



ORIGINAL RESEARCH PAPER

Prediction of Total Soil Nitrogen Variations Using Three Machine Learning Approaches and Remote Sensing Data

Sh. Felegari¹, k. Moravej^{*1}, A. Sharifi², A. Golchin¹, P. Karami³¹ Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran² Department of Geomatics and Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran³ Department of Range and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Kurdistan, Iran

ABSTRACT

Received: 01 January 2024

Reviewed: 18 March 2024

Revised: 13 April 2024

Accepted: 30 April 2024

KEYWORDS:

Boosted Regression Trees
Landsat-9
Random Forest
Sentinel-1
Support Vector Machine
Total Soil Nitrogen

* Corresponding author

✉ kmoravej@znu.ac.ir

☎ (+9824) 33052606

Background and Objectives: Every country relies on soil as a vital natural resource that significantly contributes to environmental conservation and food production. Preparation of soil nutrient distribution map serves as a valuable tool for managers to make decisions. Due to the time-consuming and expensive nature of laboratory analysis for these variables on a large scale, efforts have been made to explore soil nitrogen through remote sensing. The current research deals with the application of remote sensing methods along with regression and random forest models to predict total soil nitrogen in Gilan province. This study aimed to answer two main questions: (1) Can SAR data be used to quantify total soil nitrogen (TSN) (2) How do SVM, BRT and RF algorithms perform in predicting soil nitrogen content?

Methods: This study focused on evaluating the data capabilities of Landsat-9 and Sentinel-1 satellites individually and in combination, using advanced algorithms such as Support Vector Machine (SVM), Boosted Regression Tree (BRT), and Random Forest (RF). The purpose of this evaluation was strategic, aiming to showcase the diverse conditions of the study area based on land cover/land use, climatic, and topographical parameters. Various variables, including climate parameters, topographic components, and remote sensing subscale indices, were investigated in conjunction with SAR data and optical images. Nonlinear machine learning algorithms, specifically SVM, RF, and BRT, were employed to predict total soil nitrogen status by modeling complex relationships between soil properties and environmental variables. R software, utilizing the CARET package for parameter input, was employed to implement the algorithm.

Findings: The results indicated the following: RF and BRT algorithms outperformed SVM and were effective in monitoring total soil nitrogen values. Multi-temporal SAR images showed higher accuracy in monitoring total soil nitrogen content compared to optical remote sensing data, facilitating more realistic predictions in paddy soils. The integration of environmental variables led to an increase in the accuracy of algorithms, where remote sensing variables played a crucial role, contributing to 61% and 51% effects in RF and BRT algorithms, respectively. The comparison of SVM and RF algorithms revealed that RF ranked second after the BRT algorithm, and the accuracy of total soil nitrogen estimation was not achieved with the SVM algorithm. However, both BRT and RF algorithms were able to monitor changes in total soil nitrogen. BRT performed better, accurately recording 58% of changes, as evidenced by a higher R^2 value (0.58) and lower RMSE (0.25 mg/kg) and MAE (0.19 mg/kg) values.

Conclusion: In conclusion, the following key points were extracted from this research: 1) RF and BRT algorithms outperformed SVM in effectively monitoring total soil nitrogen levels; 2) multi-temporal SAR images demonstrated higher accuracy in tracking total soil nitrogen compared to optical remote sensing, enabling precise predictions in paddy soils; 3) the incorporation of environmental variables enhanced algorithmic accuracy; and 4) remote sensing variables contributed 61% and 51% to RF and BRT algorithms, respectively.



NUMBER OF REFERENCES

30



NUMBER OF FIGURES

4



NUMBER OF TABLES

5

مقاله پژوهشی

پیش‌بینی تغییرات نیتروژن کل خاک با استفاده از سه رویکرد متفاوت یادگیری ماشین و داده‌های سنجنش از دور

شیلان فعله‌گری^۱، کامران مروج^{۱*}، علیرضا شریفی^۲، احمد گلچین^۱، پرویز کریمی^۳^۱ گروه مهندسی علوم خاک، دانشگاه زنجان، زنجان، ایران^۲ گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران^۳ گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه کردستان، کردستان، ایران

چکیده

پیشینه و اهداف: هر کشوری به خاک به‌عنوان یک منبع طبیعی حیاتی، متکی است که به‌طور قابل توجهی در حفظ محیط زیست و تولید مواد غذایی کمک می‌کند. تهیه نقشه توزیع عناصر غذایی خاک به‌عنوان یک ابزار ارزشمند برای تصمیم‌گیری مدیران عمل می‌کند. با توجه به ماهیت زمان‌بر و پرهزینه آنالیز آزمایشگاهی برای این متغیرها در مقیاس بزرگ، تلاش‌هایی برای کاهش نیتروژن خاک از طریق سنجنش از دور انجام شده است. پژوهش حاضر، به کاربرد روش‌های سنجنش از دور همراه با مدل‌های رگرسیون و تصادفی جنگل برای پیش‌بینی نیتروژن کل خاک در استان گیلان می‌پردازد. این مطالعه، با هدف پاسخ به دو سوال اصلی انجام شد: (۱) آیا می‌توان از داده‌های SAR برای تعیین مقدار نیتروژن کل خاک استفاده کرد؟ (ب) الگوریتم‌های ماشین بردار پشتیبانی (SVM)، درخت رگرسیون تقویت‌شده (BRT) و جنگل تصادفی (RF) در پیش‌بینی محتوای نیتروژن خاک چگونه عمل می‌کنند؟

روش‌ها: این مطالعه بر ارزیابی قابلیت‌های داده‌های ماهواره‌های Landsat-9 و Sentinel-1 به‌صورت جداگانه و ترکیبی، با استفاده از الگوریتم‌های پیشرفته مانند SVM، RF و BRT متمرکز شد. نمونه‌برداری هدفمند به‌صورت استراتژیک انجام شد تا شرایط متنوع منطقه مورد مطالعه را بر اساس پوشش زمین/ کاربری زمین، پارامترهای اقلیمی و توپوگرافیکی نشان دهد. متغیرهای مختلف، از جمله پارامترهای آب‌وهوایی، اجزای توپوگرافی، و شاخص‌های زیرگروه سنجنش از دور، در ارتباط با داده‌های SAR و تصاویر نوری مورد بررسی قرار گرفتند. الگوریتم‌های یادگیری ماشین غیرخطی، به‌ویژه SVM، RF و BRT برای پیش‌بینی وضعیت نیتروژن کل خاک با مدل‌سازی روابط پیچیده بین خواص خاک و متغیرهای محیطی مورد استفاده قرار گرفتند. نرم افزار R، با استفاده از بسته CARET برای ورودی پارامتر، برای پیاده‌سازی الگوریتم استفاده شد.

یافته‌ها: نتایج، حاکی از موارد زیر بود: عملکرد الگوریتم‌های RF و BRT از SVM پیشی گرفت و در پایش مقادیر نیتروژن کل خاک مؤثر بود، تصاویر SAR چند زمانی دقت بالاتری را در نظارت بر محتوای نیتروژن کل خاک در مقایسه با داده‌های سنجنش از دور نوری نشان دادند و پیش‌بینی‌های واقعی‌تر را در خاک‌های شالیزارهای تسهیل کردند، ادغام متغیرهای محیطی منجر به افزایش دقت الگوریتم‌ها شد. متغیرهای سنجنش از دور، نقش مهمی را در این پژوهش ایفا کردند و به ترتیب، تأثیر ۶۱ و ۵۱ درصدی در الگوریتم‌های RF و BRT گزارش شد. مقایسه الگوریتم‌های SVM و RF نشان داد که RF پس از اجرای الگوریتم BRT در رتبه دوم قرار دارد و دقت برآورد نیتروژن کل خاک با الگوریتم SVM به‌دست نیامد. با این حال، الگوریتم‌های BRT و RF قادر به نظارت بر تغییرات نیتروژن کل خاک بودند و BRT با دقت دقیق ۵۸٪ تغییرات به‌دلیل مقدار R2 بالاتر (۰/۵۸) و مقادیر RMSE (۰/۲۵ میلی‌گرم بر کیلوگرم) و MAE (۰/۱۹ میلی‌گرم بر کیلوگرم) کمتر، عملکرد بهتری داشت.

نتیجه‌گیری: در انتها، موارد کلیدی زیر به‌عنوان نتیجه‌گیری کلی از این پژوهش استخراج شد: (۱) الگوریتم‌های RF و BRT در نظارت بر سطوح نیتروژن کل خاک به‌طور مؤثر بهتر از SVM عمل کردند، (۲) تصاویر SAR چند زمانی دقت بالایی در ردیابی نیتروژن کل خاک در مقایسه با سنجنش از دور نوری ارائه می‌دهند و پیش‌بینی دقیق را در خاک‌های شالیزارهای تسهیل می‌کنند، (۳) ترکیب متغیرهای محیطی باعث افزایش دقت الگوریتمی شد و (۴) متغیرهای سنجنش از دور به ترتیب ۶۱ و ۵۱ درصد به الگوریتم‌های RF و BRT کمک کردند.

تاریخ دریافت: ۱۱ دی ۱۴۰۲
تاریخ داوری: ۲۸ اسفند ۱۴۰۲
تاریخ اصلاح: ۲۴ فروردین ۱۴۰۳
تاریخ پذیرش: ۱۱ اردیبهشت ۱۴۰۳

واژگان کلیدی:

درختان رگرسیون تقویت شده
لندست-۹
جنگل تصادفی
سنتینل-۱
ماشین بردار پشتیبانی
نیتروژن کل خاک

* نویسنده مسئول

✉ kmoravej@znu.ac.ir

① ۰۲۴-۲۳۰۵۲۶۰۶

مقدمه

سوی دیگر، در زمین‌های زراعی کنیا، SVM برای تخمین کربن آلی خاک سطحی مؤثرتر از RF بود. عملکرد متضاد الگوریتم‌ها در این دو مکان را می‌توان به انواع خاک، الگوهای آب‌وهوا، و شیوه‌های کاربری زمین در هر منطقه نسبت داد که بر توانایی‌های الگوریتم‌ها برای مدیریت تغییرات مختلف خواص خاک تأثیر می‌گذارد [۱۲].

