



Digital Mapping of Soil Surface Calcium Carbonate Equivalent in a Semi-Arid Region Using Gene Expression Programming Algorithm (Case Study: Abyek Plain, Qazvin)

Mohsen Davirand¹ | Fereydoon Sarmadian^{2✉} | Sajjad Teimouri Bardyani¹

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

Mohsen.daviran@ut.ac.ir

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

3. Department of Soil Science, Faculty of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail:

Sajjad.timuri@ut.ac.ir

Article Info	ABSTRACT
Article type: Research Article	Predicting soil calcium carbonate equivalent (CCE) is essential for effective land management in semi-arid regions characterized by high soil and topographic variability. This study evaluated the performance of Gene Expression Programming (GEP) for spatial prediction of CCE. Environmental predictors included seven Landsat 8 spectral bands, 15 derived spectral indices, nine ASTER-derived spectral indices, and 20 topographic variables derived from the digital elevation model (DEM). Recursive Feature Elimination based on Support Vector Machines (RFE-SVM) was employed for feature selection, reducing the predictor set by 84%. The optimized GEP model, consisting of three independent genes linked by an additive function, generated an explicit and interpretable mathematical equation for CCE prediction. Model performance demonstrated moderate predictive capability, with R^2 and RMSE values of 0.62 and 4.13 in the training phase, and 0.52 and 4.25 in the testing phase, respectively. However, regression slopes of 0.509 and 0.487 for the training and testing datasets indicated substantial regression toward the mean and underestimation of prediction variability. The resulting prediction map revealed a distinct north-south spatial gradient consistent with regional hydrological and topographic conditions. Overall, the integrated RFE-SVM-GEP framework provides a practical basis for digital CCE mapping in semi-arid environments. Nevertheless, improving the quality and spatial representativeness of training data, incorporating more informative environmental predictors, and integrating hybrid modeling approaches are recommended to enhance prediction accuracy, particularly for extreme CCE values.
Article history:	
Received: Oct. 23, 2025	
Revised: July. 4, 2026	
Accepted: July. 13, 2026	
Published online: Aug. 2026	
Keywords: <i>Gene Expression Programming (GEP), Calcium Carbonate Equivalent (CCE), Semi-arid regions, Digital Soil Mapping</i>	
Cite this article: Davirand, M., Sarmadian, F., Teimouri Bardyani, S., (2026) Digital Mapping of Surface Soil Calcium Carbonate Content in Arid and Semi-Arid Regions Using Gene Expression Programming (Case Study: Abyek Plain, Qazvin), <i>Iranian Journal of Soil and Water Research</i>, 57 (6), 1421-1445. http://doi.org/10.22059/ijswr.2026.404804.670039 © The Author(s). Publisher: University of Tehran Press. DOI: http://doi.org/10.22059/ijswr.2026.404804.670039	



EXTENDED ABSTRACT

Objective

The precise spatial delineation of Calcium Carbonate Equivalent (CCE) in the surface horizon (0–30 cm) is a critical prerequisite for advanced pedological research and the operationalization of sustainable land management (SLM) in water-stressed, semi-arid environments. Conventional Digital Soil Mapping (DSM) techniques, particularly those leveraging opaque 'black-box' machine learning algorithms, often preclude the necessary mechanistic understanding of pedogenic factors governing CCE distribution, thereby impeding scientific knowledge transfer. This study was conceived to rigorously assess the capacity of Gene Expression Programming (GEP)—an evolutionary computation paradigm designed to yield explicit, parsimonious mathematical models—for spatially predicting CCE content. The core aim was to derive a transparent, non-linear, and fully interpretable predictive pedotransfer function within the heterogeneous Abyek Plain, Qazvin, offering a demonstrable scientific advantage over correlative, non-mechanistic models for enhanced environmental stewardship.

Research Method

The study was executed within the Abyek Plain, Iran, an expansive pedoscape (approximately 60,000 ha) characterized by a semi-arid climate, thermic soil temperature regime, and diverse Quaternary geomorphic surfaces. A total of 240 georeferenced surface soil samples were acquired, ensuring representative coverage across the region's stratified pedological heterogeneity. Laboratory characterization confirmed the soil matrix as strongly alkaline and highly calcareous (mean CCE: 11.30%; pH: 7.88). The initial ancillary data catalogue comprised 51 covariates extracted at a 30-meter spatial resolution from multi-temporal median composite imagery (Landsat 8, ASTER) and an ALOS Digital Elevation Model (DEM). To address inherent multicollinearity and optimize the model's predictive stability, the Recursive Feature Elimination (RFE) algorithm was systematically employed. RFE successfully pruned the catalogue to the optimal subset of eight predictors, including the Brightness Index, Wind Exposition, Wetness Index, DEM, Convergence Index, and PCA3. The finalized dataset was randomly partitioned into an 80% calibration set and a 20% independent validation set. The GEP framework was configured with 30 chromosomes and 6 genes, utilizing the additive fitness function to minimize the Root Mean Square Error (RMSE) and facilitate the evolution of a multi-genic prediction model.

