالعه

برای بررسی عملکرد تصاویر پریسما و لندست ۸ در مدل‌سازی کربنات کلسیم و درصد رس با مدل جنگل تصادفی، در بخشی از اراضی شهرستان آبیک استان قزوین انجام شده است. تعداد ۲۲۹ نمونه سطحی در منطقه مورد مطالعه برداشت و مختصات مکانی نمونه‌ها ثبت شد. سپس، این نمونه‌ها به آزمایشگاه منتقل شده و خصوصیات کربنات کلسیم معادل (روش تیتراسیون) و درصد رس (روش هیدرومتری) آن‌ها اندازه‌گیری شد. داده‌های کربنات کلسیم معادل دارای چولگی ۱/۴۶ و کشیدگی ۲/۹۵، درصد رس دارای چولگی ۰/۴۲ و کشیدگی

۰/۶۱ است. پس از بررسی کشیدگی و چولگی داده‌ها تصمیم بر نرمال‌سازی داده‌ها با روش جذرگیری برای مقادیر کرنات کلسیم معادل و داده‌های درصد رس بدون نرمال‌سازی مورد استفاده قرار گرفت. شمای کلی روش تحقیق در (شکل ۲) آورده شده است.



شکل ۲. فلوجارت (نمای کلی) روش تحقیق

تصاویر ماهواره‌ای

تصویر ماهواره‌ای پرسماز وبگاه ماهواره پرسمابه نشانی Prisma.asi.it قابل دسترسی و داتلود است تصاویر ماهواره پرسمادر قالب دو سطح L1 و L2 ارائه می‌شوند که تصاویر سطح L1 شامل محصولات ابرطیفی و پانکروماتیک کالیبره شده در بالای جو، شامل فیلتر ابری، فیلتر آفتابی، فیلتر طبقه‌بندی کلی، مشخصات تصویر اعم از زمان و مختصات تصویر و داده‌های کالیبراسیون است. در سطح L2 ما داری سه نوع تصویر هستیم. این تصاویر شامل L2b، L2c، L2d هستند. (ASI., 2020)

در این تحقیق از تصاویر سطح L2d در تاریخ ۱۳۹۹/۵/۱۸ که دارای تمامی تصحیحات لازم برای نقشه‌برداری بودند استفاده شد. همچنین برای ماهواره لندست ۸ نیز از نزدیک‌ترین تصویر به تاریخ تصویربرداری ماهواره پرسماو نمونه‌برداری خاک استفاده شد. مدل رقومی ارتفاع ۳۰^۲ متر نیز از ماهواره آوس ALOS از وبگاه سازمان زمین شناسی ایالات متحده داتلود شد.

پیش پردازش تصاویر

عملیات پیش پردازش که گاهی اوقات به‌عنوان بازیابی و تصحیح تصویر شناخته می‌شود، برای تصحیح اعوجاج‌های رادیومتری و هندسی خاص حسگر و پلت فرم داده‌ها در نظر گرفته شده است. اصلاحات رادیومتری ممکن است به دلیل تغییرات در روشنایی عرصه، هندسه مشاهده عرصه، شرایط جوی، نویز و پاسخ حسگر ضروری باشد. هر یک از تصحیحات مربوط به عوامل نام‌برده شده بسته به حسگر و پلتفرم خاصی که برای به‌دست‌آوردن داده‌ها و شرایط در حین جمع‌آوری داده استفاده می‌شود، متفاوت خواهد بود. همچنین، ممکن است برای تسهیل مقایسه بین داده‌ها، تبدیل و یا کالیبره کردن داده‌ها به واحدهای تابشی یا بازتابی شناخته شده مطلوب باشد (Tu et al., 2018; Lucieer et al., 2014).

برای تصاویر ابرطیفی مبتنی بر ماهواره و فضاپیما، تصحیح هندسی و کالیبراسیون عموماً توسط ارائه‌دهندگان تصاویر انجام می‌شود و اصلاحات رادیومتری و جوی برای ماهواره لندست ۸ را با استفاده از دستور FLAASH و پیروی از مراحل استاندارد پردازش تصویر موجود در نرم‌افزار سنچس از دور (ENVI 5.6) انجام شد (Lucieer et al., 2014).

پردازش حذف باندهای دارای اختلال یک روش ضروری در داده‌های سنچس از دور به‌ویژه در مجموعه داده‌های ابرطیفی است. در حال حاضر، بیشترین روش‌های مورد استفاده برای حذف این باندها مربوط به بازرسی‌های بصری (Du and Yang., 2008) و ویژگی‌های

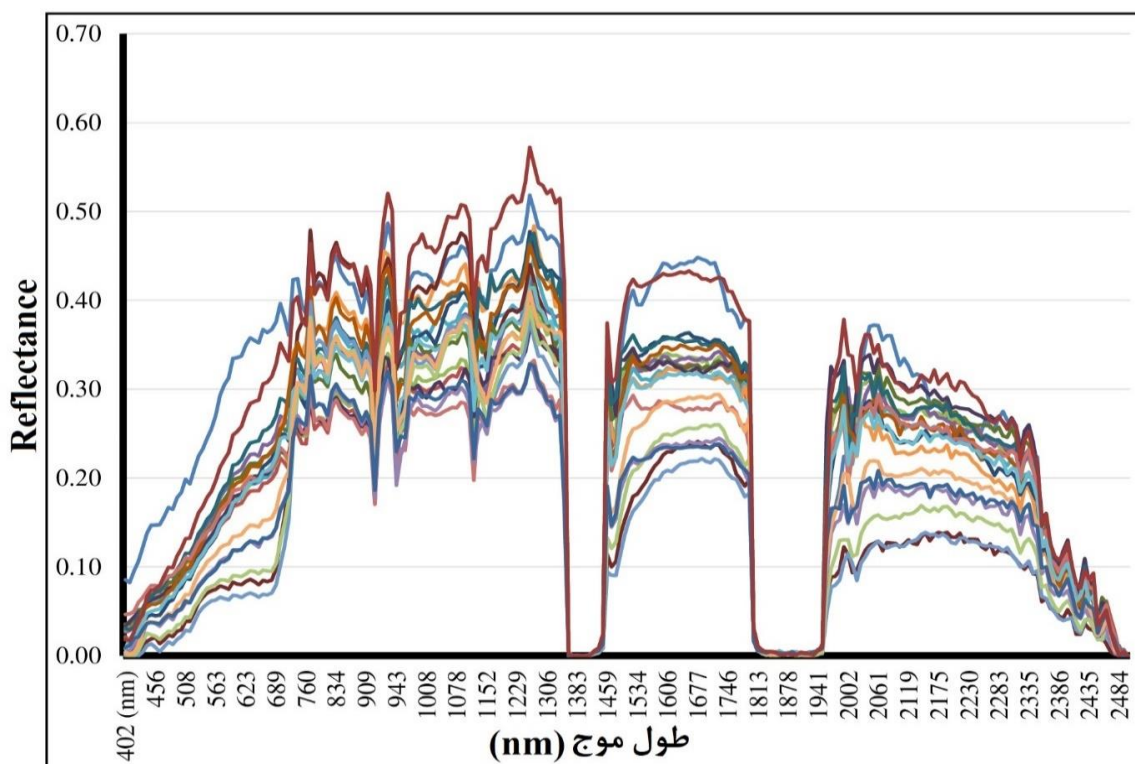
حسگر (Sun and Geng., 2014) است. باندهای دارای نویز یا خطا که در تصویر بودند به صورت بصری تشخیص داده و حذف شدند بعد از انجام این عملیات تعداد ۴۰ باند حذف شد.

در مجموعه داده‌های پریسمادو منبع اصلی نویز وجود دارد: وجود خطوط گمشده عمودی^۱ و افقی^۲ در باندهای متعدد و وجود "نویز گاوسی مانند" که به شدت قسمت انتهایی مجموعه داده مادون قرمز موج کوتاه را خراب می‌کند (Alicandro et al., 2022). همچنین تمامی باندهای تصاویر را جهت مطابقت بیشتر با خصوصیات سنجنده ماهواره پریسما بیا اندازه اصلی ماهواره (۳۰ متر) استفاده کردیم. محدوده‌های طیفی حذف شده در (جدول ۱) آورده شد. تمام این مراحل در نرم‌افزار ENVI 5.6 انجام شد.

جدول ۱. محدوده‌های طیفی حذف شده از مجموعه باندهای ماهواره پریسما (Tagliabue et al., 2022)

محدوده طیفی (نانومتر)	دلیل حذف
۱۵۱۰-۱۳۵۰	محدوده جذب آب اتمسفر
۲۰۰۰-۱۸۰۰	محدوده جذب آب اتمسفر
۲۵۰۰-۲۳۲۰	نسبت سیگنال به نویز بالا
۹۸۰-۹۴۰	همپوشانی دو محدوده مادون قرمز موج کوتاه و مرئی مادون قرمز

ماهواره ابرطیفی پریسمابه دلیل تعداد بالای باند نیاز به برخی از عملیات کاهش ابعاد داشت؛ زیرا وارد کردن تمام باندها به مدل باعث خطای بیش برآزش مدل می‌شد؛ از روش‌های تجزیه مولفه‌های اصلی^۳ و بیشترین همبستگی برای کاهش ابعاد تصویر استفاده شد. تمام منحنی‌های مشخصه خاک در نمونه‌های مختلف مورد تجزیه و تحلیل قرار گرفت (شکل ۳). بازتاب طیفی خاک با افزایش طول موج و با نوسانات مشخص افزایش یافت. تصاویر ماهواره پریسمادر دو محدوده ۱۳۳۵ تا ۱۴۵۵ نانومتر و ۱۸۰۰ تا ۱۹۶۰ نانومتر بازتابی ندارند و این عدم بازتاب مربوط به ناحیه جذب طیفی توسط بخار آب موجود در جو است.