اخیراً، محققان از ترکیب تکنیک‌های سنجش از دور و الگوریتم‌های یادگیری ماشین برای مدل‌سازی پارامترهای خاکی مانند توزیع نیتروژن کل خاک استفاده کرده‌اند. در مطالعه‌ای در منطقه خاک سیاه شمال شرقی چین از الگوریتم RF و تصاویر Sentinel-2A برای ایجاد نقشه توزیع فضایی نیتروژن کل خاک استفاده شد [۱۱]. در این مطالعه، ۱۰۴ نمونه خاک سطحی از سایت‌های زمین کشاورزی جمع‌آوری شد و از ۲۱ متغیر پیش‌بینی‌کننده برای برآورد نیتروژن کل خاک استفاده شد. الگوریتم RF در ثبت دقیق تغییرات در نیتروژن کل خاک مؤثر بود و استفاده از متغیرهای پیش‌بین مناسب عملکرد آن را بیشتر بهبود بخشید. به‌طور مشابه، در مطالعه‌ای دیگر نیتروژن کل خاک در شهر لوشون در شمال شرقی استان لیائونینگ چین ترسیم شد. در این مطالعه، از ۱۱۵ نمونه خاک و ۱۲ متغیر کمکی استفاده و سه الگوریتم رگرسیون گام به گام و RF آزمایش شد که الگوریتم RF دقیق‌ترین نتایج را ارائه داد.

همچنین، مطالعات متعددی استفاده از رادار دهانه مصنوعی (SAR) را برای بررسی خواص شیمیایی خاک، به‌ویژه محتوای نیتروژن کل، با استفاده از تصاویر ماهواره‌ای مورد بررسی قرار داده‌اند. به‌عنوان مثال، در مطالعه‌ای از داده‌های Landsat-7 و Sentinel-1 برای بررسی محتوای کربن آلی خاک استفاده شد [۱۳] و با موفقیت تغییرات کربن آلی خاک در شرق چین ترسیم شد. به‌طور مشابه، سدی و همکاران [۱۴] از تصاویر نوری و داده‌های باندهای L ALOS PALSAR برای مطالعه محتوای کربن آلی خاک در خاک‌های آمزون استفاده کرد و دریافت که ترکیب ضریب پراکندگی عقب داده‌های SAR باعث افزایش دقت پیش‌بینی می‌شود [۱۴]. عامل بین امواج الکترومغناطیسی و ذرات خاک استنتاج خواص شیمیایی خاک را از طریق داده‌های SAR امکان‌پذیر می‌کند. همان‌طور که امواج به سطح خاک نفوذ می‌کنند، تغییری در قطبش ایجاد می‌کنند که می‌تواند برای استنباط اطلاعات در مورد خواص خاک اندازه‌گیری شود. محتوای رطوبت خاک یکی از ویژگی‌هایی است که می‌توان از داده‌های SAR استنباط کرد، زیرا به شدت به ثابت دی الکتریک خاک مربوط می‌شود. هنگامی که رطوبت خاک افزایش می‌یابد، ثابت دی الکتریک خاک نیز افزایش می‌یابد که منجر به تغییر در قطبش امواج الکترومغناطیسی می‌شود. علاوه بر این، بافت خاک و محتوای ماده آلی را می‌توان به داده‌های SAR مرتبط دانست، زیرا آن‌ها به ترتیب بر توزیع ذرات خاک و خواص دی الکتریک خاک تأثیر می‌گذارند. تجزیه و تحلیل داده‌های SAR می‌تواند اطلاعات ارزشمندی در مورد ویژگی‌های خاک ارائه دهد که می‌تواند برای کاربردهای مختلف، از جمله پیش‌بینی عملکرد محصول، مدیریت کاربری اراضی، و نظارت بر محیط زیست مفید باشد [۱۵-۱۷]. ادغام این داده‌های اضافی برای بهبود دقت

خاک، در تولید غذا و حفظ و نگهداری محیط نقش مهمی دارد. افزایش و کاهش مقدار نیتروژن در خاک در اثر شیوه‌های مختلف کشاورزی، منجر به آلودگی میکروبی خاک، کاهش شدید بهره‌وری و سلامت خاک و نقصان در محصولات کشاورزی مورد نیاز انسان می‌شود [۱]. با توجه به نگرانی‌های زیست محیطی و سلامتی و نقش اساسی خاک و کیفیت آن در سلامت بوم‌سازگان، آگاهی از میزان نیتروژن کل به‌عنوان شاخص مهمی در کیفیت خاک و محصولات کشاورزی ضروری است. فناوری سنجش از دور چندین مزیت را نسبت به روش‌های سنتی برای اندازه‌گیری محتوای مواد مغذی خاک ارائه می‌کند [۲]. در واقع، داده‌های سنجش از دور با استفاده از فن‌آوری‌های نقشه‌برداری دیجیتال خاک و مدل‌های پیش‌بینی‌کننده آن فرصت‌های جدیدی برای نظارت بر ویژگی‌های خاک مثل بررسی منابع نیتروژن و چگونگی توزیع آن‌ها را از فضا فراهم می‌کند [۳]. رویکردهای نقشه‌برداری دیجیتال خاک، از جمله استفاده از الگوریتم‌های جنگل تصادفی (RF)، مدل خطی تعمیم یافته (GLM)، ماشین بردار پشتیبان (SVM)، درخت رگرسیون تقویت شده (BRT) و شبکه‌های عصبی مصنوعی (ANNs)، قابلیت پیش‌بینی تغییرات فضایی پارامترهای مختلف خاکی مانند محتوای نیتروژن در خاک را فراهم می‌کند [۴]. این رویکردها، نه تنها نیاز به کار میدانی گسترده و هزینه‌های مرتبط را کاهش می‌دهد، بلکه فرصت‌های جدیدی را برای پایش ویژگی‌های خاک فراهم می‌کنند [۵]. به‌طور کلی، بهترین الگوریتم جهانی برای پیش‌بینی خواص خاک وجود ندارد و انتخاب الگوریتم به ویژگی‌های منحصر به فرد محل مطالعه و متغیرهای در نظر گرفته شده بستگی دارد [۶-۸]. برای انتخاب مناسب‌ترین الگوریتم، محققان باید الگوریتم‌ها را از یک منظر واحد با در نظر گرفتن شرایط خاص منطقه مورد مطالعه، آزمایش و مقایسه کنند. این رویکرد، کمک می‌کند تا اطمینان حاصل شود که الگوریتم انتخاب شده به خوبی با ویژگی‌های خاص خاک و عوامل محیطی موجود در آن مکان مطابقت دارد و در نتیجه، پیش‌بینی‌های دقیق‌تری به‌دست می‌آید [۹]. به‌عنوان مثال، در مطالعه انجام شده در نیجریه، الگوریتم جنگل تصادفی (RF) نتایج قابل اعتمادتری نسبت به الگوریتم درخت رگرسیون تقویت شده (BRT) برای مطالعه ویژگی‌های خاک ارائه داد [۱۰]. این نتیجه، ممکن است به ترکیب خاصی از خواص خاک، پوشش گیاهی، و شرایط آب‌وهوایی در نیجریه نسبت داده شود که به نفع عملکرد الگوریتم RF است. به‌طور مشابه، در مطالعه اکوسیستم آلپ، الگوریتم BRT از الگوریتم RF بهتر عمل کرد [۱۱]. این نتیجه می‌تواند به ویژگی‌های منحصر به فرد منطقه آلپ، مانند انواع مختلف خاک، ترکیبات پوشش گیاهی، و توپوگرافی مرتبط باشد، که ممکن است به نفع توانایی الگوریتم BRT برای ثبت تعاملات پیچیده بین خواص خاک و سایر عوامل محیطی باشد [۱۱]. علاوه بر این، در حوضه آبخیز دریاچه Swyang در کره جنوبی، الگوریتم RF برای تخمین نیتروژن کل خاک در مقایسه با الگوریتم ماشین بردار پشتیبانی (SVM) بهتر عمل کرد. از

روش تحقیق

منطقه مورد مطالعه این تحقیق، در ناحیه مرکزی استان گیلان در شمال ایران با مختصات بین طول جغرافیایی ۴۹ درجه و ۱۹ دقیقه و ۴۰ ثانیه و ۴۹ درجه و ۳۲ دقیقه و ۲۰ ثانیه شرقی و عرض جغرافیایی ۳۷ درجه و ۱۹ دقیقه و ۲۰ ثانیه و ۳۷ درجه و ۲۴ دقیقه و ۵۵ ثانیه شمالی واقع شده است. بر اساس آمار ۲۰ ساله ایستگاه هواشناسی محلی، میانگین بارندگی سالانه در منطقه ۱۲۴۶ میلی‌متر و میانگین دمای سالانه ۱۶/۸ درجه سانتی‌گراد است. گزارش داده‌اند که زمین منطقه مورد مطالعه با مناطق پست با شیب کمتر از ۰/۵ درصد مشخص می‌شود [۱] و کاربری اصلی زمین برای شالی‌کاری است. منطقه مورد مطالعه و محل‌های نمونه‌برداری به ترتیب در شکل ۱ نشان داده شده است.

