



## Wet aggregate stability modeling based on random forest optimized with genetic algorithm

Sanaz Monavvar Sabegh<sup>1</sup>, Davoud Zare Haghi<sup>2</sup>, Saeed Samadianfard<sup>3✉</sup>, Hossein Rezaei<sup>4</sup>

1. Department of Soil Science and Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. Email: [s.monavvar@tabrizu.ac.ir](mailto:s.monavvar@tabrizu.ac.ir)
2. Department of Soil Science and Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. Email: [davoudhaghi@tabrizu.ac.ir](mailto:davoudhaghi@tabrizu.ac.ir)
3. Corresponding Author, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. Email: [s.samadian@tabrizu.ac.ir](mailto:s.samadian@tabrizu.ac.ir)
4. Department of Soil Science and Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. Email: [hosseinrezaei@tabrizu.ac.ir](mailto:hosseinrezaei@tabrizu.ac.ir)

### Article Info

**Article type:** Research Article

**Article history:**

**Received:** May. 13, 2024

**Revised:** June. 1, 2024

**Accepted:** June. 19, 2024

**Published online:** Sep. 2024

**Keywords:**

Genetic Algorithm,  
Random Forest,  
Wet Aggregate Stability.

### ABSTRACT

In order to effectively manage soil and water resources, it is imperative to investigate wet aggregate stability (WAS) as a fundamental indicator for assessing soil structure and quality. In this study, machine learning techniques, specifically random forest (RF) and random forest optimized with genetic algorithm (GA-RF), were employed. The analysis focused on determining the texture, organic matter content, and lime characteristics of 55 soil samples collected from the Arsbaran forests. Utilizing various input combinations based on correlations with WAS, modeling was performed across seven distinct scenarios. Furthermore, three performance metrics including correlation coefficient (CC), normalized root mean square error (NRMSE), and Wilcoxon coefficient (WI) were utilized to evaluate the effectiveness of the models. The findings indicated that the RF5 model exhibited superior performance among the random forest models, achieving NRMSE = 0.038, CC = 0.736, and WI = 0.789. Similarly, the GA-RF5 model, optimized through a genetic algorithm approach, demonstrated exceptional performance with NRMSE = 0.031, CC = 0.800, and WI = 0.842 when considering input percentages of sand, silt, and clay. Moreover, results from RF1 (NRMSE = 0.047, CC = 0.589, WI = 0.721) and GA-RF1 (NRMSE = 0.036, CC = 0.662, WI = 0.797) emphasized that clay content exhibited the strongest correlation with stability. Additionally, the incorporation of calcium carbonate equivalent in scenario 7 significantly enhanced model performance and positively influenced the prediction of wet aggregate stability. In summary, the hybrid model combining random forest with a genetic algorithm is recommended for precise and reliable determination of wet aggregate stability in studies focusing on soil properties.

Cite this article: Monavvar Sabegh, S., Zare Haghi, D., Samadianfard, S., & Rezaei, H., (2024) Wet aggregate stability modeling based on random forest optimized with genetic algorithm, *Iranian Journal of Soil and Water Research*, 55 (7), 1095-1111.

© The Author(s).

Publisher: The University of Tehran Press.



DOI: <https://doi.org/10.22059/ijswr.2024.376443.669712>

## مدل‌سازی پایداری خاکدانه‌های خیس بر اساس جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک

ساناز منور سابق<sup>۱</sup>، داوود زارع حقی<sup>۲</sup>، سعید صمدیان فرد<sup>۳</sup>، حسین رضائی<sup>۴</sup>۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: [s.monavvar@tabrizu.ac.ir](mailto:s.monavvar@tabrizu.ac.ir)۲. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: [davoudhaghi@tabrizu.ac.ir](mailto:davoudhaghi@tabrizu.ac.ir)۳. نویسنده مسئول، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: [s.samadian@tabrizu.ac.ir](mailto:s.samadian@tabrizu.ac.ir)۴. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران. رایانامه: [hosseinrezaei@tabrizu.ac.ir](mailto:hosseinrezaei@tabrizu.ac.ir)

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	مطالعه وضعیت پایداری خاکدانه‌های خیس (WAS)، به‌عنوان شاخصی رایج از ساختمان خاک و نیز ارزیابی کیفیت آن، برای مدیریت بهینه منابع خاک و آب، حائز اهمیت است. در پژوهش حاضر، برای مدل‌سازی پایداری خاکدانه‌های خیس از مدل‌های یادگیری ماشین جنگل تصادفی (RF) و جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک (GA-RF) استفاده شد. بدین منظور، ویژگی‌های بافت، ماده آلی و آهک ۵۵ نمونه خاک از جنگل‌های ارسباران تعیین و سپس با ترکیب‌های ورودی مختلف بر اساس مقادیر همبستگی با پارامتر WAS، مدل‌سازی با استفاده از هفت سناریو انجام شد. به‌منظور تعیین توانایی مدل‌های اجرا شده، سه شاخص عملکرد ضریب همبستگی (CC)، جذر میانگین مربعات خطای نرمال شده (NRMSE) و ضریب وایلموت (WI) مورد استفاده قرار گرفت. نتایج نشان داد که مدل RF5 در بین مدل‌های جنگل تصادفی با $NRMSE = 0.038$ ، $CC = 0.736$ ، $WI = 0.789$ و مدل GA-RF5 در بین مدل‌های جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک با $NRMSE = 0.031$ ، $CC = 0.800$ ، $WI = 0.842$ با ورودی درصد شن و سیلت و رس، بهترین عملکرد را داشتند. علاوه‌براین نتایج RF1 ( $NRMSE = 0.047$ )، $GA-RF1$ ( $NRMSE = 0.036$ )، $CC = 0.589$ ، $WI = 0.721$ ) و $GA-RF1$ ( $NRMSE = 0.036$ )، $CC = 0.662$ ، $WI = 0.797$ ) نشان داد که درصد رس بالاترین درجه همبستگی را با پایداری خاکدانه‌ها دارد. همچنین، با اضافه شدن کربنات کلسیم معادل در سناریو ۷، بهبود عملکرد و تأثیر مثبت این ویژگی در پیش‌بینی پایداری خاکدانه‌های خیس مشاهده گردید. بنابراین، مدل جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک برای تعیین دقیق و مناسب پایداری خاکدانه‌های خیس در مطالعات مربوط به خصوصیات خاک توصیه می‌گردد.
تاریخ دریافت: ۱۴۰۳/۲/۲۴	
تاریخ بازنگری: ۱۴۰۳/۳/۱۲	
تاریخ پذیرش: ۱۴۰۳/۳/۳۰	
تاریخ انتشار: مهر ۱۴۰۳	
واژه‌های کلیدی: الگوریتم ژنتیک، جنگل تصادفی، پایداری خاکدانه‌های خیس.	

استناد: منور سابق؛ ساناز، زارع حقی؛ داوود، صمدیان فرد؛ سعید، رضائی؛ حسین، (۱۴۰۳) مدل‌سازی پایداری خاکدانه‌های خیس بر اساس جنگل تصادفی بهینه‌شده با

الگوریتم ژنتیک، مجله تحقیقات آب و خاک ایران، ۵۵ (۷)، ۱۰۹۵-۱۱۱۱.

ناشر: مؤسسه انتشارات دانشگاه تهران.

© نویسندگان.

DOI: <https://doi.org/10.22059/ijswr.2024.376443.669712>

## مقدمه

ساختمان از جمله ویژگی‌های فیزیکی مهم خاک بوده (Amézketa, 1999) که شاخص مهمی برای مدیریت بهینه منابع خاک و آب است. این شاخص به‌گونه‌ای مستقیم، بر بسیاری از ویژگی‌های فیزیکی خاک همچون مقدار آب، هدایت آبی، گرما، تهویه، جرم مخصوص ظاهری و تخلخل خاک اثر می‌گذارد. همچنین اثرات فراوانی بر فرآیندهای شیمیایی و بیولوژیکی، توانایی خاک در حمایت از رشد گیاه، چرخه کربن، فراهمی ریزمغذی‌ها دارد. به‌طور کلی ساختمان خاک به‌دلیل رابطه مستقیمی که با کارایی و عملکرد خاک دارد، بسیار با اهمیت است. لیکن به‌دلیل پیچیدگی ساختمان خاک، توصیف کمی آن دشوار است (محمدیان خراسانی و همکاران، ۱۳۹۴). پایداری ساختمان خاک به توانایی آن در نگهداری آرایش ذرات و منافذ در هنگام مواجهه با تنش‌های متفاوت مربوط می‌شود (Angers & Carter, 2020). Tongway & Hindley (2004) پایداری ساختمان خاک را توانایی خاک در برابر نیروهای فرساینده و بازسازی آن پس از تخریب تعریف می‌کنند.

خاکدانه به‌عنوان واحد ساختمانی خاک، بیان‌گر مجموعه‌ای از ذرات اولیه است که پیوستگی‌شان به یکدیگر بیش از پیوستگی آن‌ها به ذرات خاک اطراف این مجموعه می‌باشد (بردسیری زاده و همکاران، ۱۳۹۶). خاکدانه‌ها نخست طی فرآیندهای فیزیکی تشکیل می‌شوند و فرآیندهای شیمیایی و زیستی در پایداری آن‌ها مؤثر هستند (Martinez-Mena et al, 2008). پایداری خاکدانه‌ها یکی از شاخص‌های رایج برای مطالعه و بیان وضعیت ساختمان خاک، است (Amézketa, 1999).

پایداری خاکدانه، توانایی خاک در نگهداری و حفظ ترتیب فاز جامد و فضای خلل و فرج بعد از اعمال تنش‌های مکانیکی یا نیروهای مخرب است (Diaz-Zorita et al, 2002). هنگامی که تنش‌های اعمال شده بیش از نیروهای پیوندی باشد، خاکدانه‌های ضعیف شکسته شده و تخریب ساختمان خاک روی می‌دهد (An et al, 2010; Horn et al, 1994).

تخریب ساختمان خاک زمانی شروع می‌شود که اتصال خاکدانه‌های بزرگ‌تر به خاکدانه‌های کوچک‌تر شکسته شوند. این امر منجر به کاهش تخلخل خاک و افزایش جرم مخصوص ظاهری می‌شود. طبق تئوری سلسله مراتبی خاکدانه‌ها، خاکدانه‌های ریز ( $< 0.25$  میلی‌متر) به‌دلیل نیروهای متصل‌کننده قوی‌تر، استحکام ساختاری بیشتری نسبت به خاکدانه‌های درشت ( $> 0.25$  میلی‌متر) دارند و بنابراین کمتر تحت تأثیر قرار می‌گیرند (Liu et al, 2014). پایداری خاکدانه‌های خیس (WAS) مستقیماً تحت تأثیر عوامل مختلف و برهمکنش آن‌ها است. در عین حال تشکیل، تثبیت و تخریب خاکدانه‌های پایدار در آب نیز برخی از پیچیده‌ترین فرآیندهایی است که در خاک رخ می‌دهد و درک کامل آن‌ها را دشوار می‌کند (Are et al, 2018). بهبود درک ما از این فرآیندها برای انجام اقداماتی برای حفظ پایداری خاکدانه‌های درشت و ریز در خاک، اساسی است (Wang et al, 2016).

