



## Hine learning approaches in downstream lands of Azad dam (case study: Kurdistan province)

Maryam Osat<sup>✉1</sup> | Shahrokh Fatehi<sup>2</sup> | Zeinab Zeinodini<sup>3</sup>

1. Corresponding Author, Horticulture Crop Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran. E-mail: [Maryam.ousat@gmail.com](mailto:Maryam.ousat@gmail.com)
2. Soil and Water Research Department, Kermanshah agricultural and natural resources research and education center, AREEO, Kermanshah, Iran. E-mail: [Shahrokh.fatehi@gmail.com](mailto:Shahrokh.fatehi@gmail.com)
3. Department of Science and Soil Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran. E-mail: [z.zeynoldini1990@gmail.com](mailto:z.zeynoldini1990@gmail.com)

---

### Article Info

**Article type:** Research Article

**Article history:**

**Received:** July. 5, 2023

**Revised:** Nov. 21, 2023

**Accepted:** Nov. 27, 2023

**Published online:** Jan. 21, 2024

**Keywords:**

Artificial Neural Networks,  
Decision Tree,  
Multinomial Logistic Regression,  
Random Forest.

---

### ABSTRACT

Digital Soil Mapping (DSM) encompasses a variety of methodologies that can yield precise spatial information about soil by establishing quantitative relationships between environmental covariates (predictors) and soil classes or properties. In this study, Artificial Neural Networks (ANNs), Decision Tree (DT), Multinomial Logistic Regression (MLR), and Random Forest (RF) algorithms were used to predict the soil map of downstream lands of Azad dam with an area of approximately 178.3 ha in the northwest of Sanandaj city in Kurdistan province. A random soil sampling method was used to determine the location and distribution of the 84 soil profiles in the study area. After recording soil morphological attributes, sampling of all horizons was conducted for required laboratory analysis. Afterward, the soil profiles were classified up to the family taxonomic level based on US classification system. Based on the soil taxonomy classification system, Inceptisols and Entisols order were observed by frequency, two Suborder, three Great groups, five Subgroups, and Family. To calculate the predictor variables, a digital elevation model (DEM) with a 10 m spatial resolution and Sentinel 2-B satellite images were used in the study area. To check the prediction accuracy of the models the Overall accuracy (OA), Kappa Index (K), and Brier Score (BS) were used. The best result was obtained by the ANN model (OA=0.65, K=0.53, and BS=0.16, respectively). The weakest predictions were found by DT model with OA, K, and BS of 0.38, 0.22, and 0.87, respectively.

---

Cite this article: Osat, M., Fatehi, Sh., & Zeinodini, Z. (2024) Hine learning approaches in downstream lands of Azad dam (case study: Kurdistan province), *Iranian Journal of Soil and Water Research*, 54 (11), 1697-1714.

<https://doi.org/10.22059/ijswr.2023.361410.669526>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2023.361410.669526>



## نقشه برداری رقومی کلاس‌های خاک با استفاده از رویکردهای یادگیری ماشین در بخشی از اراضی پایاب سد آزاد (مطالعه موردی: استان کردستان)

مریم اوسا<sup>۱</sup> | شاهرخ فاتحی<sup>۲</sup> | زینب زین الدینی<sup>۳</sup>

۱. نویسنده مسئول، بخش تحقیقات علوم زراعی-باغی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کردستان، ایران.

رایانامه: [Maryam.ousat@gmail.com](mailto:Maryam.ousat@gmail.com)

۲. بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، ایران. رایانامه: [Shahrokh.fatehi@gmail.com](mailto:Shahrokh.fatehi@gmail.com)

۳. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران. رایانامه: [z.zeynoldini1990@gmail.com](mailto:z.zeynoldini1990@gmail.com)

### اطلاعات مقاله

### چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۴/۱۴

تاریخ بازنگری: ۱۴۰۲/۸/۳۰

تاریخ پذیرش: ۱۴۰۲/۹/۶

تاریخ انتشار: ۱۴۰۲/۱۱/۱

### واژه‌های کلیدی:

جنگل تصادفی،

درخت تصمیم،

رگرسیون لاجستیک،

شبکه عصبی مصنوعی.

نقشه‌برداری رقومی خاک با برقراری روابط کمی میان متغیرهای محیطی و کلاس‌ها یا ویژگی‌های خاک قادر به پیش‌بینی مکانی ویژگی موردنظر است. در این پژوهش از الگوریتم‌های شبکه عصبی مصنوعی، درخت تصمیم، رگرسیون لاجستیک چند جمله‌ای و جنگل تصادفی برای پیش‌بینی نقشه خاک اراضی پایاب سد آزاد شهر سنندج با وسعت حدود ۲۱۷۸/۳ هکتار استفاده شد. در سال ۱۳۹۶ تعداد ۸۴ خاک‌رخ با الگوی تصادفی در منطقه مطالعاتی حفر، تشریح و نمونه‌برداری گردید. بر اساس ویژگی‌های ریخت‌شناختی و داده‌های آزمایشگاهی هر یک از خاک‌رخ‌ها تا سطح خانواده رده‌بندی شدند. بر اساس سیستم رده‌بندی جامع آمریکایی، دو رده اینسپتی‌سول و انتی‌سول، دو زیر رده، سه گروه بزرگ و پنج زیرگروه و خانواده مشاهده شد. برای محاسبه متغیرهای پیش‌بینی کننده، از مدل رقومی ارتفاع با قدرت تفکیک ۱۰ متر و تصویر ماهواره سنتیل ۲-B استفاده شد. برای بررسی صحت پیش‌بینی مدل‌ها از صحت عمومی نقشه، شاخص کاپا و درجه برابر استفاده شد که بهترین نتایج (به ترتیب ۰/۶۵، ۰/۵۳ و ۰/۱۶) برای الگوریتم شبکه عصبی مصنوعی به دست آمد. ضعیف‌ترین پیش‌بینی مربوط به مدل درخت تصمیم با صحت عمومی ۰/۳۸، شاخص کاپای ۰/۲۲ و درجه برابر ۰/۸۷ بود.

استناد: اوسا؛ مریم، فاتحی؛ شاهرخ، زین الدینی؛ زینب، (۱۴۰۲) نقشه برداری رقومی کلاس‌های خاک با استفاده از رویکردهای یادگیری ماشین در بخشی از اراضی پایاب سد آزاد (مطالعه موردی: استان کردستان)، مجله تحقیقات آب و خاک ایران، ۵۴ (۱۱)، ۱۶۹۷-۱۷۱۴.



<https://doi.org/10.22059/ijswr.2023.361410.669526>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2023.361410.669526>

## مقدمه

به منظور مدیریت دقیق و استفاده پایدار از اراضی، به درک درستی از پراکنش مکانی خاکها در زمین نیاز است (Angelini et al., 2016; Siqueira et al., 2015). شناسایی یا نقشه برداری خاک روشی برای تعیین الگوی پراکنش خاک، توصیف و نمایش آن به شکل قابل فهم و تفسیر برای کاربران مختلف بوده و اطلاعات پایه خاک را برای مدل سازی های محیطی فراهم می کند (Bouma et al., 1996). نقشه های خاک، منابع رایج اطلاعات خاک برای ارزیابی اراضی و برنامه های آمایش سرزمین هستند و تهیه آن ها همواره از مهم ترین بخش های مطالعات خاک محسوب می شود (Buol et al., 2003). خاک سامانه پیچیده ای است که عوامل مؤثر بر تغییر و تحولات آن به خوبی قابل شناسایی نیست؛ بنابراین تهیه نقشه هایی از خاک که بتواند پراکنش و تغییرات خاک را به خوبی نشان دهد، یکی از دغدغه های مهم خاک شناسان است (صالحی و خادمی، ۱۳۸۷). نقشه خاک با مقیاس ۱:۱۰۰۰۰۰۰ منبع اصلی اطلاعات خاک در ایران است که در سال ۲۰۰۰ میلادی تهیه و کامل گردید (بنائی، ۱۳۸۰). با توجه به وسعت زیاد ایران و همچنین با گذشت حدود دو دهه از تهیه این نقشه ها، نیاز به تهیه نقشه هایی با مقیاس منطقه ای و صحت و تفکیک بیشتر است. از مشکلات اصلی برای به هنگام کردن این نقشه ها، می توان به وسعت بزرگ کشور و همچنین کار میدانی وسیع که جزء هزینه بر این پروژه است، اشاره کرد.

در گذشته اغلب از روش معمول برای تهیه نقشه خاک و بررسی تغییرات ویژگی های آن استفاده می شد (Rossiter, 2005). به طور کلی نقشه برداری خاکها به صورت نقشه های معمول دارای کاستی های زیادی است که از جمله این موارد می توان به عدم نمایش تغییر پذیری خاک به صورت پیوسته (Scull et al., 2003; Fernandes Coelho et al., 2021)، عدم وجود اطلاعات در مورد دقت و صحت نقشه (Burrough et al., 1971)، وابستگی به مقیاس و حذف اطلاعات با تلفیق واحدهای کوچک نقشه در واحدهای بزرگ تر و مهم تر از همه وابستگی آن به مدل ذهنی کارشناس اشاره نمود. از سوی دیگر، روش های معمول نقشه برداری خاک نیاز به شمار زیادی نمونه و تجزیه های آزمایشگاهی دارند و از این نظر بسیار زمان بر و پرهزینه هستند (Zhu et al., 2001; Minasny & McBratney, 2019; Padarian et al., 2016). بدین منظور نقشه برداری رقومی خاک در پاسخ به نقد نقشه برداری معمول و به عنوان جایگزین بهتر تکنیک های معمول نقشه برداری توسعه یافت.

## پیشینه پژوهش

مک براتنی و همکاران (McBratney et al., 2003) با مروری بر مدل های مختلف تخمین و برآورد مکانی خاک و با توجه به نکات کلیدی مذکور اقدام به ارائه یک چارچوب تعمیم یافته برای تخمین و برآورد مکانی خاک نمودند؛ چارچوبی که بعدها تحت عنوان نقشه برداری رقومی خاک نامیده شد. آنان با داخل نمودن موقعیت مکانی و دیگر مشخصه های خاک، اقدام به بسط و توسعه مدل چند عاملی ینی، به منظور برآورد و تخمین یک مشخصه یا کلاس خاک نمودند. به دیگر سخن، آنان مدل پنج عاملی ینی را به یک مدل هفت عاملی از طریق کمی سازی و بهره مندی از مدل های رقومی ارتفاع، سنجش از دور، حس گرهای سنجنده خاک و دیگر تکنیک های پهنه بندی، بسط و تعمیم دادند (McBratney et al., 2003). مدل اسکورپن در حال حاضر پایه و اساس رویکرد نوین نقشه برداری خاک را تشکیل می دهد و به طور وسیعی در تحقیقات نقشه برداری رقومی خاک در حال استفاده است (Rossiter, 2004). یکی از اجزاء اصلی نقشه برداری رقومی خاک (اسکورپن) مدل ارتباط دهنده بین متغیرهای محیطی و خاک است که در این راستا می توان از انواع روش های داده کاوی بهره گرفت. روش های زیادی در نقشه برداری رقومی خاک به کار می رود که عمدتاً در گروه روش های ماشین یادگیری<sup>۱</sup> قرار دارند. ماشین یادگیری یک اصطلاح کلی برای مجموعه ای گسترده از مدل هایی است که به منظور تحلیل الگوهای موجود در داده ها و انجام پیش بینی ها، به کار می روند (Witten et al., 2011). اگرچه این روش ها اغلب به منظور داده کاوی در یک پایگاه داده ای بزرگ به کار می روند، در عین حال ابزاری جالب برای آموزش مدل های پیش بینی مکانی کلاس های خاک هستند؛ زیرا با روش معمول، روابط بین کلاس های خاک و متغیرهای کمکی محیطی، اغلب تا حد کمی قابل درک و فهم است (Grunwald, 2006).