داده‌های خاک

۱۰۰ نمونه خاک سطحی از اعماق ۰ تا ۲۰ سانتی‌متر در شالیزارهای منطقه مورد مطالعه برداشت شد. زمان برداشت نمونه‌های خاک هفته آخر شهریور ۱۳۹۹ بود. به منظور اعتبارسنجی و کالیبراسیون الگوریتم‌ها جمع‌آوری شد. با توجه به دشواری و ماهیت زمان‌بر بودن نمونه‌برداری خاک در منطقه مورد مطالعه، نمونه برداری هدفمند در مکان‌هایی انجام شد که معرف شرایط منطقه مورد مطالعه بر اساس پارامترهای پوشش زمین/ کاربری زمین، اقلیمی و پارامترهای توپوگرافی بود. در هر محل نمونه‌برداری، یک گودال یک متری برای طبقه‌بندی و بررسی خاک زیرسطحی حفر شد. مختصات GPS برای هر نقطه نمونه‌برداری ثبت شد. پس از هوا خشک شدن نمونه‌های خاک و عبور از الک ۲ میلی‌متری، نیتروژن کل خاک به روش کج‌لدال اندازه‌گیری شد [۱۸].

پیش‌بینی الگوریتم‌هایی مانند RF و رگرسیون گام به گام چندگانه ضروری است. با در نظر گرفتن مجموعه گسترده‌تری از متغیرهای پیش‌بینی‌کننده، مدل‌ها می‌توانند برهمکنش‌ها و الگوهای پیچیده‌تری را در داده‌های خاک ثبت کنند که منجر به نتایج دقیق‌تری می‌شود. با توجه به تأثیر بالقوه فعالیت‌های کشاورزی بر سطوح نیتروژن خاک، مطالعه‌ای به منظور بررسی تغییرات زمانی و مکانی نیتروژن خاک در مزارع برنج شمال ایران انجام شد. این مطالعه، با هدف پاسخ به دو سوال اصلی انجام شد: (۱) آیا می‌توان از داده‌های SAR برای تعیین مقدار نیتروژن کل خاک استفاده کرد؟ (ب) الگوریتم‌های SVM، BRT و RF در پیش‌بینی محتوای نیتروژن خاک چگونه عمل می‌کنند؟ برای پرداختن به این سوالات، این مطالعه از متغیرهای پیش‌بینی‌کننده مانند باندها و شاخص‌های سنجش از دور، پارامترهای توپوگرافی، کاربری زمین/ پوشش زمین (LULC)، داده‌های اقلیمی و ۱۰۰ نمونه خاک استفاده شد. شکاف پژوهشی در این مطالعه حول محور نیاز به روش‌های کارآمد برای پیش‌بینی نیتروژن کل خاک است. تجزیه و تحلیل آزمایشگاهی برای نقشه‌های توزیع عناصر در خاک هم‌زمان بر و هم‌گران است. بنابراین، نیاز به کشف روش‌های جایگزین مانند سنجش از دور برای مقابله با این چالش وجود دارد. به‌طور خلاصه در این پژوهش تلاش شد اطلاعاتی درباره مدل‌های پیش‌بینی‌کننده نیتروژن خاکی و رویکردهای نقشه‌برداری دیجیتالی خاک به خواننده داده شود. در بخش مواد و روش‌ها به واکاوی سه الگوریتم یادگیری غیرخطی SVM، RF و BRT برای پیش‌بینی وضعیت نیتروژن خاک به کمک نرم‌افزار R پرداخته شد. در بخش نتایج، به بررسی خروجی‌های حاصل از سه الگوریتم یادگیری غیرخطی SVM، RF و BRT پرداختیم و بر اساس نتایج حاصل از مقادیر R2 برتری هر کدام از مدل‌ها نسبت به دیگری شرح داده شد.



شکل ۱: موقعیت منطقه مورد مطالعه و نقاط نمونه برداری
Fig. 1: Location of the study area and sampling points

متغیرهای محیطی

امکان ارجاع مکانی دقیق را فراهم می‌کند. علاوه بر این، فیلتر لکه‌های برای کاهش نویز ذاتی موجود در تصاویر SAR استفاده شد. این تکنیک فیلتر کردن، با استفاده از یک فیلتر پنجره 5×5 ، به صاف کردن نویز لکه‌ها و حفظ ویژگی‌های مهم تصویر کمک می‌کند [۱۵]. تمام این مراحل پردازش با استفاده از نرم‌افزار SARscape 5.2 انجام شد، ابزاری تخصصی که برای تجزیه و تحلیل و دستکاری داده‌های SAR طراحی شده است. این نرم‌افزار مجموعه‌ای جامع از عملکردهای متناسب با وظایف پردازش داده‌های SAR را ارائه می‌دهد.

تصاویر با استفاده از Aster GDEM به مقیاس دسی بل تبدیل شدند. شاخص NDVI با استفاده از باند ۴ و ۵ برای شناسایی نوع و شدت پوشش گیاهی در منطقه مورد مطالعه محاسبه شد. روش تصحیح هندسی چند جمله‌ای برای تصحیح تمام تصاویر استفاده شد. اهداف تحقیق با استفاده از باندهای ۴، ۵ و ۶ لندست-۹، که به ترتیب نشان دهنده رشد گیاهی، پوشش و زیست توده هستند، با طول موج‌های ۰/۶۴ تا ۰/۶۷ میکرومتر برای باند ۴، ۰/۸۵ تا ۰/۸۸ میکرومتر برای باند ۵ و ۱/۵۷ تا ۱/۶۵ میکرومتر برای باند ۶ انجام شد. در مقابل ۸ متغیر محیطی به دست آمده از تصاویر سنجش از دور شامل BC-1، BC-2، BC-3، BC-4، Band-4، Band-5، Band-6، NDVI بودند. لازم به ذکر است که تصاویر به دست آمده کمتر از ۱۰ درصد پوشش ابری داشتند.

جدول ۱: اطلاعات مربوط به تصاویر سنتینل-۱
Table 1: information related to Sentinel-1 images

تاریخ Date	پرتو Beam	قطبی شدن Polarization	زاویه وقوع Incident Angle (°)	جهت Direction
28 June 2021	WI	VV	39.36	Ascending
28 July 2021	WI	VV	39.02	Descending
28 August 2021	WI	VV	39.02	Descending
28 September 2021	WI	VV	39.35	Ascending

متغیرهای آب و هوا

پارامترهای اقلیمی برای منطقه مورد مطالعه با استفاده از داده‌های وب سایت CARD و یک ایستگاه هواشناسی واقع در منطقه مورد مطالعه به دست آمد. مجموعه داده‌های MAT و MAP برای دوره مطالعه نیز به دست آمد. یک دوره ۲۰ ساله (۱۴۰۰-۱۳۸۰) در نظر گرفته شد و میانگین دو داده MAP و MAT برای تعیین متغیر پیش‌بینی محاسبه شد. برای درونیابی داده‌ها، روش مدل‌سازی سطح HASM انتخاب شد زیرا نشان داده شده است که نتایج بهتری نسبت به سایر روش‌های درونیابی ارائه می‌دهد.

داده‌های کاربری/پوشش زمین (Land Use/Land Cover)

در این مطالعه، از یک رویکرد طبقه‌بندی نظارت شده برای تولید داده‌های کاربری/پوشش زمین (Land Use/Land Cover) استفاده شد. به‌طور خاص ما از الگوریتم طبقه‌بندی جنگل تصادفی به دلیل اثربخشی آن در مدیریت مجموعه داده‌های بزرگ و پیچیده و همچنین توانایی آن در ارائه نتایج طبقه‌بندی قابل اعتماد استفاده کردیم. الگوریتم جنگل

این مطالعه، متغیرهای مختلفی مانند پارامترهای آب و هوایی، اجزای توپوگرافی، شاخص‌های زیرگروه سنجش از دور و LULC را بررسی کرد. در مجموع ۱۵ متغیر، از جمله نیتروژن کل خاک، ارتفاع، شیب، جهت، شاخص رطوبت توپوگرافی (TWI) و تصاویر داندودی از باند ۴، ۵، ۶ و لندست-۹ و همچنین NDVI، میانگین بارش سالانه (MAP) و میانگین دمای سالانه (MAT) مورد بررسی قرار گرفتند. تغییرات این پارامترها با استفاده از نرم‌افزار QGIS با تفکیک مکانی ۳۰ متر ترسیم شد، و همچنین با استفاده از این نرم‌افزار نقشه نهایی از مقادیر متغیرهای مورد بررسی و مقادیر اندازه‌گیری نیتروژن کل خاک تهیه شد. برای مدل‌سازی، مقدار پیکسل متغیرهای پیش‌بینی‌کننده به محل نمونه-برداری از خاک اختصاص داده شد. همچنین در این پژوهش شاخص گیاهی تفاوت نرمال شده (NDVI: Normalized Difference Vegetation Index) نیز به‌عنوان یکی از پارامترهای ورودی مورد نیاز محاسبه شد. مقدار شاخص NDVI بین -۱ تا +۱ متغیر است. در واقع، این شاخص معیاری از فعالیت فتوسنتزی گیاه است. هرچه مقدار NDVI بیشتر باشد، پوشش‌های گیاهی متراکم با سلامت بیشتر انتظار می‌رود. مقادیر منفی شاخص NDVI عموماً پوشش‌های آبی است. همچنین مقادیر بین صفر تا ۰/۲ را به اراضی فاقد پوشش گیاهی می‌توان نسبت داد [۱۸]. شاخص NDVI، تفاضل نرمال شده بازتاب در باندهای نزدیک مادون قرمز (NIR) و قرمز به فرم رابطه زیر است [۲۴]:

$$NDVI = \frac{R_{NIR} - R_{red}}{R_{NIR} + R_{red}}$$

در این فرمول، NIR و RED مقادیر بازتاب اندازه‌گیری شده در باند طیفی نزدیک به مادون قرمز و قرمز را نشان می‌دهند [۲۴].