روش‌های مختلفی برای اندازه‌گیری پایداری خاکدانه وجود دارد که اساس آن‌ها بر خرد شدن نمونه‌های خاکدانه پس از اعمال تنش‌های مکانیکی است (Amézketa, 1999). برای ارزیابی پایداری ساختمان خاک در برابر فرسایش بادی، از روش الک خشک استفاده می‌شود. حال آنکه برای ارزیابی فرسایش آبی، روش الک تر به‌کار گرفته می‌شود (Kemper & Rosenau, 1986). عوامل مؤثر بر پایداری خاکدانه در ارزیابی میزان خاک فرسایش یافته، پتانسیل خاک به پوسته شدن و سله بستن، قابلیت نفوذپذیری خاک، درصد جوانه‌زنی و پیش‌بینی ظرفیت خاک در تولیدات گیاهی مهم است. عوامل مؤثر در تشکیل خاکدانه‌ها به سه دسته عوامل بیولوژیکی، عوامل فیزیکی و ترکیبات خاک تقسیم می‌گردند (Mbagwu, 2004). پایداری خاکدانه‌ها از خصوصیات اصلی مؤثر بر فرسایش‌پذیری و ویژگی‌های اصلی هیدرولیکی و فیزیکی خاک است. آگاهی از مؤلفه‌های خاک کنترل‌کننده پایداری خاکدانه‌ها برای حفظ ساختمان خاک بسیار مهم است. همچنین شناسایی خصوصیات بحرانی خاک مؤثر بر پایداری خاکدانه‌ها در حوزه آبخیزی که فرسایش خاک ناشی از الگوهای طبیعی پایداری خاکدانه‌ها است، پایه و اساس استراتژی‌های حفاظتی است (آرمین و همکاران، ۱۳۹۵).

در هر حال، پایداری خاکدانه‌ها دربردارنده اطلاعات کلیدی در مورد عملکرد و کیفیت خاک است (Chrenková et al, 2014; Wu et al, 2017). امروزه روش‌های غیرمستقیم مانند توابع انتقالی خاک و توابع تخمین برای بررسی پایداری خاک مورد استفاده قرار می‌گیرند. در این توابع برخی ویژگی‌های فیزیکی و مکانیکی دریافت خاک با استفاده از داده‌های سهل‌الوصول مانند توزیع اندازه ذرات خاک، ماده آلی (OM)<sup>۱</sup>، کربنات کلسیم معادل (CCE)<sup>۲</sup>، و ویژگی‌های توپوگرافیکی و پوشش گیاهی تخمین زده می‌شوند (Besalatpour et al,



2012). استفاده از توابع انتقالی در برآورد ویژگی‌های دیریافت خاک از جمله پایداری خاکدانه‌ها، هزینه و زمان لازم برای اندازه‌گیری مستقیم این ویژگی‌ها را کاهش می‌دهد (علی‌جانپور شلمانی و همکاران، ۱۳۹۰).

نتایج تحلیل رگرسیون پژوهش تاجیک (۱۳۸۳) بیان‌گر آن بود که در کرمانشاه و مازندران تغییرات WAS به‌طور عمده متأثر از مقدار ماده آلی (با مقدار ضریب تبیین  $R^2$  ۰/۷۳۲ در سطح احتمال ۰/۰۰۰۱) است درحالی‌که در گلستان مقدار رس و در آذربایجان غربی مقدار شن بیشترین تأثیر را داشته‌اند. در بین ویژگی‌های مؤثر بر پایداری ساختمان خاک در پژوهش خزائی و همکاران (۱۳۸۷) بیشترین نقش مربوط به ماده آلی معرفی شد. به‌گونه‌ای که برای زمان ۵ دقیقه الک کردن خاکدانه‌ها در آب به‌منظور اعمال تنش‌های مکانیکی-آبی، تأثیرگذاری مقدار ماده آلی بر پایداری خاکدانه‌ها که با میانگین وزنی قطر (MWD)<sup>۱</sup> تعیین شد، حدود ۱۸ برابر رس و ۲۶ برابر کربنات کلسیم (CaCO<sub>2</sub>) بود. طبق ارزیابی نیکپور و همکاران (۱۳۹۰) اثر ویژگی‌های ذاتی خاک بر پایداری ساختمان خاک، مقدار ماده آلی خاک بیشترین تأثیر را بر شاخص MWD خاکدانه‌ها با استفاده از روش یودر و شاخص دوت-دلینپیر دارد و پس از ماده آلی، رس، رس ریز و کربنات کلسیم تأثیرگذار بودند. همچنین، نتایج پژوهش بسالت‌پور و همکاران (۱۳۹۴) نشان داد که کارایی مدل‌های مبتنی بر ماشین‌های بردار پشتیبان (SVM)<sup>۲</sup> در تخمین پایداری خاکدانه‌ها از طریق محاسبه میانگین هندسی قطر خاکدانه‌ها (GMD)<sup>۳</sup> بسیار بیشتر از روش رگرسیون خطی چندگانه (MLR)<sup>۴</sup> مرسوم است. مقادیر میانگین مربعات خطا (MSE) و ضریب همبستگی (CC)<sup>۵</sup> مدل SVM طراحی شده به‌ترتیب، برابر ۰/۰۰۵ و ۰/۸۶ و درصد خطا برای تخمین پایداری خاکدانه‌ها برابر ۱۰/۷ بود درحالی‌که مقدار درصد خطا برای مدل رگرسیونی برازش داده شده، حدود ۱۵/۷ بود. نتایج تحقیق آرمن و همکاران (۱۳۹۵) در خاک‌های حوزه آبخیز طالقان با استفاده از ضریب همبستگی و روش گام به گام حاکی از وجود همبستگی منفی پایداری تر خاکدانه‌ها با درصد آهک، رس و رطوبت اشباع و همبستگی مثبت با درصد شن ریز و شن خیلی ریز بود. سعادت و همکاران (۱۳۹۸) براساس یافته‌های مطالعه‌ای با هدف ارزیابی روش‌های مناسب و استاندارد تعیین پایداری خاکدانه که قادر به تشخیص شرایط کیفی فیزیکی خاک‌های دارای بافت متوسط در هر دو منطقه خشک و مرطوب باشد، بیان کردند که آزمون‌هایی مثل کمپر-رزنا با اشباع سریع و همچنین لایزسونایس در حالت مرطوب شدن سریع می‌توانند مورد استفاده واقع شوند.

بر اساس نتایج تحقیق کوچمی ساردو و همکاران (۱۳۹۹) از تحلیل حساسیت، کربنات کلسیم معادل، ذرات شن و ماده آلی به‌عنوان فاکتورهای کلیدی برای پیش‌بینی پایداری خاکدانه‌ها در شهرستان رفسنجان (دشت داوران) معرفی شدند. با توجه به نتایج به‌دست آمده از برآورد شاخص MWD، مقادیر محاسبه شده  $R^2$ ، میانگین درصد خطای مطلق (MAEP)<sup>۶</sup> و جذر میانگین مربعات خطا (RMSE)<sup>۷</sup> برای عملکرد الگوریتم ترکیبی ژنتیک-شبکه عصبی مصنوعی (GA-ANN)<sup>۸</sup> به‌ترتیب برابر با ۰/۹۴، ۲۱/۳۹ و ۰/۰۷۵ درصد بودند. Ben-Hur et al, (1985) نشان دادند که با افزایش میزان کربنات کلسیم در خاک، پایداری خاکدانه‌ها افزایش یافته و در نتیجه از پراکندگی خاکدانه‌ها کاسته می‌شود. Minhas & Sharma (1986) تأثیر مثبت کربنات کلسیم بر پایداری خاکدانه‌ها را گزارش کردند. Besalatpour et al, (2013) در یک مطالعه مقایسه‌ای در حوزه آبخیز بازفت واقع در جنوب غربی ایران، از سه مجموعه مختلف ویژگی‌های سهل‌الوصول به عنوان ورودی استفاده کردند، اولین مجموعه شامل ویژگی‌های خاک از جمله میزان رس، کربنات کلسیم معادل و محتوای مواد آلی خاک بود. مجموعه دوم شامل ویژگی‌های توپوگرافی (شیب و جهت) و تفاوت شاخص گیاهی نرمال شده (NDVI)<sup>۹</sup> و مجموعه سوم ترکیبی از خواص خاک، شیب و NDVI بود. براساس نتایج ایشان، ANN و سیستم استنتاج عصبی فازی تطبیقی (ANFIS)<sup>۱۰</sup> پایداری خاکدانه‌های خیس را دقیق‌تر از مدل‌های خطی تعمیم‌یافته (GLM)<sup>۱۱</sup> و MLR پیش‌بینی کردند. برآورد MWD با استفاده از مجموعه دوم داده‌ها منجر به کمترین مقادیر کارایی مدل‌ها شد. مقادیر کارایی مدل (MEF)<sup>۱۲</sup> مشاهده‌شده برای مدل‌های توسعه‌یافته MLR، GLM، ANN و ANFIS

1. Mean Weight Diameter
2. Support Vector Machines
3. Geometric Mean Diameter
4. Multiple Linear Regression
5. Correlation Coefficient
6. Mean Absolute Percentage Error
7. Root Mean Square Error
8. Genetic Algorithm-Artificial Neural Networks
9. Normalized Difference Vegetation Index
10. Adaptive Neuro-Fuzzy Inference System
11. Generalized Linear Model
12. Model Efficiency Factor

