

انبوهش‌زدایی واحدهای نقشه خاک با استفاده از مدل دسمارت: ترکیب مدل‌های مبتنی بر سیستم درختی و داده‌های جدید خاک‌رخی

چکیده

نقشه‌های مرسوم خاک بصورت واحدهای چندضلعی می‌باشند که در آن‌ها، واحدهای خاک با مرزهای مشخص از یکدیگر تفکیک شده‌اند اما تغییرات کلاس‌های خاک در واحدها مشخص نمی‌باشد. با توجه به نیاز به اطلاع از تغییرات کلاس‌های خاک در واحدهای نقشه، هدف این مطالعه انبوهش‌زدایی واحدهای نقشه خاک با استفاده از روش دسمارت (DSMART) در منطقه آبیگ می‌باشد. مدل دسمارت براساس مدل‌های درخت C5.0، جنگل تصادفی و تقویت گرادیان افراطی در دو سناریو انجام شد: (۱) استفاده از اطلاعات واحدهای نقشه خاک موروثی یک میلیونیم کشور و (۲) با اطلاعات ۲۳۰ خاک‌رخ جدید در سطح زیرگروه‌های خاک. عملکرد مدل‌ها و عدم قطعیت آن‌ها با شاخص‌های کمی ارزیابی شدند. در سناریوی اول، میزان صحت کلی نقشه‌ها بین ۰/۲۹ تا ۰/۳۷ و مقدار کاپا بین ۰/۱۷ تا ۰/۲۹ متغیر بود که بهترین نتایج از مدل تقویت گرادیان افراطی با شاخص درهمی ۰/۷۴ بدست آمد. در سناریوی دوم، در مدل جنگل تصادفی صحت کلی نقشه‌ها از ۰/۵۱ تا ۰/۶۳ و کاپا از ۰/۴۴ تا ۰/۶۰ افزایش یافت و شاخص درهمی به ۰/۶۵ کاهش یافت. مقایسه این نقشه‌ها با توزیع مکانی زیرگروه‌های خاک منطقه بیانگر تطبیق خوب نقشه‌ها با هم بود، به‌طوری‌که در سناریوی دوم به میزان ۴۳ درصد افزایش نشان داد. در سناریوی اول، متغیرهای توپوگرافی، و در سناریوی دوم، میانگین بارندگی و شاخص پوشش گیاهی عمودی بیشترین اهمیت را در مدل‌سازی داشتند. ترکیب داده‌های جدید خاک‌رخی صحت مدل‌سازی را تا ۲۶ درصد افزایش داد. این نتایج کارایی روش دسمارت با اطلاعات خاک‌رخی اضافی در انبوهش‌زدایی واحدهای نقشه خاک موروثی را تأیید می‌کنند.

کلمات کلیدی: دسمارت، ریزمقیاس‌سازی، نقشه‌برداری رقومی خاک، یادگیری ماشین

Disaggregation of Soil Map Units Using the DSMART Model: Integrating Tree-Based Models and New Soil Profile Data

Abstract

Conventional soil maps consist of polygon units where soil types are delineated with clear boundaries, yet intra-unit variability of soil classes remains undefined. To address this, the present study aims to disaggregate the inherited one-millionth soil map units of Iran using the DSMART model in the Abyek region. The DSMART was applied using three tree-based algorithms: C5.0, Random Forest (RF), and Extreme Gradient Boosting (XGBoost), across two scenarios: (1) utilizing legacy soil map information, and (2) integrating 230 new soil profiles at the subgroup level. Model performance and uncertainty were evaluated using overall accuracy, Kappa coefficient, and confusion index. In Scenario 1, map accuracy ranged from 0.29 to 0.37, with Kappa values between 0.17 and 0.29. The highest performance was achieved by XGBoost, showing a confusion index of 0.74. In Scenario 2, accuracy improved to 0.51–0.63 and Kappa to 0.44–0.60, with the best results from the RF model, although confusion index slightly dropped to 0.65. Spatial consistency with observed soil subgroup distribution improved significantly—by 43% in Scenario 2. Topography proved most influential in Scenario 1, while mean annual rainfall and vertical vegetation index dominated in Scenario 2. Incorporating new soil profile data enhanced model performance by up to 26%. These findings underscore the effectiveness of DSMART, particularly when enriched with new soil data, in refining legacy soil map units for more precise soil class delineation.

Keywords: DSMART, Downscaling, Digital soil mapping, Machine learning

Extended Abstract

Introduction

Legacy soil data are vital for preserving national resources and serve as a foundational reference for digital soil mapping and environmental modeling. However, many conventional soil maps—such as those produced at a scale of 1:1,000,000—lack the spatial detail required for modern land management applications. These maps often rely heavily on expert judgment and broad physiographic units, resulting in polygonal representations that blend multiple soil classes without describing their internal variability. Consequently, such aggregated maps limit our understanding of soil transitions and spatial patterns critical to agricultural planning, ecological assessments, and policy decisions.

Spatial disaggregation has emerged as a solution for enhancing map resolution and extracting more detailed, pixel-level soil information from generalized units. Nevertheless, relying solely on inherited map units without refining them with updated data can propagate uncertainties and reduce the predictive power of resulting maps. Therefore, integrating new, geo-referenced soil profile observations is essential to improve model calibration, capture local variability, and reduce epistemic uncertainty. This study aims to disaggregate and update Iran's national 1:1,000,000 legacy soil map in the Abyek region of the Qazvin Plain using DSMART (Disaggregation of Soil Map Units Through Resampled Classification Trees) in conjunction with three tree-based algorithms in two different scenarios. The primary goal is to evaluate the combined effect of algorithm choice and input data richness on the spatial delineation of soil classes.

Materials and Methods

The study was conducted in the Abyek region of the Qazvin Plain, spanning 58,000 hectares with varied topography and semi-arid to arid climate conditions. To support disaggregation, eight key environmental covariates representing soil-forming factors—topography, vegetation, salinity, and climate—were selected using a robust feature selection method. Disaggregation of legacy soil map units was performed using the DSMART model under two scenarios: (1) relying solely on existing map information, and (2) incorporating 230 new soil profiles classified at the USDA subgroup level. Three tree-based machine learning algorithms (C5.0, Random Forest, and XGBoost) were evaluated. Model outputs included probabilistic soil class maps. Model validation was performed using confusion matrices, overall accuracy, Kappa coefficients, the confusion (entropy) index, and spatial concordance via Cramér's V.

Results and Discussion

Scenario 1 yielded modest accuracy, with overall prediction rates ranging from 0.29 to 0.37. Among the three models, XGBoost achieved the highest Shannon diversity and composite performance index, demonstrating strength in capturing soil variability. However, high confusion indices in central and southern areas pointed to unresolved overlaps between closely related soil classes, likely due to limited input diversity and high-class imbalance.

Scenario 2, enhanced with 230 new profiles, resulted in substantial performance gains. Overall accuracy rose to 0.67 and Kappa reached 0.60, with agreement between predicted and observed soil subgroups improving by 43%. Random Forest achieved the highest composite accuracy-diversity score (1.34), while XGBoost maintained the lowest confusion index (0.60), reflecting more confident classifications. The inclusion of new field data improved the model's ability to predict rare subgroups and redistributed class probabilities more realistically across polygons. Notably, this also shifted the influence of environmental predictors: while topographic indices (e.g., elevation, MvBF, valley depth) dominated Scenario 1, climatic (precipitation) and vegetation indices became more important in Scenario 2, suggesting improved model understanding of soil-environment relationships.

These results confirm that DSMART's disaggregation capability is highly sensitive to the quality and quantity of input data. More accurate soil class delineation—particularly for underrepresented or complex classes—requires sufficient profile observations and landscape-driven covariates. The findings align with recent studies emphasizing the role of disaggregated models and enriched data inputs for refining legacy maps in diverse agroecological settings.

Conclusion

This study demonstrated the applicability of the DSMART model for disaggregating Iran's one-millionth legacy soil map units in the Abyek region. Among the tree-based models tested, Random Forest and Extreme Gradient Boosting outperformed C5.0 in modeling soil subgroups. Incorporating 230 new soil profiles substantially improved model performance, underscoring the importance of high-quality field data. Topographic variables were most influential in the baseline model, while climatic and vegetation indices became more prominent with data enrichment. The results affirm DSMART's adaptability for refining outdated soil maps and support future use of additional soil observations and soil-landscape relationships to enhance spatial prediction accuracy in complex terrains.

Keywords: DSMART, Downscaling, Digital soil mapping, Machine learning

Author Contributions

Conceptualization, Z.R., F.S. and A.J.; methodology, Z.R.; software, Z.R.; validation, Z.R.; formal analysis, Z.R.; investigation, Z.R., and A.J.; resources, Z.R.; data curation, F.S.; writing—original draft preparation, Z.R.; writing—review and editing, Z.R., F.S. and A.J.; visualization, Z.R., and A.J.; supervision, F.S.; project administration, Z.R. and F.S.; funding acquisition, Z.R. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Not applicable.

Acknowledgements

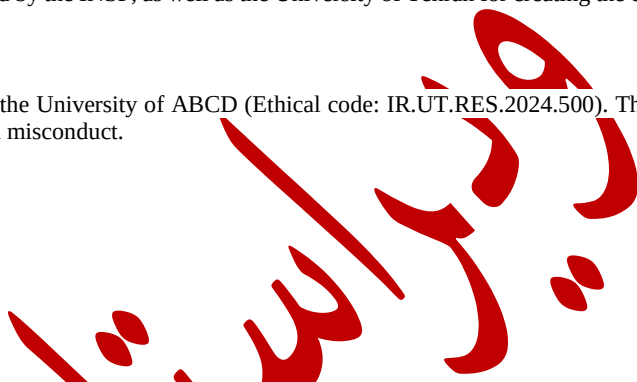
This work is based on research funded by the Iran National Science Foundation (INSF) under project No. 4024343. The first author gratefully acknowledges the financial support provided by the INSF, as well as the University of Tehran for creating the conditions necessary to carry out and complete this study.

Ethical considerations

The study was approved by the Ethics Committee of the University of ABCD (Ethical code: IR.UT.RES.2024.500). The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.



مقدمه

استفاده از داده‌ها و اطلاعات موروثی خاک نقش مهمی در حفظ سرمایه‌های ملی و همچنین منابع کشور ایفا می‌کند. مطالعات مختلف نشان داده‌اند که این داده‌ها می‌توانند کیفیت نقشه‌های رقومی خاک را با بهره‌گیری از روش‌های مختلف بهبود بخشند (Rasaei et al., 2020؛ Lázaro-López et al., 2021). با این حال، بسیاری از مطالعات مرسوم بدون در نظر گرفتن اطلاعات دقیق موقعیت مکانی خاک‌ها تهیه شده‌اند، به‌طوری که مدل ذهنی کارشناسان متخصص، تعیین کننده واحدهای خاک بوده است. در نتیجه، این واحدها بصورت چندضلعی‌هایی متشکل از چندین کلاس خاک منفرد ارائه شده‌اند. چنین محدودیت‌هایی می‌تواند باعث کاهش قابلیت اعتماد این نقشه‌ها شود، به‌ویژه هنگامی که تفکیک دقیق واحدهای خاک انجام نشود (Lázaro-López et al., 2021).

انبوهش‌زدایی مکانی واحدهای نقشه خاک، یکی از راهکارهای ارائه شده برای حل این مشکل است. این فرآیند به افزایش مقیاس اطلاعات خاک و استخراج جزئیات دقیق‌تر از نقشه‌های سنتی خاک با قدرت تفکیک مکانی بالاتر اشاره دارد. با توجه به اهمیت قابلیت اعتماد اطلاعات استخراج شده از این نقشه‌های موروثی کوچک مقیاس و همچنین نیاز به دریافت داده‌های دقیق‌تر درباره نوع و تغییرات خاک در منطقه، این فرآیند باید با دقت بالا و با در نظر گرفتن ویژگی‌های منطقه و نظر کارشناس انجام پذیرد. در غیر این صورت، ممکن است حجم زیادی از اطلاعات جدید اما با عدم قطعیت بالا تولید شود که نه تنها موجب سردرگمی کارشناسان، بلکه باعث اتخاذ تصمیم‌های نادرست توسط مدیران و مسئولان مربوطه خواهد شد (Thompson et al., 2010).

انبوهش‌زدایی واحدهای نقشه خاک با استفاده از روش‌های کلاس‌بندی نظارت نشده و نظارت شده انجام می‌شود. در روش نظارت نشده تفکیک اجزای واحدهای نقشه قدیمی بر پایه اطلاعات نقشه خاک، قواعد خاک-زمین نما و نظر کارشناس و مدل‌های مختلف ریاضیاتی انجام می‌شود (Zeraatpisheh et al., 2019؛ Flynn et al., 2019؛ Minai et al., 2020). با این حال، این فرآیند عمدتاً بصورت نظارت شده و با بهره‌گیری از مدل‌های یادگیری ماشین انجام می‌شود. در میان مطالعات انجام شده، اخیراً درخت‌های طبقه‌بندی بطور گسترده‌ای برای تفکیک اجزای واحدهای نقشه‌های مرسوم خاک مورد استفاده قرار گرفته‌اند و کارایی آن در مطالعات

Ellili- و Flynn et al. (2019), Machado et al. (2018), Vincent et al. (2018), Malone et al. (2017), Holmes et al. (2015) Bargaoui et al. (2020) تایید شده است.