Findings

The GEP-derived model exhibited commendable performance and robustness in spatially modeling the CCE concentration. The calibration phase yielded a Coefficient of Determination of $R^2 = 0.62$ and a minimal RMSE of 4.13%. Crucially, the model maintained its predictive efficacy on the independent validation set, achieving a statistically acceptable $R^2 = 0.52$ and an RMSE of 4.25%. This performance affirms the model's generalized applicability across the pedoscape. The explicit mathematical architecture generated by GEP facilitated the quantifiable assessment of predictor contributions:

The Brightness Index emerged as the single most influential covariate, contributing 23% of the model's predictive power. This strong positive correlation ($r=0.48$) is mechanistically linked to the high spectral albedo associated with surficial salt and carbonate accumulation zones, particularly in poorly-drained, evaporative depressions. Wind Exposition followed with a substantial 18% contribution, suggesting that wind-driven sediment transport (aeolian processes) and micro-relief significantly regulate the redistribution of CCE. The Wetness Index demonstrated a significant negative correlation ($r=-0.44$), consistent with the hydrogeomorphic mechanism of CCE leaching in well-drained or topographically lower accumulation areas.

Conclusion

This investigation successfully demonstrated the utility of the GEP algorithm as a superior, mechanistically transparent alternative for the Digital Soil Mapping of CCE in complex semi-arid environments. The synthesis of RFE for predictor selection and GEP for model derivation resulted in an explicit, yet highly accurate, predictive model ($R^2 = 0.52$ validation). The resultant multi-genic pedotransfer function not only provides high-resolution spatial predictions but also offers unparalleled interpretability into the geo-environmental controls—such as spectral response linked to surficial salinization/carbonation and topographic features—that govern CCE distribution. This transparent modeling output constitutes an invaluable evidence base for precision agricultural intervention (e.g., targeted amendments, optimized irrigation scheduling) and informed environmental policy in calcareous, water-scarce regions, thereby supporting the tenets of sustainable land management. Future research should prioritize integrating time-series covariates to capture the dynamic spatio-temporal fluxes influencing CCE distribution.

Funding

The study was funded by the University of Tehran, Country Iran.

Authorship contribution

Conceptualization, Fereydoon Sarmadian; methodology, Fereydoon Sarmadian; software, Mohsen Davirand; validation, Fereydoon Sarmadian; formal analysis, Fereydoon Sarmadian; investigation, Mohsen Davirand, Sajjad Teimouri; resources, Mohsen Davirand ; data curation, Fereydoon Sarmadian; writing—original draft preparation, Mohsen Davirand; writing—review and editing, Mohsen Davirand; visualization, Mohsen Davirand; supervision, Fereydoon Sarmadian; project administration, Fereydoon Sarmadian; funding acquisition, Fereydoon Sarmadian.

All authors have read and agreed to the published version of the manuscript.

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that any AI and AI-assisted software was not used for writing process.

Data availability statement

Data available on request from the authors.

Acknowledgements

The authors sincerely appreciate the support and facilities provided by University of Tehran.

Ethical considerations

The authors avoided data fabrication, falsification, and plagiarism, and any form of misconduct.

Conflict of interest

The authors declare no conflict of interest.

نقشه‌برداری رقومی مقادیر کربنات کلسیم خاک سطحی در مناطق خشک و نیمه‌خشک با استفاده از الگوریتم بیان ژن (مطالعه موردی: دشت آبیک، قزوین)

محسن دوبرند^۱ | فریدون سرمیدان^۲ | سجاد تیموری بردیانی^۳

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه:

Mohsen.daviran@ut.ac.ir

۲. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه:

fsarmad@ut.ac.ir

۳. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران. رایانامه:

Sajjad.timuri@ut.ac.ir

چکیده

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۸/۱

تاریخ بازنگری: ۱۴۰۵/۴/۱۳

تاریخ پذیرش: ۱۴۰۵/۴/۲۲

تاریخ انتشار: شهریور ۱۴۰۵

واژه‌های کلیدی:

الگوریتم برنامه‌نویسی بیان ژن

(GEP)

کربنات کلسیم معادل (CCE)