با استفاده از اولین مجموعه داده ها به ترتیب ۶۰/۷۶، ۶۲/۹۸، ۷۷/۶۸ و ۷۷/۱۵ بود. نتایج مطالعه (Besalatpour et al, 2014) نشان داد مدل ANN در پیش بینی GMD در مقایسه با مدل های MLR و SVM دقت بیشتری دارد. مقدار خطای به دست آمده در پیش بینی GMD با استفاده از مدل ANN ۶/۹ درصد بود در حالی که برای مدل های MLR و SVM به ترتیب ۱۵/۷ و ۱۰/۶ درصد بود. انتخاب ویژگی با استفاده از PGA دقت پیش بینی تمام تکنیک های بررسی شده را بهبود بخشید. در نتیجه، مدل ANN مبتنی بر PGA می تواند به عنوان جایگزینی برای مدل های رگرسیون معمولی برای پیش بینی GMD در نظر گرفته شود. (Bhattacharya et al, 2021) با هدف مقایسه دو رویکرد یادگیری ماشین مبتنی بر هوش مصنوعی (AI) در پیش بینی پایداری خاکدانه مرطوب نشان دادند ANN عملکرد بهتری در پیش بینی MWD نسبت به MLR دارد، در حالی که SVM بهترین مدل با کمترین مقادیر RMSE (۰/۱۰۳ و ۰/۱۲ برای آموزش و آزمایش) بود. (Bouslihim et al, 2021) با هدف مقایسه MLR و جنگل تصادفی (RF) در پیش بینی MWD عملکرد قابل قبول مدل RF شامل خصوصیات خاک به تنهایی با  $R^2 = ۰/۶$  و  $RMSE = ۰/۲۶۱$  را نشان دادند و افزودن شاخص های سنجش از دور به خواص خاک، سبب افزایش دقت نشد. در این بررسی، ماده آلی مرتبط ترین متغیر برای پیش بینی در هر دو مدل شناسایی شد. (Zeraatpisheh et al, 2021) شاخص های پایداری خاکدانه ها را با استفاده از نقشه برداری دیجیتال خاک و مدل های ML با استفاده از تغییرات محیطی به عنوان رویکردهای مقرون به صرفه، پیش بینی کردند. مدل های ML از جمله RF، k-نزدیک ترین همسایه (kNN)، SVM، ANN ارزیابی شدند. نتایج مدل سازی نشان داد که مدل RF برای MWD ( $R^2 = ۰/۷۴$ ،  $NRMSE = ۲۴/۲۸$ )، برای GMD ( $R^2 = ۰/۷۵$ ،  $NRMSE = ۱۲/۷۲$ ) و برای WAS ( $R^2 = ۰/۵۸$ ،  $NRMSE = ۱۰/۴۰$ ) عملکرد بهتری داشتند. پژوهش (Zhai et al, 2022) با مقایسه ضریب تبیین و جذر میانگین مربعات خطا نشان داد که روش ماشین بردار پشتیبان نسبت به روش شبکه عصبی مصنوعی در مدل سازی پایداری خاکدانه های خیس برتری دارد. علاوه بر این، تجزیه و تحلیل اهمیت نسبی نشان می دهد که کربن آلی، سیلت و رس خاک پارامترهای ورودی اصلی هستند. با توجه به زمان بر و پرهزینه بودن تعیین مقدار WAS با استفاده از روش های آزمایشگاهی، پیش بینی آن به طور غیرمستقیم از داده های در دسترس تر، سودمند خواهد بود. تکنیک های یادگیری ماشین (ML) را می توان رویکردی کم هزینه با تحمل عدم دقت، عدم قطعیت، تقریب و جلوگیری از مشکلات بیش برآزش دانست. این روش با توجه به قابلیت خود در تجزیه و تحلیل داده ها در مقیاس بزرگ و در نتیجه حل مسائلی که روش های خطی مرسوم تاکنون قادر به حل آن ها به طور مقرون به صرفه و رضایت بخشی نبوده اند، بسیار کارا می باشد (Chau et al, 2005; Wang et al, 2009; Huang et al, 2010). هدف از پژوهش حاضر مدل سازی پایداری خاکدانه های خیس با استفاده از روش ML جنگل تصادفی در حالت منفرد و ادغام آن با الگوریتم ژنتیک جهت بهینه سازی پارامترهای RF می باشد که وجه تمایز پژوهش حاضر در زمینه پیش بینی WAS است.

## روش شناسی پژوهش

### منطقه مورد مطالعه و تجزیه های خاک

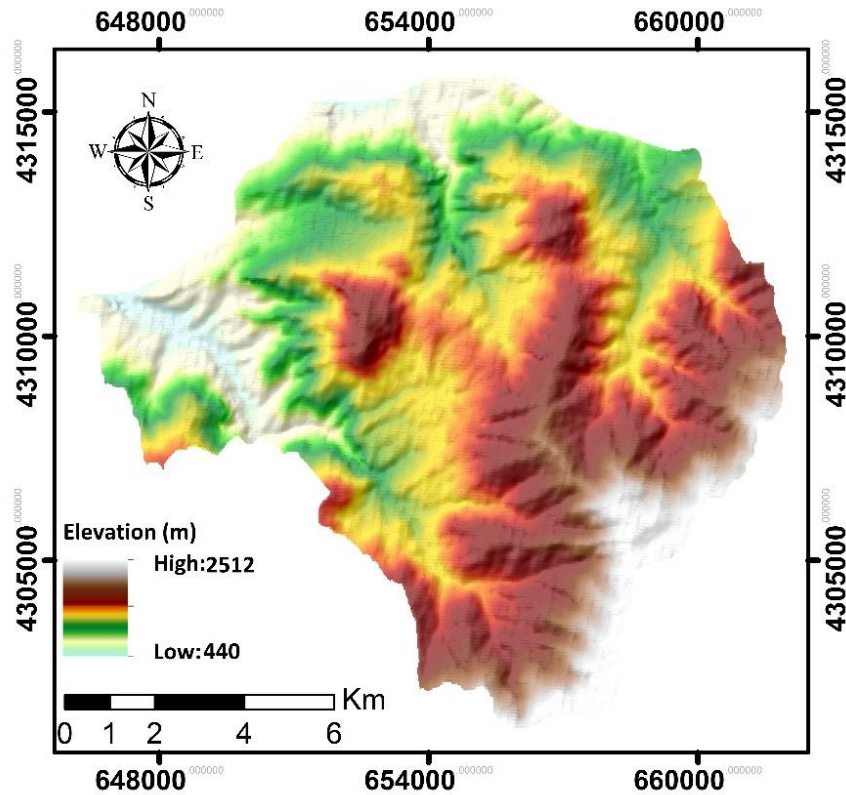
با توجه به هدف تحقیق و به منظور حذف اثرات دخالت انسانی در پایداری خاکدانه ها و با توجه به بکر بودن خاک های جنگلی بخشی از اراضی جنگلی منطقه ارسباران با مختصات موصوف در شکل ۱ به عنوان محدوده مورد مطالعه انتخاب شد. طبق گزارش های موجود در این منطقه که تحت توده های جنگلی بلوط و ممرز است، خانواده هایی از خاک های انتی سول، اینسپتی سول، مالی سول و آلفی سول با تنوعی از مقادیر و شکل های ماده آلی، بافت و آهک مشاهده می شود (منابع طبیعی آذربایجان شرقی، ۱۳۸۲؛ رضائی و همکاران، ۱۳۹۵؛ رضائی و همکاران، ۱۳۹۹). همچنین این منطقه با فیزیوگرافی کوهستان دارای اقلیم مدیترانه ای نیمه مرطوب معتدل است (سازمان هواشناسی استان آذربایجان شرقی، ۱۴۰۰).

در راستای انجام مطالعه ۵۵ نمونه از خاک سطحی محدوده مورد مطالعه واقع در شرایط مختلف محیطی تهیه و پس از انتقال به آزمایشگاه، اجزای بافت خاک شامل در صد شن، سیلت و رس به روش هیدرومتری (Gee & Bauder 1986) درصد ماده آلی به روش اکسیداسیون تر (Nelson & Sommers 1996) و مقادیر کربنات کلسیم معادل به روش تیتراسیون برگشتی (Allison & Moodie 1965) و پایداری خاکدانه های خیس طبق رابطه ۱ ارائه شده توسط (kemper & rosenau 1986) تعیین شدند.

1. Parallel genetic algorithm
2. Random forest
3. K-nearest neighbors
4. Machine learning

$$WAS\% = \frac{W_{ds} - W_{sand}}{W_{dry} - W_{sand}} \times 100\% \quad \text{رابطه (۱)}$$

که در آن  $W_{ds}$  جرم خاک متلاشی شده در محلول،  $W_s$  جرم ذرات شن و  $W_{dry}$  جرم خاکدانه‌های اولیه هوا خشک شده (۱-۲ میلی‌متر) است.



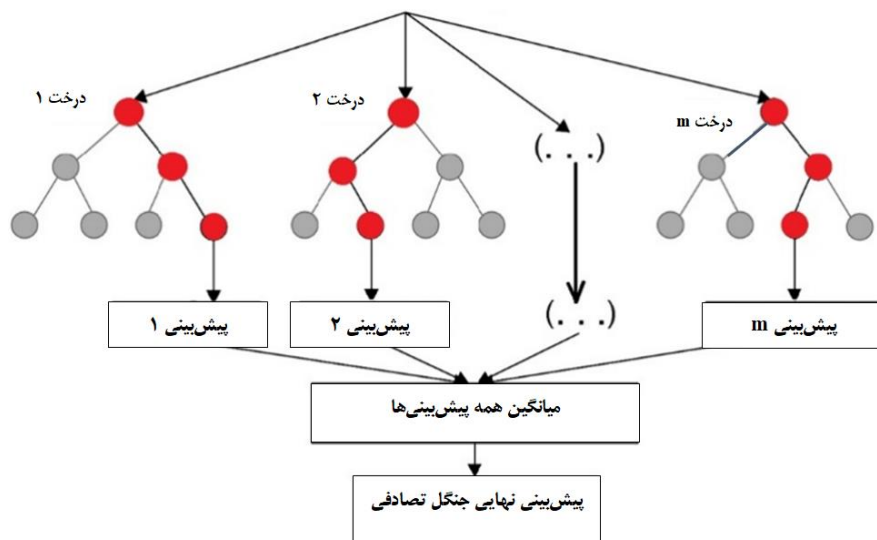
شکل ۱. نقشه سه بعدی مدل رقومی ارتفاع محدوده مورد مطالعه

## مدل‌سازی پایداری خاک

### جنگل تصادفی

طبقه‌بندی کننده RF از درختان تصمیم‌گیری متعدد (ntree) استفاده می‌کند که از نمونه‌های بوت استرپ (۷۰ درصد) از کل جمعیت نمونه ( $n$ ) رشد می‌کنند (Breiman, 2001). نمونه‌برداری بوت استرپ باعث می‌شود، RF در مقایسه با درخت‌های تصمیم‌گیری حساسیت کمتری نسبت به بیش‌برازش داشته باشد. طی این روش در هر تقسیم دودویی، پیش‌بینی‌کننده‌ای که بهترین تقسیم را ایجاد می‌کند از یک زیرمجموعه تصادفی  $m_{try}$ ، از کل مجموعه پیش‌بینی‌کننده انتخاب می‌شود. تعداد پیش‌بینی‌کننده‌های آزمایش شده در هر تقسیم  $m_{try}$  توسط کاربر تعریف می‌گردد. در نتیجه،  $m_{try}$  به‌عنوان پارامتر اصلی تنظیم RF شناخته می‌شود و بنابراین باید بهینه‌سازی (Svetnik et al, 2003, 2004) روند رشد درخت به‌صورت بازگشتی انجام شود تا اندازه گره به حداقل  $K$  برسد (Hastie et al, 2009). نمونه  $X_i$  از ۳۰ درصد باقی مجموعه داده آموزشی که در رشد درخت تصمیم استفاده نشده‌اند (نمونه‌های خارج از کیسه (OOB))، وارد درخت تصمیم می‌شود و یک کلاس پیش‌بینی شده به هر نمونه OOB اختصاص داده می‌شود ( $YOOB(X_i)$ ). خروجی حاصل از RF یک مدل واحد می‌باشد و نرخ خطای کلی ( $ER_{OOB}$ )، با استفاده از رابطه ۲ به دست می‌آید.