در ایران نیز این مدل در راستای تفکیک اجزای واحدهای نقشه چندضلعی خاک توسط فاتحی و همکاران (۱۳۹۷)، Zeraatpisheh et al. (2019) و et al. (2019) مورد استفاده قرار گرفته است. فاتحی و همکاران (۱۳۹۷) با استفاده از درخت طبقه‌بندی C5.0 مدل دسمارت ارائه شده توسط Odgers et al. (2014) و با کمک متغیرهای محیطی موفق شدند نقشه خاک سنتی منطقه‌ای در کرمانشاه را به خوبی تفکیک کرده و نقشه رقومی زیرگروه‌های خاک را با قدرت تفکیک مکانی ۱۰ متر تهیه کنند. نتایج این مطالعه بیانگر دقت پایین مدل دسمارت به دلیل تعداد نامناسب داده‌های اعتبارسنجی بود. Zeraatpisheh et al. (2019) با استفاده از مدل‌های نظارت نشده (میانگین‌های کا معمولی و فازی) و نظارت شده (درخت طبقه‌بندی C5.0 دسمارت) نقشه خاک یک میلیونیم کشور را در قسمتی از منطقه بروجن با قدرت تفکیک مکانی ۳۰ متر به اجزای واحدهای خاک در سطح گروه بزرگ و زیرگروه تفکیک کردند. نتایج این مطالعه بیانگر دقت بالاتر روش نظارت شده دسمارت نسبت به روش نظارت نشده بود. با این حال، دقت نقشه بدست آمده از مدل دسمارت نسبت به نقشه مرسوم خاک کمتر بود، که این امر به وجود واحدهای خاک مرکب متشکل از بیش از یک کلاس خاک نسبت داده شد. Jamshidi et al. (2019) نیز از الگوریتم درخت طبقه‌بندی C5.0 مدل دسمارت برای انبوهش‌زدایی نقشه مرسوم خاک منطقه سعادت‌شهر به نقشه رقومی کلاس‌های خاک با قدرت تفکیک مکانی ۱۰ متر استفاده کردند. نتایج این پژوهش نشان داد که دقت نقشه رقومی سطوح احتمال کلاس‌های مرتبه اول خاک در فرآیند انبوهش‌زدایی ۶۸ درصد بود، اما در سطوح احتمال بالاتر، کاهش معناداری (دقت ۱۷ درصد در سطح دوم و ۰/۵ درصد در سطح سوم) مشاهده شد.

با این حال، مطالعات انجام شده برای تفکیک اجزای واحدهای نقشه خاک با روش‌های مختلف و با بکارگیری الگوریتم‌های مختلف یادگیری ماشین قابل اجرا می‌باشد. به عنوان مثال، Flynn et al. (2019) از الگوریتم‌های خطی مختلف با کارایی ۰/۱۷ تا ۰/۴۲ در این زمینه استفاده کردند. نتایج مطالعه Gagkas and Lilly (2024) نشان داد که الگوریتم جنگل تصادفی در تفکیک مکانی واحدهای نقشه همگون و مرکب عملکرد موفق‌تری داشته و دقت پیش‌بینی قابل قبولی ارائه کرده است. آنان این امر را به توانایی بالای این الگوریتم در تشخیص و تمایز خاک‌های مشابه در مناطق با تنوع مکانی زیاد نسبت دادند. Easher et al. (2023) از سه مدل درخت تصمیم، جنگل تصادفی و نزدیک‌ترین همسایه همراه با روش نمونه‌برداری هاپرکیوب برای انبوهش‌زدایی واحدهای چندضلعی نقشه خاک یک زیرحوضه در سطح گروه بزرگ و سری خاک در کانادا استفاده کردند.

با این حال، میزان و نوع اطلاعات در دسترس از نقشه و واحدهای خاک در انتخاب مدل و کیفیت نتایج نهایی بسیار موثر می‌باشد. فرآیند انبوهش‌زدایی نقشه موروثی نیازمند دسترسی به اطلاعات نقشه است (Zeraatpisheh et al., 2019) و محدودیت در داده‌های موروثی یکی از چالش‌های اصلی در اجرای این فرآیند محسوب می‌شود (Malone et al., 2017). از این رو، در راستای استفاده از نقشه‌های موروثی و مرسوم خاک؛ مفهوم دیگری تحت عنوان به روز رسانی نقشه‌ها مطرح شده است. در این زمینه، می‌توان به مطالعه Zeraatpisheh et al. (2019) در ایران اشاره کرد. ایشان از اطلاعات نیم‌رخ جدید خاک برای به روز رسانی نقشه موروثی یک میلیونیم کشور استفاده کردند. Ellili-Bargaoui et al. (2020) از داده‌های جدید خاک و همچنین روابط خاک-زمین نما برای بهبود کارایی الگوریتم دسمارت در انبوهش‌زدایی نقشه موروثی ۲۵۰ هزار در فرانسه استفاده کردند. فاتحی و همکاران (۱۳۹۷) از اطلاعات نقشه نیمه‌تفصیلی خاک در روند انبوهش‌زدایی واحدهای نقشه خاک در کرمانشاه استفاده کردند.

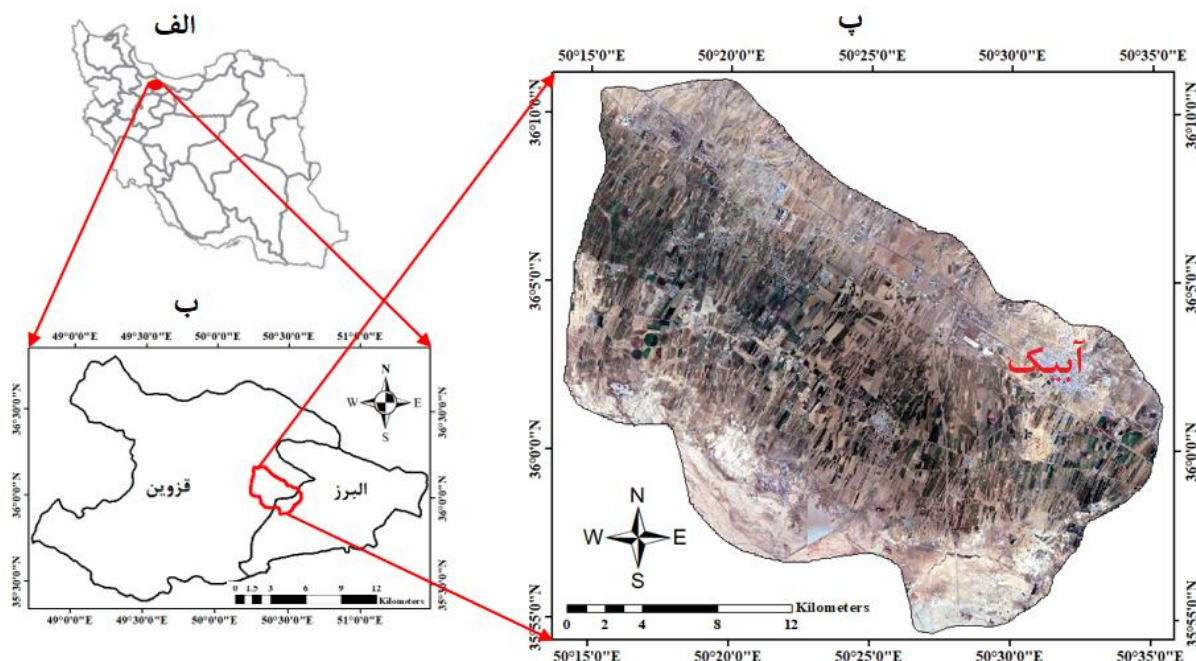
دشت قزوین یکی از مناطق کلیدی در تولید محصولات کشاورزی کشور است. با وجود ظرفیت بالای خاک‌های این منطقه، بهره‌برداری از اراضی به طور بهینه و هماهنگ با ویژگی‌های خاک انجام نمی‌شود. تهیه نقشه‌های دقیق پراکنش خاک می‌تواند نقش مهمی در ارتقای مدیریت پایدار منابع خاک و بهینه‌سازی بهره‌برداری از اراضی ایفا کند. مطالعات متعددی به منظور تهیه نقشه‌های پراکنش مکانی کلاس‌های خاک در این منطقه انجام شده است (موسوی و همکاران، ۱۳۹۸؛ رحمانی و همکاران، ۱۴۰۱؛ ممتازی بروجنی و همکاران، ۱۴۰۱؛ رسائی و همکاران، ۱۴۰۳؛ Neyestani et al., 2021). هرچند که چالش‌هایی مانند تعداد محدود نمونه‌های خاک و هزینه‌های بالای مطالعات صحرایی دقت مدل‌های پیش‌بینی‌کننده را کاهش داده‌اند (رحمانی و همکاران، ۱۴۰۱). در نتیجه، استفاده از داده‌ها و اطلاعات قدیمی خاک به صورت نقشه‌های رقومی خاک می‌تواند به درک تغییرات خاک کمک کند. با این حال، بهره‌برداری از این نقشه‌های قدیمی مستلزم استخراج اطلاعات از واحدهای نقشه می‌باشد. این پژوهش با هدف انبوه‌شدایی واحدهای نقشه موروثی خاک یک میلیونیم کشور و به‌روزرسانی آن در منطقه مطالعاتی آبیک دشت قزوین انجام شده است. منطقه مورد نظر دارای اطلاعات خاک‌شناسی دقیق و ارزشمندی است (Mousavi et al., 2022؛ Neyestani et al., 2021؛ Mousavi et al., 2023؛ موسوی و همکاران، ۱۴۰۰ الف، ب، پ؛ رحمانی و همکاران، ۱۴۰۱؛ خاموشی و همکاران، ۱۴۰۱؛ رضائی و همکاران، ۱۴۰۲؛ خاموشی و همکاران، ۱۳۹۷؛ موسوی و همکاران، ۱۳۹۸؛ ممتازی بروجنی و سرمدیان، ۱۴۰۱) که می‌تواند به تحقق اهداف این مطالعه کمک کند. برای این منظور، از مدل پایه‌ای دسمارت (درخت تصمیم C.50) به همراه دو الگوریتم درخت محور جنگل تصادفی و تقویت گرادیان افراطی، برای اولین بار در کشور استفاده شد. با توجه به اینکه نوع و میزان اطلاعات وارد شده به مدل در کارایی آن اهمیت بسزایی دارد، مطالعه حاضر در دو سناریو (۱) با استفاده از اطلاعات واحدهای نقشه و (۲) با وارد کردن اطلاعات جدید خاک‌رخ‌ها در واحدهای نقشه اجرا گردید. بدین ترتیب، ابتدا عملکرد مدل‌های مختلف در تفکیک واحدهای نقشه خاک با استفاده از اطلاعات راهنمای نقشه بررسی گردید، سپس با تلفیق داده‌های جدید خاک‌رخ‌ها در واحدهای نقشه، امکان ارزیابی تأثیر داده‌های ورودی در بهبود نتایج مدل‌سازی فراهم شد. بنابراین در این مطالعه اثر نوع مدل مورد استفاده در انبوه‌شدایی اطلاعات نقشه خاک یک میلیونیم در منطقه مورد مطالعه و همچنین اثر اطلاعات ورودی مدل بر کارایی مدل و بهبود کیفیت نقشه‌های خروجی مورد بررسی قرار گرفت. این مطالعه نه تنها به بهینه‌سازی روش‌های تولید نقشه‌های خاک دقیق‌تر کمک می‌کند، بلکه می‌تواند راهکارهای عملی برای ارتقای پایگاه‌های داده مکانی و بهبود تصمیم‌گیری‌های مرتبط با مدیریت خاک و اکوسیستم‌های طبیعی ارائه دهد.

مواد و روش‌ها

منطقه مورد مطالعه

به منظور اجرای این مطالعه، قسمتی از اراضی دشت قزوین واقع در مرز استان‌های البرز و قزوین (منطقه آبیک) انتخاب گردید (شکل ۱). منطقه مورد بررسی در بین عرض‌های جغرافیایی ۳۵ درجه و ۵۵ دقیقه تا ۳۶ درجه و ۱۰ دقیقه شمالی و طول‌های جغرافیایی ۵۰ درجه و ۱۵ دقیقه تا ۵۰ درجه و ۳۵ دقیقه شرقی قرار دارد که مساحت حدود ۵۸۰۰۰ هکتار را در بر می‌گیرد. بیشتر قسمت‌های منطقه دشت می‌باشند (ارتفاع: ۱۱۴۴ متر در جنوب منطقه تا ۱۷۵۲ متر بالای سطح دریا در شمال منطقه). بطور کلی از نظر فیزیوگرافی منطقه از شمال به جنوب شامل واحدهای تپه، فلات، دشت‌های دامنه‌ای و آبرفتی و اراضی پست می‌باشد. پارامترهای اقلیمی دما و بارندگی نیز از شمال به جنوب منطقه تغییرات محسوسی را نشان می‌دهند. میانگین دمای سالانه ۱۳ تا ۱۵ درجه سانتی‌گراد و میانگین بارندگی ۳۱۴ تا ۲۵۳ میلی‌متر است که با در نظر گرفتن سیستم طبقه‌بندی اقلیمی کوپن، منطقه مورد بررسی در اقلیم خشک و نیمه‌خشک واقع شده است (موسوی و همکاران، ۱۴۰۰؛ رضائی و همکاران، ۱۴۰۲). رژیم حرارتی خاک در

سرتاسر منطقه ترمیم می‌باشد و رژیم رطوبتی شامل زریک خشک، اریدیک ضعیف و آکوئیک می‌باشد (Soil Survey Staff, 2022). جنس مواد مادری تشکیل دهنده خاک‌ها در شمال و مرکز، رسوبات آبرفتی و کوهرفتی (سنگ‌های آذرین بازالتی آهکی کواترنری) و در جنوب بصورت کفه‌های گلی و نمکی می‌باشد. حاصل این تغییرات فاکتورهای خاک‌سازی، تشکیل خاک‌های متنوع در سطح منطقه است. کشاورزی دیم، کشاورزی آبی، باغ و مراتع شور و غیرشور کاربری‌های مهم اراضی در منطقه می‌باشند (رسانی و همکاران، ۱۴۰۳). منطقه مورد مطالعه یکی از قسمت‌های مهم کشاورزی در دشت قزوین می‌باشد که در طول زمان به دلیل تغییرات اقلیمی و همچنین اجرای مدیریت‌های مختلف دستخوش تغییرات زیادی شده است بطوریکه، در حال حاضر عمده قسمت‌های جنوبی منطقه در معرض خطر جدی شور و قلیائی شدن می‌باشند (موسوی و همکاران، ۱۴۰۰؛ رسانی و همکاران، ۱۴۰۲) و سایر قسمت‌ها نیز با توجه به استعداد خاک‌ها مورد استفاده قرار نمی‌گیرند. به دلیل اهمیت و موقعیت منطقه، در مطالعات مختلف خاک‌شناسی مورد توجه قرار گرفته است که حاصل آن نقشه‌های قدیمی و مرسوم خاک می‌باشد که در نقشه خاک یک میلیونیم که برای سرتاسر کشور تهیه شده است، خلاصه شده‌اند (Roozitalab et al., 2018).