داده‌های سنجش از دور

متغیرهای سنجش از دور مورد استفاده در این مطالعه شامل داده‌های SAR و تصاویر نوری بودند. تصاویر سنتینل-۱ از آژانس فضایی اروپا (ESA) و داده‌های لندست-۹ از سازمان زمین‌شناسی ایالات متحده (USGS) داندود شد. حسگر سنتینل-۱ از دو صورت فلکی ماهواره‌ای Sentinel-1A و Sentinel-1B تشکیل شده است و چهار تصویر پیچیده تک‌ظاهری در حالت تداخل سنجی گسترده (WI) استفاده شده است [۱۵]. در این پژوهش از ۴ (۲ تصویر از سنتینل-۱ و ۲ تصویر از لندست-۹) تصویر استفاده شد. جدول ۱ اطلاعات دقیقی در مورد داده‌های SAR ارائه می‌دهد. پردازش داده‌های SAR شامل چندین مرحله حیاتی برای اطمینان از دقت و قابلیت استفاده است. کالیبراسیون، که برای ویژگی‌های خاص حسگر و عوامل محیطی تنظیم می‌شود، برای افزایش قابلیت اطمینان داده‌ها انجام شد. رادیومتری، یکی دیگر از فرآیندهای ضروری، با هدف تصحیح تغییرات در شدت سیگنال ناشی از عواملی مانند زمین و زاویه تابش است و در این پژوهش بر روی داده‌ها اعمال شد. ژئوکدینگ، بخشی جدایی‌ناپذیر از پردازش داده‌های SAR، شامل اختصاص مختصات جغرافیایی به هر پیکسل در تصویر است که

RF و BRT را می‌توان برای تقویت مزایای درختان ترکیب کرد و تغییرات متغیرهای پاسخ را از متغیرهای پیش‌بینی تحلیل کرد. برای تقویت الگوریتم، باید یک رویکرد تکراری اتخاذ شود و به‌طور مداوم درختان را به آن اضافه کرد. ویژگی اصلی الگوریتم BRT افزایش دقت پیش‌بینی و تطبیق نامعقول داده‌ها است که می‌تواند به طور کاذب گرادبان تصادفی را افزایش دهد.

تجزیه و تحلیل آماری

نرم افزار R برای تولید نقشه‌های نیتروژن کل خاک با استفاده از بسته‌های مختلف، از جمله RF-Package (BRT)، GBM-Package (SVM) و MAP TOOLS-Package (SVM)، KERN-Lab-Package (SVM) استفاده شد. علاوه بر این، برای انجام تجزیه و تحلیل آماری و تجزیه و تحلیل متغیرهای محیطی از SPSS نسخه ۲۴ استفاده شد. برای ارزیابی اثربخشی داده‌های سنتینل-۱ در نقشه‌برداری نیتروژن کل خاک نیتروژن کل خاک، از ترکیب‌های مختلفی از متغیرهای محیطی استفاده شد، همان‌طور که در جدول ۲ نشان داده شده است. عملکرد این ترکیب‌های مختلف با استفاده از سه الگوریتم BRT، SVM و RF پیش‌بینی شد.

جدول ۲: تمام متغیرهای مورد نیاز برای نیتروژن کل خاک
Table 2: All variables required for mapping of TSN

شماره	مدل	توصیف
ID	Model	Description
I	Model A	ترکیبی از متغیرهای آب و هوایی، توپوگرافی، متغیرهای به‌دست آمده از Landsat-9 و کاربری و پوشش زمین A combination of climate variables, topography, variables obtained from Landsat-9 and Land Use and Land Cover
II	Model B	متغیر به‌دست آمده از لندست-۹ Variable obtained from Landsat-9
III	Model C	متغیر به‌دست آمده از سنتینل-۱ Variable obtained from Sentinel-1
IV	Model D	متغیر مربوط به سنجش از دور (مانند متغیرهای به‌دست آمده از لندست-۹ و سنتینل-۱) The variable related to remote sensing (such as the variables obtained from Landsat-9 and Sentinel-1)
V	Model E	ترکیبی از سنجش از دور (لندست-۹ و سنتینل-۱)، آب‌وهوا، توپوگرافی، کاربری اراضی و متغیرهای پوشش زمین A combination of remote sensing (Sentinel1 and landsat9), climate, topography, land use and land cover variables

از معادلات (۱) تا (۳) [۲۰] برای تعریف میانگین خطای مطلق (MAE)، ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین (R2) استفاده شد که به‌عنوان شاخص در تجزیه و تحلیل داده‌ها استفاده شد:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad [20] \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad [20] \quad (2)$$

تصادفی یک تکنیک یادگیری ماشین است که با ساخت تعداد زیادی درخت تصمیم در طول مرحله آموزش و سپس ترکیب پیش‌بینی‌های آن‌ها برای تولید خروجی طبقه‌بندی نهایی عمل می‌کند. این رویکرد یادگیری گروهی، دقت و استحکام مدل طبقه‌بندی را افزایش می‌دهد، به‌ویژه در سناریوهایی با داده‌های ورودی با ابعاد بالا و عدم تعادل کلاس. در طول مرحله آموزش، ما از مجموعه‌ای از نمونه‌های آموزشی حاوی اطلاعات طیفی و برچسب‌های حقیقی زمینی که کلاس‌های مختلف پوشش زمین را نشان می‌دهند، استفاده کردیم. این نمونه‌ها برای آموزش مدل طبقه‌بندی جنگل تصادفی برای تشخیص و تمایز بین انواع پوشش‌های زمین بر اساس امضای طیفی آن‌ها استفاده شد. هنگامی که طبقه‌بندی کننده جنگل تصادفی آموزش داده شد، آن را روی تصاویر ماهواره‌ای که منطقه مورد مطالعه را پوشش می‌دهند، اعمال کردیم تا هر پیکسل را در یکی از کلاس‌های پوشش زمین از پیش تعریف شده طبقه‌بندی کنیم. خروجی این فرآیند به لایه‌های نهایی نقشه خروجی اضافه شد. این لایه، علاوه بر این که توزیع فضایی طبقات پوشش زمین را در سراسر منطقه مورد مطالعه به تصویر می‌کشد، تمرکز اصلی آن بر روی نواحی با کشت برنج (شالیزارها) می‌باشد. در این پژوهش، با توجه به اینکه منطقه مورد مطالعه در بخش مرکزی استان گیلان در شمال ایران واقع شده است، کلاس‌های طبقات پوشش زمین شامل: جنگل‌ها، علفزارها، زمین‌های زراعی، تالاب‌ها، مناطق شهری، بدنه‌های آبی (مانند رودخانه‌ها و دریاچه‌ها) و زمین‌های بایر است. بر اساس هدف این پژوهش تمرکز ما بر کلاس زمین‌های زراعی و در این کلاس توجه به اراضی شالیکاری می‌باشد.

الگوریتم‌های پیش‌بینی

الگوریتم‌های یادگیری ماشین غیرخطی، از جمله SVM، RF و BRT، در این مطالعه برای پیش‌بینی وضعیت نیتروژن کل خاک با مدل‌سازی روابط پیچیده بین خواص خاک و متغیرهای محیطی مورد استفاده قرار گرفت. نرم افزار R برای وارد کردن پارامترهای مختلف به الگوریتم‌ها با استفاده از رویکرد جستجوی شبکه با بسته CARET استفاده شد. الگوریتم SVM که از یادگیری آماری سرچشمه می‌گیرد، برای طبقه‌بندی تمام داده‌ها به‌صورت جداگانه استفاده شد [۱۹]. برای بهبود دقت مدل SVM، از توابع پایه شعاعی، چند جمله‌ای، سیگموئیدی و خطی استفاده شد. مدل الگوریتم SVM در تهیه نقشه‌های نیتروژن کل خاک استفاده شد که RBF به‌عنوان هسته عملکرد آن در نظر گرفته شد. الگوریتم دیگری که مبتنی بر یادگیری ماشین است، RF است که برای مطالعات رگرسیون و طبقه‌بندی استفاده می‌شود. الگوریتم RF از درختان زیادی برای دستیابی به پیش‌بینی هدف استفاده می‌کند، جایی که هر درخت با استفاده از بوت استرپ داده‌های آموزشی ساخته می‌شود. نمونه‌ها به دو گروه نمونه‌های داخل کیسه و نمونه‌های بیرون کیسه تقسیم می‌شوند و میزان خطا در نمونه‌های بیرون کیسه قابل تخمین است. فلسبرگ و همکاران [۲۵] توضیح دادند که الگوریتم‌های

و مقدار متوسط نیتروژن ۰/۹۲ میلی گرم بر کیلوگرم گزارش شد. توزیع محتوای نیتروژن تغییر جزئی با مقدار متوسط بالاتر از انحراف استاندارد نشان داد. توزیع چولگی، با ضریب چولگی ۰/۴ ثبت شد. پنج مدل مختلف (مدل‌های A، B، C، D و E) با ترکیب متغیرهای محیطی مختلف برای نقشه برداری نیتروژن کل خاک ایجاد شدند. جدول ۳ تمام اطلاعات مربوط به مقدار نیتروژن کل خاک و متغیرهای پیش‌بینی کننده نمونه‌های خاک را نشان می‌دهد. در توصیف مدل‌های طراحی شده برای این پژوهش می‌توان اشاره کرد که مدل A فاقد اطلاعات و داده‌های نوری بود، در حالی که مدل‌های B و C اثرات SAR و داده‌های نوری را بررسی کردند. مدل D از اطلاعات و داده‌های تصاویر Sentinel-1 و Landsat-9 استفاده کرد و مدل E تمام متغیرهای مربوط به داده‌های سنجش از دور را بدون محدودیت در نظر گرفت. عملکرد این مدل‌ها با استفاده از سه الگوریتم (RF، SVM و BRT) ارزیابی شده و در جدول ۴ ارائه شده است.