شکل ۳. بازتاب فراطیفی خاک بر روی زمین در نمونه‌های خاک منطقه مورد مطالعه

در این تحقیق همچنین از شاخص‌های طیفی و شاخص‌های مشتق شده از مدل رقومی ارتفاع برای افزایش دقت نقشه‌ها استفاده شد. پس از انتخاب ورودی، مدل‌سازی انجام و نتایج حاصل ثبت شد. ورودی‌های مدل به شرح زیر بودند:

شاخص‌های طیفی (Spectral Index)

تجزیه مولفه‌های اصلی (PCA)

شاخص‌های طیفی و تحلیل مولفه‌های اصلی (Spectral Index + PCA)

شاخص‌های طیفی و شاخص‌های مستخرج از مدل رقومی ارتفاع (Spectral Index + DEM Index)

تجزیه مولفه‌های اصلی و شاخص‌های مستخرج از مدل رقومی ارتفاع (Spectral Index + PCA)

تجزیه مولفه‌های اصلی (PCA)، شاخص طیفی، شاخص‌های مشتق شده از مدل رقومی ارتفاع (Spectral Index + DEM Index)

(+ PCA)

متغیرهای کمکی

به طور معمول، متغیرهای محیطی، لایه‌های رستری جامع و جغرافیایی هستند، از جمله نقشه‌های زمین‌شناسی و مدل‌های رقومی ارتفاع، تصاویر ماهواره‌ای و مشتقات مدل‌های رقومی ارتفاع را شامل می‌شوند و هیچ پیش‌نیازی در مورد نوع مدلی که ارتباط بین این متغیرها را مدل‌سازی کند وجود ندارد (Grunwald., 2010). در اصل متغیرهای کمکی بیان‌کننده پارامترهای دخیل در فرایند خاک‌سازی هستند. در تحقیق حاضر، از سه منبع اصلی اطلاعات مکانی شامل شاخص‌های مستخرج از تصاویر سنجش از دور (RS)، مدل رقومی ارتفاع و تجزیه به مولفه‌های اصلی به‌عنوان متغیرهای کمکی استفاده شد. به طور خاص، از ماهواره شامل پریسما (قدرت تفکیک ۳۰ متر) و لندست ۸ (قدرت تفکیک ۳۰ متر) استفاده شد. علاوه بر این، شاخص‌های طیفی (به‌عنوان مثال، NDVI؛ شاخص گیاهی تفاوت نرمال شده) معیارهایی هستند که برای تحلیل داده‌های طیفی تصاویر ماهواره‌ای استفاده می‌شوند. این شاخص‌ها با استفاده از ترکیب بین دو یا چند باند مختلف در تصویر، ویژگی‌های زمینی مختلف را تحلیل می‌کنند (Bouzekri et al., 2015; Mishra et al., 2022). از این دو تصویر سنجش از دوری محاسبه شد. فرمول و مراجع برای برآورد هر شاخص سنجش از دوری در (جدول ۲) گزارش شده است.

در استفاده از تصاویر ماهواره‌ای برای مدل‌سازی خصوصیات خاک، برخی از متغیرهای مورد استفاده ممکن است دارای اهمیت کمتری باشند و این موضوع می‌تواند زمان مدل‌سازی را افزایش دهد. جهت تحلیل تنوع فضایی تصاویر ابرطیفی، ما از روش تجزیه و تحلیل مؤلفه‌های اصلی استفاده کردیم. این روش به ما این امکان را می‌دهد که مجموعه بزرگی از متغیرهای همبسته را با استفاده از تعداد کمتری از متغیرهای معرف (مؤلفه‌های اصلی) خلاصه کنیم. مؤلفه‌های اصلی توانایی توصیف بیشترین تنوع را در مجموعه داده اصلی دارند. با استفاده از تجزیه و تحلیل مؤلفه‌های اصلی، می‌توان مقدار زیادی از متغیرهای پیش‌بینی‌کننده مرتبط (مانند داده‌های ابرطیفی) را با تعداد کمتری از مؤلفه‌های اصلی شرح داد. این خصوصیت از تجزیه و تحلیل مؤلفه‌های اصلی زمانی مفید است که تعداد زیادی از متغیرهای پیش‌بینی‌کننده مرتبط وجود دارد (Diek and De jong., 2016; Fang et al., 2022). در این تحقیق بعد از انجام تحلیل مؤلفه‌های اصلی برای ماهواره پریسما در هر دو محدوده طیفی مادون قرمز موج کوتاه و مرئی مادون قرمز تعداد ۳، تجزیه و تحلیل مؤلفه‌های اصلی اول که بیشتر از ۹۸ درصد اطلاعات را در خود داشتند استفاده شد. برای ماهواره لندست ۸ نیز از ۵، تجزیه و تحلیل مؤلفه‌های اصلی اول که بیشتر از ۹۸ درصد اطلاعات را در خود داشتند استفاده شد.

تعداد ۲۰ شاخص در این تحقیق از مدل رقومی ارتفاع به‌وسیله نرم‌افزار SAGA GIS 8.0.1 استخراج شد. در بین این شاخص‌هایی که دارای همبستگی بالایی با داده‌های آزمایشگاهی بودن انتخاب و در مدل‌سازی به کار برده شدند. این شاخص‌ها در (جدول ۳) آورده شده‌اند. علاوه بر این شاخص‌های از مدل رقومی زمین^۱ نیز استفاده شد.

مدل‌سازی

در مدیریت مجموعه داده‌های پیچیده خاک جهت افزایش دقت در برآورد خصوصیات خاک، انتخاب یک استراتژی مدل‌سازی مناسب بسیار حائز اهمیت است. از آنجایی که تکنیک‌های یادگیری ماشینی، قادر به مدل‌کردن روابط غیرخطی بین متغیرهای کمکی و متغیرهای هدف خاک هستند، بنابراین می‌توان از آن‌ها برای این منظور استفاده کرد. استفاده از تکنیک‌های یادگیری ماشینی، می‌تواند منجر به افزایش دقت و صحت در برآورد خصوصیات خاک شود (Padarian et al., 2020; Khaledian and Miller., 2020). مدل جنگل تصادفی در

سراسر جهان برای پیش‌بینی پارامترهای خاک استفاده شده است و پیش‌بینی قابل قبولی به نسبت سایر روش‌ها از خود در نقشه‌برداری خواص خاک را نشان داده‌است (Zolfaghari Nia et al., 2022). جنگل تصادفی شامل چندین درخت تصمیم مستقل است که به صورت هم‌زمان آموزش داده می‌شوند. ورودی الگوریتم، متغیرهای کمکی محیطی هستند که در هر درخت به‌عنوان متغیرهای توضیحی برای تقسیم گره‌های تصمیم استفاده می‌شوند. در گره‌های ریشه، متغیرهایی با اهمیت بیشتر استفاده می‌شوند. برای تعیین اهمیت متغیرها، از روش عمق حداقل استفاده می‌شود که بر اساس موقعیت متغیر در درخت‌ها اهمیت آن‌ها را تعیین می‌کند. در مسائل رگرسیون، برای تخمین متغیر هدف، برآورد میانگین خروجی درخت‌ها استفاده می‌شود و در مسائل طبقه‌بندی، کلاس با اکثریت رای به‌عنوان پیش‌بینی نهایی در نظر گرفته می‌شود. استفاده از الگوریتم جنگل تصادفی، امکان تعیین اهمیت نسبی متغیرهای کمکی محیطی را به شکلی ساده فراهم می‌کند (Ishwaran and Kogalur., 2010; Mallah et al., 2022). در این تحقیق نیز اهمیت پارامترها توسط خود مدل تعیین شد. مدلسازی‌های انجام شده در این تحقیق با استفاده از نرم افزار R و کتابخانه‌های H2O، caret، raster انجام شده است.