$$ER_{OOB} = n^{-1} \sum_{i=1}^n I[Y_{OOB}(X_i) \neq Y_i] \quad \text{رابطه (۲)}$$



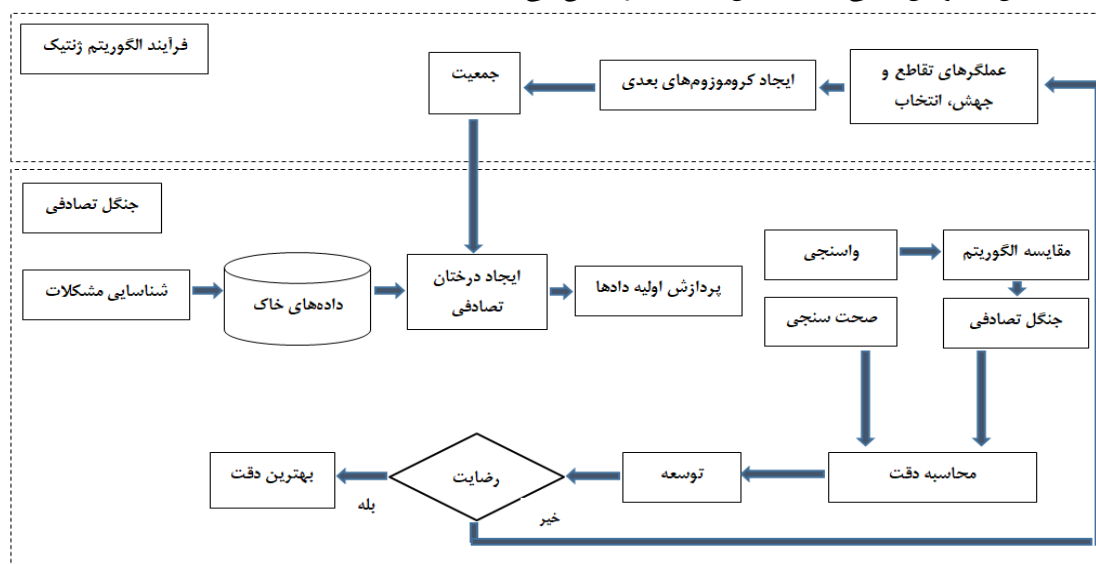
شکل ۲. روند کار RF (Zeini et al, 2023)

### الگوریتم ژنتیک

GA رویکرد اکتشافی و بهینه سازی جستجو است که از تکامل طبیعی الهام گرفته شده است. در GA، راه حل های پیشنهادی مسائل عمدتاً به صورت دودویی خطی به صورت رشته هایی از ۰ و ۱ کدگذاری می شوند. این رشته ها با طول های ثابت کروموزوم نامیده می شوند. در واقع، هر کروموزوم به عنوان یک راه حل خاص بر اساس تابع تناسب برای تعیین مناسب بودن کروموزوم ارزیابی در نظر گرفته می شود (Kouchami-Sardoo et al, 2019). عملگر تقاطع ویژگی های والدین انتخاب شده برای ایجاد دو فرزند را ترکیب می کند. عملگر جهش یک یا چند جزء از فرد انتخاب شده را تغییر می دهد تا از هر گونه رکودی که ممکن است در طول فرآیند جستجو رخ دهد جلوگیری کند. پس از چند نسل در تکامل، زمانی که معیار توقف برآورده شد، افرادی که در جمعیت زنده مانده اند، راه حل های بهینه در نظر گرفته می شوند (Alqahtani et al, 2019).

### مدل RF مبتنی بر GA

حداقل تعداد نمونه ها برای هر تقسیم و تعداد درختان در جنگل دو پارامتر کلیدی RF هستند که می توانند بر دقت طبقه بندی تأثیر بگذارند. بنابراین انتخاب پارامتر بهینه یک مشکل واقعی در طبقه بندی مبتنی بر RF می باشد. ایده پشت مدل RF مبتنی بر GA، بهینه سازی طبقه بندی کننده RF با انتخاب مقادیر پارامترهای مناسب و بهبود نرخ پیش بینی WAS با استفاده از RF بهینه شده است. GA می تواند با تولید مقادیر تصادفی برای پارامترهای خاص RF، یک مرز تصمیم گیری جدید ایجاد کند که دارای بالاترین مقدار تابع تناسب GA است (Assiri, 2020). شکل ۳ مراحل اصلی ساخت مدل GA-RF را نشان می دهد.



شکل ۳. نمودار مدل GA-RF (Ramdhani et al, 2023)

### توابع عملکرد

در این مطالعه مقادیر خطای بین روش‌های محاسباتی توسط پارامترهای آماری جذر میانگین مربعات خطای نرمال شده (NRMSE)<sup>۱</sup>، ضریب همبستگی (CC) و ضریب ویلموت (WI)<sup>۲</sup> و با استفاده از روابط ۳ تا ۵ مورد بررسی قرار گرفت.

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}}{\bar{O}_i} \quad \text{رابطه ۳}$$

$$CC = \frac{(\sum_{i=1}^n O_i P_i - \frac{1}{n} \sum_{i=1}^n O_i \sum_{i=1}^n P_i)}{(\sum_{i=1}^n O_i^2 - \frac{1}{n} (\sum_{i=1}^n O_i)^2) (\sum_{i=1}^n P_i^2 - \frac{1}{n} (\sum_{i=1}^n P_i)^2)} \quad \text{رابطه ۴}$$

$$WI = 1 - \left[ \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \right] \quad \text{رابطه ۵}$$

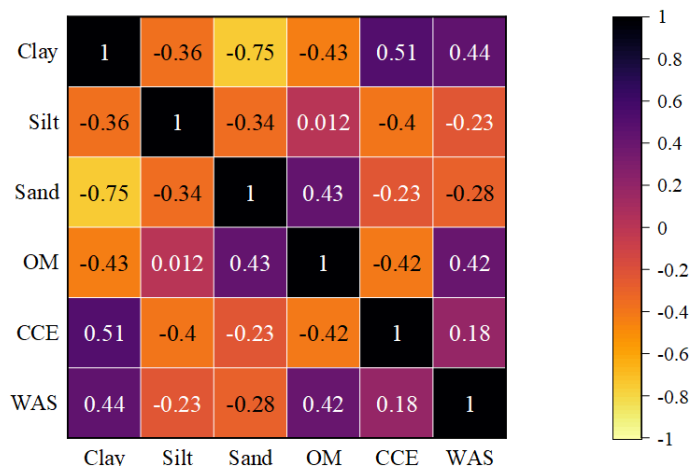
در روابط فوق، n تعداد داده‌های مشاهده شده،  $P_i$  مقادیر پیش‌بینی شده،  $O_i$  مقادیر مشاهده شده و  $\bar{O}_i$  میانگین مقادیر مشاهده شده می‌باشند.

### یافته‌های پژوهش

با هدف پیش‌بینی WAS، ویژگی‌های مدنظر در اجرای تحقیق حاضر از خاک‌های مورد مطالعه شامل درصد شن، سیلت، رس، OM و CCE تعیین شد که توصیف آماری آن‌ها در جدول ۱ ارائه شده است. با بررسی ضرایب همبستگی بین ویژگی‌های خاک و WAS حاصل از تجزیه‌های آزمایشگاهی که در شکل ۴ با نمودار نقشه حرارتی ارائه شده است، مشاهده شد که ویژگی‌های انتخابی رابطه‌ای قوی با متغیر هدف دارند.

جدول ۱. ویژگی‌های آماری داده‌های مورد استفاده برای مرحله واسنجی

پارامتر	واحد	اختصار	بیشترین	کمترین	میانگین	خطای استاندارد	انحراف معیار
رس	%	C	۶۲/۵۰	۵	۳۰/۷۳	۱۲/۷۹	۰/۴۲
سیلت	%	Si	۵۷/۵۰	۲/۵۰	۲۶/۷۳	۸/۹۵	۰/۳۴
شن	%	S	۷۰	۱۰	۴۲/۵۵	۱۲/۶۹	۰/۳۰
ماده آلی	%	OM	۳۳/۵۰	۱/۶۸	۱۳/۶۴	۷/۹۹	۰/۵۹
کربنات کلسیم معادل	%	CCE	۵۷/۷۵	-۰/۵۰	۲۲/۴۸	۱۶/۷۸	۰/۷۵
پایداری خاکدانه	%	WAS	۹۷/۸۹	۶۶/۸۱	۸۹/۸۴	۶/۲۴	۰/۰۷



شکل ۴. نمودار نقشه حرارت ضرایب همبستگی بین ویژگی‌های خاک و WAS



در ادامه ۷ سناریوی پیش‌بینی مقدار WAS با استفاده از ویژگی‌های خاک برای مدل‌سازی طبق جدول ۲ مورد بررسی قرار گرفت. تکمیل تحقیق با مدل‌های RF و GA-RF با ترکیب ورودی‌های مختلف برای پیش‌بینی WAS انجام شد. برای این منظور ۷۰ درصد داده‌ها برای واسنجی و ۳۰ درصد برای صحت‌سنجی مدل‌های مذکور به کار رفت. جدول ۲ پارامترهای به کار رفته برای توسعه مدل‌ها یعنی تعداد درختان، عمق حداکثر، اطمینان، حداقل اندازه برگ، حداقل اندازه برای تقسیم، تعداد جایگزین‌های پیش‌هرس و نسبت زیر مجموعه را نشان می‌دهد.