شکل ۱- موقعیت قرارگیری منطقه مورد مطالعه در ایران (الف)، مرز استان‌های قزوین و البرز (ب) و بر روی تصویر توپوگراف (پ)

متغیرهای محیطی کمکی

به منظور انبوه‌شدن دایمی واحدهای نقشه موروثی خاک یک میلیونیم در منطقه مورد مطالعه، بر اساس تابع اسکورپین (McBratney et al., 2003)، ۸ متغیر محیطی مهم توصیف کننده توزیع مکانی کلاس‌های خاک با وضوح مکانی ۱۲/۵ متر با استفاده ترکیبی از نتایج مدل بروتا^۱ (Kursa and Rudnicki, 2010) اجرا شده در نرم‌افزار R و همچنین نظر کارشناس انتخاب شدند (جدول ۱). مدل

1. Boruta

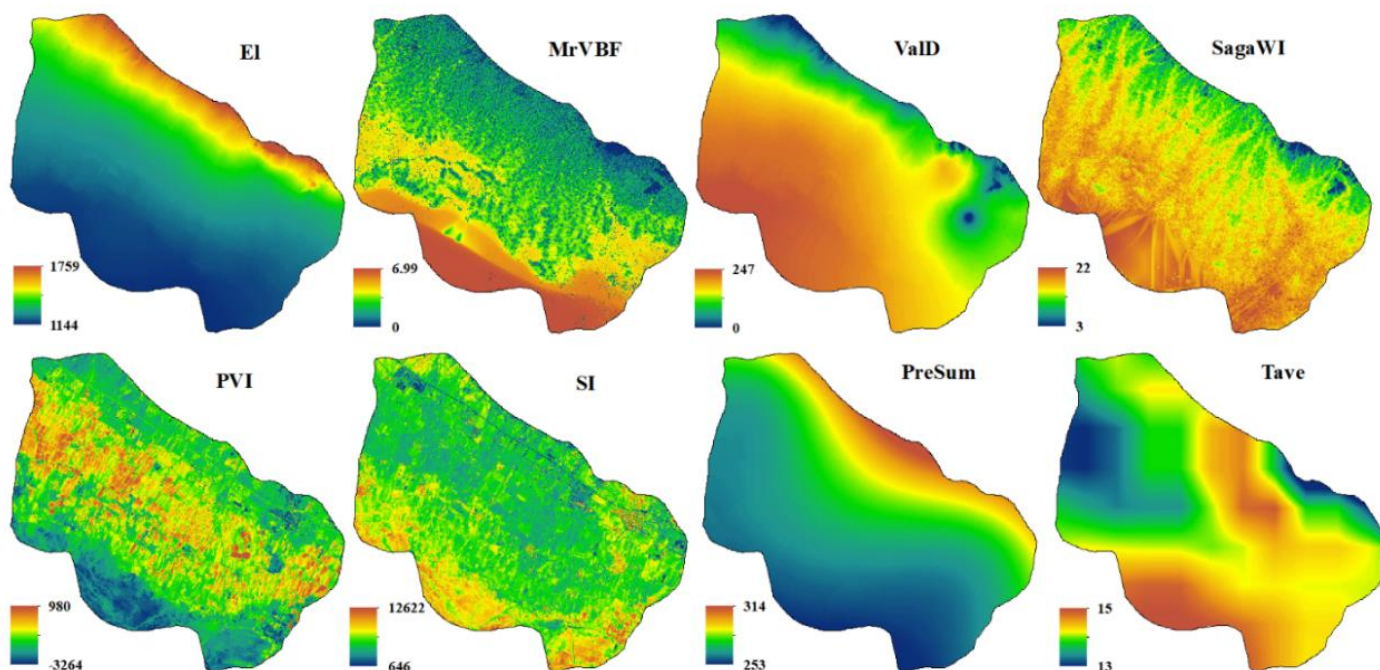
2. Expert opinion

بروتا مبتنی بر مدل جنگل تصادفی بوده که یک روش خودکار برای انتخاب متغیرهای مهم می‌باشد. این مدل به صورت تکرار شونده میزان اهمیت هر متغیر را ارزیابی کرده و مجموعه‌ای بهینه از متغیرها را ارائه می‌دهد و در نهایت تنها متغیرهای تأثیرگذار را حفظ کرده و متغیرهای غیر مؤثر را حذف می‌کند. از مزایای کلیدی این مدل می‌توان به مقاومت در برابر هم‌خطی متغیرها، حفظ متغیرهای کم تأثیر اما مرتبط، بهبود قابلیت اطمینان مدل به همراه افزایش دقت پیش‌بینی و کاهش پیچیدگی آن اشاره کرد.

جدول ۱- متغیرهای محیطی انتخاب شده برای انبوه‌ش‌زدایی واحدهای نقشه خاک موروثی

علامت اختصاری	متغیر محیطی (واحد)	عامل اسکورین	منبع	وضوح مکانی (متر)	مرجع
El	ارتفاع (متر)	پستی و بلندی	مدل رقومی ارتفاع	۱۲/۵	https://asf.alaska.edu/
MrVBF	شاخص همواری دره با قدرت تفکیک بالا (-)	پستی و بلندی	مدل رقومی ارتفاع	۱۲/۵	Gallant and Dowling (2003)
ValD	عمق دره (متر)	پستی و بلندی	مدل رقومی ارتفاع	۱۲/۵	Abdel-Kader (2011)
SagaWI	شاخص خیزی ساگا (-)	پستی و بلندی	مدل رقومی ارتفاع	۱۲/۵	Conrad et al. (2015)
PVI	شاخص پوشش گیاهی عمودی (-)	موجودات زنده خاک	تصاویر ماهواره‌ای	۱۰	Richardson and Wiegand (1977)
SI	شاخص شوری (-)	مواد معدنی	تصاویر ماهواره‌ای	۱۰	Allbed and Kumar (2013)
PreSum	میانگین بارندگی (میلی‌متر)	اقلیم	داده‌های اقلیمی	۲۵۰	Krivoruchko and Gribov (2019)
Tave	میانگین دما (درجه سانتی‌گراد)	اقلیم	داده‌های اقلیمی	۲۵۰	Krivoruchko and Gribov (2019)

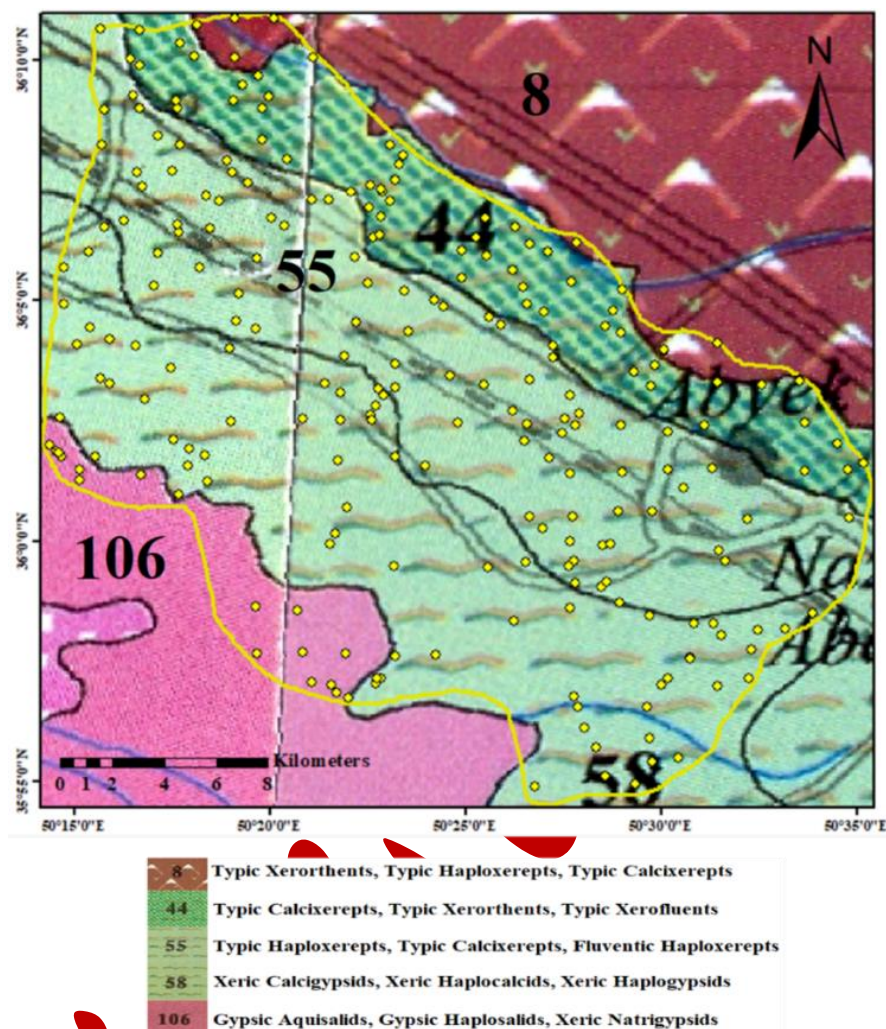
متغیرهای ژئومورفومتریک بدست آمده از مدل رقومی ارتفاع به عنوان نماینده عامل خاک‌سازی پستی و بلندی، شاخص‌های بدست آمده از تصاویر ماهواره‌ای سنتینل ۲ (میانگین‌گیری شده برای یک دوره ۶ ساله از سال ۲۰۱۷ تا سال ۲۰۲۳) به عنوان نماینده عوامل خاک‌سازی ارگانیک یا موجودات زنده و خاک، نقشه‌های میانگین سالانه دما و بارندگی (Mousavi et al., 2022) به عنوان نماینده عامل خاک‌سازی اقلیم مورد استفاده قرار گرفتند. شکل ۲ توزیع مکانی این متغیرهای محیطی در منطقه مورد مطالعه را نشان می‌دهد.



شکل ۲- متغیرهای محیطی با وضوح مکانی ۱۲/۵ متر (اطلاعات مربوط به هر نقشه در جدول ۱ ارائه شده است).

نقشه موروثی خاک

نقشه خاک یک میلیونیم کشور (مقیاس ۱:۱۰۰۰۰۰۰) در اوایل دهه ۸۰ بر اساس اطلاعات مطالعات پیشین خاک‌شناسی و خاک‌رخ‌ها با استفاده از داده‌های کمکی از جمله تصاویر ماهواره‌ای، ژئومورفولوژی، توپوگرافی، پوشش گیاهی و اقلیم در سطح کشور تهیه شده است (Roostitalab et al., 2018). واحدهای نقشه خاک بصورت چندضلعی‌های با بیش از یک نوع خاک می‌باشند که با مرزهای مشخص تفکیک شده‌اند و فرض می‌شود که خاک‌ها با نسبت‌های مشخص در هر واحد وجود دارند. از مجموع ۸۶ واحد خاک شناسایی و تفکیک شده در این نقشه، منطقه مورد مطالعه چهار واحد خاک و یک واحد متفرقه پالایا را در بر می‌گیرد. توزیع این واحدها در محدوده مورد مطالعه به همراه راهنمای نقشه در منطقه مورد مطالعه در شکل ۳ نشان داده شده است.



شکل ۳- واحدهای نقشه خاک مرسوم یک میلیونیم کشور در منطقه مورد مطالعه به همراه راهنمای آن. موقعیت مکانی ۲۳۰ خاکرخ استفاده شده در آموزش مدل‌ها و آزمون کارایی آن‌ها نیز بر روی واحدهای نقشه نشان داده شده است.