جدول ۲. شاخص‌های طیفی استفاده شده برای ماهواره لندست-۸ و پریسما

نام شاخص	پریسما		لندست-۸	
	فرمول	منبع	فرمول	منبع
DVI	$\frac{R_{800}}{R_{680}}$	Galvao et al., 2005	$\frac{\text{Near_Infrared}}{\text{red}}$	Pu et al., 2008
EVI	$2.5 \times \frac{R_{1094} - R_{700}}{(R_{1094} + 6R_{700} - 7.5R_{447}) + 1}$	Galvao et al., 2011	$2.5 \times \frac{\text{Near_Infrared} - \text{Red}}{(\text{Near_Infrared} + 6\text{Red} - 7.5\text{Blue}) + 1}$	Li et al., 2018
GNDVI	$\frac{R_{1094} - R_{560}}{R_{1094} + R_{560}}$	Tan et al., 2013	$\frac{\text{Near_Infrared} - \text{Green}}{\text{Near_Infrared} + \text{Green}}$	Lymburn et al., 2000
MNDVI	$\frac{R_{1094} - R_{2200}}{R_{1094} + R_{2200}}$	Colovic et al., 2022	$\frac{\text{Near_Infrared} - \text{SWIR}_2}{\text{Near_Infrared} + \text{SWIR}_2}$	Jurgens., 1997
NDVI	$\frac{R_{800} - R_{670}}{R_{800} + R_{670}}$	Inoue et al., 2008	$\frac{\text{Near_Infrared} - \text{Red}}{\text{Near_Infrared} + \text{Red}}$	Roy et al., 2016
OSAVI	$\frac{(1 + 0.16) * (R_{800} - R_{670})}{(R_{800} + R_{670} + 0.16)}$	Wu et al., 2008	$\frac{(1 + 0.16) * (\text{Near_Infrared} - \text{Red})}{(\text{Near_Infrared} + \text{Red} + 0.16)}$	Anderson et al., 2004
RDVI	$\frac{R_{800} - R_{670}}{\sqrt{R_{800} + R_{670}}}$	Haboudane et al., 2004	$\frac{\text{Near_Infrared} - \text{Red}}{\sqrt{\text{Near_Infrared} + \text{Red}}}$	Pôças et al., 2015
RVI	$\frac{R_{800}}{R_{670}}$	Broge et al., 2001	$\frac{\text{Near_Infrared}}{\text{Red}}$	Lymburn et al., 2000
SAVI	$\frac{(1 + L) * (R_{800} - R_{670})}{(R_{800} + R_{670} + L)}$	Haboudane et al., 2004	$(1 + L) \times \frac{\text{Near_Infrared} - \text{Red}}{\text{Near_Infrared} + \text{Red} + L}$	Pôças et al., 2015
Brightness	$0.3037(R_{487}) + 0.2793(R_{560}) + 0.4743(R_{660}) + 0.5585(R_{833}) + 0.5082(R_{1650}) + 0.1863(R_{2213})$	Thenkaibil et al., 2002	$0.3029\text{Blue} + 0.2786\text{Green} + 0.4733\text{Red} + 0.5599\text{Near_Infrared} + 0.508\text{SWIR}_1 + 0.1872\text{SWIR}_2$	Baig et al., 2014
Greenness	$-0.2848(R_{487}) - 0.2435(R_{560}) - 0.5436(R_{660}) + 0.7243(R_{833}) + 0.084(R_{1650}) - 0.18(R_{2213})$	Schlerf et al., 2005	$-0.2941\text{Blue} - 0.243\text{Green} - 0.5424\text{Red} + 0.7276\text{Near_Infrared} + 0.0713\text{SWIR}_1 - 0.1608\text{SWIR}_2$	Baig et al., 2014
Wetness	$0.1509(R_{487}) + 0.1973(R_{560}) + 0.3279(R_{660}) + 0.3406(R_{833}) - 0.7112(R_{1650}) - 0.4572(R_{2213})$	Thenkaibil et al., 2002	$0.1511\text{Blue} + 0.1973\text{Green} + 0.3283\text{Red} + 0.3407\text{Near_Infrared} - 0.7117\text{SWIR}_1 - 0.4559\text{SWIR}_2$	Baig et al., 2014

جدول ۳. شاخص های مستخرج از مدل رقومی ارتفاع

نام شاخص	فرمول محاسبه شاخص	منبع
Mean Ruggedness of Valley Factor (MRVBF) Bottom	$\frac{(ELEV - MIN_ELEV)}{(MAX_ELEV - MIN_ELEV)}$	(Huang et al., 2017)
Wind Exposition	$\frac{\cos((Wind\ Direction - Aspect) * \pi)}{180}$	(Panahi et al, 2020).
Relative Slope Position	$\frac{(Slope - Min_Slope)}{(Max_Slope - Min_Slope)}$	(Liu et al, 2016)
Topographic Position Index	$ELEV - MEAN_ELEV$	(Skentos, 2018).
Topographic Wetness Index	$\ln\left(\frac{A}{S \cdot \tan(Slope)}\right)$	(Riihimäki et al, 2021)

ELEV: ارتفاع نقطه، MIN_ELEV: کمینه ارتفاع، MAX_ELEV: بیشینه ارتفاع، MEAN_ELEV: میانگین ارتفاع منطقه، Aspect: جهت شیب، Wind Direction: جهت باد منطقه، pi: ثابت عددی ۳٫۱۴۲، Slope: شیب

برای اعتبارسنجی عملکرد لندست ۸ و پریسمادر مدلسازی خصوصیات خاک با مدل جنگل تصادفی آماره های ریشه میانگین مربعات خطا، نسبت انحراف عملکرد^۱، نسبت دامنه بین چارکی^۲ ضریب تبیین از مقادیر اندازه گیری شده خصوصیات خاک و مقادیر ویژگی های خاک پیش بینی شده محاسبه شد (معادلات ۱ تا ۴). همچنین کلاس بندی مقادیر دو شاخص نسبت انحراف عملکرد و نسبت دامنه بین چارکی به ترتیب در جدول های ۴ و ۵ آورده شده است. در فرآیند مدلسازی از ۸۰ درصد داده ها (۱۸۳ نمونه) برای واسنجی و همچنین بهینه سازی پارامترهای مدل جنگل تصادفی با استفاده از روش اعتبارسنجی متقابل ۱۰ گام با ۱۰ تکرار در بسته "caret" و از ۲۰ درصد باقی مانده داده ها (۴۶ نمونه) برای اعتبارسنجی نقشه نهایی پیش بینی شده مقادیر کرنات کلسیم معادل و درصد رس استفاده شد.

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} \quad \text{رابطه ۱}$$

$$RMSE_{fo} = \left[\sum_{i=1}^N \frac{(Z_{fi} - Z_{oi})^2}{N} \right]^{\frac{1}{2}} \quad \text{رابطه ۲}$$

$$RPD = RMSE \times \sqrt{\frac{N}{(N - 1)}} \quad \text{رابطه ۳}$$

$$RPIQ = \frac{Q_{75} - Q_{25}}{SEP} \quad \text{رابطه ۴}$$

جدول ۴. کلاس بندی شاخص نسبت انحراف عملکرد (McDowell et al., 2012)

نسبت انحراف عملکرد مقادیر	کلاس
X < 1	خیلی ضعیف
1 < X < 1.4	ضعیف
1.4 < X < 1.8	متوسط
1.8 < X < 2	خوب
2 < X < 2.5	خیلی خوب
X > 2.5	عالی

جدول ۵. کلاس‌بندی شاخص نسبت انحراف عملکرد (De Santana And Daly., 2022)

نسبت انحراف عملکرد مقادیر	کلاس
$X < 2.02$	خیلی ضعیف
$2.02 < X < 2.70$	ضعیف
$2.70 < X < 3.38$	متوسط
$3.38 < X < 4.05$	خوب
$4.05 < X$	عالی

نتایج و بحث

خلاصه آماری داده‌های آموزش و اعتبارسنجی خصوصیات سطحی ۲۲۹ نمونه خاک دشت قزوین در جدول زیر ارائه شده است. همان‌طور که مشاهده می‌گردد، داده‌های واسنجی و اعتبارسنجی دارای مقادیر حداقل و حداکثر و میانگین دارای اختلاف کمی نسبت به هم بودن مقادیر این آماره‌ها را در (جدول ۶) ذکر شده است.

جدول ۶. خلاصه آماره داده‌های آموزش و اعتبارسنجی

نمونه‌های اعتبارسنجی		نمونه‌های آموزش		خصوصیت آماره
مقدار رس	کربنات کلسیم معادل	مقدار رس	کربنات کلسیم معادل	
۲۱/۸۵	۱۲/۵۷	۲۴/۵۵	۱۲/۶۲	میانگین
۰	۱۰/۷۵	۰	۱۱	مد
۲۰	۸/۳۴	۲۳	۱۱/۲۷	میانه
۷۰	۵۷/۲۸	۷۱	۳۹	بیشینه
۰	۱/۱۳	۰	۱/۱	کمینه
۲۰۶/۱۷	۵۲/۴	۱۹۹/۵	۵۱/۷	واریانس
۹/۴۳	۹/۲۷	۸/۱۲	۴/۰۹	انحراف از معیار

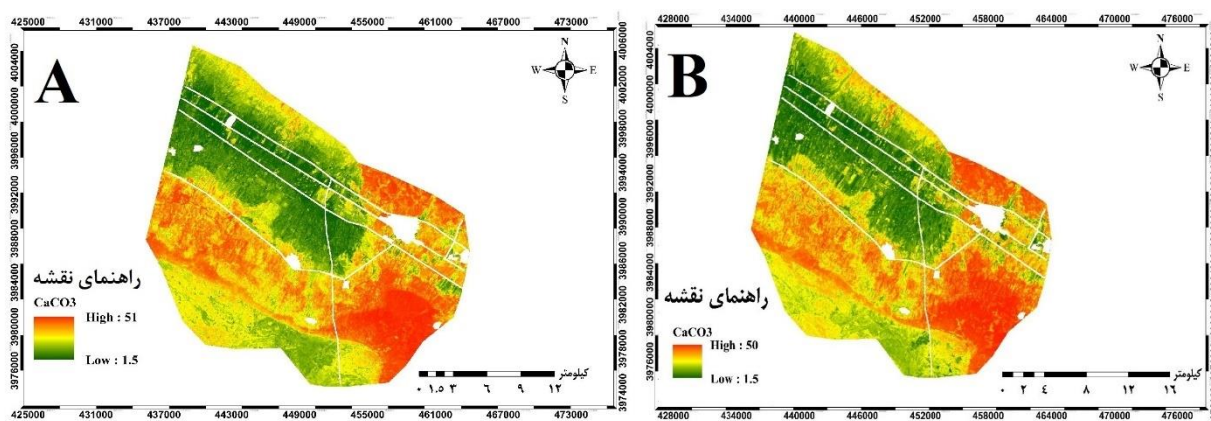
آماره‌های خصوصیات خاک در منطقه مورد مطالعه نشان‌دهنده تغییرات زیاد مقدار این خصوصیات در منطقه است. توزیع فراوانی داده‌های واسنجی و اعتبارسنجی مدل نشان‌دهنده انتخاب بهینه دودسته داده مستقل از خصوصیات خاک در این پژوهش است. نتایج ورودی‌های مختلف برای تهیه نقشه کربنات کلسیم معادل خاک با استفاده از داده‌های دو ماهواره لندست ۸ و پریسما با مدل جنگل تصادفی در (جدول ۷) آورده شده است.