از جمله روش های ماشین یادگیری که در زمینه های مختلف علوم خاک به کار رفته اند می توان به رگرسیون لاجستیک<sup>۲</sup>، جنگل



تصادفی<sup>۱</sup>، شبکه عصبی<sup>۲</sup> و درخت تصمیم‌گیری<sup>۳</sup> اشاره کرد. در زمینه استفاده از روش‌های مختلف ماشین یادگیری برای پیش‌بینی کلاس‌های خاک مطالعات گوناگونی در سطح کشور و جهان صورت گرفته است. نتایج مطالعات صورت گرفته نشان داده است که استفاده از مدل‌های پیچیده‌تر مانند جنگل تصادفی در مقایسه با مدل‌های خطی تصمیم یافته دقت بیشتری دارند (Camera et al., 2017). Brungard et al., (2015) از روش‌های مختلف ماشین یادگیری از جمله کای نزدیکترین همسایه<sup>۴</sup>، ماشین بردار پشتیبان<sup>۵</sup>، رگرسیون لاجستیک، جنگل تصادفی و شبکه عصبی مصنوعی برای پیش‌بینی کلاس‌های خاک در سه منطقه نیمه‌خشک در ایالات متحده آمریکا استفاده و جنگل تصادفی را به‌عنوان بهترین مدل معرفی نمودند. (Heung et al., 2016) از الگوریتم‌های ماشین بردار پشتیبان، درخت تصمیم، رگرسیون لاجستیک و شبکه عصبی مصنوعی برای پیش‌بینی کلاس‌های خاک در سطح گروه بزرگ پایین دره فریز در کانادا استفاده کردند و ماشین بردار پشتیبان را به‌عنوان یک مدل با دقت بالا در پیش‌بینی کلاس‌های خاک معرفی کردند. جنگل تصادفی یکی از مدل‌های بسیار پرطرفدار در پیش‌بینی کلاس‌ها و ویژگی‌های خاک است که کارایی آن‌ها توسط محققان بسیاری تایید شده است (Khaledian & Miller, 2020; Monteiro et al., 2023). تقی‌زاده و همکاران (Taghizadeh-Mehrjardi et al., 2019) نشان دادند که استفاده از مدل‌های ترکیبی کارایی بالاتری در پیش‌بینی کلاس‌های خاک نسبت به مدل‌های انفرادی دارد.

از جمله مطالعات انجام‌گرفته در کشور می‌توان به نقشه‌برداری رقومی کلاس‌های خاک با استفاده از نقشه خاک قدیمی در منطقه خشک جنوب شرق ایران (عباس‌زاده افشار و همکاران، ۱۳۹۶)، نقشه‌برداری رقومی کلاس‌های خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه‌ی اردکان استان یزد (تقی‌زاده مهرجردی و همکاران، ۱۳۹۳) و نقشه‌برداری رقومی افق‌های مشخصه و گروه‌های بزرگ خاک در منطقه زرنند کرمان (جعفری و همکاران، ۱۳۹۱) اشاره نمود. فرقانی و همکاران (۱۳۹۹) از رگرسیون درختی توسعه یافته در نقشه‌برداری رقومی کلاس‌های خاک در اراضی سیلابی با پستی و بلندی کم استفاده کرده و کارایی این روش را نشان دادند. مقصودی و همکاران (۱۳۹۹) به‌منظور نقشه‌برداری رقومی کلاس‌های فامیل خاک از مدل‌های جنگل تصادفی و رگرسیون درختی توسعه یافته استفاده کردند و برتری جنگل تصادفی را نشان دادند. رحیمی مشکله و همکاران (۱۴۰۲) از الگوریتم‌های جنگل تصادفی، درخت تصمیم و رگرسیون لجستیک چند جمله‌ای در پیش‌بینی مکانی زیرگروه‌های خاک در بخشی از اراضی استان زنجان استفاده کرده و نشان دادند که مدل رگرسیون لجستیک چند جمله‌ای برآوردهای قابل قبول‌تری ارائه می‌دهد.

در زمینه تهیه نقشه کلاس‌های خاک با استفاده از روش‌های الگوریتم ماشین تاکنون مطالعه‌ای در استان کردستان صورت نگرفته است. بر این اساس و با توجه به نیاز کشور به داده‌های رقومی اطلاعات خاک، این پژوهش با هدف به‌کارگیری الگوریتم‌های مختلف ماشین یادگیری در پیش‌بینی مکانی کلاس‌های مختلف خاک در سطح خانواده و ارزیابی توان برآورد این روش‌ها در پیش‌بینی کلاس‌های خاک در شهرستان سنندج استان کردستان انجام گرفت.

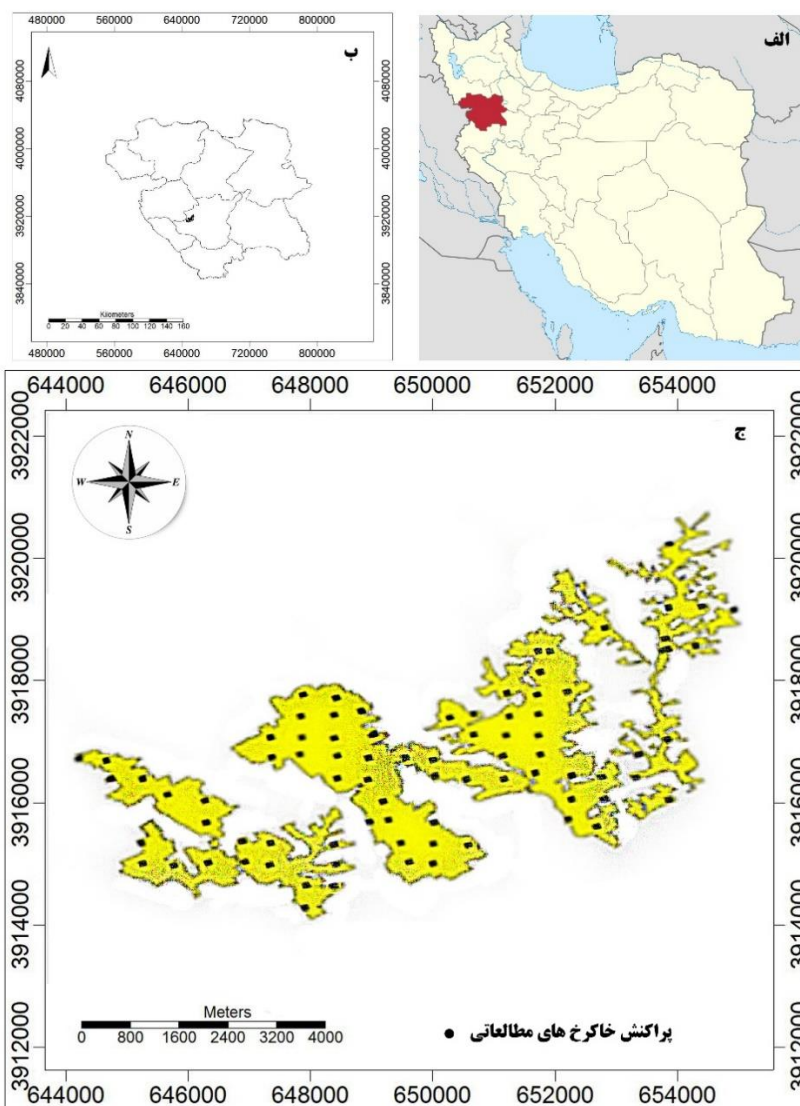
## روش‌شناسی پژوهش

### منطقه مطالعاتی

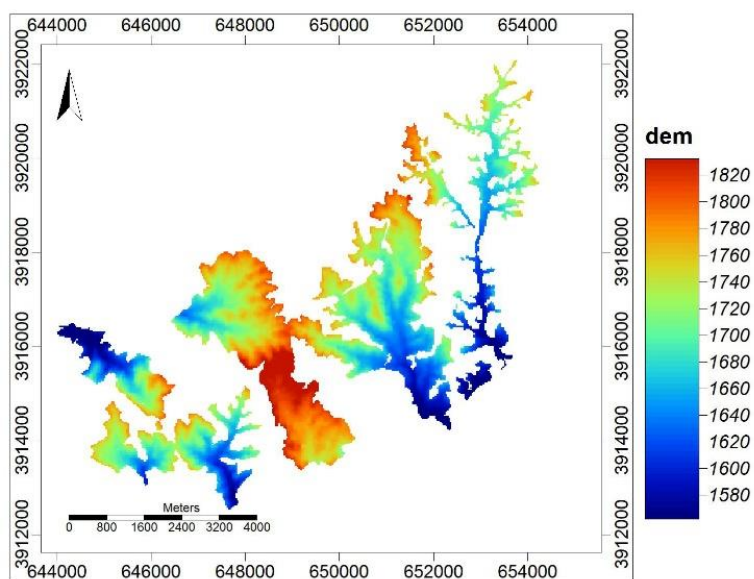
منطقه مورد مطالعه محدوده اراضی پایاب سد آزاد با وسعت حدود ۲۱۷۸/۳ هکتار است که در ۳۳ کیلومتری شمال غرب شهرستان سنندج واقع در استان کردستان و حدفاصل مختصات جغرافیائی "۴۶° ۳۵' ۵/۴۲" تا "۴۶° ۴۱' ۵۷/۲۵" طول شرقی و "۴۴/۰۶' ۲۰" تا "۳۵° ۴۸/۹" عرض شمالی واقع شده است (شکل ۱).

شکل ۲ مدل رقومی ارتفاع منطقه را نشان می‌دهد. میانگین دمای سالانه‌ی هوا ۱۱/۹ درجه سلسیوس، میانگین بارندگی سالانه ۵۰۰ میلی‌متر (اداره کل هواشناسی استان کردستان) و رژیم طوبتی و حرارتی خاک‌ها با استفاده از برنامه NSM<sup>۶</sup>، به ترتیب زیریک و مزیک است. بر اساس طبقه‌بندی اقلیمی دمارتن اصلاح‌شده منطقه دارای اقلیم مدیترانه‌ای سرد است. سازندهای زمین‌شناسی محدوده مطالعاتی بیشتر از جنس شیل آهکی سیلتی و رسی است (سازمان زمین‌شناسی و اکتشافات معدنی کشور) و واحد فیزیوگرافی منطقه تپه است (Mahler, 1970). شکل ۳ لندفرم‌های شناسایی‌شده در منطقه را نشان می‌دهد.