جدول ۲. پارامترهای ورودی سناریوهای پیش‌بینی WAS

سناریو		پارامترهای ورودی	
۱	C		
۲	OM		
۳	C	OM	
۴	C	OM	S
۵	C	Si	S
۶	C	OM	Si
۷	C	OM	Si

جدول ۳. پارامترهای مدل‌های RF و GA-RF

مدل	تعداد درختان	عمق حداکثر	اطمینان	پارامتر			نسبت زیرمجموعه
				حداقل اندازه برگ	حداقل اندازه برای تقسیم	تعداد جایگزین‌های پیش‌هرس	
RF	۱۰۰	۱۰	۰/۱۰۰	۲	۴	۳	۰/۲۰۰
GA-RF1	۲۷	۲۷	۰/۳۷۱	۴۱	۲۰	۳	۰/۸۸۹
GA-RF2	۸۱	۵	۰/۳۷۰	۴۱	۲۰	۵۵	۰/۱۲۶
GA-RF3	۶۳	۸۶	۰/۲۸۲	۴۱	۲۰	۳	۰/۹۶۸
GA-RF4	۹۴	۵	۰/۳۶۷	۴۱	۲۰	۳	۰/۱۵۵
GA-RF5	۲۷	۲۷	۰/۳۸۲	۵۵	۵۱	۷۷	۰/۹۰۱
GA-RF6	۹۴	۵	۰/۳۴۷	۹۸	۱۳	۶۸	۰/۱۴۲
GA-RF7	۲۷	۲۷	۰/۳۸۴	۵۵	۵۱	۷۷	۰/۸۸۸

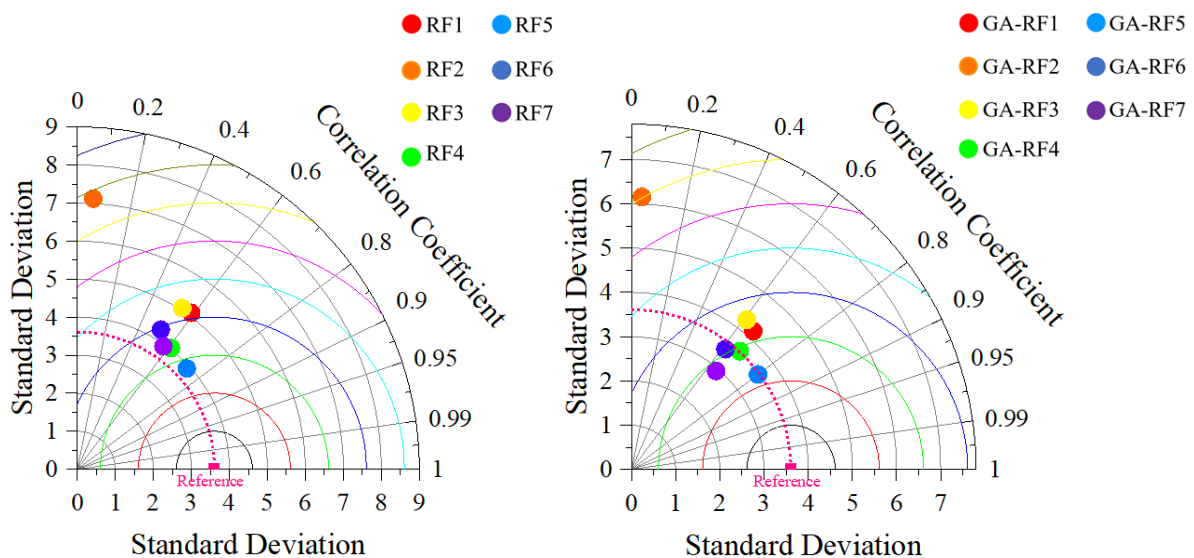
با توجه به مدل‌سازی‌های صورت گرفته بر اساس هر سناریو، نتایج معیارهای ارزیابی برای مدل‌های RF و GA-RF در جدول ۴ ارائه شده است. مدل RF5 در بین مدل‌های جنگل تصادفی با پارامترهای NRMSE برابر با ۰/۰۳۸، CC معادل با ۰/۸ و WI برابر با ۰/۷۸۹ و مدل GA-RF5 در بین مدل‌های جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک با مقادیر پارامترهای یادشده به ترتیب برابر با ۰/۰۳۱، ۰/۸۰۰، ۰/۸۴۲ بهترین عملکرد را داشتند. در دیاگرام‌های تیلور (شکل ۵) نیز فاصله شعاعی کمتر سناریو ۵ در هر دو حالت RF، از نقطه مرجع قابل مشاهده است. لازم به ذکر است در این دیاگرام‌ها، فاصله نشان‌دهنده مقدار جذر میانگین مربعات خطای روش‌های مورد مطالعه می‌باشد. مقایسه نتایج مدل‌های RF5 و GA-RF5 حاکی از آن است که GA سبب بهبود عملکرد RF در پیش‌بینی پایداری خاکدانه‌ها شده است. زیرا مقدار R و WI به ترتیب ۸ و ۶/۷۲ درصد افزایش داشتند و NRMSE ۱۸/۴۲ درصد کاهش یافت. لازم به یادآوری است که سناریو ۵ که به‌عنوان گزینه برتر شناخته شده، مدلی است که بر مبنای محتوای ذرات شن و سیلت و رس اجرا شده است.

با توجه به نمودارهای راداری اعتبارسنجی مدل‌های ارائه‌شده در شکل ۶ (RF1، NRMSE = ۰/۰۴۷، CC = ۰/۵۸۹، WI = ۰/۷۲۱) و GA-RF1 (NRMSE = ۰/۰۳۶، CC = ۰/۶۶۲، WI = ۰/۷۹۷) می‌توان چنین استنباط نمود که درصد رس مهم‌ترین ویژگی سهل‌الوصول خاک برای پیش‌بینی پایداری خاکدانه‌ها می‌باشد که می‌تواند به‌عنوان یک ماده سیمانی عمل کند که ذرات را در خاکدانه کنار هم نگه می‌دارد (Boix-Fayos et al, 2001). تحقیقات (Lado et al, 2004) نشان داد که پایداری خاکدانه با افزایش میزان رس در محدوده ۹ تا ۶۳ درصد، افزایش می‌یابد. در خاک‌های مورد آزمایش، محدوده محتوای رس ۵ تا ۶۲/۵ درصد است و نتایج تأیید می‌کند که رس خاک عامل مهمی برای خاکدانه‌سازی و پایداری خاکدانه‌ها می‌باشد. همچنین Alekseeva et al, (2009) در خاک‌های نواحی حاره و نیمه‌حاره وجود بافت ریز و رس غالب کائولینیت را عامل اصلی خاکدانه‌سازی گزارش کرده‌اند. علاوه بر این، از نتایج آماری مدل‌های مربوط به

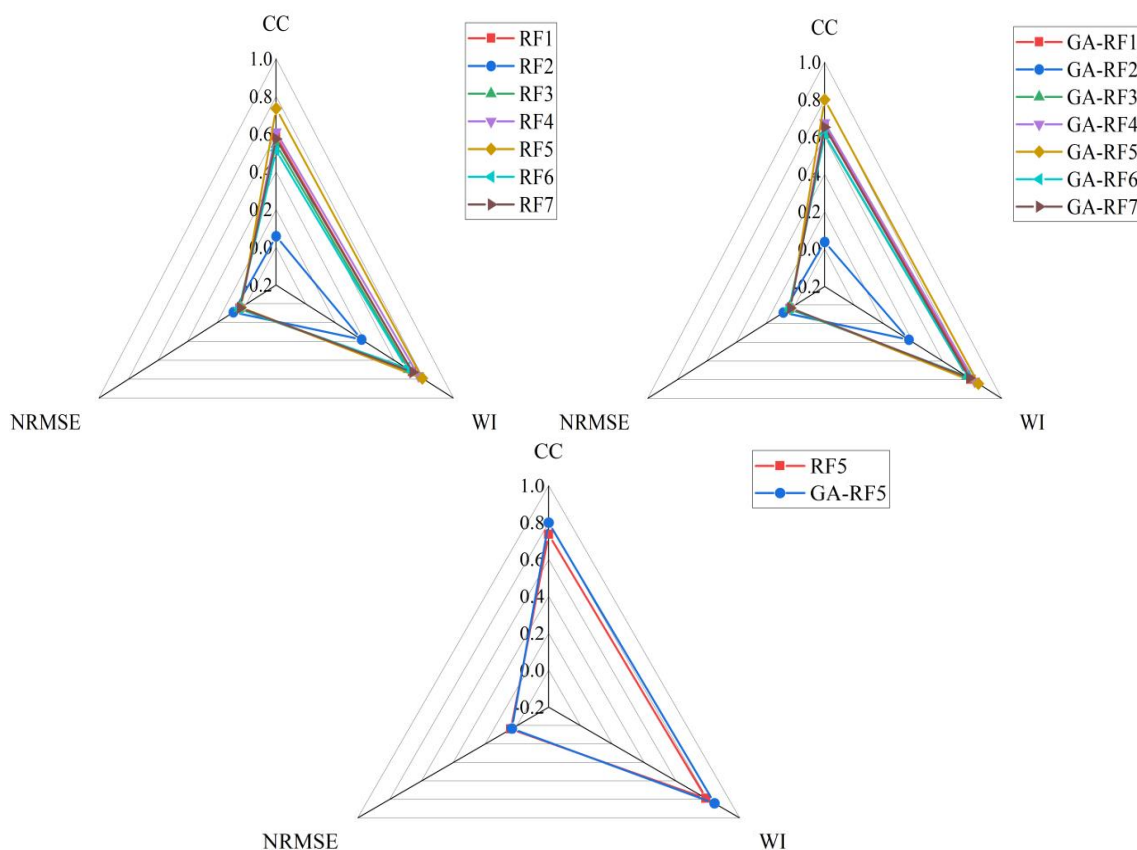
سناریو ۲ می‌توان نتیجه گرفت که در خاک‌های مورد بررسی، ماده آلی در خاکدانه‌سازی تأثیر اندکی داشته است. (Falsone et al, (2006 نیز نتایج مبنی بر عدم تأثیر ماده آلی بر پایداری خاکدانه‌های درشت را ارائه نموده‌اند. نتایج مشابه از همبستگی بالای بین پایداری خاکدانه‌ها و میزان رس و ارتباط آن با OM در سطوح پایین توسط Alaboz et al, (2022 نیز گزارش شده است. همچنین نتایج سناریو ۷ نشان می‌دهد اضافه شدن CCE به پارامترهای ورودی در مدل‌های RF7 ( $WI = 0.724$ ,  $CC = 0.575$ ,  $NRMSE = 0.43$ ) و GA-RF7 ( $WI = 0.780$ ,  $CC = 0.652$ ,  $NRMSE = 0.33$ ) تأثیر مثبتی در پیش‌بینی پایداری خاکدانه‌ها دارد، چراکه سبب کاهش  $10/42$  و  $10/81$  درصدی NRMSE نسبت به RF6 و GA-RF6 شده است. در این زمینه Bouajila & Gallal (2008) نشان دادند که حضور آهک در خاک بر میزان فعالیت ماده آلی در جهت تغییرات پایداری خاکدانه مؤثر است بدین نحو که با افزایش درصد آهک خاک اثر ماده آلی در پایداری کم‌رنگ می‌شود.