جدول ۷. نتایج مدلسازی کربنات کلسیم معادل در خاک با داده‌های ماهواره‌های لندست ۸ و پریسما با استفاده از مدل جنگل تصادفی

نسبت انحراف عملکرد	نسبت انحراف عملکرد	ریشه میانگین مربعات خطا نرمال شده	ریشه میانگین مربعات خطا	ضریب تبیین	ورودی‌های مدل	ماهواره
۰/۷۳	۱/۱۱	۰/۲۳	۰/۸۳	۰/۴۳	PCA	لندست ۸
۰/۷۶	۱/۶۳	۰/۲۱	۰/۷۵	۰/۵۸	PCA + DEM Index	
۰/۷۳	۱/۰۶	۰/۲۵	۰/۸۷	۰/۳۶	Spectral Index	
۰/۷۷	۱/۷۸	۰/۲۱	۰/۷۳	۰/۶۱	Spectral Index + DEM Index	
۰/۷۳	۱/۱۱	۰/۲۴	۰/۸۴	۰/۴۳	Spectral Index + PCA	
۰/۸۷	۱/۸۶	۰/۲۱	۰/۷۳	۰/۶۱	Spectral Index + DEM Index + PCA	
۰/۷۳	۱/۰۹	۰/۲۴	۰/۸۴	۰/۴۱	PCA	پریسما
۰/۹۲	۱/۲۷	۰/۲۰	۰/۷۲	۰/۶۲	PCA + DEM Index	
۰/۷۱	۱/۰۳	۰/۲۵	۰/۸۹	۰/۲۵	Spectral Index	
۰/۷۶	۱/۲۱	۰/۲۲	۰/۷۷	۰/۵۵	Spectral Index + DEM Index	
۰/۷۳	۱/۰۹	۰/۲۴	۰/۸۵	۰/۳۹	Spectral Index + PCA	
۰/۷۷	۱/۲۶	۰/۲۰	۰/۷۳	۰/۶۱	Spectral Index + DEM Index + PCA	

نقشه‌های تهیه شده به وسیله تصاویر ماهواره لندست ۸ برای مقدار کربنات کلسیم معادل (شکل ۴A) که با ورودی‌های مدل رقومی ارتفاع، تجزیه مولفه‌های اصلی و شاخص‌های طیفی تهیه شدند دارای بالاترین مقدار ضریب همبستگی و کمترین مقدار ریشه میانگین مربعات خطا برای کربنات کلسیم معادل خاک بودند. مدل جنگل تصادفی با آماره‌های (ضریب تبیین: ۰/۶۱؛ ریشه میانگین مربعات خطا: ۰/۷۳؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۱؛ نسبت دامنه بین چارکی: ۰/۸۷؛ نسبت انحراف عملکرد: ۱/۸۶) بود. نقشه‌های تهیه شده با تصاویر ماهواره پریسما با برای مقدار کربنات کلسیم معادل (شکل ۴B) با ورودی‌های شاخص‌های مستخرج از مدل رقومی ارتفاع و تجزیه مولفه‌های اصلی دارای شاخص‌های آماری (ضریب تبیین: ۰/۶۲؛ ریشه میانگین مربعات خطا: ۰/۷۲؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۰؛ نسبت دامنه بین چارکی: ۰/۹۲؛ نسبت انحراف عملکرد: ۱/۲۷) بود. نتایج این مدل برای تصاویر پریسما نشان داد که استفاده از تحلیل مولفه‌های اصلی برای بهبود کیفیت نقشه‌های تولید شده کربنات کلسیم معادل موثر تر از استفاده از شاخص‌های طیفی است. استفاده هم‌زمان از شاخص‌های طیفی و تحلیل مولفه‌های اصلی همراه با شاخص‌های مدل رقومی ارتفاع باعث افزایش دقت مدل می‌شود. این نتایج توسط رحمانی و همکاران در سال ۱۳۹۹، Shan و همکاران ۲۰۲۱ و Lagacherie و همکاران ۲۰۰۸ نیز تایید شده است.

همچنین نتایج نشان می‌دهد استفاده از شاخص‌های مدل رقومی ارتفاع تأثیر زیادی در بهبود نتایج دارد. مدلی که از شاخص‌های مستخرج از مدل رقومی استفاده کرده‌اند به طور معمول ریشه میانگین مربعات خطا کمتر و ضریب تبیین بیشتری دارند در مقایسه با مدل‌هایی که تنها از شاخص‌های طیفی و تحلیل مولفه‌های اصلی استفاده کرده‌اند. رحمانی و همکاران، ۱۳۹۹ در تحقیقی به منظور ارزیابی نقش ویژگی‌های ژئومورفومتری در نقشه برداری رقومی خاک برای رس و کربنات کلسیم معادل به نتایج مشابهی رسیدند. استفاده از مدل رقومی ارتفاع در تحقیق (Marques et al., 2018) باعث افزایش دقت نقشه ویژگی‌های خاک شد که تایید کننده نتایج این تحقیق است. با توجه به دو شاخص نسبت دامنه بین چارکی و نسبت انحراف عملکرد در مدل جنگل تصادفی تفاوتی بین ورودی‌های مختلف وجود ندارد و همه در محدوده عملکرد متوسط قرار می‌گیرند. در شکل ۴ نقشه‌های حاصل از تصاویر دو ماهواره لندست ۸ و پریسما است.



شکل ۴. A: نقشه کربنات کلسیم معادل با مدل جنگل تصادفی و تصاویر ماهواره لندست ۸، B: نقشه کربنات کلسیم معادل با مدل جنگل تصادفی و تصاویر ماهواره پریسما

با توجه به مقادیر ضریب تبیین اختلاف دقت در بین تصاویر پریسما و لندست ۸ برای نقشه برداری مقادیر کربنات کلسیم معادل ناچیز بوده و این اختلاف معنی دار نمی‌باشد و می‌توان گفت که کارایی داده‌های ماهواره پریسما به نسبت داده‌های لندست ۸ در تهیه نقشه کربنات کلسیم معادل تفاوتی ندارد این نتایج معیار با نتایجی است که توسط Angelopoulou و همکاران در سال ۲۰۲۲ بیان شد. شاخص‌های طیفی به تنهایی عملکرد چندان خوبی نداشته‌اند؛ اما وقتی با اطلاعات مدل رقومی ارتفاع که معرف شرایط سطح است ترکیب شدند توانستند نقشه بهتری ارائه دهند. بهبود دقت و عملکرد مدل‌سازی با استفاده از تمام ویژگی‌های مستخرج از تصاویر ماهواره‌ای و مدل رقومی ارتفاع در تحقیق (Dindaroglu et al., 2022) نیز تأیید شد. محدوده مورد مطالعه رسوبی بوده و به‌طور کلی، از نهشته‌های سیلابی و رودخانه‌ای جور نشده با تغییرات کم شیب می‌باشد. با در نظر گرفتن شرایط حاکم در منطقه و بررسی مطالعات مشابه در نقشه برداری رقومی اراضی سیلابی و با اراضی دارای پستی و بلندی کم، شاخص‌هایی مانند شاخص همواری دره با درجه تفکیک بالا

شاخص صافی کف دره با وضوح چندگانه^۱، شاخص موقعیت شیب نسبی^۲، شاخص میانگین ردیابی شعاعی موقعیت توپوگرافی^۳، شاخص همگرایی^۴، فاصله شبکه کانالی^۵ و آنالیز سایه روشن^۶ در نشان دادن فرآیند انتقال رسوب به ویژه جریان رسوبات و مواد در حال حرکت موثر گزارش شدند (پهلوان راد و همکاران، ۱۳۹۹; Mosleh et al., 2016).