1. Random forest, RF
2. Neural networks, NN
3. Decision tree, DT
4. K-nearest neighbors, KNN
5. Support vector machines, SVM
6. Newhall Simulation Model

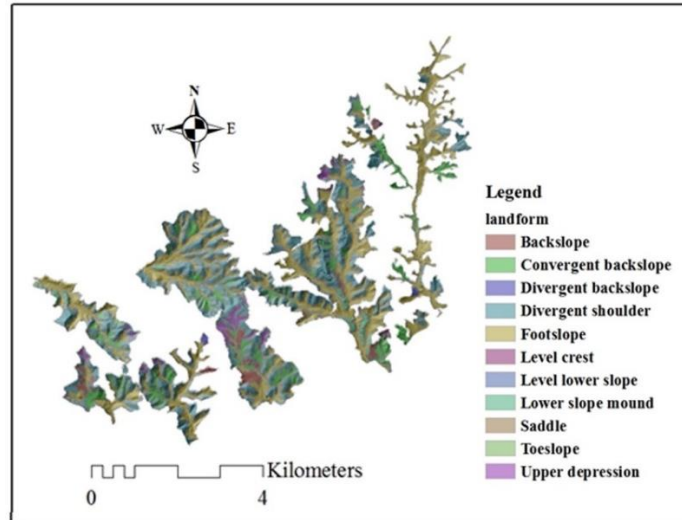


شکل ۱. موقعیت منطقه مطالعاتی و پراکنش خاکرخ های مطالعاتی



شکل ۲. مدل رقومی ارتفاع منطقه مطالعه شده

لندفرم‌های اصلی منطقه با استفاده از نرم‌افزار LandMapR شناسایی شدند (MacMillan, 2003). مهم‌ترین و وسیع‌ترین لندفرم‌های این ناحیه شانه واگرا<sup>۱</sup>، شیب پشتی همگرا<sup>۲</sup> و پای شیب<sup>۳</sup> می‌باشند (MacMillan et al., 2000). کاربری اصلی در منطقه کشاورزی است که عمده محصولات شامل گیاهان زراعی، گندم، جو و نخود و باغ‌های زردآلو است.



شکل ۳. نقشه لندفرم‌های شناسایی شده در منطقه

#### عملیات میدانی

مطالعات میدانی این پژوهش در سال ۱۳۹۶ انجام گرفت. با استفاده از بسته SP در محیط R به صورت شبکه منظم با فواصل ۵۰۰ متر، موقعیت مکانی ۸۴ خاک‌رخ (شکل ۱) تعیین و نسبت به حفر و تشریح خاک‌رخ‌ها با استفاده از راهنمای تشریح و نمونه‌برداری خاک (Schoeneberger et al., 2012) اقدام گردید. طی مطالعه صحرایی، ویژگی‌های مورفولوژیک مهم افق‌های مختلف خاک، از جمله رنگ، بافت، ساختمان، سنگریزه، پایداری و سایر موارد مشخص شد و از حدود ۲۵۲ افق ژنتیکی نمونه‌برداری شد. بر اساس نتایج ریخت‌شناختی در صحرا و نتایج تجزیه‌های آزمایشگاهی، خاک‌ها بر اساس روش جامع رده‌بندی خاک آمریکایی (Soil Survey Staff, 2022) تا سطح خانواده طبقه‌بندی شدند.

#### تجزیه‌های آزمایشگاهی

پس از هوا خشک شدن نمونه‌های خاک و عبور از الک ۲ میلی‌متری، ویژگی‌های فیزیکی و شیمیایی مهم خاک از جمله قابلیت هدایت الکتریکی، واکنش خاک، بافت خاک، مقدار مواد خنثی شونده و مقدار کربن آلی بر اساس روش‌های متداول مندرج در نشریات شماره‌های ۸۹۳ و ۱۰۲۴ (علی‌احیایی و بهبهانی‌زاده، ۱۳۷۱) مؤسسه تحقیقات خاک و آب تعیین شد.

#### محاسبه متغیرهای محیطی

استخراج متغیرهای محیطی با هدف تعیین متغیرهای پیش‌بینی کننده در مدل اسکورین انجام شد. در این مطالعه برای پیش‌بینی کلاس‌های خاک از متغیرهای محیطی به‌دست‌آمده از مدل رقومی ارتفاع<sup>۴</sup> با قدرت تفکیک مکانی ۱۰ متر استفاده شد. انتخاب متغیرهای کمی با توجه به وضعیت منطقه و دانش کارشناس و از میان متغیرهای کمی متداول در نقشه‌برداری رقومی خاک‌ها صورت گرفته است. برای محاسبه ویژگی‌های مربوط به عوارض زمینی/توپوگرافی از جمله شیب، ارتفاع، شاخص همواری دره و سایر ویژگی‌های مهم از نرم‌افزار SAGA 3.0.0 و بر اساس روش ارائه‌شده توسط Hengl et al., (2003) استفاده شد. همچنین از تصویر ماهواره سنتیل 2-B در محیط نرم‌افزار SAGA-GIS برای تهیه شاخص گیاهی تفاضلی نرمال شده<sup>۵</sup>، شاخص رس<sup>۶</sup>، شاخص تعدیل خاک پوشش گیاهی<sup>۷</sup>، شاخص اندازه

1. Divergent shoulder
2. Convergent Backslope
3. Footslope
4. Digital Elevation Model, DEM
5. Normalized Difference Vegetation Index (NDVI)
6. Clay Index
7. Soil Adjusted Vegetation Index (SAVI)



ذرات<sup>۱</sup>، شاخص اکسید آهن<sup>۲</sup> و شاخص قرمزی خاک<sup>۳</sup> (Andronikov et al., 1991; Olaya, 2004) بهره گرفته شد. متغیرهای کمکی استفاده شده برای تمامی مدل ها یکسان می باشد.

### مدل سازی

برای پیش بینی کلاس های خاک در سطح خانواده، مقادیر متغیرهای کمکی محیطی و کلاس های خاک شناسایی شده در منطقه مورد مطالعه وارد محیط نرم افزار (R-Studio 0.97.551) (R Development Core Team, 2013) گردید و با استفاده از بسته های نرم افزاری جنگل تصادفی، درخت تصمیم، شبکه عصبی مصنوعی و رگرسیون لاجستیک چندجمله ای با مراحل گوناگون مدل سازی انجام گرفت.

### جنگل تصادفی

جنگل تصادفی نوعی از روش های ماشین یادگیری است که شامل مجموعه ای از درختان رده بندی و رگرسیونی (CART) تصادفی است (Breiman, 2001). در این الگوریتم شمار زیادی درخت تصمیم گیری ساخته شده که برای یک پیش بینی معین با هم ترکیب می شوند. در رگرسیون، پیش بینی انجام شده میانگینی از خروجی های هر درخت جدا است، در حالی که در طبقه بندی، با توجه به همگرایی بیشتر درختان تصمیم گیری، درست بودن رده بندی سنجش می شود. در مرحله آموزش، الگوریتم جنگل تصادفی چندین درخت شبیه CART<sup>۴</sup> می سازد که هر یک از این درخت ها بر یک نمونه گیری قابل جایگزین (نمونه گیری همراه با جایگذاری) از داده های آموزشی استوار است. این مدل شامل سه پارامتر از جمله تعداد متغیرهای مورد استفاده برای ساخت هر درخت، تعداد درختان و حداقل گره انتهایی است که توسط کاربر تعریف می شود (Rostaminia et al., 2021). در این مدل سازی ntree=1000 و mtry=5 در نظر گرفته شده است. این مدل در نرم افزار R و با استفاده از بسته Random Forest اجرا شد (Kuhn, 2018).

### درخت تصمیم

مدل درخت تصمیم گیری روش موفقی برای پیش بینی و تبیین رابطه میان برخی متغیرهای کمکی و متغیر هدف است. درخت تصمیم گیری برگرفته از منطق، مدیریت و آمار بوده و از ابزارهای بسیار مؤثر در ماشین یادگیری جهت تشخیص الگو و استخراج اطلاعات است (Moran & Bui, 2002). درخت تصمیم گیری برای دسته بندی و پیش بینی بر پایه استنتاج استقرائی<sup>۵</sup> به کار می رود و به صورت خودکار یک سری از قواعد<sup>۶</sup> را از روی داده های آموزشی می سازد. همچنین تصمیمات پیچیده را به چند تصمیم ساده قابل درک و ساده تقسیم می کند. در این پژوهش از بسته نرم افزاری C5.0 برای ساخت درخت تصمیم گیری و در نهایت پیش بینی کلاس های خاک استفاده شد. بسته نرم افزاری C5.0 یکی از کارآمدترین الگوریتم های استقرائی است (Quinlan, 2001). بسته نرم افزاری C5.0 از نسبت بهره اطلاعات برای ارزیابی تقسیمات در هر گره استفاده می کند. به دیگر سخن تقسیم داده ها را بر پایه افزودن بهره اطلاعات انجام می دهد. هر شاخه توسط تقسیم اول تعریف می شود و خود دوباره تقسیم می شود و این روند تا زمانی ادامه می یابد که دیگر نتوان داده ها را تقسیم کرد (Adu-Poku, 2012).

### رگرسیون لاجستیک چندمتغیره

مدل رگرسیون لاجستیک شکل خاصی از مدل خطی تعمیم یافته است که به دو صورت قابل اجرا است: مدل لاجستیک دوتایی و لاجستیک چندجمله ای. متغیر وابسته دارای چند کلاس یا طبقه است، مثل کلاس های خاک در یک منطقه. به منظور تبدیل مجموعه ای از مدل های رگرسیون لاجستیک دو جمله ای به مدل های چندجمله ای تعمیم یافته، از معادله ۱ استفاده شد:

$$P_i = \frac{\exp(P_i)}{\exp(P_1) + \exp(P_2) + \dots + \exp(P_n)} \quad \text{رابطه ۱}$$

Pi نشان دهنده احتمال رخداد برای کلاس i و مخرج معادله نشان دهنده مجموع احتمالات رخداد برای n کلاس است. در طبقه بندی

نهایی، هر نقطه داده به کلاسی تعلق می گیرد که بیشترین احتمال رخداد را دارد.