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O}_i)}{\sum_{i=1}^n (P_i - \bar{O}_i)} \quad [20] \quad (3)$$

معادلات ۱ تا ۳ به ترتیب MAE، RMSE و R^2 را تعریف می‌کنند، جایی که P_i و O_i محتوای نیتروژن کل خاک مشاهده شده و پیش‌بینی شده را نشان می‌دهند و n تعداد نقاط نمونه است. لازم به ذکر است که واحدهای RMSE و MSE در هر پژوهش با واحدهای پارامتر مورد بحث در آن پژوهش مطابقت دارند. در این مطالعه، تلاش شد تغییرات نیتروژن کل خاک بر حسب میلی گرم بر کیلوگرم اندازه‌گیری شود، پس منطبق بر واحد نیتروژن کل خاک واحد دو پارامتر آماری RMSE و MSE نیز میلی گرم بر کیلوگرم (در این پژوهش) می‌باشد.

نتایج و بحث

مقادیر نیتروژن کل خاک نمونه‌ها از ۰/۲۱ تا ۱/۴ میلی گرم بر کیلوگرم متغیر بود، همچنین مقدار انحراف استاندارد ۰/۴۴، مقدار میانگین ۰/۷۹

جدول ۳: داده‌های آماری پارامترهای محیطی و مقادیر نیتروژن کل خاک.
Table 3: Statistical data of Environmental parameters and TSN values.

	Minimum	Maximum	Mean	Median	Standard Deviation	Skewness
نیتروژن کل خاک (گرم بر کیلوگرم) TSN (g/Kg)	.21	1.40	.79	.92	.44	-.40
ارتفاع (متر) Elevation (m)	500	1700	1050	987	231.36	.68
جهت (درجه) Aspect (Deg)	0.0	345	210	134	101	.41
شیب (درجه) Slope (Deg)	.28	15.32	12.01	4.65	4.68	.94
TWI	6.21	16.35	9.24	9.45	3.74	.81
BC-1 (db)	-24.13	1.37	-14.26	-14.56	3.79	1.27
BC-2 (db)	-16.45	-4.21	-10.31	-9.41	2.37	-.23
BC-3 (db)	-16.87	-2.31	-10.23	-9.4	1.84	-.28
BC-4 (db)	-25.14	-6.94	-17.36	-17.54	2.64	.21
باند ۴ (عدد دیجیتالی) Band-4 (digital number)	324.13	3145.36	1084.51	927.45	841.36	.70
باند ۵ (عدد دیجیتالی) Band-5 (digital number)	2047.31	4861.58	3316.74	3612.34	641.52	-.23
باند ۶ (عدد دیجیتالی) Band-6 (digital number)	1362.47	4123.54	2145.78	3478.31	674.42	.80
شاخص گیاهی تفاوت نرمال شده NDVI	.04	.76	.50	.63	.37	-.31
میانگین بارندگی سالیانه (میلی‌متر) MAP (mm)	92	149	124	91.5	19.6	.48
میانگین درجه حرارت سالیانه (درجه سانتی‌گراد) MAT (°C)	4.97	7.32	6.54	7.0	.60	-.70

جدول ۴: کارایی مدل‌های مورد استفاده در رابطه با الگوریتم‌های RF، SVM و BRT
Table 4: The efficiency of the used models in relation to the RF, SVM, and BRT algorithms

	Model	MAE (mg/kg)	RMSE (mg/kg)	R2
RF	Model A	0.20	0.26	0.53
	Model B	0.23	0.30	0.38
	Model C	0.23	0.28	0.45
	Model D	0.21	0.26	0.51
	Model E	0.20	0.25	0.56
SVM	Model A	0.21	0.27	0.48
	Model B	0.23	0.29	0.37
	Model C	0.23	0.29	0.41
	Model D	0.22	0.28	0.45
	Model E	0.21	0.26	0.51
BRT	Model A	0.20	0.26	0.53
	Model B	0.24	0.30	0.38
	Model C	0.24	0.29	0.41
	Model D	0.20	0.25	0.56
	Model E	0.19	0.25	0.58

مقادیر متغیرهای محیطی مشابه بین دو الگوریتم متفاوت است. متغیرهای محیطی در الگوریتم RF به شرح زیر بود: Band-4، Band-6، NDVI، BC-3 و LULC، اما متغیرهای محیطی در الگوریتم BRT شامل «ارتفاع، NDVI، BC-2، BC-3 و LULC» بود. هر دو الگوریتم RF و BRT برای سه متغیر کاربری زمین، پوشش زمین و BC-3 مقادیر تقریباً برابری داشتند. در الگوریتم RF، متغیرهای آب‌وهوایی و توپوگرافی، داده‌های Sentinel-1A و LULC به ترتیب ۱۰، ۱۳، ۱۶ و ۲۸ درصد در برآورد تغییرات نیتروژن کل خاک را به خود اختصاص دادند. با این حال، متغیرهای Landsat-9 با سهم ۳۳٪ در نظارت بر تغییرات نیتروژن کل خاک مهم‌ترین پارامترهای مورد بررسی بودند. از سوی دیگر، متغیرهای سنجش از دور سهم بیشتری در الگوریتم BRT به ترتیب با مقادیر ۵۱٪ و ۶۱٪ داشتند. متغیرهای اقلیمی و توپوگرافی، داده‌های LULC، Sentinel-1A و Landsat-9 به ترتیب ۷، ۱۷، ۲۵، ۳۱ و ۲۰ درصد در تخمین تغییرات نیتروژن کل خاک توسط الگوریتم BRT موثر بودند.

پیش‌بینی توزیع فضایی نیتروژن کل خاک

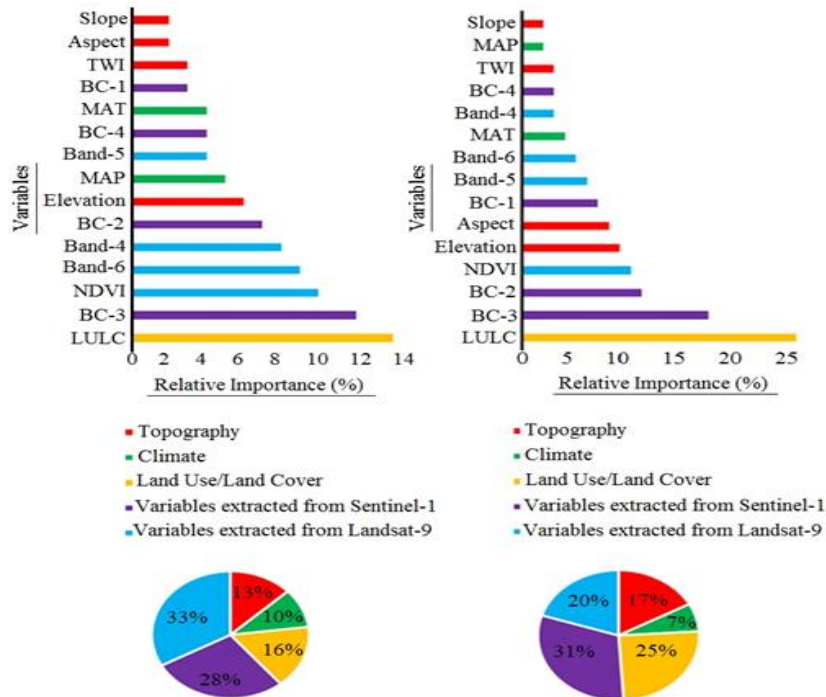
نقشه‌های پیش‌بینی توزیع فضایی نیتروژن کل خاک در شکل ۳ (مدل E) و شکل ۴ (مدل D) ارائه شده است. این نقشه‌ها توسط الگوریتم‌های BRT، RF و SVM تهیه شده است. مقادیر نیتروژن کل خاک گزارش شده در مدل E به ترتیب ۰/۸۲ (±۰/۲۵)، ۰/۸۴ (±۰/۳۲) و ۰/۸۵ (±۰/۲۱) گرم بر کیلوگرم بود (جدول ۵). نتایج نشان داد که مقدار داده‌های نیتروژن کل خاک مشاهده شده (واقعی) بیشتر از نیتروژن کل خاک پیش‌بینی شده توسط الگوریتم‌های BRT، RF و SVM بود. برای بررسی تغییرات ناشی از سطح نیتروژن کل خاک دو مدل E و D از الگوریتم‌های BRT و RF استفاده شد. دلیل اصلی این انتخاب این بود که هر دو الگوریتم ذکر شده، مبتنی بر درخت هستند.

دقت پنج مدل با استفاده از شاخص‌های R^2 ، RMSE و MAE ارزیابی شد. مقدار R^2 بالاتر و مقادیر کمتر RMSE و MAE نشان‌دهنده دقت بالاتر الگوریتم BRT بالاترین دقت را در پیش‌بینی توزیع نیتروژن کل خاک (با مقدار $R^2=0/58$ ، مقدار $RMSE=0/25$ mg/kg و مقدار $RME=0/19$) نشان داد. در بین سه الگوریتم، مدل C (شامل داده‌های SAR چند زمانی) از نظر دقت از مدل B (شامل متغیرهای Landsat-9) بهتر عمل کرد. این نتایج نشان داد که ترکیب داده‌های SAR در نقشه‌برداری نیتروژن کل خاک بسیار موفق بود و به‌طور بالقوه می‌تواند جایگزین تصاویر نوری شود. بر اساس نتایج، مدل C با الگوریتم RF دارای بالاترین مقدار R^2 و کمترین مقادیر MAE و RMSE به ترتیب ۰/۲۳، ۰/۲۳ میلی‌گرم بر کیلوگرم و ۰/۲۸ میلی‌گرم بر کیلوگرم بود که نشان‌دهنده دقت ۴۵٪ در پایش تغییرات نیتروژن کل خاک است (جدول ۴).