مناطق نیمه‌خشک،

نقشه‌برداری رقومی خاک

پیش‌بینی کربنات کلسیم معادل (CCE) خاک برای مدیریت اراضی در مناطق نیمه‌خشک، به‌ویژه در مناطقی با تنوع بالای خاک و توپوگرافی، از اهمیت ویژه‌ای برخوردار است. این پژوهش با هدف ارزیابی توانایی الگوریتم برنامه‌نویسی بیان ژن (GEP) در مدل‌سازی مکانی CCE در منطقه‌ای نیمه‌خشک انجام شد. بدین منظور، مجموعه‌ای از متغیرهای محیطی شامل ۷ باند طیفی لندست ۸ و ۱۵ شاخص طیفی مشتق‌شده از آن، ۹ شاخص طیفی مستخرج از تصاویر ماهواره ASTER، همراه با ۲۰ متغیر توپوگرافیکی DEM و مشتقات آن، به‌عنوان ورودی مدل در نظر گرفته شدند. برای کاهش ابعاد داده و انتخاب متغیرهای بهینه، از روش حذف بازگشتی ویژگی‌ها مبتنی بر ماشین بردار پشتیبان (RFE-SVM) استفاده گردید که منجر به کاهش ۸۴ درصدی متغیرها شد. مدل GEP با ترکیب سه ژن مستقل و تابع پیونددهنده جمع، معادله‌ای ریاضی و تفسیرپذیر برای پیش‌بینی CCE استخراج کرد. ارزیابی عملکرد مدل در فاز آموزش ($R^2=0.62$ و $RMSE=4.13$) و در فاز آزمون ($R^2=0.52$ و $RMSE=4.25$) نشان‌دهنده توانایی نسبی مدل در یادگیری الگوها و تعمیم به داده‌های جدید است. با این حال، شیب خطوط رگرسیون در هر دو فاز (به ترتیب ۰/۵۰۹ و ۰/۴۸۷) که به مراتب کمتر از واحد هستند، بیانگر انقباض شدید مدل به سمت میانگین و فشرده‌شدگی دامنه‌ی پیش‌بینی است. نتایج نقشه‌ی پیش‌بینی‌شده، الگوی مکانی واضحی از شمال به جنوب را نشان داد که با عوامل هیدرولوژیکی و توپوگرافی مرتبط است. به‌طور کلی، این مطالعه نشان داد که هرچند رویکرد ترکیبی RFE-SVM و GEP پایه‌ای مناسب برای نقشه‌برداری رقومی CCE در مناطق نیمه‌خشک ارائه می‌دهد، اما بهبود کیفیت و توزیع داده‌های آموزشی، گنجانیدن متغیرهای محیطی مؤثرتر و استفاده از مدل‌های ترکیبی برای افزایش دقت در پیش‌بینی حدود مقادیر ضروری است.

استناد: دوبرند؛ محسن، سرمیدان؛ فریدون، تیموری بردیانی؛ سجاد، (۱۴۰۵) نقشه‌برداری رقومی مقادیر کربنات کلسیم خاک سطحی در مناطق خشک و نیمه‌خشک با استفاده از الگوریتم بیان ژن (مطالعه موردی: دشت آبیک، قزوین)، مجله تحقیقات آب و خاک ایران، ۵۷ (۶)، ۱۴۲۱-۱۴۴۵.



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

<http://doi.org/10.22059/ijswr.2026.404804.670039>

DOI: <http://doi.org/10.22059/ijswr.2026.404804.670039>

خاک یکی از ارکان اصلی اکوسیستم‌های زمینی است که در چرخه‌های زیست‌ژئوشیمیایی، تولید غذا، حفظ حاصل‌خیزی، تنظیم اقلیم و مدیریت پایدار منابع طبیعی نقش اساسی دارد (Rillig et al., 2023; Pereira et al., 2018). در مناطق خشک و نیمه‌خشک، که حدود ۴۰ درصد از سطح خشکی‌های جهان را دربر می‌گیرند؛ ویژگی‌های فیزیکوشیمیایی خاک، به‌ویژه مقادیر کربنات کلسیم معادل (CCE)، از مهم‌ترین عوامل کنترل‌کننده کیفیت خاک، بهره‌وری کشاورزی و پایداری محیط‌زیست محسوب می‌شوند (UNCCD, 2017). مقدار CCE بر pH، قابلیت جذب عناصر غذایی، ظرفیت نگهداری آب، ساختمان خاک و فعالیت‌های زیستی اثرگذار است؛ به‌گونه‌ای که مقادیر زیاد آن می‌تواند دسترسی گیاه به برخی عناصر غذایی را محدود کند، در حالی که مقادیر متعادل آن موجب بهبود پایداری خاکدانه‌ها و ساختمان خاک می‌شود (Hartemink & Barrow, 2023; Bolan et al., 2023). از این‌رو، شناخت و نقشه‌برداری دقیق توزیع مکانی CCE نقش مهمی در مدیریت پایدار خاک، کاهش تنش‌هایی نظیر شوری و فرسایش و برنامه‌ریزی کشاورزی ایفا می‌کند (Nguyen et al., 2025).

نقشه‌برداری رقومی خاک با بهره‌گیری از متغیرهای کمکی محیطی، از جمله داده‌های طیفی و توپوگرافی، امکان پیش‌بینی پیوسته توزیع مکانی ویژگی‌های خاک را فراهم می‌کند (Wadoux et al., 2020; Zou et al., 2024). تلفیق این داده‌ها با الگوریتم‌های یادگیری ماشین، به‌ویژه در مناطق نیمه‌خشک که نمونه‌برداری میدانی محدود و پرهزینه است، دقت برآورد ویژگی‌هایی مانند کربنات کلسیم معادل (CCE) را افزایش داده است (Rahmani Siyalarz et al., 2024). پیشرفت‌های اخیر در الگوریتم‌های یادگیری ماشین و محاسبات تکاملی^۲، نیز با مدل‌سازی روابط پیچیده و غیرخطی بین متغیرهای محیطی و ویژگی‌های خاک، عملکرد نقشه‌برداری رقومی را بهبود بخشیده‌اند (Taghizadeh-Mehrjardi et al., 2021; Ramezan, 2022). همچنین، استفاده از داده‌های ماهواره‌ای و مدل‌های رقومی ارتفاع به‌عنوان متغیرهای کمکی، اطلاعات مکانی ارزشمندی درباره ویژگی‌های خاک، الگوهای مکانی و توپوگرافی ارائه می‌دهد و دقت پیش‌بینی ویژگی‌های خاک در مقیاس‌های وسیع را بهبود می‌بخشد (Tadić et al., 2023; Mahmoudabadi et al., 2017). این پیشرفت‌ها، ضرورت توسعه رویکردهای تلفیقی مبتنی بر سنجش از دور و یادگیری ماشین را برای نقشه‌برداری و پایش CCE بیش از پیش آشکار می‌سازد (Gholizzadeh et al., 2018; Teimouri Bardyani & Sarmadian, 2024).