رگرسیون لاجستیک چندجمله ای با استفاده از معادله ۲ تعریف می شود (Kempen et al., 2009):

$$\log\left(\frac{\pi_{ij}}{\pi_{ij}}\right) = \alpha_j + \beta_{jx_i}, j = 1, 2, \dots, J - 1 \quad \text{رابطه ۲}$$

1. Grain Size Index (GSI)
2. Iron Oxide (IO)
3. Soil Redness Index (SRI)
4. Classification and Regression Tree
2. Inductive Reasoning
3. Rulle



که در آن  $\alpha_j$ ، یک مقدار ثابت،  $\beta_j$  بردار ضریب رگرسیون، برای  $j = 1, 2, \dots, J - 1$  و  $x_j$  بردار متغیرهای مستقل (تشریح کننده) است. توزیع احتمال در این مدل رگرسیونی چند جمله‌ای و  $J - 1$  از معادله ۳ قابل محاسبه است.

$$P(y_i = j) = \pi_{ij} \frac{\exp(\beta_j x_j)}{1 + \sum_{j=1}^J \exp(\beta_j x_j)} \quad \text{رابطه ۳}$$

و احتمال رخداد هر کلاس با استفاده از معادله ۴ برابر است با:

$$P(y_i = 0) = \pi_{i0} = \frac{1}{1 + \sum_{j=1}^J \exp(\beta_j x_j)} \quad \text{رابطه ۴}$$

### شبکه عصبی مصنوعی

شبکه عصبی مصنوعی روشی برای تحلیل اطلاعات است که از سیستم عصبی انسان الهام گرفته و مشابه مغز انسان به پردازش و دسته‌بندی اطلاعات می‌پردازد. این سیستم پردازشی از تعداد زیادی عناصر به هم پیوسته به نام نرون‌ها تشکیل شده که برای حل یک مسئله به طور هماهنگ با هم عمل می‌کنند. شبکه عصبی مصنوعی پرسپترون چندلایه با یک لایه مخفی که دارای تابع فعال‌سازی سیگموئید در لایه مخفی و تابع فعال‌سازی خطی در لایه خروجی بود، استفاده شد. تعداد نرون‌ها بین دو تا ۱۰ نرون متغیر بوده و بهترین تعداد نرون به صورت سعی و خطا تعیین شد (Marquardt, 1963). در این مطالعه از نرم افزار R و بسته نرم‌افزاری nnet برای ساختن شبکه عصبی مصنوعی و در نهایت پیش‌بینی کلاس‌های خاک استفاده شد.

### ارزیابی عملکرد مدل‌ها

نکته بسیار مهم در تولید نقشه خاک کیفیت و صحت آن است. نقشه‌برداری رقومی خاک امکان محاسبه کمی صحت نقشه خاک را از طریق ارزیابی مدل‌های به کار برده شده در تهیه نقشه ممکن ساخته است. برای ارزیابی صحت نقشه یا به عبارتی دقت مدل ابتدا مجموعه داده‌ها به دو مجموعه داده آموزشی (۸۰ درصد) و اعتبارسنجی (۲۰ درصد) تقسیم شدند. برای محاسبه دقت پیش‌بینی از صحت عمومی<sup>۱</sup> (معادله ۵)، صحت کاربر<sup>۲</sup> (معادله ۶)، صحت تولیدکننده<sup>۳</sup> (معادله ۷)، شاخص کاپا<sup>۴</sup> (معادله ۸) (Lu et al., 2004) و شاخص درجه برابر (معادله ۹) (Johansson et al., 2010) استفاده شد.

صحت عمومی نشان می‌دهد که چه تعداد از واحدها یا طبقات جدا شده صحیح است. این کمیت عبارت است از نسبت مجموع کل درایه‌های قطری به کل درایه‌ها.

$$OA = \frac{\sum_{i=1}^n X_{ii}}{N} \quad \text{رابطه ۵}$$

OA صحت عمومی، N تعداد کل مشاهده‌ها و  $\sum P_{ii}$  جمع عناصر قطر اصلی ماتریس خطا است.

صحت کاربر نشان می‌دهد که اطلاعاتی که کاربر از نقشه دریافت می‌کند چه اندازه درست است. این کمیت عبارت از نسبت مجموع درایه‌های قطری یک ردیف به مجموع کل درایه‌های همان ردیف است.

$$UA = \frac{X_{ii}}{X_{i0}} \quad \text{رابطه ۶}$$

UA صحت کاربر،  $X_{ii}$  درایه‌های قطری ردیف i ام،  $X_{i0}$  مجموع ستون‌ها برای ردیف i ام.

صحت تولیدکننده بیان‌گر ارتباط کلاس‌های صحیح پیش‌بینی شده (tt) و مجموع کلاس‌های صحیح پیش‌بینی شده و کلاس حضور مشاهده شده (واقعی) که به غلط جزء کلاس‌های عدم حضور پیش‌بینی شده‌اند (ft) است.

$$PR = \frac{tt}{tt + ft} \quad \text{رابطه ۷}$$

شاخص کاپا شاخصی قوی است که نسبت احتمال حضور یا عدم حضور کلاس که به درستی به وسیله مدل پیش‌بینی شده‌اند را محاسبه می‌کند؛ بنابراین شاخص کاپا همیشه از مقدار صحت کل کمتر است.

$$KI = \frac{\text{observed} - \text{chance agreement}}{1 - \text{chance agreement}} \quad \text{رابطه ۸}$$

در این معادله دقت مشاهده به عنوان مجموع ارزش قطر در ماتریس و شانس خطا شامل مجموع ارزش غیر قطر ماتریس است. دامنه تغییرات شاخص کاپا بالاتر از ۰/۸، ۰/۴ تا ۰/۸ و کمتر از ۰/۴ به ترتیب نشان‌دهنده توافق قوی، متوسط و ضعیف هستند

1. Overall accuracy, OA
2. User's accuracy, UA
3. Producer's reliability, PR
4. Kappa Index, KI



(Congalton, 1991).

شاخص درجه برابر که برای محاسبه آن از معادله ۹ استفاده شد.

$$\text{Brier score} = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^n (F_{ij} - E_{ij})^2 \quad \text{(رابطه ۹)}$$

که در آن  $n \times r$  و  $F_{ij}$  به ترتیب عبارت اند از تعداد کلاس ها، تعداد مشاهده ها و احتمال تخمین اینکه مشاهده  $n_i$  به کلاس  $r_j$  متعلق باشد. اگر کلاس مشاهده شده با پیش بینی شده یکی شود،  $F_{ij}$  برابر یک و در غیر این صورت برابر با صفر در نظر گرفته می شود (Johansson et al., 2010). در نهایت مدلی با بالاترین مقدار صحت عمومی، شاخص کاپا و کمترین مقدار درجه برابر به عنوان بهترین مدل برای پیش بینی کلاس های خاک در سطح خانواده خاک در نظر گرفته شد.

## نتایج و بحث

### آشنایی با خاک منطقه

بر اساس نتایج حاصل از مطالعات صحرایی و تجزیه های آزمایشگاهی، خاک رخی های بررسی شده در منطقه در دو رده ای انتی سول و اینسپتی سول بر اساس سیستم جامع رده بندی آمریکایی قرار داشتند. برخی از ویژگی های فیزیکی و شیمیایی پروفیل های شاهد در جدول ۱ آورده شده است. رده ای اینسپتی سول با فراوانی ۶۷ خاکرخ از ۸۴ خاکرخ، در منطقه مطالعاتی گسترش بیشتری داشت. تکامل این خاکها در منطقه مطالعاتی در حدی نبود که بتوان آن ها را در رده های متکامل تری قرار داد. افق های سطحی تشکیل شده در این خاکها شامل اکریک و در زمین های کشاورزی آنتروپیک بود و در عمق نیز بیشتر افق های کلسیک و کمبیک مشاهده شد. انتی سول های منطقه فاقد تکامل خاکرخ بوده و در سطح این خاکها افق های سطحی اکریک و آنتروپیک مشاهده شد. این در حالی است که در افق های تحتانی معمولاً افق مشخصه ای تشکیل نشده است.

جدول ۱. ویژگی های فیزیکی و شیمیایی خاک رخی های شاهد به همراه طبقه بندی آن ها تا سطح خانواده

Texture	Clay	Silt	Sand %	CaCO <sub>3</sub>	OC	Gravel %	meq/l						CEC meq/100gr	EC ds/m	pH	افق
							Na <sup>+</sup>	K <sup>+</sup>	Ca <sup>2+</sup> + Mg <sup>2+</sup>	CO <sub>3</sub> <sup>2-</sup>	HCO <sub>3</sub> <sup>-</sup>	Cl <sup>-</sup>				
Fine-loamy, mixed, superactive, mesic Lithic Xerothents: P-1																
CL	۳۲	۴۰	۲۸	۲۳/۸	۳/۲	۱۵-۲۰	۱	...	۱۰	...	۸	۲	۱۹/۴	۱/۴	۷/۴۵	Ap
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	R
P-58: Fine, mixed, active, mesic Lithic Xerothents																
SiC	۴۴	۴۲	۱۴	۲۱/۳	۰/۸۲	۱۰-۱۵	۱/۱۹	۰	۳/۶	۰	۳	۱/۵	۲۴/۲	۰/۸۸	۷/۴۵	Ap
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	R
Fine, mixed, active, mesic Typic Haploxerepts: P-11																
SiC	۴۲	۴۴	۱۴	۱	۰/۷۷	۰	۱/۳۷	۰	۶	۰	۴/۵	۱	۲۱/۶	۱	۷/۷	Ap
C	۴۸	۳۸	۱۴	۰/۹۵	۰/۴۱	۰	۱/۹۴	۰	۵/۶	۰	۴	۱/۵	۲۱/۶	۰/۹۵	۷/۶	Bw1
C	۵۲	۳۸	۱۰	۰/۹	۰/۱۳	۰	۲/۴۲	۰	۴/۸	۰	۴	۰/۵	۲۲	۰/۸۵	۷/۷	Bw2
Fine, mixed, active, mesic, Typic Calcixerepts: P-48																
SiC	۴۸	۴۴	۸	۲۳	۰/۹۲	۰	۱/۳۹	...	۳/۶	۰	۳/۵	۱/۵	۳۳/۶	۰/۹۲	۷/۵	Ap
C	۵۶	۳۶	۸	۲۱	۰/۴۹	۰	۱/۱۲	...	۳/۸	۰	۲/۵	۱	۳۱/۲	۰/۹	۷/۵۵	Bw
C	۵۶	۳۸	۶	۳۹	۰/۱	۰	۱/۴۱	...	۲/۸	۰	۲	۱/۵	۲۸/۸	۰/۸	۷/۷	Bk1
C	۴۴	۴۸	۸	۵۳/۱	۰/۱	۰	۱/۲۸	...	۳/۶	۰	۲/۵	۰/۵	۱۷/۴	۰/۸۸	۷/۷	Bk2
Fine, mixed, active, mesic, Typic Haploxerepts: P-17																
SiC	۴۴	۴۸	۸	۱۸/۵	۱/۱	۰	۱/۲۱	۰	۳/۶	۰	۳/۵	۱	۱۹	۰/۸۸	۷/۵	Ap
C	۵۶	۳۸	۶	۲۲/۵	۰/۹۲	۰	۱/۴۲	۰	۲/۸	۰	۳	۱	۲۶/۴	۰/۸۷	۷/۶	Bw
SiC	۴۸	۴۲	۱۰	۲۲/۵	۰/۳۱	۰	۱/۰۴	۰	۲/۸	۰	۲/۵	۱	۲۰	۰/۸۸	۷/۷۵	Bk1
C	۵۰	۳۸	۱۲	۲۳	۰/۱۳	۰	۱/۱۷	۰	۲/۴	۰	۲/۵	۰/۵	۲۰	۰/۸۹	۷/۱۵	Bk2
Fine, mixed, active, mesic, Chromic Haploxerepts: P-24																
C	۵۲	۴۰	۸	۲/۵	۰/۳۶	۰	۱/۲۱	۰	۴	۰	۳/۵	۱/۵	۲۰/۸	۰/۸۲	۷/۱	Ap
C	۶۰	۳۰	۱۰	۱/۳	۰/۱	۰	۱/۵۴	۰	۴	۰	۳/۵	۱	۲۲/۴	۰/۸۲	۷	Bss1
C	۶۲	۲۸	۱۰	۲	۰/۱۳	۰	۱/۲۳	۰	۳/۲	۰	۲/۵	۱	۳۱/۶	۰/۸۵	۷/۳	Bss2

بر اساس سیستم جامع رده‌بندی خاک‌ها، به ترتیب دو زیررده، سه گروه بزرگ، پنج زیرگروه و پنج خانواده در منطقه مورد مطالعه شناسایی شد که نام و فراوانی هر یک از آن‌ها در جدول ۲ ارائه شده است.