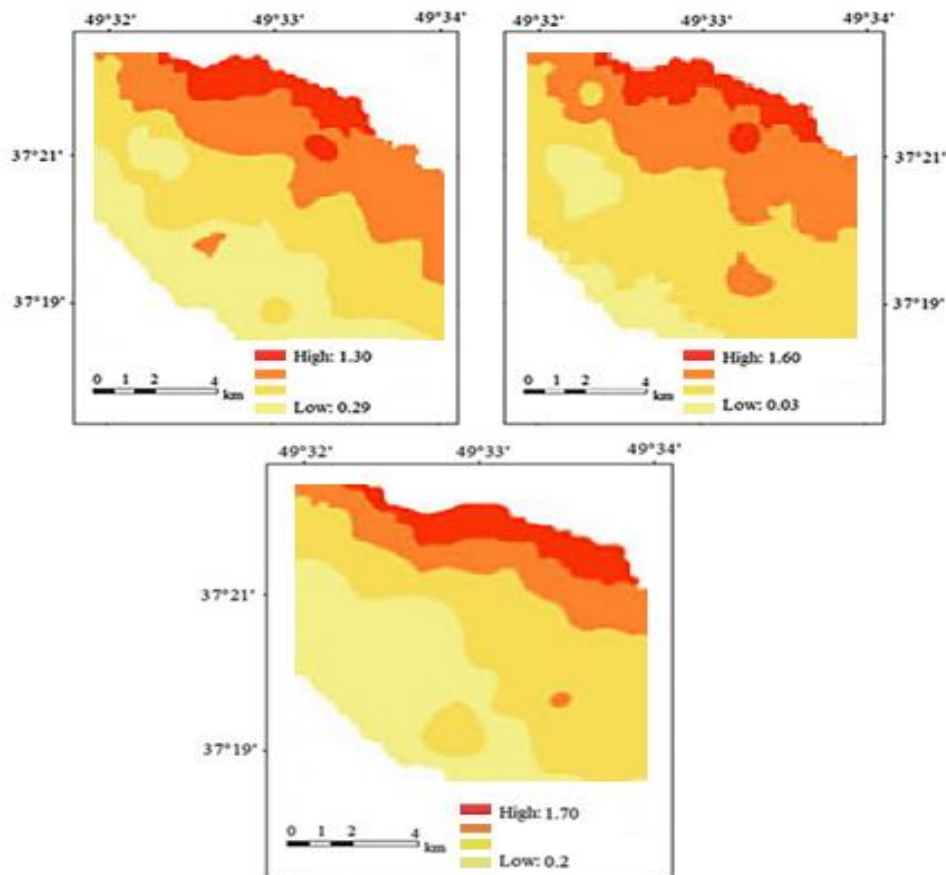
موفقیت مدل C را می‌توان به استفاده از داده‌های چند زمانی Sentinel-1A و Landsat-9 نسبت داد. الگوریتم BRT با تصاویر SAR نیز دقت پیش‌بینی نیتروژن کل خاک را به ترتیب ۵۰، ۱۸ و ۱۸ درصد از نظر R^2 ، RMSE و MAE بهبود بخشید. دقت مدل A نیز با ترکیب داده‌های Sentinel-1A افزایش یافت که منجر به مقادیر $R^2=0/58$ برای الگوریتم BRT و $R^2=0/56$ برای الگوریتم RF شد. در نهایت، مدل E با الگوریتم‌های BRT و RF که شامل چندین متغیر سنجش از دور، آب‌وهوا، توپوگرافی و LULC بود، بالاترین عملکرد را در پیش‌بینی تغییرات نیتروژن کل خاک داشت، که به ترتیب قادر به توضیح ۵۸ و ۵۶ درصد از تغییرات نیتروژن کل خاک بودند.

اهمیت نسبی داده‌های زیست محیطی

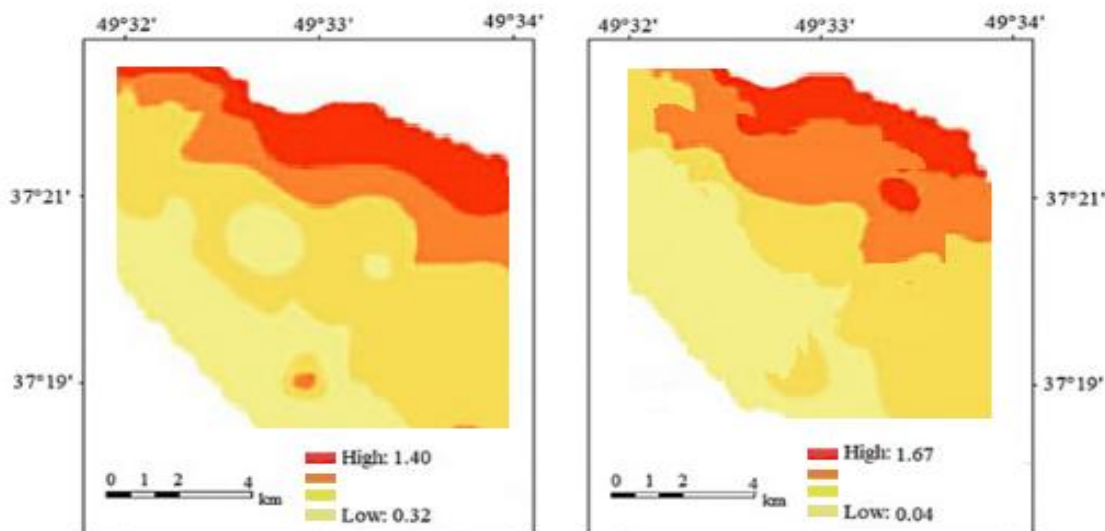
برای نتایج بهتر، نرمال‌سازی متغیرها در این مطالعه انجام شد. شکل ۲ مقادیر هر متغیر محیطی را در مدل E تحت تاثیر الگوریتم‌های BRT و RF نشان می‌دهد.



شکل ۲: مقدار میانگین هر متغیر پیش‌بینی با استفاده از مدل E برای تکنیک‌های RF (چپ) و BRT (راست)
 Fig. 2: Mean value of each predictor variable using E model for RF (left) and BRT (right) techniques



شکل ۳: توزیع مقادیر نیتروژن کل خاک در خاک‌های شالیزار شمال ایران با استفاده از مدل E با RF (چپ)، BRT (راست) و SVM (پایین). در مدل E تمامی پارامترهای اقلیمی، توپوگرافی و سنجش از دور پوشش داده شد
 Fig. 3: Distribution of STN values in paddy soils of northern Iran using E model with RF (left), BRT (right) and SVM (down). In model E, all climate parameters, topography and remote sensing were covered



شکل ۴: نقشه پراکنش مقادیر STN در خاکهای شالیزاری شمال ایران با استفاده از مدل D با RF (چپ)، BRT (راست). در مدل D، پارامترهای سنجش از دور پوشش داده شد. به دلیل عدم دقت قابل قبول SVM از ترسیم نقشه مربوط به این مدل خودداری شد.

Fig. 4: Distribution map of STN values in paddy soils of northern Iran using model D with RF (left), BRT (right). In model D, remote sensing parameters were covered. Due to the lack of acceptable accuracy of SVM, drawing the map related to this model was avoided

جدول ۵: مقادیر نیتروژن کل خاک گزارش شده در مدل E با روش های SVM،

RF و BRT

Table 5: Total soil nitrogen values reported in model E with SVM, BRT and RF methods

Algorithm	Minimum	Maximum	Mean	Standard Deviation (SD)
SVM	0.22	1.41	0.85	0.21
BRT	0.11	1.68	0.84	0.32
RF	0.28	1.35	0.82	0.25

نقشه توزیع فضایی مقادیر نیتروژن کل خاک توسط Zhang و همکاران [۱۱] ایجاد شد. پژوهش آن‌ها با استفاده از تصاویر ماهواره‌ای چند طیفی از Sentinel-2A و مقایسه الگوریتم‌های مختلف با استفاده از RF انجام شد. آن‌ها برای تخمین توزیع فضایی نیتروژن کل خاک از ۲۱ متغیر کمکی استفاده کردند که شامل باندها و شاخص‌های طیفی اصلی و همچنین متغیرهای محیطی بود. نتایج نشان داد که الگوریتم RF بر اساس تکنیک‌های سنجش از دور قادر به ثبت دقیق تغییرات در نیتروژن کل خاک است [۱۱]. در مطالعه دیگری توسط وانگ و همکاران [۲۴]، نیتروژن کل خاک در شهر لوشون در شمال شرقی استان لیائونینگ، چین، با استفاده از الگوریتم‌های RF و رگرسیون چندگانه گام به گام نقشه‌برداری شد. برای نقشه‌برداری نیتروژن کل خاک در مناطق جنگلی، آن‌ها پیشنهاد کردند که داده‌های سنجش از دور و متغیرهای محیطی باید به‌عنوان پیش‌بینی‌کننده‌های اصلی استفاده شوند [۲۴].

ارزش متغیرهای پیش‌بینی‌کننده

استفاده از تصاویر SAR برای پیش‌بینی دقیق محتوای نیتروژن کل خاک بسیار مهم است. داده‌های رادار می‌توانند اطلاعاتی فراتر از سطح خاک و پوشش گیاهی را ضبط کنند و با استفاده از سیگنال‌های پس‌پراکنده

همچنین عملکرد قابل قبول‌تری نسبت به الگوریتم SVM برای نظارت بر تغییرات نیتروژن کل خاک دارند. الگوریتم‌های BRT و RF عملکرد رضایت‌بخشی را با مقادیر متوسط -0.03 و -0.01 در پیش‌نیتروژن کل خاک با استفاده از مدل‌های D و E نشان دادند. با این حال، شایان ذکر است که تفاوت معنی‌دار در مقادیر پیش‌بینی شده نیتروژن کل خاک بین مدل‌های D و E تنها در مناطقی با زمین برهنه و پوشش گیاهی مشاهده شد.