الگوریتم برنامه‌نویسی بیان ژن^۳ (GEP) به عنوان یک رویکرد داده‌محور کارآمد در محاسبات تکاملی، نارسایی روش‌های جعبه سیاه^۴ (شبکه‌های عصبی عمیق، جنگل تصادفی، ماشین بردار پشتیبان و ...) را در مدل‌سازی ویژگی‌های خاک جبران می‌کند. این الگوریتم قادر است روابط غیرخطی و پیچیده بین ویژگی‌های خاک و متغیرهای پیش‌بین را کشف و مدل‌ها را در قالب فرمول‌های ریاضی صریح و قابل تفسیر ارائه دهد. این شفافیت و قابلیت تفسیر، GEP را نسبت به رگرسیون‌های خطی و سایر مدل‌های پیچیده در موقعیت بهتری برای ثبت روابط غیرخطی قرار می‌دهد و آن را به انتخابی منطقی و علمی برای مدل‌سازی در مناطق وسیع و متنوع تبدیل می‌کند (Algaifi et al., 2021؛ Emamgolizadeh et al., 2023؛ Fayyaz et al., 2024؛ Waschkowski et al., 2025).

با وجود گسترش مطالعات نقشه‌برداری رقومی CCE، پژوهش‌های انجام‌شده در دشت قزوین و سایر مناطق نیمه‌خشک ایران عمدتاً بر الگوریتم‌های جعبه سیاه مانند جنگل تصادفی، کوبیست و درخت تصمیم متمرکز بوده‌اند و کاربرد الگوریتم‌های شفاف و تفسیرپذیر نظیر GEP کمتر مورد توجه قرار گرفته است. از این رو، هدف این پژوهش ارزیابی توانایی الگوریتم GEP در پیش‌بینی و نقشه‌برداری مکانی CCE در دشت آبیگ با استفاده از داده‌های Landsat 8، ASTER و مدل رقومی ارتفاع ALOS و استخراج یک معادله ریاضی صریح برای تبیین روابط غیرخطی بین CCE و متغیرهای محیطی است. همچنین عملکرد مدل با استفاده از شاخص‌های آماری R^2 و RMSE ارزیابی شد تا قابلیت کاربرد آن در مدیریت پایدار خاک و کشاورزی دقیق در مناطق نیمه‌خشک بررسی شود.

با وجود پیشرفت‌های قابل توجه در نقشه‌داری (قومی خاک و توسعه‌ی الگوریتم‌های یادگیری ماشین) (Taghizadeh-Mehrjardi et al., 2020)

1. Calcium Carbonate Equivalent
2. Evolutionary Computation
3. Gene Expression Programming
4. Black-box Models



(al., 2024)، کاربرد الگوریتم GEP در پیش‌بینی مقادیر CCE خاک در مناطق نیمه‌خشک ایران، به‌ویژه در شرایط پدولوژیکی متنوع، کمتر مورد بررسی قرار گرفته است. مطالعات نشان داده‌اند که ترکیب داده‌های طیفی و توپوگرافیکی با الگوریتم‌های محاسبات تکاملی، دقت پیش‌بینی ویژگی‌های خاک را، به‌ویژه در مناطق با تنوع بالای خاک و توپوگرافی، به‌طور قابل توجهی بهبود می‌بخشد (Zeraatpisheh et al., 2019). با این حال، شکاف‌هایی در انتخاب متغیرهای بهینه و ارزیابی دقت الگوریتم GEP برای پیش‌بینی CCE در مقیاس‌های منطقه‌ای در مناطق نیمه‌خشک وجود دارد و اغلب مطالعات بر مدل‌های جعبه سیاه متمرکز بوده‌اند (Shahabi et al., 2022; Zhou et al., 2024).