جدول ۴. نتایج کلی مدل‌ها

مدل	پارامترهای آماری		
	NRMSE	WI	CC
RF1	0.47	0.721	0.589
RF2	0.89	0.379	0.59
RF3	0.51	0.967	0.546
RF4	0.39	0.764	0.612
RF5	0.38	0.789	0.736
RF6	0.48	0.682	0.514
RF7	0.43	0.724	0.575
GA-RF1	0.36	0.797	0.662
GA-RF2	0.80	0.371	0.38
GA-RF3	0.41	0.763	0.611
GA-RF4	0.33	0.816	0.676
GA-RF5	0.31	0.842	0.800
GA-RF6	0.37	0.765	0.616
GA-RF7	0.33	0.780	0.652



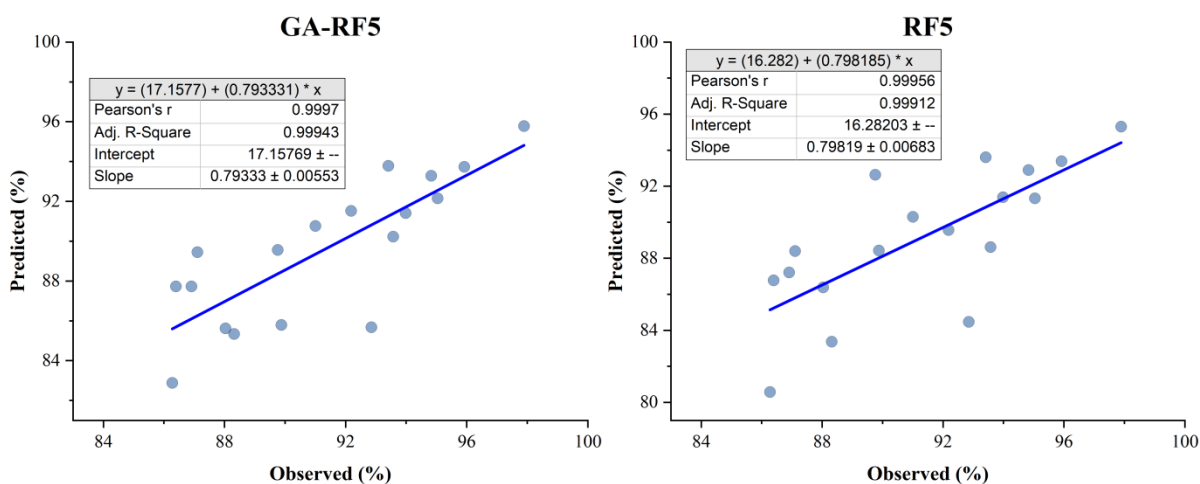
شکل ۵. دیاگرام‌های تیلور مدل‌ها



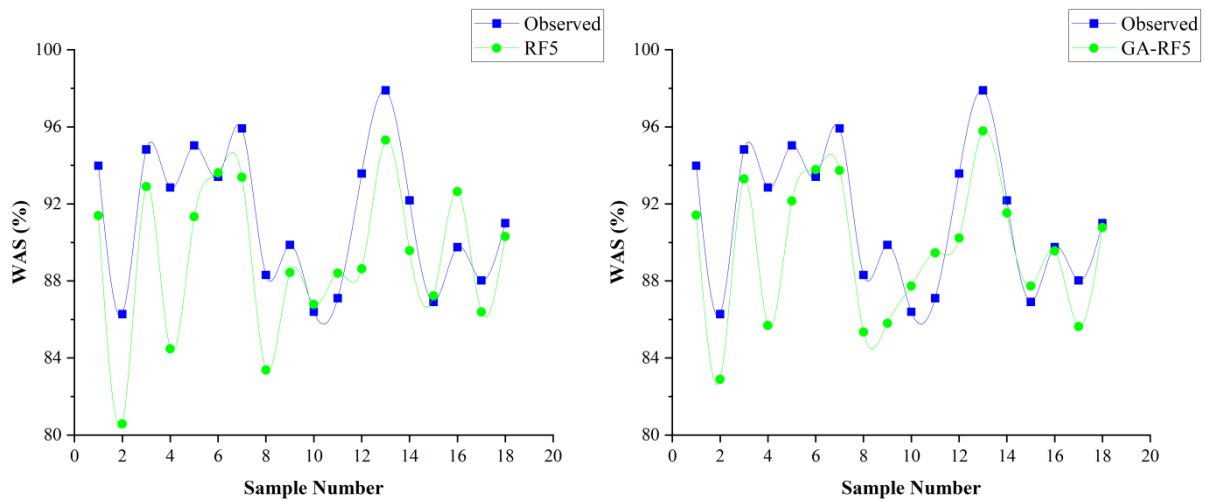
شکل ۶. نمودارهای راداری اعتبارسنجی مدل ها

شکل ۷ نمودارهای پراکنش مقادیر WAS مشاهداتی و پیش بینی شده از مدل های RF5 و GA-RF5 را نشان می دهد که بهترین نتایج آماری را داشتند. مقایسه مقادیر WAS مشاهداتی و پیش بینی شده از این مدل ها در شکل ۸ ارائه شده است که نشان می دهد ادغام RF با GA سبب بهبود دقت مدل سازی شده است.

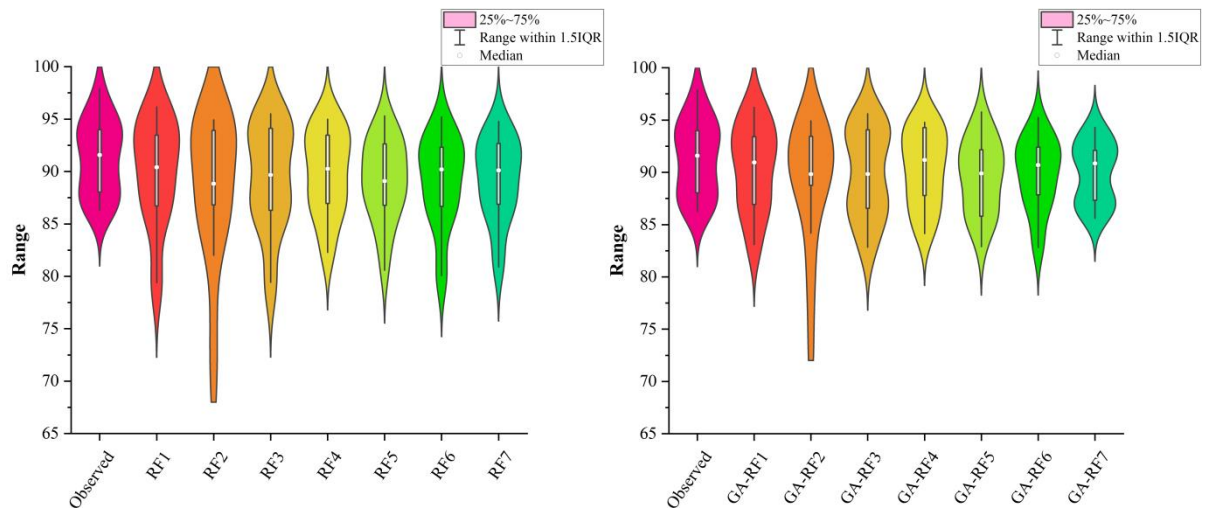
شکل ۹ عملکرد مدل های مورد مطالعه را با استفاده از نمودارهای ویولن و جعبه ای نشان می دهد. با توجه به پراکندگی و دامنه توزیع داده ها مدل های جنگل تصادفی بهینه شده نتایج دقیق تری را در پیش بینی WAS نسبت به مدل های جنگل تصادفی منفرد ارائه می دهند. توزیع داده های سناریوهای ۴، ۵ و ۷ پیش بینی های GA-RF بیشترین تشابه را با مقادیر مشاهده شده دارند. همچنین میانه این مدل ها تقریباً مطابق میانه پایداری خاکدانه های مشاهده شده می باشند.



شکل ۷. پراکنش مقادیر WAS مشاهداتی و پیش بینی شده از مدل های RF5 و GA-RF5



شکل ۸. مقایسه مقادیر WAS مشاهداتی و پیش‌بینی شده از مدل‌های RF5 و GA-RF5



شکل ۹. نمودارهای ویولن و جعبه‌ای مدل‌ها

## نتیجه‌گیری

پایداری خاکدانه‌ها یک ویژگی اساسی است که بر فرسایش‌پذیری خاک و ویژگی‌های هیدرولیکی تأثیر می‌گذارد. یک رویکرد جایگزین برای اندازه‌گیری فیزیکی این ویژگی، برآورد WAS بر اساس پارامترهای سهل‌الوصول خاک شامل درصد ذرات شن، سیلت و رس، ماده آلی و کربنات کلسیم معادل با استفاده از مدل‌سازی است. نتایج حاصل از تحلیل‌های آماری نشان داد مدل‌های RF5 و GA-RF5 شامل ورودی‌های رس، سیلت و شن، بهترین عملکردها را داشتند. مقایسه نتایج این دو مدل حاکی از آن است که GA سبب بهبود عملکرد RF در پیش‌بینی پایداری خاکدانه‌ها شده است. همچنین، با توجه به نتایج به‌دست آمده از مدل‌های RF1 ( $CC = 0/589$ ,  $NRMSE = 0/047$ )، RF2 ( $WI = 0/721$  و  $GA-RF1$  ( $CC = 0/662$ ,  $NRMSE = 0/036$ ) و  $GA-RF7$  ( $WI = 0/797$ ،  $CC = 0/662$ ،  $NRMSE = 0/036$ ) اهمیت مقدار رس در این خاک‌ها آشکار است که سایر فرآیندها و عوامل خاکدانه‌سازی را پوشش می‌دهد. بررسی نتایج سناریوهای ۶ و ۷ کاهش ۱۰/۴۳ و ۱۰/۸۱ درصدی NRMSE را در دو حالت منفرد و بهینه‌شده نشان داد که بیانگر تأثیر مثبت آهک در پیش‌بینی WAS بود. به‌طور کلی استفاده از روش یادگیری ماشین GA-RF دقت قابل قبول برای پیش‌بینی WAS از ویژگی‌های خاک ارائه نمود و در این پژوهش درصد رس به‌عنوان پارامتر خاک کلیدی در این امر تعیین شد.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