با توجه به نتایج به دست آمده برای مقادیر نسبت انحراف عملکرد و نسبت انحراف عملکرد که به دست آمد. داده‌های ماهواره‌های پریسماعملکرد مشابه با لندست ۸ دارد، زیرا مقادیر نسبت انحراف عملکرد و نسبت انحراف عملکرد مدل مبتنی بر پریسمما همانند ماهواره لندست ۸ بوده و مقادیر ریشه میانگین مربعات خطا نرمال شده و ریشه میانگین مربعات خطا آنها نیز برابر است. که نشان از عدم تفاوت در عملکرد بین دو ماهواره دارد. هنگامی که از ترکیب ورودی‌های مختلف استفاده می‌شود، تفاوت عملکرد مدل اجرا شده با داده‌های پریسمما و لندست ۸ نیز برابر است.

استفاده از ورودی‌های مختلف باعث بهبود بیشتر عملکرد مدل‌های اجرا شده با داده‌های پریسمما به نسبت ماهواره لندست ۸ نمی‌شود. بنابراین به طور کلی می‌توان نتیجه گرفت که پریسماعملکرد بهتری نسبت به لندست ۸ در نقشه برداری مقادیر کلسیم خاک در این منطقه ندارد. حتی هنگامی که از ترکیب ورودی‌های مختلف استفاده می‌شود تفاوتی دیده نمی‌شود.

نتایج نشان داد استفاده از مدل رقومی ارتفاع و شاخص‌های مستخرج از آن باعث افزایش عملکرد مدل می‌شود که علت این امر تأثیرپذیری زیاد مقادیر کربنات کلسیم معادل خاک از توپوگرافی منطقه است. با توجه به نتایج نمی‌توان گفت که کدام یک از نقشه‌های حاصل از تجزیه مولفه‌های اصلی تصاویر ماهواره‌ای یا نقشه‌های حاصل از شاخص‌های طیفی مستخرج از تصاویر ماهواره‌ای دارای صحت و دقت بیشتری است. همچنین نتایج نشان نداد که کدام یک از مشتقات تصاویر لندست ۸ یا مشتقات تصاویر ماهواره پریسمما برای تهیه نقشه‌های مقادیر کربنات کلسیم معادل خاک کارایی بالاتری دارد.

مقادیر کربنات کلسیم معادل پیش‌بینی شده با استفاده از طیف‌های ماهواره پریسمما و لندست ۸ با روند تغییرات کربنات کلسیم معادل خاک در منطقه مطابقت داشت عملکرد بالای ماهواره پریسمما در مدل‌سازی کربنات توسط Casa و همکاران در ۲۰۲۳ نیز تایید شد. مقادیر بالای کربنات کلسیم معادل در قسمت‌های شمالی منطقه به دلیل رسوبات آهکی میوسن، سنگ‌های آهکی و فرسایش خاک اراضی بالادست این ناحیه است که عموماً دارای مواد مادری آهکی هستند که این مورد با نقشه‌های تهیه شده مطابقت داشت، مقادیر کم کربنات کلسیم معادل به طور منطقی با مناظر خاکی مطابقت دارد که با آبشویی کربنات به دلیل فرایندهای پدوژنتیک و نقش خاک‌ورزی عمیق را در حذف نامنظم انباشته‌های کربنات عمیق، به‌ویژه در منظر خاکی مرکزی منطقه مورد مطالعه مستند می‌کند، با مطالعه موسوی و همکاران ۱۴۰۰ مطابقت دارد.

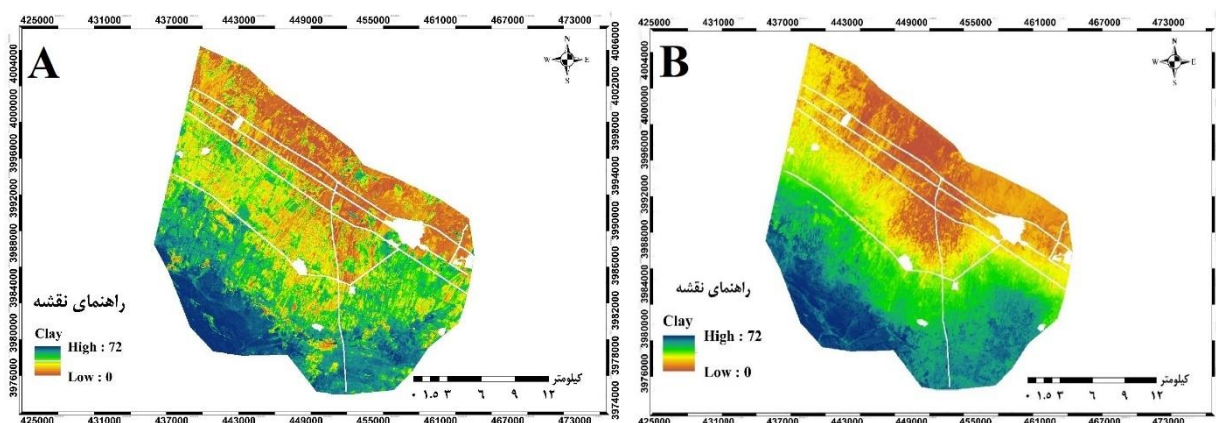
نقشه‌های درصد رس که به‌وسیله تصاویر ماهواره پریسمما برای منطقه مورد مطالعه با ورودی‌های مدل رقومی ارتفاع، تجزیه مولفه‌های اصلی و شاخص‌های طیفی تهیه شدند اطلاعات کامل نقشه‌های تهیه شده در (جدول ۸) آورده شده است، بهترین مدل جنگل تصادفی با شاخص‌های آماری (ضریب تبیین: ۰/۸۹؛ ریشه میانگین مربعات خطا: ۴/۹۸؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۲؛ نسبت دامنه بین چارکی: ۰/۸۰؛ نسبت انحراف عملکرد: ۲/۱۱) بهترین نتایج با ورودی‌های تحلیل مولفه‌های اصلی و شاخص‌های مستخرج از مدل رقومی ارتفاع حاصل شد (شکل ۵A). در ماهواره لندست ۸ با مدل جنگل تصادفی بهترین مدل دارای شاخص‌های آماری (ضریب تبیین: ۰/۸۶؛ ریشه میانگین مربعات خطا: ۵/۲۳؛ ریشه میانگین مربعات خطا نرمال شده: ۰/۲۴؛ نسبت دامنه بین چارکی: ۰/۷۹؛ نسبت انحراف عملکرد: ۲/۱۸) بود که ورودی‌های این مدل شامل تجزیه مولفه‌های اصلی و شاخص‌های مستخرج از مدل رقومی ارتفاع بود (شکل ۵B). نسبت انحراف عملکرد و نسبت انحراف عملکرد نشان‌دهنده قدرت تبیینی مدل هستند. چون نسبت انحراف عملکرد و نسبت انحراف عملکرد مدل مبتنی بر داده‌های پریسمما کمی بیشتر است پس این مدل قدرت تبیینی به نسبت بیشتری دارند. همچنین ریشه میانگین مربعات خطا نرمال شده و ریشه میانگین مربعات خطا نشان‌دهنده خطا و دقت مدل هستند. چون ریشه میانگین مربعات خطا نرمال شده و ریشه میانگین مربعات خطا نرمال شده و دقت مدل کمتر است پس این مدل خطای کمتر و دقت بیشتری به نسبت لندست ۸ دارند. و ضریب تبیین نشان‌دهنده میزان توافق میان پیش‌بینی‌ها و مقادیر واقعی است. چون برای برخی موارد ضریب تبیین مدل

1. Multi-resolution Valley Bottom Flatness index
2. Relative Slope Position
3. Mean Ray-Tracing Topographic Position Index
4. Convergence Index
5. Channel Network Distance
6. Analytical Hillshading

مبتنی بر داده‌های پریسما بیشتر است، پس پیش‌بینی‌های این مدل با مقادیر واقعی تطابق بیشتری دارند در (شکل ۵) نقشه‌های تهیه شده برای درصد رس در خاک آورده شده است.

جدول ۸. نتایج مدل‌سازی درصد رس در خاک با ماهواره‌های لندست ۸ و پریسما

ماهواره	ورودی‌های مدل	ضریب تبیین	ریشه میانگین مربعات خطا	ریشه میانگین مربعات خطا نرمال شده	نسبت انحراف عملکرد	نسبت دامنه بین چارکی
لندست ۸	PCA	۰/۵۷	۵/۴۵	۰/۲۵	۱/۶۰	۰/۷۸
	PCA + DEM Index	۰/۸۵	۵/۱۸	۰/۲۴	۲/۲۰	۰/۸۹
	Spectral Index	۰/۶۸	۵/۳۴	۰/۲۴	۱/۳۷	۰/۶۷
	Spectral Index + DEM Index	۰/۷۱	۵/۰۲	۰/۲۳	۱/۴۲	۰/۶۸
	Spectral Index + PCA	۰/۵۹	۵/۰۳	۰/۲۳	۱/۴۲	۰/۶۸
	Spectral Index + DEM Index + PCA	۰/۸۶	۵/۲۳	۰/۲۴	۲/۱۸	۰/۷۹
پریسما	PCA	۰/۶۸	۴/۳۵	۰/۲۰	۱/۸۲	۰/۷۹
	PCA + DEM Index	۰/۸۹	۴/۹۸	۰/۲۲	۲/۱۱	۰/۸۰
	Spectral Index	۰/۵۸	۵/۱۱	۰/۲۳	۱/۴۳	۰/۷۰
	Spectral Index + DEM Index	۰/۷۱	۵/۱۳	۰/۲۳	۱/۵۳	۰/۷۰
	Spectral Index + PCA	۰/۶۰	۴/۴۵	۰/۲۰	۱/۵۱	۰/۷۲
	Spectral Index + DEM Index + PCA	۰/۹۰	۴/۱۱	۰/۱۸	۲/۲۹	۰/۹۵