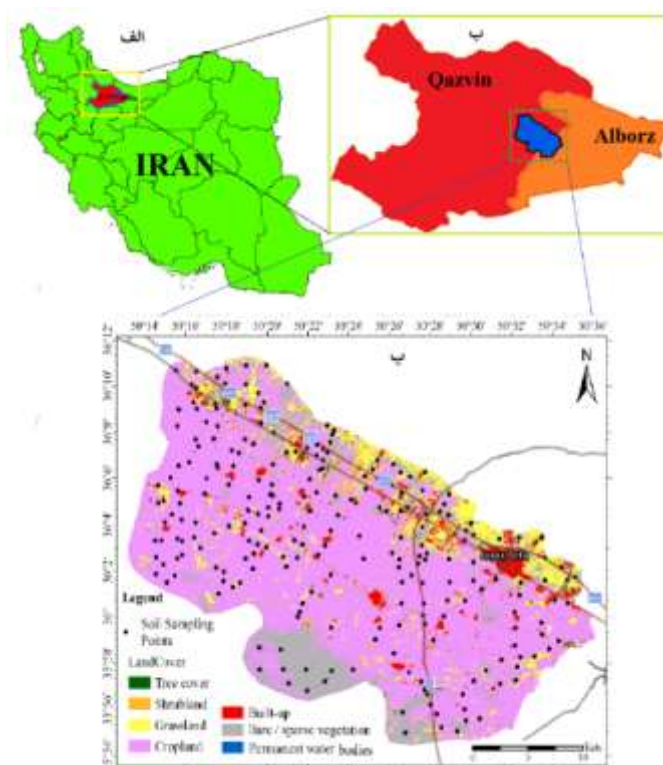
به عنوان مثال در حیطه مرور مطالعات داخلی، (Teimouri Bardyani & Sarmadian (2024) در مطالعه‌ای در دشت آبیک قزوین با بهره‌گیری از ۲۲۹ نمونه خاک سطحی و تصاویر ماهواره‌ای پریسما و لندست ۸، به نقشه‌برداری CCE و درصد رس با الگوریتم جنگل تصادفی پرداختند. متغیرهای کمکی شامل شاخص‌های طیفی، مؤلفه‌های اصلی و شاخص‌های مستخرج از مدل رقومی ارتفاع بودند. بهترین مدل برای تخمین CCE با استفاده از تصاویر پریسما و ترکیب ورودی PCA، شاخص‌های طیفی و شاخص‌های مدل رقومی ارتفاع به دست آمد که مقادیر R^2 ، RMSE، RPD و RPIQ به ترتیب ۰/۶۲، ۰/۷۲، ۰/۲۷ و ۰/۷۷ بود. این پژوهش نشان داد که تصاویر ابرطیفی پریسما عملکردی مشابه لندست ۸ در نقشه‌برداری CCE دارد و استفاده همزمان از سه دسته متغیر کمکی دقت مدل را بهبود می‌بخشد. در پژوهشی دیگر، (Lotfollahi et al., (2023) در منطقه نیمه‌خشک دشت چهاردولی در غرب ایران، به مدلسازی تغییرات مکانی CCE در اعماق مختلف خاک با استفاده از تکنیک‌های یادگیری ماشین شامل جنگل تصادفی و درخت تصمیم پرداختند. نتایج نشان داد که مدل جنگل تصادفی عملکرد بهتری نسبت به درخت تصمیم داشته و میانگین CCE با افزایش عمق خاک از ۳/۵٪ در سطح به ۶۳/۸٪ در عمق ۶۰-۳۰ سانتی‌متر افزایش یافته است. شاخص NDVI و سطح پایه شبکه آبراهه به ترتیب در لایه‌های سطحی و عمقی بیشترین اهمیت را داشتند. (Mousavi et al. (2021) در مطالعه‌ای در دشت قزوین با استفاده از ۲۷۸ خاکرخ و ۲۲ متغیر محیطی (اعم از توپوگرافی، سنجش از دور، اقلیم و ویژگی‌های خاک) به مدلسازی تغییرات عمقی CCE در پنج عمق با سه الگوریتم یادگیری ماشین جنگل تصادفی، رگرسیون درخت تصمیم و نزدیکترین همسایگی (k-NN) پرداختند. روش تجزیه مؤلفه‌های اصلی پنج متغیر کلیدی (MrVBF)، میانگین دمای سالانه، شاخص سبزی‌نگی و اثر باد) را انتخاب نمود. الگوریتم جنگل تصادفی با دامنه R^2 بین ۰/۸۳ تا ۰/۷۶ و RMSE بین ۳/۲۱ تا ۲/۱۴٪، بهترین عملکرد را داشت. در اعماق سطحی، درصد رس و در اعماق زیرسطحی، احتمال افق کلسیک مهم‌ترین پیش‌ران تغییرات CCE بودند. در منطقه نیمه‌خشک بروجن در استان چهارمحال و بختیاری، (Zeraatpisheh et al., (2019) به نقشه‌برداری رقومی CCE، کربن آلی و رس سطح خاک (عمق ۳۰-۰ سانتی‌متر) با استفاده از ۳۳۴ نمونه خاک و متغیرهای محیطی حاصل از مدل رقومی ارتفاع و تصاویر لندست ۸ (شاخص‌هایی نظیر NDVI، RVI و باند ۴) پرداختند. مدل‌های غیرخطی جنگل تصادفی، کوبیست (Cubist)، درخت رگرسیون و رگرسیون خطی چندمتغیره (MLR) مقایسه شدند. نتایج اعتبارسنجی متقابل ۱۰-تایی نشان داد که مدل کوبیست با R^2 برابر ۰/۳۰ و RMSE برابر ۹/۵۲، بهترین عملکرد را برای پیش‌بینی CCE داشت.