## منابع

- آرمین، محسن؛ روحی پور، حسن؛ احمدی، حسن؛ سلاجقه، علی؛ مهدیان، محمدحسین و قربان نیا خیبری، وجیهه. (۱۳۹۵). روابط بین پایداری خاکدانه‌ها و برخی خصوصیات خاک در خاک‌های حوزه آبخیز طالقان. *مرتع و آبخیزداری (منابع طبیعی ایران)*, ۲۹(۲), ۲۷۵-۲۹۵.
- بردسیری زاده، سعید؛ اسفندیارپور بروجنی، عیسی؛ بسالت پور، علی اصغر و عباس زاده دهجی، پیمان. (۱۳۹۶). تعیین موثرترین جزء خاکدانه در برآورد پایداری ساختمان خاک با بهره‌گیری از روش زمین‌آمار. *آب و خاک (علوم و صنایع کشاورزی)*, ۵۴(۲), ۵۳۳-۵۴۴.
- بسالت پور، علی اصغر، شیرانی، حسین، و اسفندیارپور بروجنی، عیسی. (۱۳۹۴). مدل سازی پایداری خاکدانه‌ها با استفاده از ماشین‌های بردار پشتیبان و رگرسیون خطی چند متغیره. *آب و خاک (علوم و صنایع کشاورزی)*, ۴۱(۲), ۴۰۶-۴۱۷.
- تاجیک، فواد. (۱۳۸۳). ارزیابی پایداری خاکدانه‌ها در برخی مناطق ایران. *علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)*, ۱۲۴-۱۰۷, ۸(۱).
- خزایی، علی؛ مصدقی، محمدرضا و محبوبی، علی اکبر. (۱۳۸۷). ارزیابی پایداری ساختمان در ۲۱ سری از خاک‌های استان همدان به روش الگ تر و رابطه آن با برخی ویژگی‌های ذاتی خاک. *پژوهش کشاورزی*, ۱۷۱-۱۸۱, ۸(الف).
- رضایی، حسین، جعفرزاده، علی اصغر، علیجانپور، احمد، شهبازی، فرزین، و ولی زاده کامران، خلیل. (۱۳۹۹). وضعیت ماده آلی خاک در توده های جنگلی ارسباران. *آب و خاک (علوم و صنایع کشاورزی)*, ۱۲۷-۱۱۵, ۳۴(۱).
- رضائی، حسین؛ جعفرزاده، علی اصغر؛ علیجانپور، احمد؛ شهبازی، فرزین و ولیزاده کامران، خلیل. (۱۳۹۵). تکامل ژنتیکی خاک‌های جنگلی ارسباران در امتداد یک نیمرخ ارتفاعی زیر حوضه کلیبرچای سفلی. *دانش آب و خاک*, ۱۶۱-۱۵۵, ۲۶(۴,۱).
- سازمان هواشناسی استان آذربایجان شرقی (۱۴۰۰). آمار هواشناسی ایستگاه سینوپتیک کلیبر. سازمان هواشناسی جمهوری اسلامی ایران، تهران.
- سعادت، سعید؛ اسمعیل نژاد، لیلا؛ رضائی، حامد؛ میرخانی، رسول و سیدمحمدی، جواد. (۱۳۹۸). مقایسه روش‌های ارزیابی پایداری خاکدانه به عنوان یکی از شاخص‌های کیفیت فیزیکی خاک. *آب و خاک*, ۳۰۳-۲۸۹, ۳۳(۲).
- علی جانپور شلمانی، عادل، شعبان پور، محمود، اسدی، حسین، و باقری، فرید. (۱۳۹۰). تخمین پایداری خاکدانه در خاک‌های جنگلی استان گیلان بوسیله شبکه عصبی مصنوعی و توابع انتقالی رگرسیونی. *دانش آب و خاک (دانش کشاورزی)*, ۱۶۲-۱۵۳, ۲۱(۳).
- کوچمی ساردو، ایرج، شیرانی، حسین؛ اسفندیارپور بروجنی، عیسی و بسالت پور، علی اصغر. (۱۳۹۹). تعیین ویژگی های مؤثر بر پایداری ساختمان خاک های مناطق خشک با استفاده از الگوریتم ترکیبی ژنتیک-شبکه عصبی مصنوعی. *تحقیقات کاربردی خاک*, ۱۴۳-۱۲۹, ۳(۸).
- محمدیان خراسانی، شیوا؛ همایی، مهدی و پذیرا، ابراهیم. (۱۳۹۴). ارزیابی پایداری خاکدانه‌ها با استفاده از مدل‌های فرکتالی و روش‌های کلاسیک. *حفاظت منابع آب و خاک (علمی-پژوهشی)*, ۵۱-۳۹, ۴(۳).
- منابع طبیعی استان آذربایجان شرقی (۱۳۸۲). طرح صیانت از جنگل‌های ارسباران شمالی (جمع‌بندی مطالعات جنگل‌های ارسباران شمالی). اداره کل منابع طبیعی استان آذربایجان شرقی، تبریز. ۵۲۸ صفحه.
- نیک پور، معصومه، محبوبی، علی اکبر، مصدقی، محمدرضا، و صفادوست، آزاده. (۱۳۹۰). بررسی اثر ویژگی‌های ذاتی خاک بر پایداری ساختمان برخی از خاک‌های استان همدان. *علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)*, ۹۶-۸۵, ۱۵(۵۸).

## REFERENCES

- Alaboz, P., Dengizb, O., & Saygın, F. (2022). Estimation of soil aggregate stability by different regression methods. *Conference Paper*.
- Alekseeva, T.V., Sokolowska, Z., Hajnos, M., Alekseev, A.O., & Kalinin, P.I. (2009). Water stability of aggregates in subtropical and tropical soils (Georgia and China) and its relationships with the mineralogy and chemical properties. *Eurasian Soil Science*, 42, 415-425. DOI: 10.1134/S1064229309040085.
- Alijanpour Shalmani, A., Shabanpour, M., Asadi, H., & Bagheri, F. (2011). Estimation of Soil Aggregate Stability in Forest's Soils of Guilan Province by Artificial Neural Networks and Regression Pedotransfer Functions. *Water and Soil Science*, 21(3), 153-162. (In Persian).
- Allison, L.E., & Moodie, C.D. (1965). Carbonate. *Methods of soil analysis: part 2 chemical and microbiological properties*, 9, 1379-1396. DOI: 10.2134/agronmonogr9.2.c40.
- Alqahtani, M., Gumaei, A., Mathkour, H., & Maher Ben Ismail, M. (2019). A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. *Sensors*, 19(20), 4383. DOI: 10.3390/s19204383.
- Amézketa, E. (1999). Soil aggregate stability: a review. *Journal of sustainable agriculture*, 14(2-3), 83-151. DOI: 10.1300/J064v14n02\_08.
- An, S., Mentler, A., Mayer, H., & Blum, W.E. (2010). Soil aggregation, aggregate stability, organic carbon and nitrogen in different soil aggregate fractions under forest and shrub vegetation on the Loess Plateau, China. *Catena*, 81(3), 226-233. DOI: 10.1016/j.catena.2010.04.002.



- Angers, D.A., & Carter, M.R. (2020). Aggregation & organic matter storage in cool, humid agricultural soils. In *Structure & organic matter storage in agricultural soils*, 193-211. CRC Press.
- Are, M., Kaart, T., Selge, A., Astover, A., & Reintam, E. (2018). The interaction of soil aggregate stability with other soil properties as influenced by manure and nitrogen fertilization. DOI: 10.13080/z-a.2018.105.025.
- Armin, M., rouhipour, H., Ahmadi, H., Salajegheh, A., Mahdian, M. H., & ghorbannia kheybari, V. (2016). Relationship between Aggregate Stability and Selected Soil Properties in Taleghan Watershed. *Journal of Range and Watershed Management*, 69(2), 275-295. DOI: 10.22059/jrwm.2016.61683. (In Persian).
- Assiri, A. (2021). Anomaly classification using genetic algorithm-based random forest model for network attack detection. *Computers, Materials & Continua*, 66(1). DOI:10.32604/cmc.2020.013813.
- Bardsirizadeh, S., Esfandiarpour Borujeni, I., Besalatpour, A.A., & Abbaszadeh Dehaji, P. (2017). Use of Geostatistical Method to Determine the Most Effective Aggregate Component for Estimating Soil Structural Stability *Journal of Water and Soil*, 31(2), 533-544. DOI: 10.22067/jsw.v31i2.54438. (In Persian).
- Ben-Hur, M., Shainberg, I., Bakker, D., & Keren, R. (1985). Effect of soil texture & CaCO<sub>3</sub> content on water infiltration in crusted soil as related to water salinity. *Irrigation Science*, 6, 281-294. DOI: 10.1007/BF00262473.
- Besalatpour, A. A., Shirani, H., & ESFANDIARPOUR, B. I. (2015). Modeling of soil aggregate stability using support vector machines and multiple linear regression. *Journal of Water and Soil*, 29(2), 406-417. DOI: 10.22067/JSW.V0I0.22620. (In Persian).
- Besalatpour, A., Hajabbasi, M.A., Ayoubi, S., Afyuni, M., Jalalian, A., & Schulin, R. J. S. S. (2012). Soil shear strength prediction using intelligent systems: artificial neural networks & an adaptive neuro-fuzzy inference system. *Soil science & plant nutrition*, 58(2), 149-160. DOI: 10.1080/00380768.2012.661078.
- Besalatpour, A.A., Ayoubi, S., Hajabbasi, M.A., Jazi, A.Y., & Gharipour, A. (2014). Feature selection using parallel genetic algorithm for the prediction of geometric mean diameter of soil aggregates by machine learning methods. *Arid Land Research and Management*, 28(4), 383-394. DOI: 10.1080/15324982.2013.871599.
- Besalatpour, A.A., Ayoubi, S., Hajabbasi, M.A., Mosaddeghi, M., & Schulin, R. (2013). Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. *Catena*, 111, 72-79. DOI: 10.1016/j.catena.2013.07.001.
- Bhattacharya, P., Maity, P. P., Ray, M., & Mridha, N. (2021). Prediction of mean weight diameter of soil using machine learning approaches. *Agronomy journal*, 113(2), 1303-1316. DOI: 10.1002/agj2.20469.
- Boix-Fayos, C., Calvo-Cases, A., Imeson, A.C., & Soriano-Soto, M.D. (2001). Influence of soil properties on the aggregation of some Mediterranean soils and the use of aggregate size and stability as land degradation indicators. *Catena*, 44(1), 47-67. DOI: 10.1016/S0341-8162(00)00176-4.
- Bouajila, A., & Gallali, T. (2008). Soil Organic Carbon Fractions and Aggregate Stability in Carbonated. *Journal of Agronomy*, 7(2), 127-137.
- Bouslihim, Y., Rochdi, A., & Paaza, N.E.A. (2021). Machine learning approaches for the prediction of soil aggregate stability. *Heliyon*, 7(3). DOI: 10.1016/j.heliyon.2021.e06480.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. DOI: 10.1023/A:1010933404324.
- Chau, K.W., Wu, C.L., & Li, Y.S. (2005). Comparison of several flood forecasting models in Yangtze River. *Journal of Hydrologic Engineering*, 10(6), 485-491. DOI: 10.1061/(ASCE)1084-0699(2005)10:6(485).
- Chrenková, K., Mataix-Solera, J., Dlapa, P., & Arcenegui, V. (2014). Long-term changes in soil aggregation comparing forest & agricultural l& use in different Mediterranean soil types. *Geoderma*, 235, 290-299. DOI: 10.1016/j.geoderma.2014.07.025.
- Diaz-Zorita, M., Perfect, E., & Grove, J. H. (2002). Disruptive methods for assessing soil structure. *Soil and Tillage Research*, 64(1-2), 3-22. DOI: 10.1016/S0167-1987(01)00254-9.
- East Azerbaijan Meteorological Organization. (2021). *Meteorological Statistics of Kaleybar Synoptic Station*. Tehran: Meteorological Organization of the Islamic Republic of Iran. (In Persian).
- East Azerbaijan Natural Resources Organization. (2003). *Forest Conservation Plan for Northern Arasbaran Forests (Summary of Northern Arasbaran Forests Studies)*. Tabriz: East Azerbaijan Natural Resources General Directorate. (In Persian).
- Falsone, G., Bonifacio, E., Santoni, S., & Zanini, E. (2006). Wet aggregate stability of some Botswana soil properties. *ARID LAND RESEARCH AND MANAGEMENT*, 20(1), 15-28.
- Gee, G.W., & Bauder, J.W. (1986). Particle-size analysis. *Methods of soil analysis: Part 1 Physical and*