جدول ۲. نوع و فراوانی کلاس‌های خاک مشاهده شده در منطقه مطالعاتی

کلاس خاک	نام	تعداد مشاهده‌ها
رده	Entisols	۱۷
	Inceptisols	۶۷
زیررده	Orthents	۱۷
	Xerepts	۶۷
گروه بزرگ	Xerorthents	۱۷
	Haploxerepts	۵۰
	Calcixerepts	۱۷
	Lithic Xerorthents	۱۷
زیرگروه	Typic Haploxerepts	۲۶
	Typic Calcixerepts	۱۷
	Fluventic Haploxerepts	۱۶
	Chromic Haploxerepts	۸
	Fine-loamy, mixed, mesic Lithic Xerorthents	۱۷
خانواده	Fine, mixed, mesic Typic Haploxerepts	۲۶
	Fine, mixed, mesic Chromic Haploxerepts	۸
	Fine, mixed, mesic Fluventic Haploxerepts	۱۶
	Fine, mixed, mesic Typic Calcixerepts	۱۷

#### پیش‌بینی کلاس‌های خاک در سطح خانواده

نتایج صحت پیش‌بینی کلاس‌های خاک با استفاده از مدل‌های مختلف یادگیری ماشین در جدول ۳ ارائه شده است. مدل‌های مختلف در پیش‌بینی کلاس‌های خاک عملکرد متفاوتی داشته‌اند اما در مجموع مدل‌ها، کلاس‌هایی را که فراوانی بیشتری داشته‌اند، با صحت بیشتری پیش‌بینی کرده‌اند. همان‌طور که قبلاً اشاره شد کلاس Fine, mixed, mesic, Typic Haploxerepts بیشترین فراوانی را در منطقه مطالعاتی داشته است؛ بنابراین به دلیل تعداد زیاد داده برای پیش‌بینی، مدل‌های به‌کاررفته این کلاس خاک را با صحت تولیدکننده و کاربر بالایی پیش‌بینی نموده‌اند. مدل شبکه عصبی مصنوعی Fine, mixed, mesic, Fluventic Haploxerepts را با صحت تولیدکننده ۱۰۰ و صحت کاربر ۸۰ پیش‌بینی کرده است. مدل‌های جنگل تصادفی، رگرسیون لاجستیک و درخت تصمیم با صحت تولیدکننده بالا و مدل‌های جنگل تصادفی و رگرسیون لاجستیک با صحت کاربر بالایی این کلاس خاک را پیش‌بینی کردند. بعضی از مدل‌های به‌کاررفته از جمله مدل‌های شبکه عصبی مصنوعی و رگرسیون لاجستیک کلاس Fine, mixed, mesic, Fluventic Haploxerepts را با وجود اینکه تعداد مشاهده‌های کمتری در مقایسه با Fine, mixed, mesic, Typic Haploxerepts داشت با صحت تولیدکننده و کاربر بالاتری پیش‌بینی نمودند. دلیل صحت بالا در پیش‌بینی این کلاس خاک احتمالاً به دلیل ارتباط بالای این کلاس خاک با متغیر محیطی و همچنین گسترش این خاک‌ها در بخش‌های مختلف دشت است که خاک مشابه با آن‌ها وجود ندارد. در بین کلاس‌های خانواده خاک، کلاس Fine, mixed, mesic, Chromic Haploxerepts به دلیل تعداد کمتر مشاهده‌ها در منطقه توسط مدل‌های به‌کاررفته با صحت کاربر و تولیدکننده پایینی پیش‌بینی شد به‌نحوی که صحت تولیدکننده و کاربر پیش‌بینی این کلاس خاک توسط مدل‌های شبکه عصبی مصنوعی، رگرسیون لاجستیک و درخت تصمیم، صفر به دست آمد. در پژوهش‌های مختلف به اهمیت تعداد بالای مشاهده‌ها در صحت پیش‌بینی‌ها توسط مدل‌های مختلف اشاره شده است (Campling et al., 2002; Hengl et al., 2007). از دلایل دیگر پیش‌بینی ضعیف این کلاس خاک در منطقه احتمالاً می‌توان به مسطح بودن و تغییرات کم توپوگرافی در محل وقوع این کلاس خاک در منطقه اشاره نمود که سبب کاهش عملکرد مدل‌ها در پیش‌بینی این کلاس خاک شده است.

به‌طور کلی صحت تولیدکننده و کاربر پیش‌بینی کلاس‌های مختلف خانواده به‌جز Fine, mixed, mesic, Chromic Haploxerepts نسبتاً مناسب بود که دلیل عمده آن مربوط به تعداد مناسب مشاهده برای هر کلاس و انتخاب مناسب متغیرهای کمکی بوده است.

جدول ۳. مقادیر صحت کاربر و تولید کننده کلاس های خاک توسط مدل های مختلف

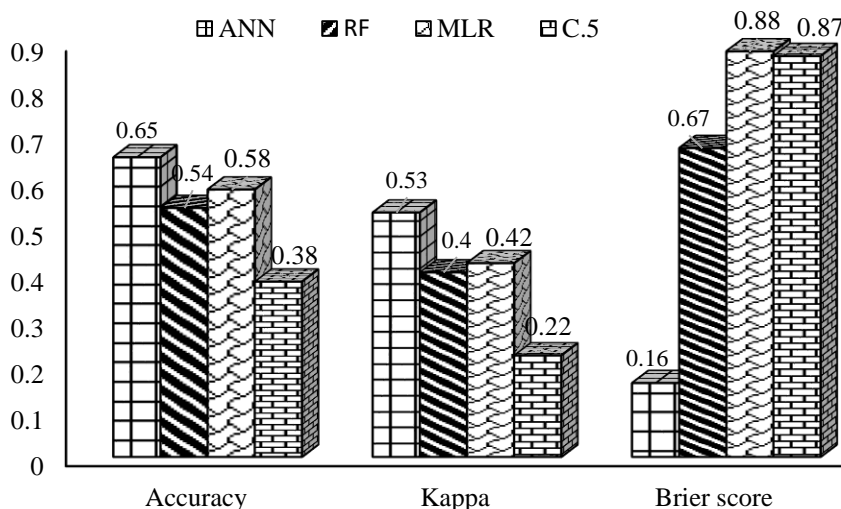
نام خاک	صحت کاربر				صحت تولید کننده			
	ANN	DT	MLR	RF	ANN	DT	MLR	RF
Fine, mixed, mesic, Chromic Haploxerepts	۰	۰	۰	۳۳/۳۳	۰	۰	۰	۳۳/۳۳
Fine, mixed, mesic, Fluventic Haploxerepts	۸۰	۲۱/۴۳	۹۰/۹۱	۷۷/۷۸	۱۰۰	۷۵	۸۳/۳۳	۶۳/۶۴
Typic Calcixerepts Fine, mixed, mesic,	۵۰	۴۰	۱۴/۲۹	۵۰	۵۰	۲۵	۵۰	۱۶/۶۷
Fine, mixed, mesic, Typic Haploxerepts	۵۰	۷۵	۶۰	۳۷/۵	۵۰	۴۲/۸۶	۶۰	۱۰۰
Fine-loamy, mixed, mesic, Lithic.xerorthents	۶۰	۶۶/۶۷	۵۰	۵۰	۶۰	۴۰	۲۵	۶۶/۶۷

### عملکرد مدل های به کار رفته

برای تعیین بهترین مدل در پیش بینی کلاس های خاک از شاخص های صحت عمومی، کاپا و درجه برابر استفاده شد که نتایج آن در شکل ۴ مشاهده می شود.

با توجه به شکل ۴ مدل های مختلف در پیش بینی کلاس خانواده خاک عملکرد متفاوتی داشتند. بر اساس مقادیر صحت عمومی پیش بینی ها، مدل شبکه عصبی مصنوعی و درخت تصمیم به ترتیب بیشترین و کمترین دقت را در پیش بینی داشتند. در مجموع با توجه به مقادیر صحت عمومی، شاخص کاپا و شاخص درجه برابر مدل های به کار رفته به جز مدل درخت تصمیم، سایر مدل ها عملکرد نسبتاً مناسبی در منطقه مطالعاتی داشتند.

از عوامل مهم تاثیرگذار در نقشه برداری رقومی که در بسیاری از مطالعات به آن تأکید شده است، می توان به تعداد نقاط مشاهداتی (Saurette et al., 2022)، و نوع متغیرهای کمکی و الگوریتم استفاده شده اشاره کرد (Kim et al., 2012; Barthold et al., 2013). تعداد کم نمونه ها اغلب با دقت پایین پیش بینی ها همراه است (Sharififar et al., 2019). رحیمی مشکله و همکاران (۱۴۰۲) نشان دادند که رابطه مستقیمی بین فراوانی کلاس های خاک در داده های آموزشی با صحت نتایج پیش بینی کلاس های خاک برقرار است. سایر عوامل سنجنده عملکرد مدل ها نشان از پیش بینی با صحت بالاتر مدل شبکه عصبی مصنوعی در مقایسه با سایر مدل ها دارد به نحوی که این مدل مقدار کاپا و صحت عمومی بیشتر از ۵۰ درصد و شاخص درجه برابر برابر ۰/۱۶ بهترین عملکرد را داشته که نشان از صحت بالا و خطای کم در پیش بینی ها دارد. در بسیاری از پژوهش ها مدل شبکه عصبی مصنوعی را به عنوان الگوریتمی قوی در پیش بینی ویژگی های خاک و کلاس های خاک دانسته اند (Rodriguez-Galiano et al., 2015; Tiwari et al., 2015) و در برخی از پژوهش ها از جمله مطالعه تقی زاده مهرجردی و همکاران (۱۳۹۳) در پیش بینی گروه بزرگ و زیرگروه خاک در منطقه ی اردکان یزد و همچنین Luoto & Hjort, (2005) به این نتیجه دست یافتند که عملکرد مدل درخت تصمیم در مقایسه با مدل رگرسیون لاجستیک و شبکه عصبی مصنوعی بهتر است؛ بنابراین عملکرد مدل های مختلف بستگی زیادی به منطقه مطالعاتی، الگوی نمونه برداری و نوع و تعداد متغیرهای پیش بینی کننده دارد.