در پژوهش‌های خارجی نیز، (AbdelRahman et al., (2025) در منطقه نیو دلتا در غرب صحرای مصر، با هدف برآورد و نقشه‌برداری شوری خاک (EC)، pH و کربنات کلسیم (CaCO_3) در لایه‌های سطحی و زیرسطحی از تصاویر ابرطیفی ماهواره پریسما و چهار مدل یادگیری ماشین شامل رگرسیون حداقل مربعات جزئی (PLSR)، جنگل تصادفی، رگرسیون تطبیقی چندمتغیره اسپالین (MARS) و رگرسیون بردار پشتیبان (SVR) استفاده کردند. پس از حذف باندهای نویزی و انتخاب باندهای حساس با الگوریتم CARS، معادلات پیش‌بینی با رگرسیون خطی چندگانه توسعه یافت. نتایج نشان داد که مدل PLSR بهترین عملکرد را برای برآورد CaCO_3 سطحی و زیرسطحی داشت. (Wilford et al., (2015) در پژوهشی در مقیاس قاره‌ای در استرالیا به مدلسازی فراوانی کربنات کلسیم خاک در دو عمق سطحی و زیرسطحی با استفاده از داده‌های ژئوشیمیایی حاصل از طرح نقشه‌برداری ژئوشیمیایی ملی استرالیا (NGSA) شامل ۱۳۱۱ ایستگاه نمونه‌برداری و ۴۹ متغیر محیطی پیش‌بینی‌کننده (شامل اقلیم، سنگ مادر، توپوگرافی، پوشش گیاهی، داده‌های طیفی ماهواره‌ای و رادیومتری هوابرد) پرداختند. مدل‌سازی با استفاده از الگوریتم درخت تصمیم Cubist انجام شد و کربنات کلسیم خاک به صورت هنجاری از غلظت اکسید کلسیم (CaO) و گوگرد محاسبه گردید. نتایج اعتبارسنجی نشان داد که میانگین R^2 در داده‌های اعتبارسنجی حدود ۰/۵۰ است. نقشه تولید شده با دقت مکانی ۹۰ متر، توزیع کربنات کلسیم را به صورت پیوسته و با عدم قطعیت همراه ارائه داد که با نقشه‌های سنتی خاک و زمین‌شناسی تطابق خوبی داشت.

مواد و روش

منطقه مورد مطالعه

مطالعه‌ی حاضر در دشت آبیک، واقع در مرز استان‌های قزوین و البرز ایران، با مساحتی حدود ۶۰ هزار هکتار صورت پذیرفت (شکل ۱). متوسط ارتفاع منطقه برابر ۱۲۰۹ متر نسبت به سطح دریا قرار دارد و از نظر فیزیوگرافی و پستی و بلندی شامل اشکال اصلی سرزمین مانند فلات‌ها و یا تراس‌های فوقانی، دشت‌های آبرفتی دامنه‌ای و اراضی پست است. این ویژگی‌های فیزیوگرافیک نقش تعیین‌کننده‌ای در تنوع مکانی خاک و الگوهای پراکنش آن، به‌ویژه در فرایندهای تشکیل و توزیع CCE ایفا می‌کند (Yu et al, 2024).



شکل ۱. موقعیت قرارگیری منطقه مورد مطالعه در ایران (الف)، در مرز استان‌های البرز و قزوین (ب) و نقاط نمونه‌برداری (پ)

بر اساس آمارهای اقلیمی گردآوری‌شده از ایستگاه‌های باغ کوثر (دوره ۲۲ ساله) و بوبین زهرا (دوره ۱۷ ساله)، میانگین دمای سالانه ۱۳/۸۸ درجه سانتی‌گراد و میانگین بارش سالانه ۲۷۸ میلی‌متر ثبت شده است (Karbakhsh-ravari & Sarmadian, 2025). رژیم حرارتی خاک‌های منطقه، ترمیک بوده که برای کشت محصولات بهاره و پاییزه مناسب می‌باشد. از نظر رژیم رطوبتی، در قسمت‌های شمالی منطقه رژیم زیریک و در قسمت‌های جنوبی رژیم اریدیک حاکم است که نشان‌دهنده غلبه شرایط خشکی در طول سال می‌باشد. با این حال، در بخش‌هایی از نواحی جنوبی، رژیم رطوبتی اکوتیک نیز مشاهده می‌شود که معمولاً ناشی از محدودیت‌های زهکشی و اشباع زیرسطحی از آب در دوره‌های خاصی از سال است (Soil Survey Staff, 2022). بخش عمده‌ای از زمین‌ها شامل زراعت آبی هستند که عمدتاً تحت کشت محصولات زراعی یکساله با اهمیت اقتصادی بالا قرار دارند. محصولات اصلی کشت شده شامل گندم، جو، ذرت و یونجه است. علاوه بر این، زراعت دیم نیز در مناطقی که پتانسیل نگهداری رطوبت خاک بالاتر است، انجام می‌شود. سایر کاربری‌های مهم شامل مراتع شور و مراتع غیر شور است که نقش حیاتی در حفظ پوشش گیاهی طبیعی و حمایت از خدمات اکوسیستمی منطقه ایفا می‌کنند (Mousavi et al., 2021).