شکل ۵. A: نقشه کربنات کلسیم معادل با مدل جنگل تصادفی و تصاویر ماهواره لندست ۸، B: نقشه کربنات کلسیم معادل با مدل جنگل تصادفی و تصاویر ماهواره پریسما

همچنین ریشه میانگین مربعات خطا نرمال شده و ریشه میانگین مربعات خطا نشان‌دهنده خطا و دقت مدل هستند. چون ریشه میانگین مربعات خطا نرمال شده و ریشه میانگین مربعات خطا مدل اجرا شده با تصاویر ماهواره‌ای پریسما کمتر است پس این مدل خطای کمتر و دقت بیشتری دارند. و ضریب تبیین نشان‌دهنده میزان توافق میان پیش‌بینی‌ها و مقادیر واقعی است. چون برای برخی موارد ضریب تبیین مدل مبتنی بر داده‌های پریسما بیشتر است، پس پیش‌بینی‌های این مدل با مقادیر واقعی تطابق بیشتری دارند. بر اساس نتایج به دست آمده، استفاده از ترکیب تجزیه و تحلیل مولفه‌های اصلی، شاخص‌های مستخرج از مدل رقومی ارتفاع و شاخص‌های طیفی بهترین عملکرد را در مدل‌سازی خصوصیات خاک با ماهواره لندست ۸ نشان می‌دهد. استفاده از تجزیه و تحلیل مولفه‌های اصلی به عنوان یک روش تجزیه فضایی-زمانی و ترکیب آن با شاخص‌های مستخرج از مدل رقومی که داده‌های ارتفاعی را در برمی‌گیرد و شاخص‌های طیفی باعث بهبود می‌شوند زیرا اطلاعات طیفی مربوط به خاک و پوشش زمین را در برمی‌گیرند و این اطلاعات مفید برای تحلیل و پیش‌بینی ویژگی‌های خاک بسیار حائز اهمیت هستند و می‌تواند به تشخیص الگوهای پیچیده‌تر و ارتباطات بین ویژگی‌های خاک کمک کند.

استفاده از ترکیب تجزیه و تحلیل مولفه‌های اصلی، شاخص‌های مستخرج از مدل رقومی و شاخص‌های طیفی در مدل‌سازی خصوصیات

خاک با استفاده از داده‌های ماهواره پریسما، عملکرد بهتری ارائه می‌دهد. با استفاده از روش تجزیه و تحلیل مولفه‌های اصلی به عنوان یک روش تجزیه متعادل وزنی و ترکیب آن با شاخص‌های مستخرج از مدل رقومی و استفاده از شاخص‌های طیفی باعث افزایش دقت مدل می‌شود. با استفاده از شاخص‌های طیفی که با نشان دادن اطلاعات طیفی مربوط به خاک و پوشش زمین توانایی تحلیل و تجزیه داده‌های با رزولوشن بالا بهبود می‌دهد و ارتباطات مهم بین ویژگی‌های خاک و بازتاب‌های طیفی شناسایی می‌شوند. این اطلاعات مفید برای تحلیل و پیش‌بینی ویژگی‌های خاک بسیار مهم است. در نتیجه، استفاده از این ترکیب و روش‌ها در مدلسازی با داده‌های ماهواره پریسما، بهبود قابل توجهی در عملکرد مدلسازی و دقت پیش‌بینی خصوصیات خاک به همراه دارد.

نتایج نشان داد استفاده از مدل رقومی ارتفاع و شاخص‌های مستخرج از آن باعث افزایش عملکرد مدل می‌شود که علت این امر تأثیرپذیری زیاد درصد رس خاک از توپوگرافی منطقه است. همچنین نتایج نشان داد که نقشه‌های حاصل از تجزیه مولفه‌های اصلی تصاویر ماهواره‌ای دارای دقت کمتری نسبت به نقشه‌های حاصل از شاخص‌های طیفی مستخرج از تصاویر ماهواره‌ای است و مشتقات تصاویر لندست ۸ به نسبت مشتقات تصاویر ماهواره پریسمانقشه‌های درصد رس خاک با دقت کمتری مدل‌سازی می‌کنند. بهترین نسبت انحراف عملکرد برای برآورد مقادیر رس خاک با استفاده از داده‌های ماهواره پریسمابه دست آمد که برابر $0/95$ بود، درحالی‌که نتایج به دست آمده با استفاده از تصویر ماهواره لندست ۸ در بهترین حالت $0/89$ بود. این نتایج همسو با تحقیق Castaldi و همکاران در سال ۲۰۱۶ است. مقادیر درصد رس پیش‌بینی شده با استفاده از طیف‌های ماهواره پریسما و لندست ۸ با روند تغییرات درصد رس خاک در منطقه مطابقت داشت. مقادیر درصد رس در قسمت‌های جنوبی منطقه به نسبت مناطق شمالی بیشتر است. دلیل این اختلاف بافت در مناطق جنوبی به نسبت مناطق شمالی کاهش مقدار شیب از شمال به جنوب است چون با کاهش شیب میزان انرژی جریان کاهش یافته و ذرات با قطر بیشتر در زودتر ته‌نشین شده و ذرات با قطر کمتر به مناطق پست‌تر انتقال پیدا می‌کنند (Gabriels et al., 1978).

نتیجه‌گیری

هر دو ماهواره لندست ۸ و پریسما توانایی بالایی برای تهیه نقشه دو خصوصیت کربنات کلسیم معادل و درصد رس از خود نشان دادند. مقادیر ضریب تبیین نشان می‌دهد که ورودی‌های مدل، 85% از تغییرات رس و بیش از 60% از تغییرات کربنات کلسیم معادل در منطقه را به درستی نشان دادند. داده‌های ماهواره پریسما نسبت به داده‌های ماهواره لندست ۸ در تهیه نقشه این دو خصوصیت خاک با استفاده از مدل جنگل تصادفی تفاوت معنی‌داری نشان نداد. یکی از دلایل عدم تفاوت در عملکرد ماهواره‌های لندست ۸ و پریسما این باشد که هر دو ماهواره از تکنولوژی پیشرفته تصویربرداری استفاده می‌کنند و توانایی برآورد و اندازه‌گیری خصوصیات سطح خاک را دارند. همچنین، این ماهواره‌ها در شرایط مشابه در تاریخ‌های نزدیک به یکدیگر تصاویری از مناطق مشابه زمین را تهیه کرده‌اند، که می‌تواند به شبیه بودن عملکرد آن‌ها کمک کند. ترکیب ورودی‌های مختلف مانند شاخص‌های طیفی، تجزیه مولفه‌های اصلی و مدل رقومی ارتفاع، به طور قابل توجهی در بهبود عملکرد مدل موثر بوده است. با این حال، استفاده از تعداد زیاد ورودی‌ها می‌تواند محدودیت‌هایی در دقت مدل ایجاد کند. آنچه در این تحقیق مورد بررسی قرار گرفت شناسایی اهمیت سه دسته متغیر کمکی بود که در خاک بسیار پرکاربرد هستند. با این حال، این موضوع به این معنی نیست که اهمیت انتخاب بهترین ویژگی‌های هر دسته از شاخص‌های طیفی، شاخص‌های مدل رقومی ارتفاع و تجزیه مولفه‌های اصلی نادیده گرفته شود. برعکس، بهینه‌سازی انتخاب ویژگی‌ها از هر دسته، به ما این امکان را می‌دهد تا بهترین ویژگی‌های هر دسته را شناسایی کرده و از آن‌ها در مدل استفاده کنیم.

"هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد"

منابع

- پهلوان راد، محمدرضا، فرقانی، س، اسفندیاری، م و محمدی ترکاشوند، علی. ۱۳۹۹. نقشه‌برداری رقومی کلاس‌های خاک اراضی سیلابی و با پستی و بلندی کم (مطالعه موردی: شهرستان هیرمند). حفاظت منابع آب و خاک (علمی - پژوهشی). ۹(۴). ۱۲۰-۱۰۷.
- رحمانی، اصغر، سرمیدیان، فریدون و عارفی، حسین. (۱۴۰۱). نقشه‌برداری رقومی ضخامت خاک سطحی و عدم قطعیت وابسته به آن با استفاده از رویکرد یادگیری ماشین در بخشی از اراضی خشک و نیمه‌خشک دشت قزوین. تحقیقات آب و خاک ایران، ۵۳(۳)، ۵۸۵-۶۰۲.
- رحمانی، اصغر، سرمیدیان، فریدون، موسوی، سیدروح اله، و خاموشی، سیدعرفان. (۱۳۹۹). کاربرد ویژگی‌های ژئومورفومتر در نقشه برداری رقومی خاک با استفاده از منطق فازی و یادگیری ماشین. مرتع و آبخیزداری (منابع طبیعی ایران)، ۷۳(۱)، ۱۰۵-۱۲۴.