همان‌طور که مشاهده می‌شود، هر واحد نقشه از ترکیب سه خاک تشکیل شده است که با توجه به اطلاعات نقشه، خاک‌ها به ترتیب ۵۰، ۳۰ و ۲۰ درصد از هر واحد را تشکیل می‌دهند (Roozitalab et al., 2018; Zeraatpisheh et al., 2019). به عنوان مثال، واحد نقشه شماره ۵۵ را در نظر بگیرید. با توجه به راهنمای نقشه زیرگروه Typic Haploxerepts ۵۰ درصد از خاک‌های این واحد را تشکیل می‌دهد و خاک‌های Typic Calcixerepts و Fluventic Haploxerepts به ترتیب ۳۰ و ۲۰ درصد از این واحد نقشه را شامل می‌شوند. با توجه به قدیمی بودن اطلاعات نقشه، خاک‌های واحدهای نقشه با توجه به آخرین تغییرات سیستم طبقه‌بندی آمریکایی خاک‌ها (Soil Survey Staff, 2022) و همچنین اطلاعات در دسترس، طبق نظر کارشناس اصلاح شدند. به عنوان مثال، واحد شماره هشت به دلیل قرار گرفتن روی واحدهای مرتفع در بالا دست منطقه، عمدتاً بصورت رخنمون سنگی گزارش شده است (یعنی ۵۰ درصد واحد نقشه) که با خاک Typic Xerorthents جایگزین شد. واحد نقشه شماره ۱۰۶ که در لندفرم پلایا قرار دارد بصورت ترکیب سه خاک مجزا با نسبت‌های ۵۰، ۳۰ و ۲۰ درصد در نظر گرفته شد (شکل ۳).

داده‌های جدید خاک

اطلاعات ۲۳۰ خاک‌رخ در دسترس در سطح منطقه به منظور اجرای روند مدل‌سازی‌ها مورد استفاده قرار گرفت. این خاک‌رخ‌ها در ۱۱ زیرگروه خاک سیستم رده‌بندی آمریکایی خاک‌ها (Soil Survey Staff, 2022) قرار دارند که با توزیع زیرگروه‌های خاک در واحدهای نقشه خاک هم‌خوانی دارند (رسائی و همکاران، ۱۴۰۳). تعداد این خاک‌رخ‌ها در هر واحد نقشه در جدول ۲ ارائه شده است.

جدول ۲- ویژگی‌های واحدهای خاک نقشه موروثی به همراه تعداد زیرگروه‌های تاکسونومیکی ۲۳۰ خاک‌رخ در دسترس در منطقه مورد مطالعه

تعداد	زیرگروه‌های خاک	ویژگی مشخصه واحد نقشه	واحد نقشه
۷	Typic Calcixerepts	اراضی شیب‌دار، تپه	۸
۱	Typic Xerorthents		
۲	Typic Haploxerepts		
۳۵	Typic Calcixerepts		
۱۱	Typic Xerorthents	اراضی با شیب کم و دارای پستی و بلندی، فلات	۴۴
۷	Typic Haploxerepts		
۴	Typic Xerofluvents		
۳	Fluventic Haploxerepts		
۳۱	Typic Haploxerepts	اراضی مسطح پنجه شیب، دشت دامنه‌ای	۵۵
۲۳	Typic Calcixerepts		
۱۸	Fluventic Haploxerepts		
۵	Typic Xerorthents		
۲	Typic Xerofluvents	اراضی مسطح، دشت آبرفتی	۵۸
۱	Xeric Haplocalcids		
۱۷	Xeric Calcigypsis		
۱۵	Xeric Haplocalcids		
۱۳	Xeric Haplogypsis	اراضی پست	۱۰۶
۸	Typic Calcixerepts		
۵	Typic Haploxerepts		
۳	Gypsic Aquisalids		
۶	Xeric Natrigypsis	اراضی پست	۱۰۶
۶	Gypsic Aquisalids		
۶	Gypsic Haplosalids		
۱	Xeric Haplogypsis		

مدل‌سازی و انبوه‌شدایی واحدهای نقشه خاک

مدل‌سازی برای انبوهش‌زدایی واحدهای نقشه خاک مرسوم یک میلیونیم منطقه مورد مطالعه به کمک اطلاعات محیطی در دسترس از متغیرهای محیطی نماینده فاکتورهای خاک‌سازی (شکل ۲)، بر اساس مدل دسمارت (Odgers et al., 2014) انجام شد.

مدل دسمارت یکی از معروف‌ترین و پرکاربردترین روش‌ها برای انبوهش‌زدایی نقشه مرسوم چندضلعی خاک است. در این روش، مدل با در نظر گرفتن سهم نسبی هر خاک در واحد نقشه و همچنین ارتباط ریاضی بین متغیرهای محیطی و کلاس‌های خاک، تعداد نمونه تصادفی به روش نمونه‌برداری مجدد از داخل هر واحد نقشه خاک برمی‌دارد و یک نوع خاک به هر نمونه اختصاص داده و بر اساس نسبت خاک‌ها وزن‌دهی می‌کند. در نهایت، با در نظر گرفتن تعداد دفعاتی که یک پیکسل به عنوان کلاس خاصی پیش‌بینی شده است احتمال وقوع هر کلاس خاک را تعیین می‌کند و نقشه نهایی انبوهش‌زدایی شده توزیع کلاس‌های خاک را بصورت رستری در سه سطح احتمال (مطابق با فراوانی سه خاک اصلی واحدهای نقشه خاک مرسوم) ارائه می‌دهد. بدین معنی که، خروجی مدل دسمارت به صورت نقشه‌های احتمالی چندسطحی می‌باشد. این نقشه‌ها شامل سطح احتمال اول (کلاس خاکی با بیشترین احتمال در هر پیکسل)، سطح احتمال دوم (کلاس خاکی با دومین احتمال بالا)، و سطح احتمال سوم (کلاس خاکی با سومین احتمال) هستند. این رویکرد، با در نظر گرفتن درصد توزیع کلاس‌های خاک در واحدهای نقشه (بطور مثال ۵۰٪، ۳۰٪ و ۲۰٪)، امکان تحلیل عدم قطعیت در پیش‌بینی و بررسی جزئیات مکانی خاک را فراهم می‌کند.

بدین ترتیب، مدل تنها با اکتفا به اطلاعات واحدهای نقشه مورثی اقدام به انبوهش‌زدایی و تفکیک خاک‌های هر واحد می‌کند. با این حال، این مدل قابلیت تعلیم با وارد کردن اطلاعات خاک‌های اضافی را نیز دارد. همچنین، روش دسمارت با کمک مدل‌های مختلف یادگیری ماشین قابل اجرا است. بنابراین، مطالعه حاضر با استفاده از سه مدل یادگیری ماشین درخت محور از جمله درخت طبقه‌بندی C5.0، جنگل تصادفی و تقویت گرادیان افراطی، در دو سناریو جداگانه انجام شد. در سناریوی اول، انبوهش‌زدایی واحدهای خاک تنها با استفاده از اطلاعات واحدهای نقشه مورثی و اطلاعات متغیرهای محیطی کمکی انجام شد. در سناریوی دوم، علاوه بر اطلاعات متغیرهای محیطی، از اطلاعات خاک‌های در دسترس در هر واحد نقشه نیز برای تعلیم مدل‌ها بهره گرفته شد.

مدل درخت طبقه‌بندی C5.0 (Quinlan, 1993) ساده‌ترین روش درخت تصمیم می‌باشد که به عنوان تابع پیش فرض در مدل دسمارت مورد استفاده قرار می‌گیرد. مدل جنگل تصادفی (Breiman, 2001) با استفاده از ترکیبی از چندین درخت تصمیم‌گیری، عملکرد بهتری نسبت به مدل‌های تکی دارد و به این ترتیب، دقت و پایداری نتایج خود را افزایش می‌دهد. در این مدل، هر درخت تصمیم‌گیری با استفاده از یک نمونه تصادفی از داده‌ها و مجموعه‌ای تصادفی از ویژگی‌ها ساخته می‌شود. سپس نتایج تمامی درخت‌ها با هم ترکیب شده و نتیجه نهایی بر اساس میانگین یا اکثریت آرا تعیین می‌گردد. این روش به کاهش احتمال بیش‌برازش مدل کمک می‌کند و باعث می‌شود مدل عملکرد بهتری داشته باشد. مدل تقویت گرادیان افراطی^۱ (Chen and Guestrin, 2016) نیز یکی از الگوریتم‌های پیشرفته درخت محور می‌باشد که بر پایه تکنیک تقویت گرادیان^۲ کار می‌کند و با ایجاد مجموعه‌ای از درخت‌های تصمیم‌گیری ضعیف و ترکیب نتایج آن‌ها، خطای مدل را به حداقل می‌رساند. (Flynn et al., 2019) از مدل تقویت گرادیان تصادفی^۳ برای انبوهش‌زدایی واحدهای نقشه خاک استفاده کردند.

در این مطالعه مدل جنگل تصادفی با تعداد ۵۰۰ درخت (ntree = 500) و تعداد ۲ ویژگی تصادفی در هر تقسیم‌بندی (mtry = 2) بهینه گردید. همچنین، مدل تقویت گرادیان افراطی نیز با تعداد ۵۰۰ درخت اجرا شد. مدل‌سازی‌ها با استفاده از تابع dsmart

1. Extreme Gradient Boosting
2. Gradient Boosting
3. Stochastic Gradient Boosting

(Odgers et al., 2014) در نرم افزار R انجام شد. بدین منظور، از هر واحد نقشه ۲۰ نمونه تصادفی در ۲۰ تکرار برداشت شد. در نهایت، نقشه توزیع مکانی زیرگروه‌های خاک با بیشترین احتمال فراوانی تهیه شد.

ارزیابی دقت مدل‌ها

ارزیابی دقت مدل‌ها برای پیش‌بینی زیرگروه‌های خاک و همچنین انبوهش‌زدایی واحدهای نقشه خاک با تشکیل ماتریس خطا و محاسبه شاخص‌های آماری استخراج شده از آن انجام گردید (فرمول‌های ۱ تا ۲). بدین منظور، داده‌های استفاده شده در مدل در مقابل داده‌های تست (به عنوان واقعیت‌های زمینی) در هر کلاس خاک در مقابل هم قرار داده می‌شوند. تعداد خاک‌های هر کلاس که بصورت درست یا نادرست در آن کلاس قرار گرفته‌اند، بیانگر دقت مدل در تخصیص خاک‌ها با کلاس درست می‌باشند (Rossiter, 2001). در سناریوی اول از اطلاعات ۲۳۰ خاک‌رخ برای ارزیابی دقت مدل‌ها استفاده شد. در سناریوی دوم، مدل‌ها با ۷۰ درصد داده‌ها آموزش داده شدند و با توجه به اطلاعات ۳۰ درصد نمونه‌های باقیمانده مورد آزمون قرار گرفتند.

$$OA = \sum_{i=1}^n \frac{X_{ii}}{N} \quad \text{فرمول ۱}$$

صحت کلی مدل آبیانگر نسبت تعداد پیش‌بینی‌های درست به کل تعداد کلاس‌های خاک می‌باشد. دقت کلی بالاتر نشان دهنده عملکرد بالای مدل است. که در آن n تعداد سطر و ستون‌های ماتریس خطا، X_{ii} تعداد در قطر ماتریس است که در آن سطر و ستون i به هم می‌رسند (یعنی طبقه‌بندی صحیح) و N تعداد کل مشاهدات است.

$$\hat{k} = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} X_{+i})} \quad \text{فرمول ۲}$$

که در آن، X_{i+} و X_{+i} به ترتیب تعداد کل ردیف‌ها و تعداد کل ستون‌ها است. شاخص کاپا یک شاخص قوی است که احتمال طبقه‌بندی تصادفی یک کلاس را در نظر می‌گیرد و نسبت تمام موارد احتمالی حضور یا عدم حضور را که به درستی توسط یک مدل پیش‌بینی شده است را پس از محاسبه پیش‌بینی‌های شانسی اندازه‌گیری می‌کند. شاخص کاپا بالاتر نشان دهنده عملکرد بالاتر مدل است.

همچنین، میزان توافق بین نقشه خروجی مدل دسمارت با توزیع مکانی زیرگروه‌های خاک با محاسبه ضریب همبستگی کرامر محاسبه گردید. بدین منظور، نقشه انبوهش‌زدایی شده هر مدل با نقشه توزیع زیرگروه‌های همان مدل مورد ارزیابی قرار گرفتند. بدین معنی که برای محاسبه ضریب کرامر نقشه انبوهش‌زدایی شده مدل C5.0، این نقشه با نقشه پراکنش زیرگروه‌های خاک پیش‌بینی شده با مدل C5.0 تطبیق داده شد. این روند برای مدل‌های جنگل تصادفی و تقویت گرادیان افراطی نیز انجام شد.

-
1. Confusion matrix
 2. Overall accuracy
 3. K-index
 4. Chance predictions
 5. Cramer's V Index

ارزیابی عدم قطعیت مدل‌ها

عدم قطعیت، یک بخش جدایی ناپذیر مدل‌سازی پیش‌بینی کلاس‌های خاک می‌باشد. در نقشه‌برداری رقومی خاک همواره فرآیندهای تخمین و مدل‌سازی با خطا همراه هستند که باعث بروز عدم قطعیت در نقشه نهایی می‌شوند. بررسی و کمی‌سازی عدم قطعیت مدل‌های پیش‌بینی کننده از خود فرآیند پیش‌بینی مهم‌تر است. عدم قطعیت نتیجه نبود اطمینان از واقعیت است که از منابع مختلفی رخ می‌دهند. شاخص‌های آماری مختلفی برای کمی‌سازی میزان عدم قطعیت مدل‌سازی و نقشه‌های پیش‌بینی شده خروجی ارائه شده‌اند. در این مطالعه، عدم قطعیت نقشه‌های رقومی کلاس‌های به دست آمده از مدل‌های اجرا شده با استفاده از شاخص درهمی ارائه شده در فرمول ۳ محاسبه گردید.