- mineralogical methods*, 5, 383-411. DOI: 10.2136/sssabookser5.1.2ed.c15.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
- Horn, R., Taubner, H., Wuttke, M., & Baumgartl, T. (1994). Soil physical properties related to soil structure. *Soil and Tillage Research*, 30(2-4), 187-216. DOI: 10.1016/0167-1987(94)90005-1.
- Huang, Y., Lan, Y., Thomson, S. J., Fang, A., Hoffmann, W.C., & Lacey, R.E. (2010). Development of soft computing and applications in agricultural and biological engineering. *Computers and electronics in agriculture*, 71(2), 107-127. DOI: 10.1016/j.compag.2010.01.001.
- Kemper, W.D., & Rosenau, R.C. (1986). Aggregate stability and size distribution. *Methods of soil analysis: Part 1 Physical and mineralogical methods*, 5, 425-442. DOI: 10.2136/sssabookser5.1.2ed.c17.
- Khazaee, A., Mosaddeghi, M.R., & Mahboubi, A.A. (2008). Structural stability assessment using wet sieving method and its relations with some intrinsic properties in 21 soil series from Hamadan province. *Agricultural Research*, 8(1 (A)), 171-181. (In Persian).
- Kouchami-Sardoo, I., Shirani, H., & Besalatpour, A.A. (2020). Determining the Features Influencing the Structural Stability of Soils of Arid Regions Using a Hybrid GA-ANN Algorithm. *Applied Soil Research*, 8(3), 129-143. (In Persian).
- Lado, M., Ben-Hur, M., & Shainberg, I. (2004). Soil wetting and texture effects on aggregate stability, seal formation, and erosion. *Soil Science Society of America Journal*, 68(6), 1992-1999. DOI: 10.2136/sssaj2004.1992.
- Liu, M.Y., Chang, Q.R., Qi, Y.B., Liu, J., & Chen, T. (2014). Aggregation and soil organic carbon fractions under different land uses on the tableland of the Loess Plateau of China. *Catena*, 115, 19-28. DOI: 10.1016/j.catena.2013.11.002.
- M. Nikpur, M., Mahboubi, A.A., Mosaddeghi, M. R., & Safadoust, A. (2012). Assessment of Soil Intrinsic Properties Effects on Soil Structural Stability of Some Soils in Hamadan Province. *JWSS*, 15(58), 85-96. DOI: 20.1001.1.24763594.1390.15.58.6.0. (In Persian).
- Martinez-Mena, M., Lopez, J., Almagro, M., Boix-Fayos, C., & Albaladejo, J. (2008). Effect of water erosion & cultivation on the soil carbon stock in a semiarid area of South-East Spain. *Soil & Tillage Research*, 99(1), 119-129. DOI: 10.1016/j.still.2008.01.009.
- Mbagwu, J.S. (2004). Aggregate stability and soil degradation in the tropics. In Proceedings of the conference report on the lecture given at the College on Soil Physics, Trieste, Italy, 3-21 March 2003, 246-252.
- Minhas, P.S., & Sharma, D.R. (1986). Hydraulic conductivity & clay dispersion as affected by application sequence of saline & simulated rain water. *Irrigation Science*, 7, 159-167. DOI: 10.1007/BF00344071.
- Mohammadian Khorasani, Sh., Homaei, M., & Pazira, M. (2015). Evaluating soil aggregate stability using classical methods and fractal models. *Journal of Water and Soil Resources Conservation*, 4(3), 39-51. DOI: 20.1001.1.22517480.1394.4.3.4.1. (In Persian).
- Nelson, D.W., & Sommers, L.E. (1996). Total carbon, organic carbon, and organic matter. *Methods of soil analysis: Part 3 Chemical methods*, 5, 961-1010. DOI: 10.2136/sssabookser5.3.c34.
- Ramdhani, Y., Putra, C., & Alamsyah, D. (2023). Heart failure prediction based on random forest algorithm using genetic algorithm for feature selection. *International Journal of Reconfigurable and Embedded Systems (IJRES)*, 12, 205. DOI: 10.11591/ijres.v12.i2.pp205-214.
- Rezaei, H., Jafarzadeh, A., Alijanpour, A., Shahbazi, F., & Valizadeh Kamran, K. (2020). Soil Organic Matter Condition in Forest Stands of Arasbaran. *Water and Soil*, 34(1), 115-127. DOI: 10.22067/JSW.V34I1.80633. (In Persian).
- Rezaei, H., Jsfarzadeh, A. A., Alijanpour, A., Shahbazi, F., & Valizadeh Kamran, K. (2017). Genetically evolution of Arasbaran forests soils along altitudinal transects of Kaleybar Chai Sofla Sub-Basin. *Water and Soil Science*, 26(4.1), 151-166. (In Persian).
- Saadat, S., esmaeelnejad, L., rezaei, H., mirkhani, R., & seyedmohammadi, J. (2019). Comparing Aggregate Stability Tests as One of the Soil Physical Quality Indicators. *Water and Soil*, 33(2), 289-303. Doi: 10.22067/JSW.V33I2.73916. DOI: 10.22067/JSW.V33I2.73916. (In Persian).
- Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004. Proceedings 5* (pp. 334-343). Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-25966-4\_33.
- Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., & Feuston, B.P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences*, 43(6), 1947-1958. DOI: 10.1021/ci034160g.



- Tajik, F. (2004). Evaluation of soil aggregate stability in some regions of iron. *JWSS*, 8(1), 107-123. DOI: 20.1001.1.24763594.1383.8.1.9.0. (In Persian).
- Tongway, D., & Hindley, N. (2004). Landscape function analysis: a system for monitoring rangeland function. *African journal of range and forage science*, 21(2), 109-113. DOI: 10.2989/10220110409485841.
- Wang, J.G., Yang, W., Yu, B., Li, Z. X., Cai, C.F., & Ma, R.M. (2016). Estimating the influence of related soil properties on macro-and micro-aggregate stability in ultisols of south-central China. *Catena*, 137, 545-553. DOI: 10.1016/j.catena.2015.11.001.
- Wang, W.C., Chau, K.W., Cheng, C.T., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology*, 374(3-4), 294-306. DOI: 10.1016/j.jhydrol.2009.06.019.
- Wu, X., Wei, Y., Wang, J., Wang, D., She, L., Wang, J., & Cai, C. (2017). Effects of soil physicochemical properties on aggregate stability along a weathering gradient. *Catena*, 156, 205-215. DOI: 10.1016/j.catena.2017.04.017.
- Zeini, H.A., Al-Jeznawi, D., Imran, H., Bernardo, L.F. A., Al-Khafaji, Z., & Ostrowski, K.A. (2023). Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil. *Sustainability*, 15(2), 1408. DOI: 10.3390/su15021408.
- Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M.R., & Xu, M. (2021). Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Regional*, 27, e00440. DOI: 10.1016/j.geodrs.2021.e00440.
- Zhai, R., Wang, J., Yin, D., & Shangguan, Z. (2022). Wet aggregate stability modeling based on support vector machine in multiuse soils. *International Journal of Distributed Sensor Networks*, 18(6). DOI: 10.1177/15501329221107573.



# Wet aggregate stability modeling based on random forest optimized with genetic algorithm

## EXTENDED ABSTRACT

### Introduction

In order to effectively manage soil and water resources, it is imperative to investigate wet aggregate stability (WAS) as a fundamental indicator for assessing soil structure and quality. Given the labor-intensive and expensive nature of determining WAS values through traditional laboratory techniques, there is a clear advantage in indirectly predicting them using readily available data. Machine learning (ML) techniques present a viable alternative for this purpose. The efficacy of ML stems from its capacity to analyze data on a large scale, enabling the resolution of challenges that conventional linear methods struggle to address economically and satisfactorily. The primary objective of this study is to develop a predictive model for WAS utilizing ML, specifically the random forest (RF) method in standalone mode, and its hybrid with a genetic algorithm (GA-RF) to optimize RF parameters. This unique approach distinguishes the research in the domain of WAS prediction.

### Material and Methods

The study area selected for investigation was a portion of forested land within the Arsbaran region. A total of 55 soil samples were collected from diverse environmental conditions and subsequently analyzed in the laboratory to determine soil texture, organic matter content, and calcium carbonate equivalent levels. Wet aggregate stability, as assessed by the Kemper and Rosenau test, served as the basis for calibrating machine learning (ML) models. Seven scenarios were explored for predicting wet aggregate stability using soil characteristics through the application of the random forest method in standalone mode and with optimization through a genetic algorithm. The dataset was partitioned such that 70% of the data was allocated for training the models, while the remaining 30% was reserved for testing. Subsequently, the accuracy of the predictive models was evaluated by calculating error metrics, including normalized root mean square error (NRMSE), correlation coefficient (CC), and Wilmot coefficient (WI).

### Results and Discussion

Upon scrutinizing the correlation coefficients between soil attributes and WAS derived from laboratory analysis, a robust relationship between the selected characteristics and the target variable was evident. Among the various random forest models assessed, the RF5 model exhibited notable performance with NRMSE parameters at 0.038, CC at 0.8, and WI at 0.789. Furthermore, the GA-RF5 model, optimized using a genetic algorithm, surpassed the RF5 model with improved metrics of 0.031 NRMSE, 0.800 CC, and 0.842 WI, showcasing enhanced predictive capabilities for WAS. A comparative analysis between the RF5 and GA-RF5 models revealed that the genetic algorithm significantly enhanced the predictive accuracy of RF by elevating R and WI values by 8% and 6.72%, respectively, while reducing NRMSE by 18.42%. Notably, scenario 5 emerged as the optimal model, predicated on the composition of sand, silt, and clay particles.

The findings from RF1 (NRMSE = 0.047, CC = 0.589, WI = 0.721) and GA-RF1 (NRMSE = 0.036, CC = 0.662, WI = 0.797) underscored the pivotal role of clay content in soil structure and its influence on WAS prediction. Clay content was identified as a critical soil property impacting WAS, as it functions as a binding agent that cohesively holds soil particles together. The clay content in the analyzed soils ranged from 5% to 62.5%. Contrarily, organic matter was found to have no discernible effect on WAS, as indicated by the statistical outcomes of scenario 2 models. Moreover, scenarios 6 and 7 demonstrated a substantial reduction of 10.43% and 10.81% in NRMSE in both standalone and optimized modes, highlighting the beneficial impact of lime in enhancing WAS prediction accuracy.

### Conclusion

Wet aggregate stability stands as a fundamental soil attribute crucial in determining soil erodibility and hydraulic characteristics. Understanding the key soil components governing WAS is imperative for preserving soil structure integrity. An innovative approach to quantifying WAS involves utilizing easily accessible soil parameters for predictive modeling. The statistical analysis conducted revealed that the RF5 and GA-RF5 models, incorporating soil texture variables, exhibited superior predictive performance. A comparative assessment between these models highlighted the enhanced predictive capabilities of the GA-RF model in forecasting WAS. Furthermore, scenarios 1 and 3 underscored the pivotal role of clay content in soil composition, encapsulating various soil formation processes and factors. Overall, the utilization of the GA-RF machine learning technique yields satisfactory accuracy in predicting WAS based on soil attributes. Notably, organic matter (OM) was found to have negligible impact on WAS, while the inclusion of lime demonstrated a positive effect on improving WAS prediction accuracy.

**Keywords:** Genetic Algorithm, Random Forest, Wet Aggregate Stability.