موسوی، سیدروح اله، سرمیدیان، فریدون، امید، محمود، و بوگارت، پاتریک. (۱۴۰۰). مدل سازی تغییرات عمقی کربنات کلسیم معادل خاک با استفاده از الگوریتم‌های یادگیری ماشین در دشت قزوین. آب و خاک (علوم و صنایع کشاورزی)، ۳۵(۵)، ۷۱۹-۷۳۴.

REFERENCES

- Akpa, S. I., Odeh, I. O., Bishop, T. F., and Hartemink, A. E. (2014). Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, 78(6), 1953-1966.
- Alicandro, M., Candigliota, E., Dominici, D., Immordino, F., Masin, F., Pascucci, N., ... and Zollini, S. (2022). Hyperspectral PRISMA and Sentinel-2 Preliminary Assessment Comparison in Alba Fucens and Sinuessa Archaeological Sites (Italy). *Land*, 11(11), 2070.
- Anderson, M. C., Neale, C. M. U., Li, F., Norman, J. M., Kustas, W. P., Jayanthi, H., and Chavez, J. O. S. E. (2004). Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. *Remote sensing of environment*, 92(4), 447-464.
- Angelopoulou, T., Chabrillat, S., Pignatti, S., Milewski, R., Karyotis, K., Brell, M., ... and Zalidis, G. (2023). Evaluation of airborne hypspx and spaceborne PRISMA hyperspectral remote sensing data for soil organic matter and carbonates estimation. *Remote Sensing*, 15(4), 1106.
- ASI Prisma Products Specification Document Issue 2.3. [(accessed on 12 March 2020)]; Available online: http://prisma.asi.it/missionselect/docs/PRISMA%20Product%20Specifications_Is2_3.pdf:
- Baig, M. H. A., Zhang, L., Shuai, T., and Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, 5(5), 423-431.
- Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., and Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote sensing of environment*, 113, S38-S55.
- Bouzekri, S., Lasbet, A. A., and Lachehab, A. (2015). A new spectral index for extraction of built-up area using Landsat-8 data. *Journal of the Indian Society of Remote Sensing*, 43, 867-873.
- Broge, N. H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote sensing of environment*, 76(2), 156-172.
- Carre, F., McBratney, A. B., Mayr, T., and Montanarella, L. (2007). Digital soil assessments: Beyond DSM. *Geoderma*, 142(1-2), 69-79.
- Casa, R., Pignatti, S., Pascucci, S., Castaldi, F., and Marrone, L. (2023). Estimation of agronomic soil properties from multitemporal PRISMA satellite imaging spectroscopy. In *Precision agriculture'23* (pp. 839-845). Wageningen Academic.
- Casa, R., Pignatti, S., Pascucci, S., Castaldi, F., and Marrone, L. (2023). Estimation of agronomic soil properties from multitemporal PRISMA satellite imaging spectroscopy. In *Precision agriculture'23* (pp. 839-845). Wageningen Academic.
- Castaldi, F., Casa, R., Castrignanò, A., Pascucci, S., Palombo, A., and Pignatti, S. (2014). Estimation of soil properties at the field scale from satellite data: a comparison between spatial and non-spatial techniques. *European Journal of Soil Science*, 65(6), 842-851.
- Colovic, M., Yu, K., Todorovic, M., Cantore, V., Hamze, M., Albrizio, R., and Stellacci, A. M. (2022). Hyperspectral vegetation indices to assess water and nitrogen status of sweet maize crop. *Agronomy*, 12(9), 2181.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- Dash, P. K., Panigrahi, N., and Mishra, A. (2022). Identifying opportunities to improve digital soil mapping in India: A systematic review. *Geoderma Regional*, 28, e00478.
- De Santana, F. B., and Daly, K. (2022). A comparative study of MIR and NIR spectral models using ball-milled and sieved soil for the prediction of a range soil physical and chemical parameters. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 279, 121441.
- Dharumarajan, S., and Hegde, R. (2022). Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use and Management*, 38(1), 135-149.
- Diek, S., Schaepman, M. E., and De Jong, R. (2016). Creating multi-temporal composites of airborne imaging spectroscopy data in support of digital soil mapping. *Remote Sensing*, 8(11), 906.
- Dindaroglu, T., Tunguz, V., Babur, E., Alkharabsheh, H. M., Seleiman, M. F., Roy, R., and Zakharchenko, E. (2022). The use of remote sensing to characterise geomorphometry and soil properties at watershed scale. *International Journal of Global Warming*, 27(4), 402-421.
- Du, Q., and Yang, H. (2008). Similarity-based unsupervised band selection for hyperspectral image analysis.



- IEEE geoscience and remote sensing letters, 5(4), 564-568.
- Duan, M., Song, X., Liu, X., Cui, D., and Zhang, X. (2022). Mapping the soil types combining multi-temporal remote sensing data with texture features. *Computers and Electronics in Agriculture*, 200, 107230.
- Fang, Y., Xu, L., Wong, A., and Clausi, D. A. (2022). Multi-temporal landsat-8 images for retrieval and broad scale mapping of soil copper concentration using empirical models. *Remote Sensing*, 14(10), 2311.
- Galvao, L. S., dos Santos, J. R., Roberts, D. A., Breunig, F. M., Toomey, M., and de Moura, Y. M. (2011). On intra-annual EVI variability in the dry season of tropical forest: A case study with MODIS and hyperspectral data. *Remote Sensing of Environment*, 115(9), 2350-2359.
- Galvao, L. S., Formaggio, A. R., and Tisot, D. A. (2005). Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote sensing of Environment*, 94(4), 523-534.
- Gasmi, A., Gomez, C., Chehbouni, A., Dhiba, D., and El Gharous, M. (2022). Using PRISMA hyperspectral satellite imagery and GIS approaches for soil fertility mapping (FertiMap) in Northern Morocco. *Remote sensing*, 14(16), 4080.
- Grunwald, S. (2010). Current state of digital soil mapping and what is next. In digital soil mapping (pp. 3-12). Springer, Dordrecht.
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sensing of environment*, 90(3), 337-352.
- Huang, S.Y., Zhang, S.H., Liu, L.L., Zhu, W.Q. and Cheng, Y.M., (2021). Efficient slope reliability analysis and risk assessment based on multiple Kriging metamodels. *Computers and Geotechnics*, 137, p.104277.
- Inoue, Y., Peñuelas, J., Miyata, A., and Mano, M. (2008). Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO₂ flux measurements in rice. *Remote Sensing of Environment*, 112(1), 156-172.
- Ishwaran, H., and Kogalur, U. B. (2010). Consistency of random survival forests. *Statistics and probability letters*, 80(13-14), 1056-1064.
- Jurgens, C. (1997). The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *International Journal of Remote Sensing*, 18(17), 3583-3594.
- Karray, E., Elmannai, H., Toumi, E., Gharbia, M. H., Meshoul, S., Aichi, H., and Ben Rabah, Z. (2023). Evaluating the Potentials of PLSR and SVR Models for Soil Properties Prediction Using Field Imaging, Laboratory VNIR Spectroscopy and Their Combination. *Comput. Model. Eng. Sci*, 136, 1399-1425.
- Khaledian, Y. and Miller, B.A., (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, pp.401-418.
- Li, P., Xiao, C., and Feng, Z. (2018). Mapping rice planted area using a new normalized EVI and SAVI (NVI) derived from Landsat-8 OLI. *IEEE Geoscience and Remote Sensing Letters*, 15(12), 1822-1826.
- Lucieer, A., Malenovsky, Z., Veness, T., and Wallace, L. (2014). HyperUAS—Imaging spectroscopy from a multirotor unmanned aircraft system. *Journal of Field Robotics*, 31(4), 571-590.
- Lymburner, L., Beggs, P. J., and Jacobson, C. R. (2000). Estimation of canopy-average surface-specific leaf area using Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66(2), 183-192.
- Mallah, S., Delsouz Khaki, B., Davatgar, N., Scholten, T., Amirian-Chakan, A., Emadi, M., ... and Taghizadeh-Mehrjardi, R. (2022). Predicting Soil Textural Classes Using Random Forest Models: Learning from Imbalanced Dataset. *Agronomy*, 12(11), 2613.
- Marques, K. P., Demattê, J. A., Miller, B. A., and Lepsch, I. F. (2018). Geomorphometric segmentation of complex slope elements for detailed digital soil mapping in southeast Brazil. *Geoderma Regional*, 14, e00175.
- Martin, J. B. (2017). Carbonate minerals in the global carbon cycle. *Chemical Geology*, 449, 58-72.
- McDowell, M. L., Bruland, G. L., Deenik, J. L., Grunwald, S., and Knox, N. M. (2012). Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma*, 189, 312-320.
- Mishra, M., Singh, K. K., Pandey, P. C., Devrani, R., Pandey, A. K., Raju, K. P., ... and Pandey, M. (2022). Spectral Indices Across Remote Sensing Platforms and Sensors Relating to the Three Poles: An Overview of Applications, Challenges, and Future Prospects. *Advances in Remote Sensing Technology and the Three Poles*, 83-116.
- Mosleh, Z., Salehi, M. H., Jafari, A., Borujeni, I. E., & Mehnatkesh, A. (2016). The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environmental monitoring and assessment*, 188, 1-13.