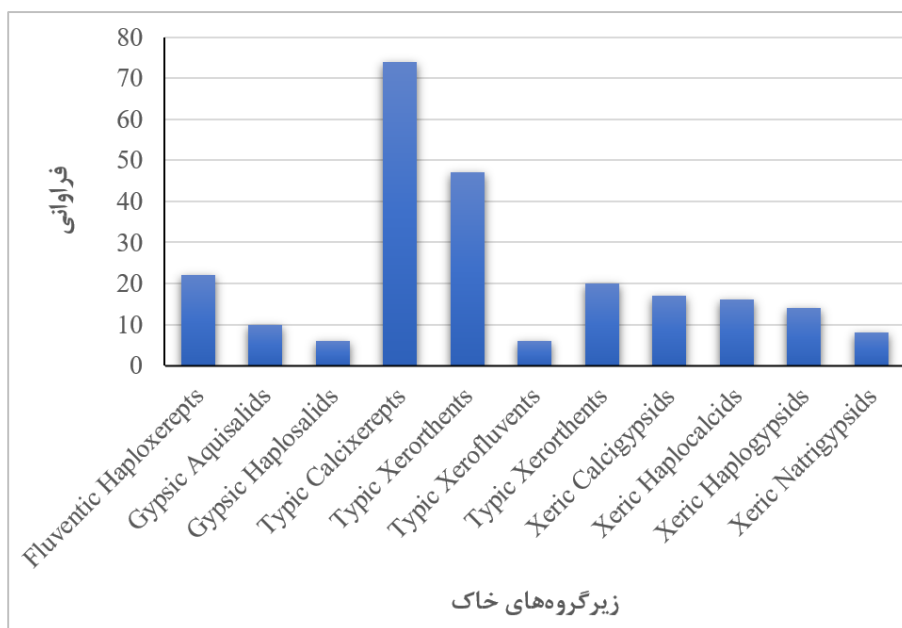
$$CI = 1 - (P_{max} - P_{max-1}) \quad \text{فرمول ۳}$$

مقادیر P_{max} و P_{max-1} به ترتیب نقشه احتمال کلاس خاک مرتبه اول و دوم در هر پیکسل می‌باشند. مقادیر شاخص درهمی از صفر تا یک متغیر است. مقادیر بیش از آن نشان دهنده عدم قطعیت بیشتر است که بازتاب کننده تنوع بالای کلاس‌های خاک می‌باشد. بدین معنی که تشخیص احتمال کلاس مرتبه اول و دوم برای مدل مشکل می‌باشد و به عبارتی کلاس خاک غالب وجود ندارد و کلاس‌های خاک در هر دو نقشه با احتمال نزدیک به هم می‌باشند. مقدار این شاخص درجه شباهت بین دو پیش‌بینی اول و دوم کلاس خاک در یک پیکسل مشخص دو نقشه را نشان می‌دهد. اگر احتمال وقوع بین محتمل‌ترین کلاس خاک و دومین کلاس خاک مشابه باشد، شاخص سردرگمی حاصل بالا خواهد بود، که این امر بیانگر سردرگمی بالای مدل در تعیین کلاس صحیح خاک است (Dolan et al., 2021; Zeraatpisheh et al., 2019).

نتایج و بحث

فراوانی زیرگروه‌های خاک در ۲۳۰ خاک‌رخ جدید مورد استفاده در شکل ۴ دیده می‌شود. بطور کلی خاک‌های منطقه به سه دسته کلی: (۱) آهکی، (۲) متأثر از املاح محلول (شور، گچی، سدیمی) و (۳) کم عمق و با تکامل خاک‌رخی کم تقسیم می‌شوند (جدول ۲). در شمال و مرکز منطقه، خاک‌ها تنوع کمتری دارند که بیشتر آن‌ها را به ترتیب خاک‌های Typic Calcixerepts و Typic Haploxerepts شامل می‌شوند. با تغییر رژیم رطوبتی و همچنین تغییر پستی و بلندی از جمله شیب و ارتفاع در جنوب منطقه، خاک‌های متنوع‌تر غنی از تجمع املاح (Xeric Natrigypsids، Xeric Haplogypsids، Xeric Haplocalcids، Gypsic Haplosalids، Gypsic Aquisalids) تشکیل شده‌اند، بطوری‌که در فواصل کم خاک‌های بسیار متفاوت از یکدیگر دیده می‌شوند (رسانی و همکاران، ۱۴۰۳).

همان‌طور که از اطلاعات نقشه خاک (شکل ۳) استنباط می‌شود، از شمال به جنوب منطقه به ترتیب خاک‌های Typic Xerorthents، Typic Calcixerepts، Typic Haploxerepts، Xeric Calcigypsids و Gypsic Aquisalids بیشترین فراوانی را در واحدهای خاک دارند. این در حالی است که خاک‌های Xeric Natrigypsids، Xeric Haplogypsids، Fluventic Haploxerepts، Typic Xerofluvents و Xeric Natrigypsids کم‌ترین فراوانی را در کل واحدهای نقشه دارند. با این حال، تنوع خاک‌ها در واحدهای نقشه بیشتر از اطلاعات راهنمای نقشه می‌باشد که این امر به دلیل دقت کم و مقیاس کوچک نقشه موروثی خاک می‌باشد. این یافته‌ها با اظهارات Yang et al. (2011) و Zeraatpisheh et al. (2019) هم‌خوانی دارد.



شکل ۴- فراوانی زیرگروه‌های خاک در ۲۳۰ خاک رخ جدید مورد استفاده در انبوهش‌زدایی واحدهای نقشه موروثی خاک در سناریوی دوم و همچنین در ارزیابی کمی مدل‌ها

انبوهش‌زدایی واحدهای نقشه موروثی خاک

جدول ۳ نتایج ارزیابی کمی نقشه خاک انبوهش‌زدایی شده با مدل‌های مختلف را نشان می‌دهد. در سناریوی اول، صحت کلی مدل دسمارت معمولی (با درخت تصمیم C5.0) برای نقشه احتمال مرتبه اول زیرگروه‌های خاک ۰/۳۷ و کاپا ۰/۱۷ می‌باشد. اگرچه این نتایج با یافته‌های سایر محققان هم‌خوانی دارد، اما علت پایین بودن صحت کلی نقشه در این مدل را می‌توان به این امر نسبت داد که این مدل، نقشه‌های پیش‌بینی تفکیک شده را در سه سطح احتمال ارائه می‌دهد (Malone et al., 2017). Zeraatpisheh et al. (2019) صحت کلی نقشه سطح احتمال اول زیرگروه‌های خاک نقشه خاک یک میلیونیم کشور در منطقه بروجن را ۰/۳۳ گزارش کردند. مقدار این شاخص در نقشه احتمال سطح اول در مطالعه Vincent et al. (2018)، Holmes et al. (2015)، و فاتحی و همکاران (۱۳۹۷) به ترتیب ۰/۲۳، ۴۰ و ۴۴ درصد گزارش شده است.

جدول ۳- نتایج ارزیابی کمی مدل دسمارت برای انبوهش‌زدایی واحدهای نقشه خاک با مدل‌های یادگیری ماشین درخت محور (C5)، درخت تصمیم C5.0، RF: جنگل تصادفی، XGB: تقویت گرادیان افراطی) در دو سناریوی مورد بررسی

سناریو	مدل	صحت کلی	شاخص کاپا	شاخص درهمی	شاخص ترکیبی (صحت کلی در شاخص شانون)	شاخص کرامر
سناریو ۱	C5	۰/۳۷	۰/۱۷	۰/۷۶	۰/۷۳	۰/۴۱
	RF	۰/۴۱	۰/۲۸	۰/۷۱	۰/۷۶	۰/۵۳

۰/۴۶	۰/۸۰	۰/۷۴	۰/۲۹	۰/۴۲	XGB
۰/۸۴	۱/۰۶	۰/۶۵	۰/۴۴	۰/۵۱	C5 سناریو ۲
۰/۸۱	۱/۳۴	۰/۶۵	۰/۶۰	۰/۶۷	RF
۰/۸۴	۱/۰۸	۰/۶۰	۰/۴۶	۰/۵۳	XGB

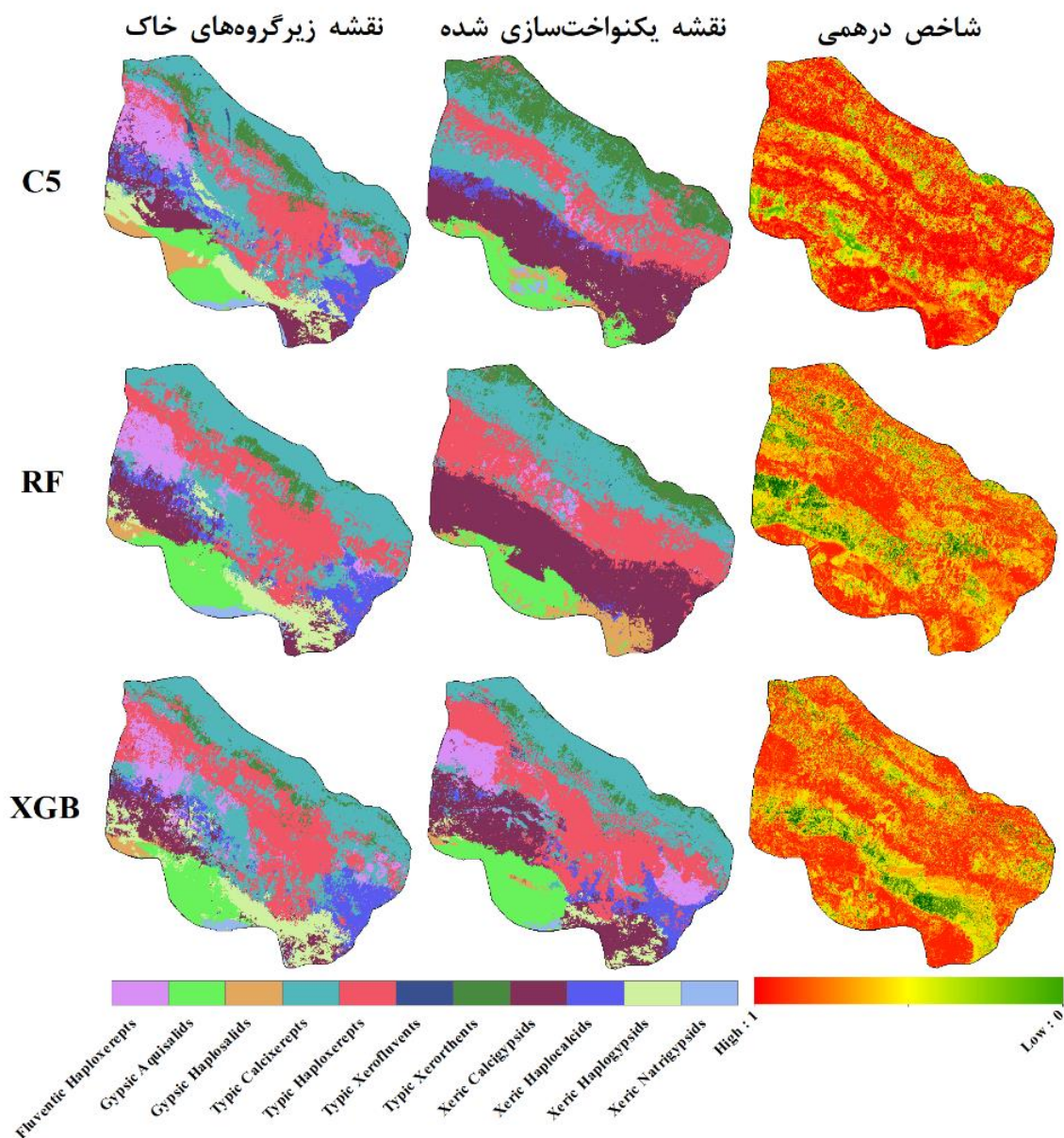
میزان صحت کلی و کاپای نقشه احتمال سطح اول زیرگروه‌های خاک در مدل‌های تقویت گرادیان افراطی و جنگل تصادفی افزایش یافته است. نتایج Easher et al. (2023) نشان داد که مدل جنگل تصادفی توانسته است در مقایسه با مدل درخت تصمیم زیرگروه‌ها و سری‌های خاک نقشه موروثی خاک کانادا را با عملکرد بهتر (با صحت کلی و کاپای بالاتر) تفکیک کند. نتایج ارزیابی انبوه‌ش‌زدایی نقشه خاک ۲۵۰ هزار اسکالند نشان داد که الگوریتم جنگل تصادفی در مقایسه با درخت تصمیم در تفکیک مکانی چند ضلعی‌های خاک موفق بوده و دقت پیش‌بینی خوبی را برای اکثر انواع خاک، به استثنای برخی از انواع خاک‌های کم وسعت که معمولاً در واحدهای نقشه پیچیده یافت می‌شوند، ارائه کرده است. ایشان علت این امر را به تمایل الگوریتم جنگل تصادفی به سوگیری بیشتر به کلاس‌های غالب و همچنین مشکل مدل در تمایز بین خاک‌های مشابه در مناطق مختلف نسبت دادند (Gagkas and Lilly, 2024). همچنین نتایج Flynn et al. (2019) نشان داد که با استفاده از مدل تقویت گرادیان تصادفی در تفکیک اجزای واحد نقشه موروثی در آفریقای جنوبی، کیفیت نقشه با صحت کلی و کاپای بیشتر در مقایسه با مدل‌های درخت تصمیم و جنگل تصادفی بهبود یافته است. با این حال، پایین بودن صحت کلی و کاپای بدست آمده از مدل‌ها را می‌توان به بالا بودن نسبی تعداد زیرگروه‌های خاک، نمونه‌برداری تصادفی واحدهای خاک، و همچنین وزن‌دهی و سوگیری آن به کلاس‌های خاک غالب در روش دسمارت نسبت داد (فاتحی و همکاران، ۱۳۹۷؛ Zeraatpisheh et al., 2019؛ Gagkas and Lilly, 2024). فاتحی و همکاران (۱۳۹۷) علت پایین بودن ضریب کاپا در مطالعه تفکیک اجزای واحدهای نقشه خاک در کرمانشاه را به گستردگی زیرگروه‌های خاک غالب در منطقه و همچنین تعداد زیاد مشاهدات واقعی مربوط به این زیرگروه در داده‌های اعتبارسنجی نسبت داده‌اند. بنابراین ایشان اظهار داشتند که این عوامل دلیلی بر پایین بودن کارایی مدل دسمارت در تفکیک زیرگروه‌های خاک نمی‌باشد.