یافته‌های تحقیق نشان داد که الگوریتم‌های RF و BRT در تخمین دقیق مقادیر نیتروژن کل خاک مؤثرتر بودند. اتوی و همکاران [۲۱] دقت الگوریتم‌های ANN، BRT و SVM را برای ارزیابی مخازن کربن آلی خاک (SOC) از طریق نقشه‌برداری دیجیتال خاک مقایسه کردند و به این نتیجه رسیدند که BRT دقت بالاتری نسبت به دو الگوریتم دیگر دارد [۲۱]. تزیاکریس و همکاران [۲۲] از RF و کریجینگ برای بررسی کربن آلی خاک با استفاده از DEM استفاده کردند و دریافتند که RF از کوکریجینگ بهتر عمل کرد [۲۲]. پنگ و همکاران [۲۳] یک نقشه توزیع نیتروژن کل خاک با استفاده از تکنیک‌های سنجش از دور تولید کرد و دقت الگوریتم‌های BRT، SVM و RF را مقایسه کرد. آن‌ها گزارش کردند که الگوریتم‌های BRT و RF از دقت بالاتری نسبت به SVM برخوردار است و ترکیب داده‌های Landsat-8 و Sentinel-2 با استفاده از الگوریتم‌های مبتنی بر درخت منجر به افزایش ۱۷ درصدی در شاخص‌های MAE و RMSE شد [۲۳]. استفاده از داده‌های Sentinel-1 و تصاویر نوری دقت بالایی را در این مطالعه تضمین کرد، که نشان می‌دهد ترکیب پارامترهای مفیدتر می‌تواند نتایج را بهبود بخشد. به‌طور خاص، نتایج نشان داد که استفاده از داده‌های Sentinel-1 به‌عنوان داده‌های SAR چند زمانی، دقت نقشه‌برداری نیتروژن کل خاک را در شالیزارهای مورد مطالعه بهبود بخشید.

نتیجه‌گیری

در این مطالعه، توزیع فضایی نیتروژن کل خاک در خاک‌های شالیزاری شمال ایران با استفاده از الگوریتم‌های SVM، BRT و RF مورد بررسی قرار گرفت. یافته‌ها نشان داد: (۱) الگوریتم‌های RF و BRT عملکرد بهتری از SVM داشتند و در نظارت بر مقادیر نیتروژن کل خاک مؤثر بودند. (۲) تصاویر SAR چند زمانی، نظارت دقیق‌تری بر محتوای نیتروژن کل خاک در مقایسه با داده‌های سنجش از راه دور نوری ارائه می‌کنند، و امکان پیش‌بینی واقعی مقادیر نیتروژن کل خاک در خاک‌های شالیزاری را فراهم می‌کنند. (۳) دقت الگوریتم‌ها با گنجاندن متغیرهای محیطی افزایش یافت. (۴) الگوریتم BRT کمترین مقادیر RMSE (۰/۲۵ میلی‌گرم بر کیلوگرم) و MAE (۰/۱۹ میلی‌گرم بر کیلوگرم) و بالاترین مقدار (۰/۵۸) را نشان داد. متغیرهای سنجش از دور به ترتیب در ۶۱٪ و ۵۱٪ تأثیر در الگوریتم‌های RF و BRT نقش داشتند. با توجه به اقلیم منحصربه‌فرد منطقه مورد مطالعه که در مجاورت دریای خزر واقع شده و در اکثر فصول سال دارای هوای ابری و مه آلود است، استفاده از داده‌های Sentinel-1 به دلیل قابلیت نفوذ ابرها و مه در مناطق دارای مه توصیه می‌شود. آب‌وهوای مشابه این مطالعه تحقیقات بیشتری را توصیه می‌کند که داده‌های جدید سنجش از دور، متغیرهای محیطی و الگوریتم‌های پیشرفته‌تر را برای دستیابی به دقت بیشتر در پیش نیتروژن خاک و سایر ویژگی‌های خاک ترکیب می‌کند.

مشارکت نویسندگان

در این مقاله، نویسندگان به نسبت سهم برابر مشارکت داشته‌اند.

تشکر و قدردانی

بدینوسیله از همکارانی که ما را در انجام این تحقیق یاری کردند، صمیمانه تشکر می‌کنیم.

تعارض منافع

«هیچ‌گونه تعارض منافع توسط نویسندگان بیان نشده است.»

منابع و مأخذ

- [1] Kosari A, Sharifi A, Ahmadi A, Khoshsima M. Remote sensing satellite's attitude control system: rapid performance sizing for passive scan imaging mode. *Aircraft Engineering and Aerospace Technology* 2020;92. <https://doi.org/10.1108/AEAT-02-2020-0030>.
- [2] Jeong G, Oeverdieck H, Park SJ, Huwe B, Ließ M. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *Catena (Amst)* 2017;154. <https://doi.org/10.1016/j.catena.2017.02.006>.
- [3] Fileborn B. Digital mapping as feminist method: critical reflections. *Qualitative Research* 2023;23. <https://doi.org/10.1177/146879412111028797>.

امکان شناسایی اهدافی مانند بافت خاک، رطوبت، شوری و زیست توده بالای زمین را فراهم کنند. فلسبرگ و همکاران [۲۵] نشان دادند که می‌توان از تصاویر رادار برای بررسی ویژگی‌های خاک با ثبت اطلاعات پوشش گیاهی استفاده کرد که به شدت بر بازتاب خاک تأثیر می‌گذارد [۲۵]. گوو و همکاران [۲۶] اثر بخشی ترکیب داده‌های دیجیتال Sentinel-1 و Sentinel-2 را برای تخمین زیست توده سطح زمین نشان داد. باندها و شاخص‌های ماهواره‌ای از جمله Sentinel-1، Sentinel-2 و Landsat در نقشه‌برداری ویژگی‌های مختلف خاک مورد استفاده قرار گرفته‌اند. نقشه‌برداری محتوای نیتروژن کل خاک به‌ویژه تحت تأثیر تغییرات کاربری زمین و پوشش زمین (LULC) است که بر بازتاب طیفی تأثیر می‌گذارد که ارتباط نزدیکی با زیست توده و تراکم گیاه دارد [۲۶]. وانگ و همکاران [۲۷] اشاره کردند که تصاویر ماهواره‌ای دیجیتال مهم‌ترین داده کمکی برای پایش تغییرات محیطی در مقایسه با سایر متغیرها هستند. متغیرهای توپوگرافی دارای قدرت پیش‌بینی بالایی برای نقشه‌برداری دیجیتالی خاک در نظر گرفته می‌شوند [۲۷]. علاوه بر ارتفاع، شاخص رطوبت، توپوگرافی و شیب متغیرهای مهمی در توزیع نیتروژن کل خاک هستند زیرا بر کانی‌سازی و سازماندهی نیتروژن و همچنین محتوای نیتروژن موجود تأثیر می‌گذارند [۲۸-۲۹]. بارندگی همچنین به‌طور غیرمستقیم بر تغییرات در چرخه نیتروژن با تأثیر بر میزان نیتروژن جذب شده توسط گیاهان و بهره‌وری پوشش گیاهی تأثیر می‌گذارد [۳۰].

در حالی که این تحقیق بینش‌های ارزشمندی را برای پیش‌بینی سطوح نیتروژن کل خاک با استفاده از روش‌های سنجش از دور و الگوریتم‌های پیشرفته ارائه می‌دهد، مهم است که چندین محدودیت را بپذیریم. اولاً، تمرکز مطالعه بر روی استان گیلان ممکن است تعمیم‌پذیری یافته‌ها را به سایر مناطق با شرایط محیطی متفاوت محدود کند. علاوه بر این، در حالی که سنجش از دور ابزار ارزشمندی برای جمع‌آوری داده‌ها فراهم می‌کند، محدودیت‌های ذاتی مانند محدودیت‌های تفکیک فضایی و طیفی وجود دارد که ممکن است بر صحت و دقت پیش‌بینی‌ها تأثیر بگذارد. علاوه بر این، اتکا به الگوریتم‌های یادگیری ماشین، سوگیری‌ها و عدم قطعیت‌های بالقوه مرتبط با انتخاب مدل و تنظیم پارامتر را معرفی می‌کند. در حالی که تلاش‌هایی برای ادغام متغیرهای مختلف زیست‌محیطی انجام شد، ممکن است هنوز عواملی وجود داشته باشد که بر سطوح نیتروژن خاک تأثیر بگذارند، مانند شیوه‌های مدیریت زمین محلی یا ویژگی‌های خاک که در تجزیه و تحلیل ثبت نشده است. در نهایت، اتکای این مطالعه به تصاویر ماهواره‌ای و پیش‌بینی‌های الگوریتمی ممکن است نیاز به صحت و اعتبار زمین را نادیده بگیرد، که می‌تواند بررسی‌های ضروری را در مورد دقت برآوردهای حاصل از سنجش از دور فراهم کند. بنابراین، در حالی که یافته‌ها بینش‌های ارزشمندی را ارائه می‌دهند، باید در برون‌یابی آن‌ها به زمینه‌های وسیع‌تر احتیاط کرد و تحقیقات آینده باید به بررسی این محدودیت‌ها برای افزایش استحکام و کاربرد یافته‌ها بپردازد.