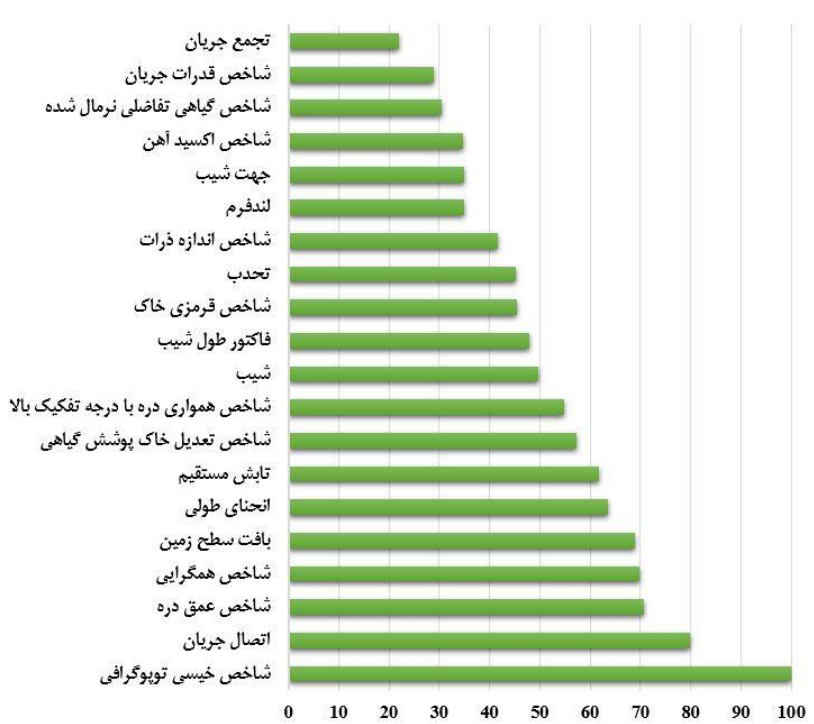
شکل ۴. مقدار صحت عمومی، شاخص کاپا و درجه برابر پیش بینی کلاس های خاک در سطح خانواده با استفاده از مدل های مختلف

### اهمیت نسبی متغیرهای محیطی

شکل ۵ اهمیت متغیرهای کمکی در پیش‌بینی کلاس‌های خاک را در مدل شبکه عصبی مصنوعی نشان می‌دهد. در منطقه مطالعه شده متغیرهای کمکی به‌دست‌آمده از مدل رقومی ارتفاع بیشترین اهمیت را در پیش‌بینی کلاس‌های خاک داشته است. این موضوع نشان می‌دهد که در این منطقه پستی و بلندی از عوامل بسیار مهم خاکسازي است. از بین متغیرهای کمکی به‌دست‌آمده از مدل رقومی ارتفاع نیز شاخص خیسی توپوگرافی بالاترین اهمیت نسبی را به خود اختصاص داده است. این شاخص که از ترکیب مساحت حوضه بالا دست و شیب به دست می‌آید (Grabs et al., 2009)، ابزاری مفید و رایج برای تعیین توزیع مکانی نفوذ و جریان آب سطحی به کار رفته و نشان دهنده تاثیر توپوگرافی بر مکان و میزان تجمع رطوبت در خاک می‌باشد (تقی‌زاده مهرجردی و همکاران، ۱۳۹۳). از شاخص خیسی توپوگرافی در مطالعات زیادی به منظور تخمین مکانی ویژگی‌های خاک استفاده شده است (Western, 2004). در مطالعات تقی‌زاده مهرجردی و همکاران (۱۳۹۳) نیز این شاخص به عنوان مهمترین متغیرهای کمکی در پیش‌بینی مکانی گروه‌های بزرگ خاک در منطقه اردکان یزد گزارش شده است.

توپوگرافی از طریق کنترل فرآیندهای فرسایش و رسوب به میزان زیادی بر تغییرات نوع خاک اثرگذار است (Chen et al., 2022). همچنین با تاثیر بر توزیع مکانی تابش خورشیدی، توزیع مجدد رطوبت و همچنین تراکم و نوع پوشش گیاهی، بر روی فرآیندهای خاکسازي و در نتیجه بر روی کلاس خاک تاثیر دارد (Han et al., 2022). در مناطق کوهستانی، توپوگرافی ارتباط نزدیکی با ویژگی‌های محیطی و خاک داشته و به‌طور گسترده‌ای در نقشه‌برداری رقومی به منظور پیش‌بینی مکانی کلاس‌های خاک مورد استفاده قرار می‌گیرند (Scull, 2005). این در حالی است که حتی در مناطق غیرکوهستانی با پستی و بلندی کم نیز این متغیرها از متغیرهای کمکی مهم و تاثیرگذار در نقشه‌برداری رقومی خاک به شمار می‌روند (Mosleh et al., 2016).

موسوی و همکاران (۱۳۹۸) نشان دادند که پارامترهای ژئومورفومتری نسبت به شاخص‌های سنجش از دور در نقشه‌برداری رقومی کلاس‌های خاک از اهمیت بالاتری برخوردار است. مطالعات بسیاری نیز بر اهمیت بالای پارامترهای ژئومورفومتری در نقشه‌برداری رقومی کلاس‌ها و ویژگی‌های خاک تاکید کرده‌اند (Taghizadeh Mehrjardi et al., 2019).

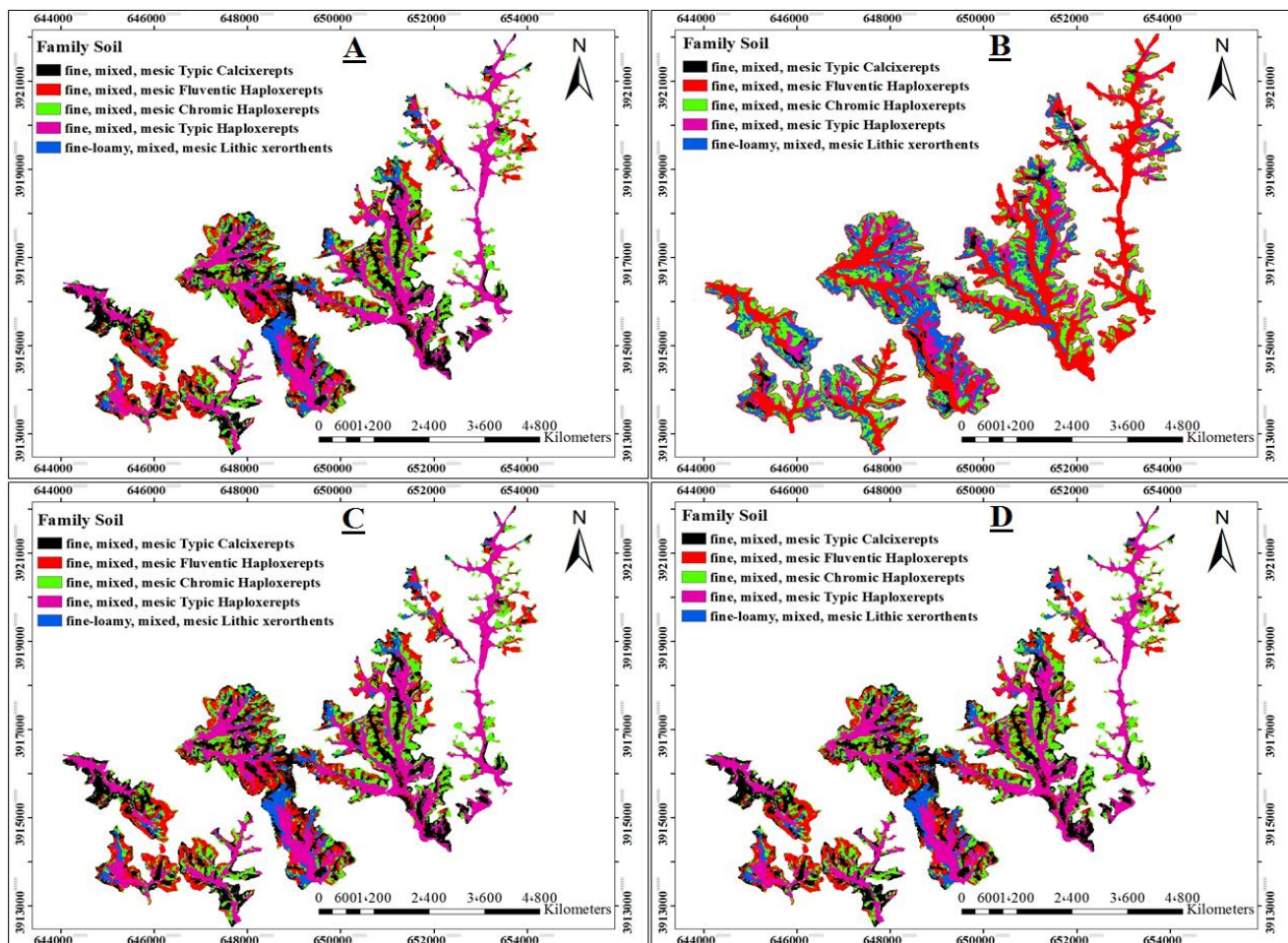


شکل ۵. اهمیت متغیرهای کمکی در پیش‌بینی کلاس‌های خاک با استفاده از مدل شبکه عصبی مصنوعی

### پراکنش مکانی کلاس‌های خاک

خاک‌های Fine, mixed, mesic, Typic Haploxerepts بیشترین پراکنش مکانی را در منطقه مطالعاتی داشت. این خاک بیشتر در اراضی

با شیب های بیش از ۲۵ درصد تشکیل شده اند. در این خاک ها با افزایش عمق مقدار کربن آلی خاک به طور منظم افزایش پیدا نکرده و در عمق ۱۲۰ سانتی متر مقدار کربن آلی کمتر از ۰/۲ درصد برآورد شده است. این خاک ها عمیق تا خیلی عمیق بوده و فاقد افق مشخصه کلسیک هستند اما افق کمبیک در آن ها به خوبی توسعه و تکامل پیدا کرده است. بافت خاک در افق های سطحی و زیرسطحی این خاک ها خیلی سنگین است. با توجه به شکل ۶ و همچنین نتایج صحت پیش بینی ها موقعیت قرارگیری این کلاس خاک توسط مدل درخت تصمیم به درستی تعیین نشده است. الگوریتم درخت تصمیم بیشتر مشاهده ها مربوط به این کلاس خاک را اشتباهاً Fine, mixed, mesic, Fluventic Haploxerepts تشخیص داده است به نحوی که با مشاهده نقشه پراکنش Fine, mixed, mesic, Fluventic Haploxerepts بیشتر از کلاس Fine, mixed, mesic, Typic Haploxerepts است که با واقعیت منطقه همخوانی ندارد. سایر مدل ها در تعیین پراکنش مکانی کلاس Fine, mixed, mesic, Typic Haploxerepts عملکرد مناسبی داشتند که با مشاهده صحرایی و واقعیت منطقه همخوانی دارد. خاک های Fine, mixed, mesic, Typic Calcixerepts در منطقه مطالعاتی بیشتر بر روی مواد مادری سنگ آهکی و در لندفرم پای شیب اراضی تپه ای تشکیل شده اند. از آنجایی که این کلاس خاک در مناطقی با ارتفاع بیشتر نسبت به پنجه شیب تشکیل شده اند، از این رو پدیده های سیلگیری و ماندابی همانند خاک های Fine, mixed, mesic Fluventic Haploxerepts در آن ها مشاهده نشده است. افق های کلسیک در این خاک ها در عمق ۱۰۰ سانتی متری از سطح خاک تشکیل شده است. در این کلاس خاک علاوه بر افق مشخصه کلسیک، افق کمبیک نیز در بیشتر خاک ها مشاهده شد.