بطور کلی، نقشه سطح احتمال اول بدست آمده از مدل تقویت گرادیان افراطی با بالاترین میزان صحت کلی و شاخص شانن، عملکرد بهتری نسبت به دو مدل دیگر در تفکیک زیرگروه‌های خاک واحدهای نقشه داشته است. همچنین، مقدار بیشتر شاخص شانن در این مدل نشان می‌دهد که توانسته است تنوع خاک‌ها را در سطح زیرگروه در منطقه به‌خوبی نمایش دهد. میزان شاخص ترکیبی بدست آمده از حاصل ضرب صحت کلی و شاخص شانن (Jafari et al., 2013) برای این مدل ۰/۸ است که تأیید کننده نتایج فوق می‌باشد.

نقشه‌های سطح اول احتمال زیرگروه‌های خاک در نقشه موروثی خاک انبوه‌ش‌زدایی شده در این سه مدل در سناریوی اول به همراه شاخص درهمی بدست آمده از تفاضل نقشه‌های سطح احتمال اول و دوم در شکل ۵ نمایش داده شده است. یافته‌ها نشان می‌دهند که با وجود اینکه مدل دسمارت نمونه‌های تصادفی را بطور جداگانه از هر واحد نقشه جمع‌آوری می‌کند، پیش‌بینی‌های دقیقی از زیرگروه‌های خاک برای کل منطقه مطالعه ارائه داده است. به عنوان مثال، واحد نقشه ۵۵ در مرکز منطقه را در نظر بگیرید؛ طبق راهنمای نقشه، این واحد از ترکیب سه خاک تشکیل شده است (شکل ۳). با این حال هر سه مدل دسمارت مورد استفاده (درخت تصمیم C5.0، جنگل تصادفی و تقویت گرادیان افراطی) علاوه بر این سه زیرگروه، خاک‌های Typic، Typic Xerorthents، Xeric Haplocalcids و Xerofluvents را نیز در این واحد تفکیک کرده‌اند که با توزیع زیرگروه‌های خاک منطقه (جدول ۲) هم‌خوانی

بالایی دارد. این توانایی مدل دسمارت در انبوهش‌زدایی واحدهای نقشه خاک، تایید کننده دقت و کارایی بالای این الگوریتم می‌باشد. همچنین، مقایسه نتایج این مدل‌ها با نقشه توزیع زیرگروه‌های خاک نشان می‌دهد که بین ۴۱ تا ۵۳ درصد انطباق بین این دو نقشه وجود دارد که این موضوع، عملکرد خوب مدل دسمارت را در انبوهش‌زدایی واحدهای نقشه خاک مطابق با واقعیت‌های زمینی تأیید می‌کند. میانگین شاخص درهمی نقشه خاک در مدل جنگل تصادفی ۰/۷۱ می‌باشد که در مقایسه با دو مدل دیگر اندکی کاهش یافته است (جدول ۳). Stoorvogel et al. (2019) بیان داشتند که الگوریتم جنگل تصادفی به دلیل ماهیت مدل‌سازی گروهی و در نتیجه انتخاب کلاس با اکثریت رای به عنوان پیش‌بینی نهایی نسبت به مدل درخت طبقه‌بندی ساده، کارایی و قدرت بیشتری در تمایز و تشخیص خاک‌ها و پیش‌بینی آن‌ها دارد. این ویژگی باعث کاهش واریانس مدل‌سازی می‌شود و تأثیر آن در نقشه پراکنش شاخص درهمی این مدل‌ها نیز دیده می‌شود (شکل ۵). بطورکلی، میزان شاخص درهمی در قسمت‌های مرکزی و جنوبی منطقه در تمام مدل‌ها زیاد می‌باشد. اما در سایر قسمت‌ها، مقدار این شاخص در مدل جنگل تصادفی نسبت به دو مدل دیگر کاهش یافته است. شاخص درهمی بالا (بیشتر از ۰/۸) بیانگر تنوع بیشتر خاک‌ها در منطقه می‌باشد که در نتیجه احتمال وجود زیرگروه خاک مرتبه اول و دوم بسیار به هم نزدیک می‌باشد (Vincent et al., 2018). این نتایج با این واقعیت که در قسمت‌های میانی و جنوبی منطقه خاک‌های متنوع‌تر با فاصله کم از یکدیگر تشکیل شده‌اند هم‌خوانی دارد (رسائی و همکاران، ۱۴۰۳).

پژوهش‌های پیشرفته



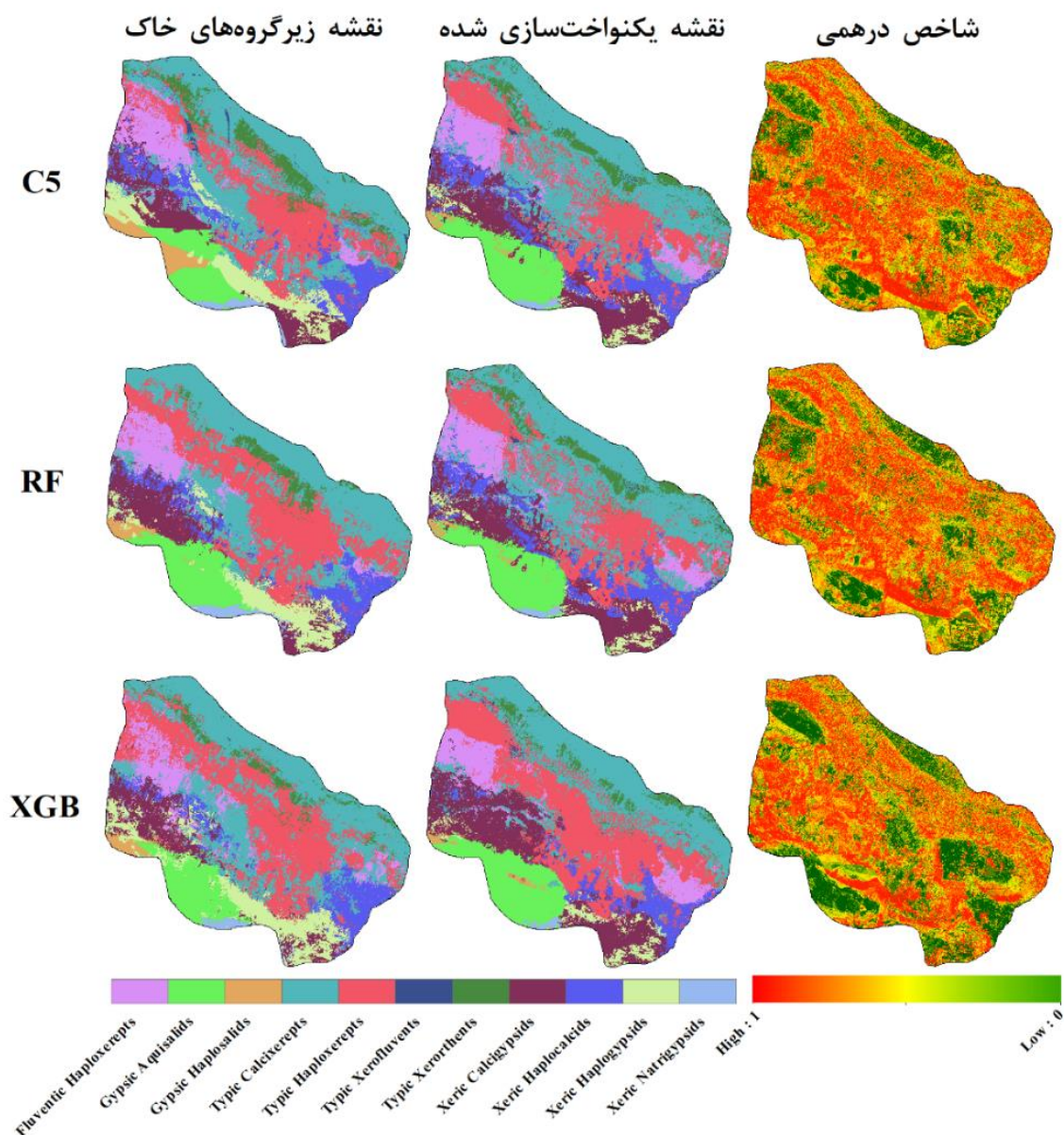
شکل ۵- پراکنش مکانی زیرگروه‌های خاک و نقشه احتمال کلاس خاک مرتبه اول حاصل از انبوهش‌زدایی نقشه خاک به همراه شاخص درهمی مربوطه در سه مدل درخت محور مورد بررسی (C5: درخت تصمیم C5.0، RF: جنگل تصادفی، XGB: تقویت گرادیان افراطی) در سناریو شماره ۱ (انبوهش‌زدایی واحدهای نقشه با استفاده از اطلاعات راهنمای نقشه موروثی خاک).

در سناریوی دوم، میزان صحت کلی نقشه‌ها از ۰/۵۱ تا ۰/۶۷ و میزان کاپا از ۰/۴۴ تا ۰/۶۰ متغیر می‌باشد. مدل جنگل تصادفی با بیشترین میزان شاخص ترکیبی (۱/۳۴)، بالاترین کارایی را در مقایسه با دو مدل دیگر نشان داده است (جدول ۳). بنابراین، این مدل با دسترسی به داده‌های جدید خاک توانسته است نقشه سطح اول احتمال زیرگروه‌های خاک را با دقت بالاتر تهیه کند. این نتایج بیانگر اهمیت میزان اطلاعات ورودی در عملکرد مدل دسمارت می‌باشد. Holmes et al. (2015) نشان دادند که با افزایش دقت داده‌های مکانی ورودی، صحت کلی نقشه‌های انبوهش‌زدایی شده تا ۷۱ درصد افزایش یافت. در مطالعات Ellili-Bargaoui et al.

(2020) و Flynn and Kostecki (2024) با وارد کردن قوانین خاک-زمین‌نمای کارشناس به مدل، صحت کلی نقشه‌های انبوهش‌زدایی شده تا حد زیادی افزایش یافته است.

شکل ۶، نقشه توزیع زیرگروه‌های خاک بدست آمده از مدل دسمارت بر پایه مدل‌های درختی مورد بررسی در سناریوی دوم، همراه با توزیع مکانی شاخص درهمی این مدل‌ها را نشان می‌دهد. مقایسه توزیع مکانی زیرگروه‌های خاک در نقشه انبوهش‌زدایی شده با نقشه زیرگروه‌های خاک پیش‌بینی شده در هر مدل نشان دهنده تطابق بالای نقشه‌ها با یکدیگر می‌باشد (شاخص کرامر بالای ۸۰ درصد). این نشان می‌دهد که ۸۰ درصد از تغییرات زیرگروه‌های خاک در نقشه‌های متناظر با هم مطابقت دارد که نشان دهنده همبستگی قوی بین دو نقشه است. میزان شاخص درهمی در قسمت‌های زیادی از منطقه، به ویژه در نواحی مرکزی زیاد می‌باشد که حاکی از عدم توانایی مدل‌ها در تعیین کلاس خاک غالب در آن سلول نقشه می‌باشد که بازتاب کننده تنوع زیاد خاک‌ها در این قسمت‌ها می‌باشد. در این میان، کمترین مقدار شاخص درهمی (۰/۶۰) مربوط به مدل تقویت گرادیان افراطی است که نشان دهنده عملکرد بهتر این مدل در تفکیک کلاس‌های خاکی با حداقل هم‌پوشانی است.