جمع‌آوری داده‌ها

برای تجزیه و تحلیل نهایی، مجموعاً ۲۴۰ پروفیل خاک مورد استفاده قرار گرفت که شامل ۲۱۱ پروفیل موجود از پایگاه داده‌های مطالعات پیشین دانشگاه تهران و ۲۹ پروفیل جدید حفر شده از طریق نمونه‌برداری هدفمند میدانی بود. این نمونه‌برداری هدفمند بر اساس طرح



تصادفی طبقه‌بندی شده انجام شد تا تنوع پدولوژیکی و ژئومورفولوژیکی منطقه به صورت جامع پوشش داده شود و از سوگیری در مجموعه‌ی داده‌ها جلوگیری شود. اگرچه هدف نهایی این پژوهش، مدل‌سازی و نقشه‌برداری از CCE خاک سطحی بود، اما به منظور رده‌بندی دقیق خاک‌ها بر اساس استانداردهای جهانی، درک فرآیندهای انتقال عمودی کربنات‌ها و نیز ایجاد یک پایگاه داده جامع قابل استفاده برای مطالعات آتی، کلیه افق‌های تشخیصی هر پروفیل (تا عمق ۱۵۰ سانتی‌متر) تشریح و نمونه‌برداری گردیدند. در این مطالعه، مقادیر CCE خاک‌های سطحی (۰-۳۰ سانتی‌متر) به دلیل ارتباط مستقیم با شوری و مدیریت خاک در مناطق نیمه‌خشک، به عنوان متغیر هدف اصلی انتخاب شد.

متغیرهای محیطی به‌عنوان پیش‌بینی‌کننده‌های اصلی محتوای CCE در الگوریتم GEP برای توضیح تغییرات مکانی آن استفاده شدند. این متغیرها شامل داده‌های طیفی و مشتقات توپوگرافی^۱ بودند که با وضوح مکانی ۳۰ متر پردازش شدند. انتخاب وضوح مکانی ۳۰ متر به دلیل تعادل بهینه بین دقت مکانی و حجم محاسبات برای مدل‌سازی در مقیاس منطقه‌ای مناسب بود (Nawar & Mouazen, 2021). این وضوح مکانی امکان ثبت تغییرات مهم خاک را بدون افزایش بیش از حد پیچیدگی مدل فراهم می‌کند، که برای کاربردهای عملی در نقشه‌برداری رقومی خاک در مناطق وسیع ضروری است (Shirazi et al., 2024; Sarmadian & Ghavami, 2020). داده‌های طیفی مورد استفاده از میانه ۱۶۷ تصویر بازتاب سطحی Landsat 8 (از ژانویه ۲۰۱۳ تا اکتبر ۲۰۲۴) با فیلتر ابرناکی کمتر از ۲۰٪ که در موتور محاسباتی گوگل ارث انجین پردازش شده‌اند، استخراج گردید. محاسبه‌ی میانه بر روی این بازه‌ی زمانی ۱۱ ساله برای کاهش اثرات ناشی از تغییرات زمانی و فصلی، مانند چرخه‌های پوشش گیاهی و رطوبت، انجام شد (Gorelick et al., 2017). میانه هفت باند طیفی اصلی Landsat 8 مورد استفاده قرار گرفت. علاوه بر این، چندین شاخص طیفی پرکاربرد به همراه شاخص‌های کانی‌شناسی، در سامانه گوگل ارث انجین استخراج شدند (جدول ۱).

برای غنی‌سازی متغیرها، داده‌های طیفی میانه از تصاویر ماهواره ASTER (۱۴ فروردین ۱۳۷۹ تا ۱۱ شهریور ۱۳۸۷) نیز پس از پردازش در سامانه گوگل ارث انجین و بازنمونه‌برداری^۲ به رزولوشن ۳۰ متر، به مجموعه داده‌ها افزوده شدند. داده‌های ASTER به دلیل داشتن باندهای طیفی متعدد در ناحیه‌ی فروسرخ حرارتی، برای استخراج شاخص‌های کانی‌شناسی خاص (جدول ۲) و بهبود پیش‌بینی CCE در مناطق خشک، بسیار ارزشمند هستند (Gholami et al., 2022).

مشتقات توپوگرافی از مدل رقومی ارتفاع ماهواره ALOS با وضوح مکانی ۳۰ متر که از پلتفرم سامانه گوگل ارث انجین قابل دسترسی بود، استخراج شدند. این مدل رقومی ارتفاع به نرم‌افزار SAGA GIS وارد شد تا ۲۱ متغیر کمکی توپوگرافی تولید شوند (جدول ۳). این متغیرها نقش کلیدی در ثبت تأثیر ناهمواری‌های زمین بر فرآیندهای هیدرولوژیکی، رسوب‌گذاری و فرسایش، و در نتیجه بر تجمع و پراکنش CCE ایفا می‌کنند (Hengl et al., 2017).

انتخاب متغیرهای محیطی

استفاده از تعداد زیاد متغیرهای پیش‌بین با همبستگی بالا می‌تواند منجر به هم‌خطی چندگانه^۳ و بیش‌برازش^۴ شود (Taghizadeh-Mehrjardi et al., 2021). برای مقابله با این چالش، الگوریتم حذف ویژگی بازگشتی^۵ به عنوان یک روش انتخابگر قدرتمند به کار گرفته شد. الگوریتم حذف ویژگی بازگشتی با استفاده از الگوریتم ماشین بردار پشتیبان به عنوان مدل پایه، متغیرهای با اهمیت کمتر را به صورت بازگشتی حذف می‌کند (Ramezan et al., 2022). این الگوریتم با ارزیابی ۵۱ زیرمجموعه‌ی متغیرها و بهره‌گیری از اعتبارسنجی متقاطع ۱۰ گام و معیار RMSE بهترین زیرمجموعه از متغیرها را برای مدل‌سازی انتخاب کرد (Kuhn & Johnson, 2013).