- learning method. *Science of the Total Environment* 2023;866. <https://doi.org/10.1016/j.scitotenv.2023.161421>.
- [16] Zhang Q, Liu M, Zhang Y, Mao D, Li F, Wu F, et al. Comparison of Machine Learning Methods for Predicting Soil Total Nitrogen Content Using Landsat-8, Sentinel-1, and Sentinel-2 Images. *Remote Sens (Basel)* 2023;15. <https://doi.org/10.3390/rs15112907>.
- [17] Denton OA, Aduramigba-Modupe VO, Ojo AO, Adeoyolanu OD, Are KS, Adelana AO, et al. Assessment of spatial variability and mapping of soil properties for sustainable agricultural production using geographic information system techniques (GIS). *Cogent Food Agric* 2017;3. <https://doi.org/10.1080/23311932.2017.1279366>.
- [18] Kjeldahl J. A new method for the determination of nitrogen in organic matter. *Zeitschrift Für Analytische Chemie* 1883.
- [19] Vapnik VN. The nature of statistical learning theory. *Statistics for Engineering and Information Science*. Springer-Verlag, New York 2000.
- [20] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 2005;30. <https://doi.org/10.3354/cr030079>.
- [21] Ottoy S, De Vos B, Sindayihebura A, Hermy M, Van Orshoven J. Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecol Indic* 2017;77. <https://doi.org/10.1016/j.ecolind.2017.02.010>.
- [22] Tziachris P, Aschonitis V, Chatzistathis T, Papadopoulou M. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena (Amst)* 2019;174. <https://doi.org/10.1016/j.catena.2018.11.010>.
- [23] Peng Y, Zhao L, Hu Y, Wang G, Wang L, Liu Z. Prediction of soil nutrient contents using visible and near-infrared reflectance spectroscopy. *ISPRS Int J Geoinf* 2019;8. <https://doi.org/10.3390/ijgi8100437>.
- [24] Wang S, Jin X, Adhikari K, Li W, Yu M, Bian Z, et al. Mapping total soil nitrogen from a site in northeastern China. *Catena (Amst)* 2018;166. <https://doi.org/10.1016/j.catena.2018.03.023>.
- [25] Felsberg A, De Lannoy GJM, Giroto M, Poesen J, Reichle RH, Stanley T. Global soil water estimates as landslide predictor: The effectiveness of smos, smap, and grace observations, land surface simulations, and data assimilation. *J Hydrometeorol* 2021;22. <https://doi.org/10.1175/JHM-D-20-0228.1>.
- [26] Guo S, Bai X, Chen Y, Zhang S, Hou H, Zhu Q, et al. An improved approach for soil moisture estimation in gully fields of the Loess Plateau using Sentinel-1A radar images. *Remote Sens (Basel)* 2019;11. <https://doi.org/10.3390/rs11030349>.
- [4] Li N, Zhao D, Arshad M, Sefton M, Triantafyllis J. Comparison of a digital soil map and conventional soil map for management of topsoil exchangeable sodium percentage. *Soil Use Manag* 2022;38. <https://doi.org/10.1111/sum.12666>.
- [5] Campolo J, Güereña D, Maharjan S, Lobell DB. Evaluation of soil-dependent crop yield outcomes in Nepal using ground and satellite-based approaches. *Field Crops Res* 2021;260. <https://doi.org/10.1016/j.fcr.2020.107987>.
- [6] Loranty MM, Abbott BW, Blok D, Douglas TA, Epstein HE, Forbes BC, et al. Reviews and syntheses: Changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions. *Biogeosciences* 2018;15. <https://doi.org/10.5194/bg-15-5287-2018>.
- [7] Jones BJD. Potassium availability: synchronizing nutrient supply and plant demand through 4R nutrient stewardship. *Crops & Soils* 2019;52. <https://doi.org/10.2134/cs2019.52.0602>.
- [8] Garrigues E, Corson MS, Angers DA, Van Der Werf HMG, Walter C. Soil quality in Life Cycle Assessment: Towards development of an indicator. *Ecol Indic* 2012;18. <https://doi.org/10.1016/j.ecolind.2011.12.014>.
- [9] Taher KI, Abdulazeez AM, Zebari DA. Data Mining Classification Algorithms for Analyzing Soil Data. *Asian Journal of Research in Computer Science* 2021. <https://doi.org/10.9734/ajrcos/2021/v8i230196>.
- [10] Akpa SIC, Odeh IOA, Bishop TFA, Hartemink AE, Amapu IY. Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma* 2016;271. <https://doi.org/10.1016/j.geoderma.2016.02.021>.
- [11] Zhang Y, Sui B, Shen H, Ouyang L. Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Comput Electron Agric* 2019;160. <https://doi.org/10.1016/j.compag.2019.03.015>.
- [12] Parham PE, Pople D, Christiansen-Jucht C, Lindsay S, Hinsley W, Michael E. Modeling the role of environmental variables on the population dynamics of the malaria vector *Anopheles gambiae* sensu stricto. *Malar J* 2012;11. <https://doi.org/10.1186/1475-2875-11-271>.
- [13] Ma Y, Minasny B, Wu C. Mapping key soil properties to support agricultural production in Eastern China. *Geoderma Regional* 2017;10. <https://doi.org/10.1016/j.geodrs.2017.06.002>.
- [14] Ceddia M, Gomes A, Vasques G, Pinheiro É. Soil Carbon Stock and Particle Size Fractions in the Central Amazon Predicted from Remotely Sensed Relief, Multispectral and Radar Data. *Remote Sens (Basel)* 2017;9. <https://doi.org/10.3390/rs9020124>.
- [15] Li Z, Liu F, Peng X, Hu B, Song X. Synergetic use of DEM derivatives, Sentinel-1 and Sentinel-2 data for mapping soil properties of a sloped cropland based on a two-step ensemble



علیرضا شریفی دانشیار گروه مهندسی نقشه‌برداری در دانشکده مهندسی عمران دانشگاه تربیت دبیر شهید رجایی می‌باشند. ایشان مدرک کارشناسی ارشد و دکتری خود را در رشته مهندسی نقشه‌برداری-سنجش از دور بترتیب در سال‌های ۱۳۸۷ و ۱۳۹۴ از دانشگاه تهران دریافت نمودند. فعالیت‌های پژوهشی اخیر وی در حوزه کاربرد یادگیری عمیق در پردازش تصویر فراطیفی و همچنین کاربردهای سنجش از دور در پایش تغییرات اقلیمی، کشاورزی، مدیریت جنگل‌ها، مدیریت کاربری و پوشش اراضی و مدل‌سازی محیطی با استفاده از کلان داده‌های سنجش از دور می‌باشد. در حال حاضر ایشان بیش از ۱۰۰ مقاله علمی در مجلات و کنفرانس‌های علمی ارائه نموده‌اند و همچنین در کمیته علمی و داوری بیش از ۴۰ مجله و کنفرانس علمی ملی و بین المللی فعالیت داشته‌اند. زمینه‌های تخصصی ایشان عبارتند از: پردازش تصاویر فراطیفی و راداری، هوش مصنوعی در اطلاعات مکانی، توسعه کاربردهای سنجش از دور.

Sharifi, A. Associate Professor at the Department of Geomatics and Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaei Teacher Training University, Tehran, Iran

✉ a_sharifi@sru.ac.ir



احمد گلچین استاد گروه مهندسی علوم خاک در دانش کده کشاورزی دانشگاه زنجان می‌باشند. فعالیت‌های پژوهشی اخیر وی در حوزه حاصلخیزی خاک، توزیع فضایی شیمی و دینامیک ماده آلی در خاک‌ها، آلودگی خاک و تغذیه گیاه می‌باشد.

Golchin, A. Professor at the Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran.

✉ agolchin2011@yahoo.com



پرویز کرمی استادیار گروه مرتعداری در دانشکده کشاورزی دانشگاه کردستان می‌باشند. فعالیت‌های پژوهشی اخیر وی در حوزه کارایی اشکال و سطوح مختلف پلات برای برآورد تولید علفزارهای نیمه استپی و شبیه سازی عملکرد اکوسیستم‌های مرتعی با مدل CENTURY می‌باشد.

Karami, P. Assistant Professor at the Department of Range and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Kurdistan, Iran.

✉ P.karami@uok.ac.ir

[27] Wang S, Adhikari K, Wang Q, Jin X, Li H. Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C:N ratio from the northeastern coastal agroecosystems in China. *Ecol Indic* 2018;84. <https://doi.org/10.1016/j.ecolind.2017.08.046>.

[28] Araya SN, Fryjoff-Hung A, Anderson A, Viers JH, Ghezzehei TA. Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. *Hydrol Earth Syst Sci* 2021;25. <https://doi.org/10.5194/hess-25-2739-2021>.

[29] Ågren AM, Larson J, Paul SS, Laudon H, Lidberg W. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. *Geoderma* 2021;404. <https://doi.org/10.1016/j.geoderma.2021.115280>.

[30] Pérez-Piqueres A, Albiach R, Domínguez A, Canet R. Effect of soil preparation techniques on the biochemical properties and microbial communities of a citrus orchard after replanting and conversion into organic management. *Applied Soil Ecology* 2017;119. <https://doi.org/10.1016/j.apsoil.2017.05.015>.

معرفی نویسندگان

AUTHOR(S) BIOSKETCHES



شیلان فعله‌گری دانشجوی مقطع دکتری مهندسی علوم خاک می‌باشد و مشغول فعالیت در حوزه آموزش و پژوهش است. زمینه‌های تخصصی و علاقمندی ایشان عبارتند از: توسعه کاربردهای سنجش از دور، پایش تغییرات اقلیمی.

Felegari, S. Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran.

✉ Shilan.Felegari@znu.ac.ir




کامران مروج دانشیار گروه مهندسی علوم خاک در دانشکده کشاورزی دانشگاه زنجان می‌باشند. فعالیت‌های پژوهشی اخیر وی در حوزه تهیه نقشه رقومی خاک، تجزیه و تحلیل‌های ژئومرفومتری و ارزیابی تناسب اراضی است. همچنین کاربردهای سنجش از دور در پایش تغییرات اقلیمی، مدیریت کاربری و پوشش اراضی و مدل‌سازی محیطی با استفاده از کلان داده‌های سنجش از دور از زمینه‌های تخصصی ایشان است.

Moravej, K. Associate Professor at the Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran.

✉ Kmoravej@znu.ac.ir

Citation (Vancouver): Felegari Sh, Moravej k, Sharifi A, Golchin A, Karami P. [Prediction of Total Soil Nitrogen Variations Using Three Machine Learning Approaches and Remote Sensing Data]. *J. RS. GEOINF. RES.* 2024; 2(1): 127-140

 <https://doi.org/10.22061/jrsg.2024.10781.1060>



COPYRIGHTS

© 2024 The Author(s). This is an open-access article distributed under the terms and conditions of the Creative Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
(<https://creativecommons.org/licenses/by-nc/4.0/>)