پژوهش‌های پیشرفته



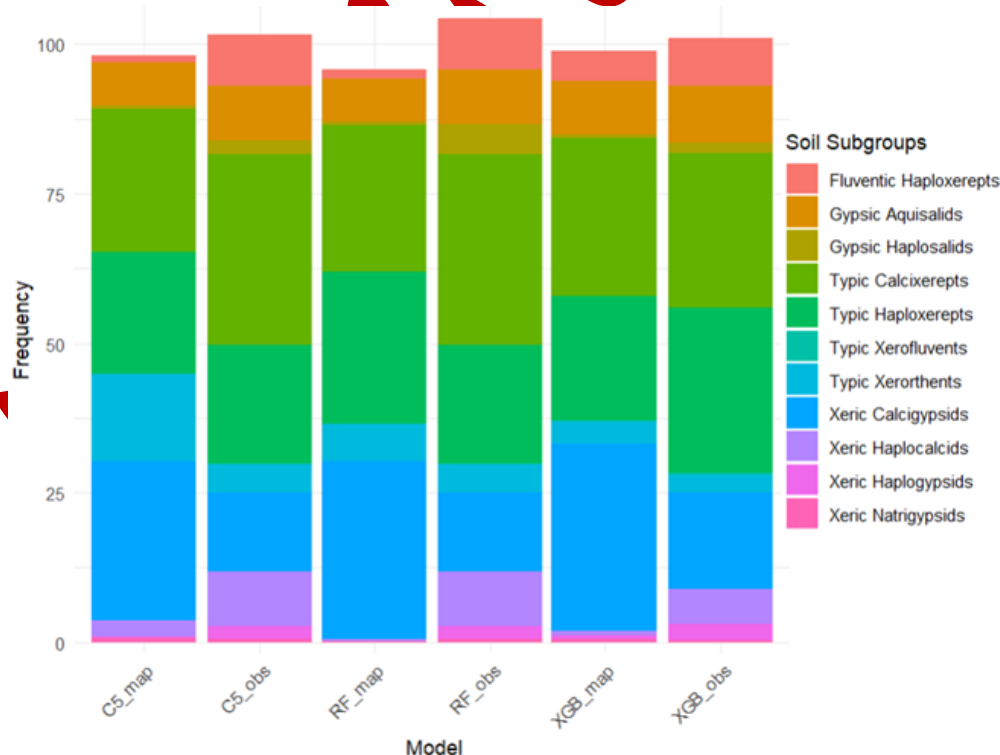
شکل ۶- پراکنش مکانی زیرگروه های خاک و نقشه احتمال کلاس خاک مرتبه اول حاصل از انبوهش زدایی نقشه خاک به همراه شاخص درهمی مربوطه در سه مدل درخت محور مورد بررسی (C5: درخت تصمیم C5.0، RF: جنگل تصادفی، XGB: تقویت گرادیان افراطی) در سناریو شماره ۲ (انبوهش زدایی واحدهای نقشه با استفاده از اطلاعات راهنمای نقشه موروثی خاک و داده های خاک رخی اضافی)

مقایسه نتایج سناریوهای اول و دوم

بررسی نتایج نشان می دهد که در سناریوی دوم، کارایی مدل های دسمارت در تفکیک اجزای واحدهای نقشه به زیرگروه های خاک افزایش قابل توجهی داشته است. بطور کلی، میزان صحت کلی تا ۲۶ درصد و شاخص کاپا تا ۳۱ درصد افزایش نشان می دهد (جدول ۳)، در حالی که شاخص درهمی تا ۱۴ درصد کاهش یافته است. علاوه بر این، افزایش ۴۳ درصدی در تطابق نقشه های حاصل از مدل های دسمارت با نقشه توزیع زیرگروه های خاک و افزایش ۵۸ درصدی شاخص ترکیبی مدل های سناریوی دوم نسبت به سناریوی

اول، بیانگر اهمیت میزان و نوع داده‌های ورودی مدل دسمارت در کارایی آن می‌باشد. (Minai et al. 2020) بیان داشتند که نتایج مدل دسمارت به شدت تحت تاثیر قدرت تفکیک مکانی متغیرهای محیطی ورودی و نوع اطلاعات واحدهای نقشه خاک قرار دارد. همچنین، نتایج (Ellili-Bargaoui et al. 2019, 2020) بر اهمیت میزان داده‌های ورودی مدل دسمارت تاکید دارد. Easher et al. (2023) نیز گزارش کردند که کارایی مدل دسمارت به روش نمونه‌برداری، نوع نمونه‌برداری و مدل مورد استفاده بستگی دارد. نتایج (Zeraatpisheh et al. 2019) نشان داد که اختلاف بین تعداد کلاس‌های خاک در سطح نقشه و داده‌های ارزیابی می‌تواند کارایی مدل را به شدت تحت تاثیر قرار دهد. (Flynn et al. 2019) تأکید کردند که الگوریتم مورد استفاده در روش دسمارت باید بر اساس در دسترس بودن مشاهدات ارزیابی انتخاب شود؛ بطوری‌که در صورت وجود تعداد زیاد کلاس‌های خاک، مدلی که سردرگمی الگوریتم دسمارت را به حداقل می‌رساند، گزینه مناسب‌تری خواهد بود.

شکل ۷ توزیع فراوانی زیرگروه‌های خاک در نقشه‌های احتمال اول مدل‌های دسمارت در سناریوهای اول و دوم را نشان می‌دهد. بررسی و مقایسه این نتایج نشان می‌دهد که در سناریوی دوم، مدل‌ها توانسته‌اند خاک‌های با کم‌ترین میزان فراوانی (Fluventic, Xeric Haplogypsis, Xeric Haplocalcids, Typic Xerofluvents, Gypsic Haplosalids, Gypsic Aquisalids, Haploxerepts, Xeric Natrigypsis) را بهتر پیش‌بینی کنند و فراوانی این زیرگروه‌های خاک در سطح منطقه در این سناریو افزایش یافته است. (Ellili-Bargaoui et al. 2020) بیان داشتند که افزودن مشاهدات خاک اضافی به مدل، امکان پیش‌بینی دقیق‌تر کلاس‌های خاصی را که در شرایط محیطی ویژه رخ می‌دهند، فراهم می‌کند. بطور کلی، نتایج این مطالعه نشان داد که مدل جنگل تصادفی با استفاده از داده‌های اضافی خاک، توانسته است نقشه سطح اول احتمال توزیع زیرگروه‌های خاک را با بالاترین میزان صحت کلی و کاپا، بهتر از سایر مدل‌ها تولید کند.

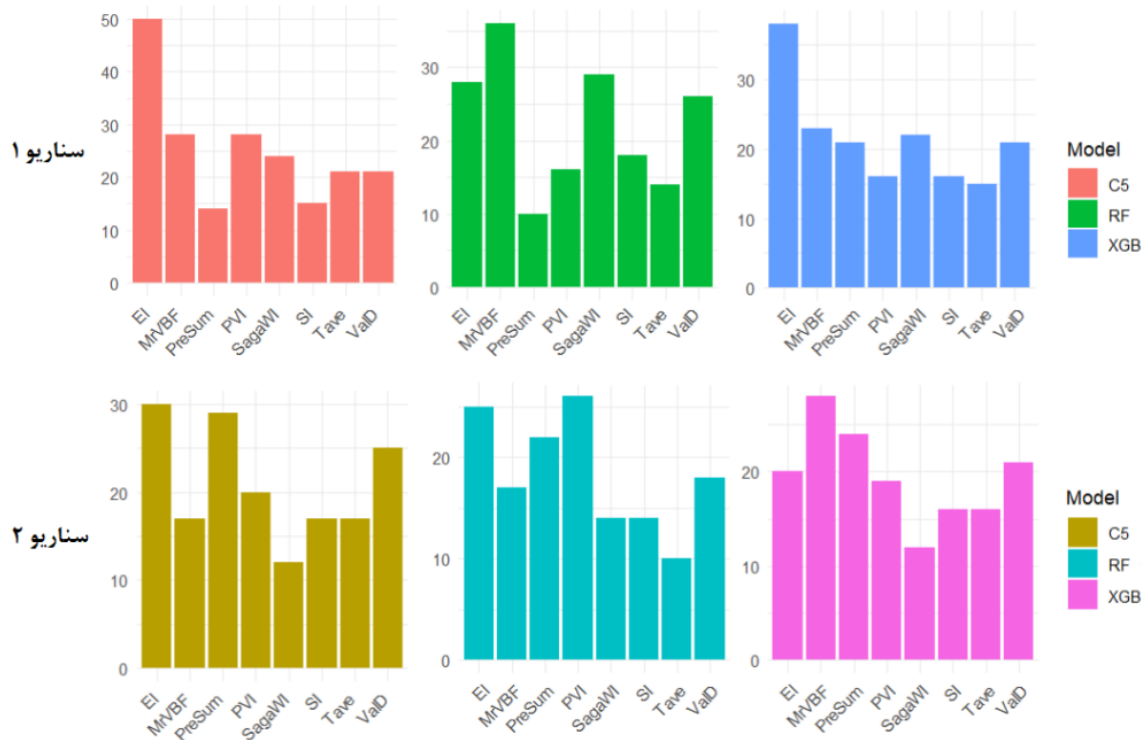


شکل ۷- توزیع فراوانی زیرگروه‌های خاک مدل‌های دسمارت در سناریوهای ۱ (با پسوند map) و ۲ (با پسوند obs)

اهمیت متغیرهای محیطی

شکل ۸ اهمیت نسبی متغیرهای محیطی استفاده شده در مدل‌های مختلف را نشان می‌دهد. با توجه به نتایج، اهمیت نسبی متغیرهای کمکی در مدل‌های مختلف و همچنین در دو سناریو متفاوت می‌باشد. با این حال، تفاوت زیاد و معنی‌داری در اهمیت متغیرها در مدل‌های مختلف دیده نمی‌شود و نمی‌توان متغیر خاصی را به‌عنوان مهم‌ترین متغیر معرفی کرد. بنابراین، مدل‌ها برای هر کدام از تکرارها از متغیرهای کمکی متفاوتی استفاده کرده‌اند. این موضوع نشان می‌دهد که مدل‌ها برای پیش‌بینی توزیع زیرگروه‌های خاک در منطقه به متغیرهای کمکی زیادی نیاز دارند. این امر ممکن است متأثر از تعداد نمونه‌های تصادفی که مدل در هر تکرار در هر واحد خاک برداشت می‌کند نیز باشد (Flynn et al., 2019).

بطور کلی، در سناریوی اول، متغیرهای توپوگرافی مثل ارتفاع، شاخص همواری دره با قدرت تفکیک بالا، عمق دره، شاخص خیزی توپوگرافی بیشترین مشارکت و اهمیت را در مدل‌سازی داشته‌اند. اما با وارد کردن مشاهدات جدید زیرگروه‌های خاک به مدل در سناریوی دوم، اهمیت متغیرهای محیطی اقلیمی مانند میانگین بارندگی و پوشش گیاهی در مدل‌سازی‌ها افزایش یافته است. این تغییر در روند مدل‌سازی منطقی به نظر می‌رسد. در سناریوی اول، مدل بر اساس اطلاعات نقشه خاک موروثی اقدام به انبوهش‌زدایی و تفکیک زیرگروه‌های خاک کرده است. بنابراین از آنجایی که این واحدها عمدتاً بر اساس تغییرات توپوگرافی در منطقه تفکیک شده‌اند متغیرهای محیطی توپوگرافی بدست آمده از مدل رقومی ارتفاع نقش بیشتری در مدل‌ها داشته‌اند. با وارد کردن اطلاعات اضافی، مدل بهتر توانسته است ارتباط بین خاک و متغیرهای محیطی را درک کند و به متغیرهای اقلیمی و پوشش گیاهی نیز در این رابطه اهمیت بیشتری داده است. این نتایج با نتایج مطالعات فاتحی و همکاران (۱۳۹۷) و Zeraatpisheh et al. (2019) هم‌خوانی دارد. اهمیت این متغیرهای محیطی در پیش‌بینی زیرگروه‌های خاک در سایر مطالعات انجام شده در سطح منطقه مورد مطالعه از جمله موسوی و همکاران (۱۳۹۸)، رحمانی و همکاران (۱۴۰۱)، ممتازی بروجنی و همکاران (۱۴۰۱)، رسائی و همکاران (۱۴۰۳) و Neyestani et al. (2021) نیز تایید شده است.



شکل ۸- میانگین اهمیت نسبی متغیرهای کمکی استفاده شده در انبوهش‌زدایی واحدهای نقشه موروثی در منطقه مورد مطالعه (برای اطلاع از متغیرهای محیطی به جدول ۱ مراجعه شود)

نتیجه‌گیری

مطالعه حاضر با هدف انبوهش‌زدایی واحدهای نقشه موروثی خاک یک میلیونیم کشور در منطقه‌ای به وسعت ۵۸۰۰۰ هکتار از اراضی دشت قزوین انجام گرفت. در این پژوهش، مدل دسمارت به‌عنوان یکی از رایج‌ترین روش‌های انبوهش‌زدایی واحدهای نقشه خاک چندضلعی در سطح زیرگروه‌های تاکسونومیکی خاک مورد استفاده قرار گرفت. به منظور بررسی اثر مدل مورد استفاده بر دقت فرآیند تفکیک، سه مدل مبتنی بر سیستم درختی شامل درخت تصمیم C5.0، جنگل تصادفی و تقویت گرادیان افراطی بر اساس اطلاعات راهنمای نقشه خاک بکار گرفته شدند. همچنین، تاثیر وارد کردن اطلاعات ۲۳۰ خاک‌رخ جدید در کارایی مدل‌های انبوهش‌زدایی واحدهای نقشه خاک مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل دسمارت توانایی مطلوبی در تفکیک زیرگروه‌های خاک از واحدهای نقشه دارد و استفاده از مدل‌های پیشرفته‌تر مانند جنگل تصادفی و تقویت گرادیان افراطی موجب افزایش دقت پیش‌بینی در مقایسه با درخت تصمیم C5.0 شد. علاوه بر این، ورود داده‌های جدید باعث بهبود کارایی مدل‌ها شد که بر اهمیت کیفیت داده‌های ورودی و آموزش مناسب مدل تأکید دارد. از میان متغیرهای محیطی مورد استفاده، عوامل مربوط به پستی و بلندی تأثیرگذارترین ویژگی‌ها در فرآیند انبوهش‌زدایی واحدهای نقشه خاک بودند. بنابراین، این مطالعه، تطبیق‌پذیری مدل دسمارت اصلاح شده برای تفکیک زیرگروه‌های خاک در واحدهای نقشه خاک قدیمی را نشان می‌دهد. استفاده از داده‌های خاک‌رخی بیشتر با تعداد بالاتری از زیرگروه‌های خاک و همچنین استفاده از روابط خاک-زمین‌نمای توصیف‌کننده واحدهای نقشه خاک برای بهبود نتایج مدل دسمارت در مطالعات آینده پیشنهاد می‌گردد.