- Mousavi, S., Sarmadian, F., Omid, M., & Bogaert, P. (2021). Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Water and Soil*, 35(5), 719-734. doi: 10.22067/jsw.2021.71748.1076. (in Persian)
- Mousavifard, S. M., Momtaz, H., Sepehr, E., Davatgar, N., and Sadaghiani, M. H. R. (2013). Determining and mapping some soil physico-chemical properties using geostatistical and GIS techniques in the Naqade region, Iran. *Archives of Agronomy and Soil Science*, 59(11), 1573-1589.
- Mzid, N., Casa, R., Pascucci, S., Tolomio, M., and Pignatti, S. (2022, July). Assessment of the Potential of PRISMA Hyperspectral Data to Estimate Soil Moisture. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5606-5609). IEEE.
- Padarian, J., Minasny, B., McBratney, A.B., (2020). Machine learning and soil sciences: a review aided by machine learning tools. *Soil* 6, 35–52. <https://doi.org/10.5194/soil-6-35-2020>.
- Panahi, M., Gayen, A., Pourghasemi, H. R., Rezaie, F., and Lee, S. (2020). Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. *Science of the Total Environment*, 741, 139937.
- Pôças, I., Rodrigues, A., Gonçalves, S., Costa, P. M., Gonçalves, I., Pereira, L. S., and Cunha, M. (2015). Predicting grapevine water status based on hyperspectral reflectance vegetation indices. *Remote sensing*, 7(12), 16460-16479.
- Pu, R., Gong, P., and Yu, Q. (2008). Comparative analysis of EO-1 ALI and Hyperion, and Landsat ETM+ data for mapping forest crown closure and leaf area index. *Sensors*, 8(6), 3744-3766.
- Rahmani, A., Sarmadian, F., & Arefi, H. (2022). Digital Mapping of Top-soil Thickness and Associated Uncertainty Using Machine Learning Approach in Some Part of Arid and Semi-arid Lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 53(3), 585-602. doi: 10.22059/ijswr.2022.338007.669195.
- Rahmani, A., Sarmadian, F., Mousavi, S. R., & Khamoshi, S. E. (2020). Application of Geomorphometric attributes in digital soil mapping by using of machine learning and fuzzy logic approaches. *Journal of Range and Watershed Management*, 73(1), 105-124. doi: 10.22059/jrwm.2020.288580.1418. (in Persian)
- Riihimäki, H., Kemppinen, J., Kopecký, M., and Luoto, M. (2021). Topographic Wetness Index as a Proxy for Soil Moisture: The Importance of Flow-Routing Algorithm and Grid Resolution. *Water Resources Research*, 57(10), e2021WR029871.
- Roy, D. P., Kovalsky, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., and Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote sensing of Environment*, 185, 57-70.
- Sahoo, R. N., Ray, S. S., and Manjunath, K. R. (2015). Hyperspectral remote sensing of agriculture. *Current Science*, 848-859.
- Schlerf, M., Atzberger, C., and Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 95(2), 177-194.
- Skentos, Athanasios. "Topographic position index-based landform analysis of Messaria (Ikaria Island, Greece)." *Acta Geobalkanica* 4.1 (2018): 7-15.
- Tagliabue, G., Boschetti, M., Bramati, G., Candiani, G., Colombo, R., Nutini, F., ... and Panigada, C. (2022). Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187, 362-377.
- Tan, C., Samanta, A., Jin, X., Tong, L., Ma, C., Guo, W., ... and Myneni, R. B. (2013). Using hyperspectral vegetation indices to estimate the fraction of photosynthetically active radiation absorbed by corn canopies. *International Journal of Remote Sensing*, 34(24), 8789-8802.
- Thenkabail, P. S., Smith, R. B., and De Pauw, E. (2002). Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric engineering and remote sensing*, 68(6), 607-622.
- Toth, C., and Józków, G. (2016). Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 22-36.
- Tu, Y. H., Phinn, S., Johansen, K., and Robson, A. (2018). Assessing radiometric correction approaches for multi-spectral UAS imagery for horticultural applications. *Remote Sensing*, 10(11), 1684.
- Vaughn Ihlen, Landsat 8 (L8) Data Users Handbook, 2019, Department of the Interior U.S. Geological Survey.
- Vaze, J., Teng, J., and Spencer, G. (2010). Impact of DEM accuracy and resolution on topographic indices. *Environmental Modelling and Software*, 25(10), 1086-1098.
- Wang, C., Li, W., Yang, Z., Chen, Y., Shao, W., and Ji, J. (2015). An invisible soil acidification: Critical role of soil carbonate and its impact on heavy metal bioavailability. *Scientific reports*, 5(1), 12735.



- Wu, C., Niu, Z., Tang, Q., and Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and forest meteorology*, 148(8-9), 1230-1241.
- Zhao, D., Wang, J., Zhao, X., and Triantafylis, J. (2022). Clay content mapping and uncertainty estimation using weighted model averaging. *Catena*, 209, 105791.
- Zolfaghari Nia, M., Moradi, M., Moradi, G., and Taghizadeh-Mehrjardi, R. (2022). Machine Learning Models for Prediction of Soil Properties in the Riparian Forests. *Land*, 12(1), 32.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445-452.
- Hengl, T., Miller, M. A., Križan, J., Shepherd, K. D., Sila, A., Kilibarda, M., ... & Crouch, J. (2021). African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Scientific Reports*, 11(1), 6130.
- Shan, M., Liang, S., Fu, H., Li, X., Teng, Y., Zhao, J., ... & Ma, Z. (2021). Spatial prediction of soil calcium carbonate content based on Bayesian maximum entropy using environmental variables. *Nutrient Cycling in Agroecosystems*, 120, 17-30.
- Lagacherie, P., Baret, F., Feret, J. B., Netto, J. M., & Robbez-Masson, J. M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, 112(3), 825-835.
- Gabriels, D., & Moldenhauer, W. C. (1978). Size distribution of eroded material from simulated rainfall: Effect over a range of texture. *Soil Science Society of America Journal*, 42(6), 954-958.

Digital Mapping of Soil Properties (Calcium carbonate and soil clay percentage) Using Landsat 8 and PRISMA Satellite Images by the Random Forest Algorithm

EXTENDED ABSTRACT

Introduction

Land management is recognized as a major challenge in solving global issues such as food, water, energy, environment, biodiversity and human health. To have accurate information about soil, soil mapping and assessment are essential. Imaging sensors including Landsat 8 and PRISMA with a wide range of spatial, temporal and spectral resolutions are important tools for soil mapping. In this research, the zoning of soil properties using PRISMA and Landsat 8 data and the effect of different inputs on the accuracy and correctness of maps have been investigated using the RF algorithm.

Material and Methods

In this study, soil properties in a region of Abik City in the Qazvin province of Iran were delineated using an RF algorithm based on laboratory data and remote sensing images. L2D surface images from PRISMA and Landsat 8 satellites, as well as a 30-meter digital elevation model, were employed for soil mapping. The remote sensing data underwent preprocessing steps, including the removal of problematic bands, geometric correction, and image calibration using ENVI 5.6 software. Additionally, 12 spectral indices, such as NDVI, and 20 indices were derived by combining two or more bands in the images with the assistance of SAGA GIS software and the digital elevation model. To analyze the spatial diversity of hyperspectral images, the principal component analysis (PCA) method was used and also the effect of different inputs on the RF algorithm were investigated. The random forest algorithm has very high accuracy in predicting soil properties due to its ability to model non-linear relationships between auxiliary variables and the target soil. This algorithm provides the possibility to determine the relative importance of environmental auxiliary variables simply. In this research, to validate the performance of the random forest model using data obtained from laboratory analysis and remote sensing images, RMSE, RPD, RPIQ and R^2 statistics were calculated from the measured values of soil properties.

Results and discussion

The results show that the simultaneous use of spectral indices and principal component analysis does not improve accuracy, and the use of Digital elevation model indices has a great effect on improving the results. In general, using principal component analysis to improve the quality of generated calcium carbonate maps is more effective than using spectral indices. But it cannot be said which of the maps obtained from the analysis of the main components of satellite images or the maps obtained from spectral indices extracted from satellite images are more accurate and precise. Also, the results did not show that the derivatives of Landsat 8 or PRISMA images are more effective for preparing maps of soil CaCO_3 values, but the predicted values using PRISMA and Landsat 8 satellite spectra were consistent with the trend of soil calcium carbonate changes in the region.

Also, the results showed that soil clay percentage maps that PRISMA-based models provide better results compared to Landsat8-based models. The accuracy of PRISMA satellite images for preparing the clay percentage map in the study area is higher than that of the Landsat8 satellite, due to the typical absorption characteristics in the SWIR spectral region. The use of the digital height model and the indicators derived from it increases the performance of the models. Also, the maps obtained from the analysis of the main components of satellite images are less accurate than the maps obtained from spectral indices extracted from satellite images. The values of clay percentage predicted using PRISMA and Landsat8 satellite spectra were consistent with the trend of changes in soil texture and soil clay percentage in the region.

Conclusion

The results showed that PRISMA images are more accurate than Landsat 8 due to their higher number of bands and greater spectral resolution. However, the high cost of PRISMA and its limited access to complete time series images are limitations of this method. The combined use of spectral indices, principal component analysis, and digital height models significantly improved the model's performance. However, a large number of inputs can limit the model's inputs, and optimizing the selection of features from each category can identify the best features and improve the model's performance.

Keywords: Calcium Carbonate, Clay Percentage, PRISMA Hyperspectral Satellite, Random Forest.