شکل ۶. پیش بینی پراکنش مکانی خانواده خاک در منطقه مطالعاتی با استفاده از مدل های الف) شبکه عصبی مصنوعی، ب) درخت تصمیم، ج) رگرسیون لاجستیک و د) جنگل تصادفی

Fine-loamy, mixed, mesic, Lithic Xerorthents خاک های کم عمق فرسایش یافته هستند. بافت آن ها لومی رسی سنگریزه دار بوده و از مواد مادری آهکی، شیلی و سیلتی زیرین منشأ گرفته اند. این خاک ها عمدتاً در مناطق مرتفع و شیب دار قرار دارند. ساختمان سنگی بخش اعظم این خاک ها را تشکیل داده است.





Fine, mixed, mesic, Chromic Haploxerepts بر روی مواد مادری شیل و سنگ‌آهک در لندفرم پای شیب تشکیل شده‌اند. موقعیت قرارگیری این کلاس خاک در نقشه‌های مدل شبکه عصبی، درخت تصمیم و رگرسیون لاجستیک احتمالاً اشتباه است و با واقعیت منطقه همخوانی ندارد. این خاک عملاً توسط مدل کم برآزش شده و در کلاس‌های دیگری قرار گرفته است. مدل جنگل تصادفی نیز این کلاس خاک را با صحت پایینی پیش‌بینی کرده است؛ بنابراین پراکنش این خاک‌ها دقیق نیست. به‌طور کلی مدل‌ها کلاس‌های Fine, mixed, mesic, Fluventic Haploxerepts و Fine, mixed, mesic, Typic Haploxerepts را با صحت بیشتری پیش‌بینی نموده‌اند که با مشاهدات میدانی و واقعیت منطقه هم همخوانی دارد.

## نتیجه‌گیری

در این مطالعه از الگوریتم‌های شبکه عصبی مصنوعی، رگرسیون لاجستیک، جنگل تصادفی و درخت تصمیم برای نقشه‌برداری رقومی کلاس‌های خاک استفاده گردید. بررسی صحت عمومی، شاخص کاپا و درجه برابر نشان دهنده عملکرد متفاوت مدل‌های مختلف می‌باشد. مدل‌های شبکه عصبی مصنوعی، رگرسیون لاجستیک و جنگل تصادفی با صحت بالایی پیش‌بینی را انجام داده‌اند اما مدل درخت تصمیم عملکرد مناسبی از خود نشان نداد. از بین مدل‌های پیشنهادی نیز شبکه عصبی مصنوعی بهترین عملکرد را در پیش‌بینی کلاس‌های خاک داشت. عملکرد متفاوت الگوریتم‌های ماشین یادگیری وابسته به منطقه مطالعاتی، نوع و تعداد متغیرهای به‌کاررفته برای پیش‌بینی است. مدل‌های به‌کاررفته کلاس‌هایی از خاک را که فراوانی بیشتری داشتند، با صحت بالاتری پیش‌بینی نمودند. از بین متغیرهای کمکی استفاده شده، پارامترهای ژئومورفومتری از اهمیت بالاتری نسبت به شاخص‌های سنجش از دور برخوردار بودند و شاخص خیزی توپوگرافی به عنوان مهم‌ترین متغیر کمکی در پیش‌بینی مکانی کلاس‌های خاک معرفی شد.

## "هیچ گونه تعارض منافع بین نویسندگان وجود ندارد"

## منابع

- بنائی؛ محمد حسین (۱۳۸۰). نقشه منابع و استعداد خاک‌های ایران. مؤسسه تحقیقات خاک و آب. ایران. تهران.
- تقی‌زاده مهرجردی؛ روح‌الله، سرمدیان؛ فریدون، امید؛ محمود، تومانیان؛ نورایر، روستا؛ محمد جواد، رحیمیان؛ محمد حسن (۱۳۹۳). نقشه‌برداری رقومی کلاس‌های خاک با استفاده از انواع روش‌های داده‌کاوی در منطقه اردکان استان یزد. *مهندسی زراعی (مجله علمی کشاورزی)*، ۳۷ (۲): ۱۱۵-۱۰۱.
- جعفری؛ اعظم، خادمی؛ حسین؛ ایوبی؛ شمس‌الله (۱۳۹۱). نقشه‌برداری رقومی افق‌های مشخصه و گروه‌های بزرگ خاک در منطقه زرنند کرمان. *علوم و فنون کشاورزی و منابع طبیعی، علوم آب و خاک*، ۱۶ (۶۲): ۱۹۳-۱۷۷.
- صالحی؛ محمد حسن، خادمی؛ حسین (۱۳۸۷). *مبانی نقشه‌برداری خاک*. جهاد دانشگاهی دانشگاه صنعتی اصفهان. ۲۱۲ صفحه.
- عباس‌زاده افشار؛ فریده، ایوبی؛ شمس‌الله، جعفری؛ اعظم (۱۳۹۶). نقشه‌برداری رقومی کلاس‌های خاک با استفاده از نقشه قدیمی خاک در منطقه خشک جنوب شرق ایران. *علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)*، ۲۱ (۱): ۲۳۹-۲۵۳.
- علی‌احیائی؛ مریم، بهبهانی‌زاده؛ علی‌اصغر (۱۳۷۱). *شرح روش‌های شیمیایی خاک*. مؤسسه تحقیقات خاک و آب (جلد اول). سازمان تحقیقات و آموزش کشاورزی. ایران. تهران.
- علی‌احیائی؛ مریم، بهبهانی‌زاده؛ علی‌اصغر (۱۳۷۱). *شرح روش‌های شیمیایی خاک*. مؤسسه تحقیقات خاک و آب (جلد دوم). سازمان تحقیقات و آموزش کشاورزی. ایران. تهران.
- رحیمی مشکله؛ مستانه، دلاور؛ محمد امیر؛ جمشیدی؛ محمد، شریفی‌فر؛ امین (۱۴۰۲). مدل‌سازی پراکنش مکانی کلاس‌های خاک با استفاده از الگوریتم‌های یادگیری ماشین در بخشی از اراضی استان زنجان. *پژوهش‌های خاک*، ۳۷ (۲): ۱۶۵-۱۴۷.
- فرقانی؛ سید جواد، پهلوان‌راد؛ محمدرضا، اسفندیاری؛ مهرداد، محمدی تراکشوند؛ علی (۱۳۹۹). نقشه‌برداری رقومی کلاس‌های خاک اراضی سیلابی و با پستی و بلندی کم (مطالعه موردی: شهرستان هیرمند). *نشریه حفاظت منابع آب و خاک*، ۹ (۱۴): ۱۲۰-۱۰۷.
- مقصودی؛ زیبا، رستمی‌نیا؛ محمود، فرامرزی؛ مرزبان، کشاورزی؛ علی، رحمانی؛ اصغر، موسوی؛ سید روح‌الله (۱۳۹۹). نقشه‌برداری رقومی کلاس فامیل خاک با استفاده از رویکرد یادگیری ماشین (مطالعه موردی: اراضی نیمه خشک غرب ایران).
- موسوی؛ سید روح‌الله، سرمدیان؛ فریدون، رحمانی؛ اصغر (۱۳۹۸). مدل‌سازی و پیش‌بینی مکانی کلاس خاک با استفاده از الگوریتم یادگیری رگرسیون درختی توسعه یافته و جنگل‌های تصادفی در بخشی از اراضی دشت قزوین. *تحقیقات آب و خاک ایران*، ۵۰ (۱۰): ۲۵۲۵-۲۵۳۸.



## REFERENCES

- Abbaszadeh Afshar, F., Ayoubi Sh., & Jafari, A. (2017). Digital Soil Mapping of Soil Classes using Conventional Soil Maps in the Arid Region Southeastern Iran. *J. Water and Soil Sci*, 21(1): 239-253. (In Persian)
- Adu-Poku, S. (2012). *Comparing classification algorithms in data mining*. (Thesis). Central Connecticut State University. pp: 24-26.
- Ali Ehiyai, M. (1993). *Description of Soil Chemical Analysis Methods (Volume 1)*. Technical Journal 893. Soil and Water Research Institute. Agricultural Research and Education Organization, Tehran, Iran. (In Persian)
- Ali Ehiyai, M., & Behbahanizadeh A. A. (1993). *Description of Soil Chemical Analysis Methods (Volume 2)*. Technical Journal 1024. Soil and Water Research Institute. Agricultural Research and Education Organization, Tehran, Iran. (In Persian)
- Andronikov, V. L. & Dorbrolov'skiy, G. V. (1991). Theory and methods for the use of remote sensing in the study of soils. *Mapping Sciences and Remote Sensing*, 28(2), 92-101.
- Angelini, M. E., Heuvelink, G. B. M., Kempen, B., Morrás, H. J. M. (2016). Mapping the soils of an Argentine Pampas region using structural equation modelling. *Geoderma*. 281, 102-118.
- Banaei, M. H. (2001). Map of Resources and Talents of Iranian Soils. Iran Soil and Water Research Institute, Tehran. (In Persian)
- Barthold F. K., Wiesmeier, M., Breuer, L., Frede, H. G., Wu, J., & Blank, F. B. (2013). Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *J. Arid Environ.* (88), 194-205.
- Bouma, J., Booltink, H., & Finke, P. (1996). Use of soil survey data for modeling solute transport in the vadose zone. *Journal of environmental quality*, 25(3), 519-526.
- Breiman L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brungard C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards, Jr, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, (239), 68-83.
- Buol, S. W., Southard, R. J., Graham, R. C., McDaniel, P. A. (2003). *Soil Genesis and Classification*. Iowa State University Press, 494 p.
- Burrough P. A., Beckett, P. H. T., & Jarvis, M. G. (1971). The relation between cost and utility in soil survey. *European Journal of Soil Science*, 22(2), 368-381.
- Camera, C., Zomeni, Z., Noller, J. S., Zissimos, A. M., Christoforou, I. C., Bruggeman, A. (2017). A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization, *Geoderma*, Volume 285, Pages 35-49, <https://doi.org/10.1016/j.geoderma.2016.09.019>.
- Campling, P., Gobin, A. & Feyen J. (2002). Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Science Society of America Journal*, (66), 1390-1401.
- Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A.C., Walter, C. (2022). Digital mapping of Global Soil Map soil properties at a broad scale: A review. *Geoderma*, 409. 10.1016/j.geoderma.2021.115567.
- Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35-46.
- Fernandes Coelho, F., Giasson, E., Campos, A. R., Tiecher, T., Ferreira Costa, J. J., Coblinski, J. A. (2021). Digital soil class mapping in Brazil: a systematic review. *Soil and Plant Nutrition*, 78 (5).
- Forghani, S. J., Pahlevanrad, M. R., Esfandiari, M., Mohammadi Torkashvand, A. (2022). Digital soil mapping of soil classes in floodplain and low relief lands (Case study: Hiranmand county). *Water and Soil Resources Conservation*, 9(4). (In Persian)
- Grabs, T., Seibert, J., Bishop, K., and Laudon, H. (2009). Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. *J. Hydrol.* 373: 15-23.
- Grunwald, S. (2006). Environmental soil-landscape Modeling, Geomorphic Information Technologies and Pedometrics. *Taylor and Francis Pub.*, USA.
- Han, X., Liu, J., Shen, X., Liu, H., Li, X., Zhang, J., Wu, P., Liu, Y. (2022). High relief yield strong topography-soil water-vegetation relationships in headwater catchments of southeastern China. *Geoderma* (428).