قدردانی

این اثر تحت حمایت مادی صندوق حمایت از پژوهشگران و فناوران کشور (INSF) برگرفته شده از طرح شماره ۴۰۲۴۳۴۳، انجام شده است. بدین وسیله نویسنده اول مقاله مراتب سپاسگزاری خود را از صندوق INSF بابت حمایت‌های مالی طرح و همچنین از دانشگاه تهران بابت فراهم آوردن امکانات اجرای آن اعلام می‌دارد.

"هیچگونه تعارض منافع بین نویسندگان وجود ندارد."

منابع

۱. خاموشی، سیدعرفان؛ سرمیدیان، فریدون و امید، محمود. (۱۴۰۱). تهیه نقشه رقومی کربن آلی ذخیره شده در خاک با استفاده از روش‌های یادگیری ماشینی. مجله تحقیقات آب و خاک ایران، ۵۳(۱۱)، ۲۶۷۱-۲۶۸۱.
۲. خاموشی، سیدعرفان؛ سرمیدیان، فریدون و کشاورزی، علی. (۱۳۹۷). تهیه نقشه رقومی خاک با مدل جنگل‌های تصادفی و ارزیابی تناسب اراضی برای بخشی از منطقه آبیک، استان قزوین. مرتع و آبخیزداری، مجله منابع طبیعی ایران، ۷۱(۴)، ۸۸۵-۸۹۹.
۳. رحمانی، اصغر؛ سرمیدیان، فریدون و عارف، حسین. (۱۴۰۱). مدل‌سازی و پیش‌بینی رقومی کلاس‌های زیرگروه خاک با استفاده از رویکرد یادگیری عمیق در بخشی از اراضی خشک و نیمه‌خشک دشت قزوین، مجله تحقیقات آب و خاک ایران، ۵۳(۱۰)، ۲۴۷۷-۲۴۹۹.
۴. رسائی، زهرا؛ سرمیدیان، فریدون و جعفری، اعظم. (۱۴۰۳). بررسی گروه‌بندی خاک با استفاده از مدل‌های خوشه‌بندی مرسوم و مدرن در بخش‌هایی از دشت قزوین. تحقیقات آب و خاک ایران، ۵۵(۸)، ۱۲۷۳-۱۲۹۵.
۵. رضایی، گردآفرین؛ سرمیدیان، فریدون؛ محمدی‌ترکاشوند، علی؛ سیدمحمدی، جواد و مرعشی‌علی‌آبادی، مریم. (۱۴۰۲). نقشه‌برداری رقومی تغییرات سطحی و زیرسطحی کربن آلی و شوری خاک در بخشی از اراضی دشت قزوین (مطالعه موردی: مناطق آبیک و نظر آباد). نشریه آب و خاک، ۳۷(۲)، ۳۱۵-۳۳۱.
۶. فاتحی، شاهرخ؛ افتخاری، کامران؛ قادری، جلال. (۱۳۹۷). انبوه‌شناسی مکانی نقشه نیمه‌تفصیلی خاک با استفاده از روش دسمارت. مجله پژوهش‌های حفاظت آب و خاک، ۲۹(۲)، ۴۹-۶۹.
۷. ممتازی بروجنی، میلاد و سرمیدیان، فریدون. (۱۴۰۱). پیش‌بینی مکانی کلاس خاک با کمک مدل درخت تصمیم‌گیری C5.0 تقویت شده با بوستینگ در بخشی از اراضی شهرستان آبیک. مرتع و آبخیزداری، مجله منابع طبیعی ایران، ۷۵(۴)، ۵۵۳-۵۷۲.
۸. موسوی، سیدروح اله؛ سرمیدیان، فریدون و 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.
۹. موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰ الف). کاربرد مدل‌های یادگیری ماشین در برآورد مکانی فسفر و پتاسیم خاک در بخشی از اراضی دشت آبیک. نشریه علمی پژوهش‌های خاک، ۳۵(۴)، ۳۹۷-۴۱۲.
۱۰. موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰ ب). مدل‌سازی تغییرات عمقی کربنات کلیم معادل خاک با استفاده از الگوریتم‌های یادگیری ماشین در دشت قزوین. نشریه آب و خاک (علوم و صنایع کشاورزی)، ۲۵(۵)، ۷۱۹-۷۳۴.
۱۱. موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰ پ). مدل‌سازی رقومی تغییرات سه بعدی شوری خاک با استفاده از الگوریتم‌های یادگیری ماشین در اراضی خشک و نیمه‌خشک دشت قزوین. تحقیقات آب و خاک ایران، ۵۲(۷)، ۱۹۱۵-۱۹۲۹.
12. Abdel-Kader, F.H., 2011. Digital soil mapping at pilot sites in the northwest coast of Egypt: a multinomial logistic regression approach. Egypt. J. Remote Sens. Space Sci. 14, 29-40.
13. Allbed, A., & Kumar, L. (2013). Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. Advances in remote sensing, 2(4), 373-385.
14. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
15. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.

16. Conrad, O., Bechtel, B., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8, 1991–2007.
17. Dolan, M. F., Ross, R. E., Albretsen, J., Skarðhamar, J., Gonzalez-Mirelis, G., Bellec, V. K., ... & Bjarnadóttir, L. R. (2021). Using spatial validity and uncertainty metrics to determine the relative suitability of alternative suites of oceanographic data for seabed biotope prediction. A case study from the Barents Sea, Norway. *Geosciences*, 11(2), 48.
18. Easher, T.H., Saurette, D., Chappell, E., Lopez, F. de J.M., Gasser, M.O., Gillespie, A., Heck, R.J., Heung, B., & Biswas, A. (2023). Sampling and classifier modification to DSMART for disaggregating soil polygon maps. *Geoderma*, 431, 116360.
19. Ellili-Bargaoui, Y., Malone, B.P., Michot, D., Minasny, B., Vincent, S., Walter, C., & Lemerrier, B. (2020). Comparing three approaches of spatial disaggregation of legacy soil maps based on the Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART) algorithm. *SOIL*, 6(2), 371–388.
20. Ellili-Bargaoui, Y., Walter, C., Michot, D., Saby, N. P., Vincent, S., & Lemerrier, B. (2019). Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma*, 356, 113907.
21. Flynn, T., & Kostecki, R. (2024). Spatial downscaling of global soil texture classes into 30 m images at the province scale. *Geomatica*, 76(2), 100028.
22. Flynn, T., Van Zijl, G., Van Tol, J., Botha, C., Rozanov, A., Warr, B., & Clarke, C. (2019). Comparing algorithms to disaggregate complex soil polygons in contrasting environments. *Geoderma*, 352, 171–180.
23. Gagkas, Z., & Lilly, A. (2024). Spatial disaggregation of a legacy soil map to support digital soil and land evaluation assessments in Scotland. *Geoderma Regional*, 38, e00833.
24. Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39, 1347–1359.
25. Holmes, K., Griffin, E., & Odgers, N. (2015). Large-area spatial disaggregation of a mosaic of conventional soil maps: Evaluation over Western Australia. *Soil Research*, 53, 865–880.
26. Jafari, A., Ayoubi, S., Khademi, H., Finke, P., Toomanian, N. (2013). Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology* 201, 86–97.
27. Jamshidi, M., Delavar, M.A., Taghizadehe-Mehrjardi, R., & Brungard, C. (2019). Disaggregation of conventional soil map by generating multi realizations of soil class distribution (case study: Saadat Shahr plain, Iran). *Environmental Monitoring and Assessment*, 191(12), 769.
28. Khamoshi, S. E., Sarmadian, F., & Keshavarzi, A. (2019). Digital soil mapping Using Random Forests and Land Suitability Evaluation for Abyek Region, Qazvin Province. *Journal of Range and Watershed Management*, 71(4), 885–899. (In Persian with English abstract).
29. Khamoshi, S. E., Sarmadian, F., & Omid, M. (2023). Predicting and Mapping of Soil Organic Carbon Stock Using Machine Learning Algorithm. *Iranian Journal of Soil and Water Research*, 53(11), 2671–2681. (In Persian with English abstract).
30. Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, 32, 100368.
31. Kursu, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(1), 1–13.
32. Lázaro-López, A., González-SanJosé, M.L., Gómez-Miguel, V., Malone, B., Lázaro-López, A., González-SanJosé, M.L., Gómez-Miguel, V., & Malone, B. (2021). Disaggregation of conventional soil maps: A review. *Soil Research*, 59(8), 747–766.
33. Machado, I.R., Giasson, E., Campos, A. R., Costa, J.J.F., da Silva, E.B., & Bonfatti, B.R. (2018). Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. *Revista Brasileira de Ciência Do Solo*, 42, e0170193.
34. Malone, B. P., Styc, Q., Minasny, B., & McBratney, A. B. (2017). Digital soil mapping of soil carbon at the farm scale: A spatial downscaling approach in consideration of measured and uncertain data. *Geoderma*, 290, 91–99.
35. McBratney, A.B., Mendonça Santos, M.L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1), 3–52.

36. Minai, J., Libohova, Z., & Schulze, D.G. (2020). Disaggregation of the 1:100,000 Reconnaissance soil map of the Busia Area, Kenya using a soil landscape rule-based approach. *Catena*, 195, 104806.
37. Momtazi Burojeni, M., & Sarmadian, F. (2023). Spatial prediction of soil classes using C5.0 boosted decision tree model in Abyek Area. *Iranian Journal of Soil and Water Research*, 75(4), 553-572. (In Persian with English abstract).
38. Mousavi, S. R. A., 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. *Journal of Water and Soil*, 50(10), 2525-2538. (In Persian with English abstract).
39. Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2021a). Application of Machine Learning Models in Spatial Estimation of Soil Phosphorus and Potassium in Some Parts of Abyek Plain. *Iranian Journal of Soil Science*, 35(4), 397-411. (In Persian with English abstract).
40. Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2021b). Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Journal of Water and Soil*, 35(5), 719-734. (In Persian with English abstract).
41. Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2021c). Digital Modeling of Three-Dimensional Soil Salinity Variation Using Machine Learning Algorithms in Arid and Semi-Arid lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 52(7), 1915-1929. (In Persian with English abstract).
42. Mousavi, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement*, 201, 111706.
43. Neyestani, M., Sarmadian, F., Jafari, A., Keshavarzi, A., & Sharififar, A. (2021). Digital mapping of soil classes using spatial extrapolation with imbalanced data. *Geoderma Regional*, 26, e00422.
44. Odgers, N., Mcbratney, A., Minasny, B., Sun, W., & Clifford, D. (2014). DSMART: An algorithm to spatially disaggregate soil map units. In *GlobalSoilMap: Basis of the Global Spatial Soil Information System—Proceedings of the 1st GlobalSoilMap Conference* (pp. 261–266).
45. Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
46. Rahmani, A., Sarmadian, F., & Arefi, H. (2023). Digital modeling and prediction of soil subgroup classes using deep learning approach in a part of arid and semi-arid lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 53(11), 2477-2499. (In Persian with English abstract).
47. Rasaei, Z., Rossiter, D.G., & Farshad, A. (2020). Rescue and renewal of legacy soil resource inventories in Iran as an input to digital soil mapping. *Geoderma Regional*, 21, e00262.
48. Rasaei, Z., Sarmadian, F., & Jafari, A. (2024) Investigating soil grouping using conventional and modern clustering models in some parts of Qazvin plain. *Iranian Journal of Soil and Water Research*, 55 (8), 1273-1295.
49. Rezaie, G., Sarmadian, F., Mohammadi Torkashvand, A., Seyedmohammadi, J., & Marashi Aliabadi, M. (2023). Digital mapping of surface and subsurface soil organic carbon and soil salinity variation in a part of Qazvin plain (Case study: Abyek and Nazarabad regions). *Journal of Water and Soil*, 37(2), 315-331. (In Persian with English abstract).
50. Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing vegetation from soil background information. *Photogrammetric engineering and remote sensing*, 43(12), 1541-1552.
51. Roozitalab, M. H., Siadat, H., & Farshad, A. (Eds.). (2018). *The Soils of Iran*. Springer International Publishing.
52. Rossiter, D.G. (2001). *Assessing the Thematic Accuracy of Area-Class Soil Maps*. Soil Science Division, ITC. Enschede Netherlands. 43p.
53. Soil Survey Staff. 2022. *Keys to Soil Taxonomy*, 13th ed. USDA-Natural Resources Conservation Service.
54. Stoorvogel, J., Mulder, V.L., & Hendriks, C. (2019). The effect of disaggregating soil data for estimating soil hydrological parameters at different scales. *Geoderma*, 347, 185–193.
55. Thompson, J., Prescott, T., Moore, A.C., Bell, J. S., Kautz, D., Hempel, F., Waltman, S., & Perry, C. H. (2010). Regional approach to soil property mapping using legacy data and spatial disaggregation techniques. *Proceedings of 19th World Congress Soil Science, Soil Solutions for a Changing World*, 1-6 August, 17–20.
56. Vincent, S., Lemerrier, B., Berthier, L., & Walter, C. (2018). Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. *Geoderma*, 311, 130–142.

57. Yang, L., Jiao, Y., Fahmy, S., Zhu, A. X., Hann, S., Burt, J. E., & Qi, F. (2011). Updating conventional soil maps through digital soil mapping. *Soil Science Society of America Journal*, 75(3), 1044-1053.
58. Zeraatpisheh, M., Ayoubi, S., Brungard, C.W., & Finke, P. (2019). Disaggregating and updating a legacy soil map using DSMART, fuzzy c-means and k-means clustering algorithms in Central Iran. *Geoderma*, 340, 249–258.

پایان کارشناسی ارشد