- Hengl, T., Rossiter, D. G., & Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Geoderma*, (120), 75–93.
- Hengl, T., Toomanian, N., Reuter, H. I., & Malakouti, M. J. (2007). Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, 140(4), 417-427.
- Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, (265), 62-77.
- Jafari, A., Khademi, H., & Ayoubi, Sh. (2013). Digital mapping of characteristic horizons and large soil groups in Zarand area of Kerman. *Journal of Soil and Water Sciences*, 16 (62): 177-193. (In Persian)
- Johansson, U., König, R., Niklasson, L. (2010). *Genetic rule extraction optimizing brier score*. In: Pelikan, M., Branke, J. (Eds.), GECCO. ACM, pp. 1007–1014.
- Kempen, B., Brus, D. J., Heuvelink, G. B., and Stoorvogel, J. J. (2009). Updating the 1: 50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3-4): 311-326.
- Khaledian, Y., Miller, B. A. (2020). Selecting appropriate machine learning method for digital soil mapping. *Applied Mathematical Modelling*, (81), 401-418.
- Kim, J., Grunwald, S., Rivero, R. G., & Robbins, R. (2012). Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Science Society of America Journal*, 76 (6), 2327-2341.
- Kuhn, M. (2018). Caret: classification and regression training. Astrophysics Source Code Library.
- Lu, D., Mausel, P., Brondizio, E. & Moran, E. (2004). Change detection techniques. *International journal of remote sensing*, 25(12), 2365-2401.
- Luoto, M., & Hjort J. (2005). Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology*, (67), 299-315.
- MacMillan, R. A. (2003). LandMapR Software Toolkit C++ Version. LandMapper Environmental Solution. User Manual.
- MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., Goddard, T. W. (2000). A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems*, 113: 81-109.
- Maghsodi, M., Rostaminia, M., Faramarzi, M., Keshavarzi, A., Rahmani, A., Mousavi, S. R. (2020). Digital mapping of soil family class using the machine learning approach (a case study: semi-arid lands in the west of Iran). *Journal of water and soil science*, Vol. 24, No. 2. (In Persian)
- Mahler. P. J. (1970). Manual of multipurpose land classification. Rep. No. 212, Soil and Water Research Institute, Iran.
- Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math* (11): 431–441.
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1), 3-52.
- Minasny, B., McBratney, A. B. (2016). Digital soil mapping: a brief history and some lessons. *Geoderma* 264, 301–311. <https://doi.org/10.1016/j.geoderma.2016.06.011>.
- Monterio, M. E. C., Avalos, F. P., Pelegrino, M. P., Vilela, R. B., Junior, F. W. A., Bueno, I. T., Li, N., Silva, S. H. G., Giasson, E., Curi, N., Menezes, M. D. (2023). Digital mapping of soil classes in southeast Brazil: environmental covariate selection, accuracy, and uncertainty. *Journal of South America Earth Sciences*.
- Moran, C. J., & Bui, E. N. (2002). Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science*, 16(6), 533-549.
- Mosleh, Z., Salehi, M. H., Jafari, A., Borujeni, I. E. and Mehnatkesh, A. (2016). The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environmental monitoring and assessment*, 188(3), 195.
- Mousavi, S. R., Sarmadian, F., Rahmani, A. (2019). Modelling and prediction of soil classes using boosting regression tree and random forests machine learning algorithms in some part of Qazvin plain. *Iranian Journal of Soil and Water Research*, 50 (10), 2525-2538. (In Persian)
- Olaya, V. I. (2004). *A gentle introduction to SAGA GIS*. The SAGA User Group eV, Gottingen, Germany, 208.
- Padarian, J., Minasny, B., McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil* 5, 79–89. <https://doi.org/10.5194/soil-5-79-2019>.
- Quinlan, J. R. (2001). *Cubist: An Informal Tutorial*. <http://www.rulequest.com>.
- Rahimi Mashkaleh, M., Delavar, M. A., Jamshidi, M., Sharififar, A. (2023). Modelling spatial distribution of soil classes using machine learning algorithms in some parts of Zanjan province. *Iranian Journal of Soil and water Research*, 37 (2). (In Persian)

- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, (71), 804-818.
- Rossiter, D. (2005). Digital soil mapping: Towards a multiple-use Soil Information System. *Análisis Geográficos (Revista del Instituto Geográfico "Augustín Codazzi")*, 32(1), 7-15.
- Rossiter, D. G. (2004). Digital soil resource inventories: status and prospects. *Soil Use and Management*, 20(3), 296-301.
- Rostaminia, M., Rahmani, A., Mousavi, R., Taghizadeh, R. (2021). Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms. *Environmental Monitoring and Assessment*, 193(185):1-17.
- Salehi, M. H., & Khademi, H. (2008). *Basics of soil mapping*. Published by Isfahan University of Technology, pp:212. (In Persian)
- Saurette, D. D., Berg, A. A., Laamrani, A., Heck, R., Gillespie, A., Voroney, P. (2022). Effects of sample size and covariate resolution on field-scale predictive digital mapping of soil carbon. *Geoderma*: 425.
- Schoeneberger, P. J., Wysocki, D. A., Benham, E. C. (2012). Soil Survey Staff. 2012. *Field book for describing and sampling soils*, Version 3.0. Lincoln: Natural Resources Conservation Service, National Soil Survey Center.
- Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography*, 27(2): 171-197.
- Scull, P., Franklin, J. & Chadwick, O. A. (2005). The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modeling*. 181: 1–15.
- Sharififar, A., Sarmadian, F., Malone, B. P., Minasny, B. (2019). Addressing the issue of digital mapping of soil classes with imbalanced class observations. *Geoderma*, 350, 84–92.
- Siqueira, D. S., Marques Jr, J., Pereira, G. T., Teixeira, D. B., Vasconcelos, V., Carvalho Junior, O. A., Martins, E. S. (2015). Detailed mapping unit design based on soil-landscape relation and spatial variability of magnetic susceptibility and soil color. *Catena*, 135, 149-162.
- Soil Survey Staff, (2022). Keys to Soil Taxonomy, 13th ed. U. S. Department of Agriculture, Natural Resources Conservation Service.
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Omid, M., Toomanian, N., Roustia, M. J., & Rahimian, M. H. (2015). Digital mapping of soil classes using different data mining techniques in Ardakan region, Yazd province. *Journal of Agricultural Engineering*. 37(2): 101-115. (In Persian)
- Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., Zeraatpisheh, M., Amirian-Chakan, A., Triantafilis, J. (2019). Digital Mapping of Soil Classes Using Ensemble of Models in Isfahan Region, Iran. *Soil System*, (3, 37). <https://doi.org/10.3390/soilsystems3020037>
- Tiwari, S. K., Saha, S. K., & Kumar, S. (2015). Prediction modeling and mapping of soil carbon content using artificial neural network, hyperspectral satellite data and field spectroscopy. *Advances in Remote Sensing*, 4(01), 63.
- Western, A. W. (2004). Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.* 286: 1-4. 113-134.
- Witten, I. H., Frank, E., Hall, M. A. (2011). *Data mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Burlington.
- Zhu, X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil Mapping Using GIS Expert Knowledge and Fuzzy Logic. *Journal of Soil Science Society of America*, (65), 1463-1472.



## Hine learning approaches in downstream lands of Azad dam (case study: Kurdistan province)

### EXTENDED ABSTRACT

#### Objective

All agro-economic and environmental activities usually need detailed information about the distribution of soil classes and their characteristics. Digital soil mapping relies on relating soil classes or properties to a particular set of covariates, which capture inherent soil spatial variation. DSM generates accurate soil spatial information through quantitative relationships between environmental covariates (predictors) and soil classes or properties. In DSM, different machine learning algorithms model The relationships between soil and auxiliary variables (e.g., remotely sensed data) and predict soil classes or properties at unknown points.

#### Research method

In this study, artificial neural network (ANN), decision tree (C.5), logistic regression (MLR), and random forest (RF) were used to predict soil class maps. The relationships between soil and the environment could be well explained by a practical collection of covariates and a good algorithm. The studied region with approximately 178.3 ha is located downstream lands of the Azad dam in northwest of Sanandaj in Kurdistan province. Based on the meteorological data, the soil temperature and moisture regimes were estimated as mesic and xeric, respectively. The main landforms of the studied area are Foothlope, Convergent Backslope, and Divergent Shoulder and the main land use is farm. A random pattern was used to determine the position and distribution of the 84 profiles in the studied area. At each observation point, a soil profile was dug and accurately described and sampled. Physico-chemical analyses like electrical conductivity, soil pH, soil texture, TNV, and organic carbon were carried out based on the standard methods. Based on morphological characteristics and laboratory data, each profile was classified up to the family level based on soil taxonomy.

#### Finding and conclusions

Two orders (Inceptisols and Entisols) two Suborder, three Great groups, five Subgroups, and a Family were identified in the studied region. Topographic attributes were derived from a digital elevation model with a 10 m cell size. All types of DEM covariates (such as curvatures, slope, aspect, elevation, convergent index, LS factor, topographic wetness index, catchment area, close depressions, valley depth, and ...) were extracted using SAGA GIS from DEM. RS covariates were extracted from atmospherically corrected sentinel-2B images. Algorithms were trained with 70% of randomly selected data and validated with the rest of 30% of the dataset. The Kappa coefficient of agreement (K), overall accuracy (OA), and Brier score were used to assess the performance of each model. The best result was obtained for the ANN model (OA=0.65, k=0.53, and BS=0.16, respectively). The weakest predictions were for the C.5 model with 0.38 Overall accuracy, 0.22 kappa index, and 0.87 Brier score. These results are due to the fact that the performance of machine learning models depends on the study area, the number of observation points, and the type and number of predictor variables.

**Keywords:** Artificial Neural Networks, Decision Tree, Multinomial Logistic Regression, Random Forest.