پیکربندی مدل GEP و مدل‌سازی

الگوریتم GEP، ترکیبی از مزایای الگوریتم ژنتیک و برنامه‌نویسی ژنتیک^۶ است که راه‌حل‌ها را به صورت کروموزوم‌های خطی کدگذاری و سپس به درخت‌های بیان^۷ تبدیل می‌کند (Ferreira, 2001). این ویژگی به الگوریتم GEP امکان می‌دهد تا روابط ریاضی پیچیده و

1. Topographic derivatives
2. Resampled
3. Multicollinearity
4. Overfitting
5. Recursive Feature Elimination
6. Genetic Algorithms
7. Genetic Programming
1. Expression Trees

غیرخطی را مدل سازی کند که برای پیش بینی ویژگی های خاک در محیط های پیچیده مناسب است (Emamgolizadeh et al., 2023). الگوریتم GEP استفاده از نرم افزار GeneXproTools 5.0 پیاده سازی شد. داده های ۲۴۰ نمونه خاک سطحی CCE به صورت تصادفی طبقه بندی شده به دو مجموعه آموزشی (۸۰٪) و آزمایش (۲۰٪) تقسیم شدند تا نمایندگی مناسبی از کل تنوع منطقه در هر دو زیرمجموعه تضمین شود.

جدول ۱. شاخص های طیفی ماهواره لندست-۸ انتخاب شده برای تهیه نقشه CCE

علامت اختصاری	متغیر	رابطه
Blue	باند ۲	Blue
Green	باند ۳	Green
Red	باند ۴	Red
NIR	باند ۵	NIR
SWIR1	باند ۶	SWIR1
SWIR2	باند ۷	SWIR2
TIRS	باند ۱۰	TIRS
Brightness	تبدیل تسلدکپ-درخشندگی	$0.3037 \times \text{Blue} + 0.2793 \times \text{Green} + 0.4743 \times \text{Red} + 0.5585 \times \text{Near Infrared} + 0.5082 \times \text{Cirrus} + 0.1863 \times \text{SWIR 2}$
EVI	شاخص پوشش گیاهی بهبود یافته	$2.5 \times \frac{\text{Near Infrared} - \text{Red}}{(\text{Near Infrared} + 6 \times \text{Red} - 7.5 \times \text{Blue}) + 1}$
Greenness	تبدیل تسلدکپ-سبزینگی	$-0.2848 \times \text{Blue} - 0.2435 \times \text{Green} - 0.5436 \times \text{Red} + 0.7243 \times \text{Near Infrared} + 0.0840 \times \text{SWIR 1} - 0.1800 \times \text{SWIR 2}$
NDVI	شاخص تفاضلی نرمال شده پوشش گیاهی	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$
SBL	خط زمینه خاک	$\text{NIR} - 2.4 \times \text{Red}$
RBNDVI	شاخص تفاضلی نرمال شده قرمز-آبی	$\frac{\text{NIR} - \text{Blue}}{\text{NIR} + \text{Blue}}$
PC1	مولفه اصلی اول	-
PC2	مولفه اصلی دوم	-
wetness	تبدیل تسلدکپ-خیسی	$0.1509 \times \text{Blue} + 0.1973 \times \text{Green} + 0.3279 \times \text{Red} + 0.3406 \times \text{Near Infrared} - 0.7112 \times \text{SWIR 1} - 0.4572 \times \text{SWIR 2}$
PC3	مولفه اصلی سوم	-
SCI	شاخص ترکیب خاک	$\frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$
NDSI	شاخص تفاضلی نرمال شده شوری	$\frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}}$
Redness Index	شاخص قرمزی	$\frac{\text{Red} - \text{Green}}{\text{Red} + \text{Green}}$
RDI	شاخص خشکی	$\frac{\text{SWIR2}}{\text{NIR}}$
Shape Index	شاخص شکل	$2 \times \text{Red} - \frac{\text{Green} - \text{Blue}}{\text{Green} - \text{Blue}}$



جدول ۲. شاخص‌های طیفی ماهواره ASTER ایجاد شده برای تهیه نقشه CCE

علامت اختصاری	متغیر	رابطه
Kaolinitic	شاخص کائولینیت	$\frac{SWIR7}{SWIR5}$
Sericite/Muscovite/Illite/Smectite	شاخص سری‌سیت/موسکوویت/ایلیت / اسمکتیت	$\frac{SWIR7}{SWIR5 + \frac{SWIR7}{SWIR6}}$
Muscovite	شاخص موسکوویت	$\frac{SWIR7}{SWIR6}$
Dolomite	شاخص دولومیت	$\frac{SWIR6 + \frac{SWIR8}{SWIR7}}{SWIR5 \times \frac{SWIR7}{SWIR6}}$
Clay	شاخص رس	$\frac{SWIR6^2}{SWIR4 + \frac{SWIR6}{SWIR5}}$
Alunite/Kaolinite/Pyrophyllite	شاخص آلونیت / کائولینیت / پیروفیلیت	$\frac{SWIR6 + \frac{SWIR8}{SWIR7}}{SWIR5 \times \frac{SWIR7}{SWIR6}}$
Amphibole / MgOH	شاخص آمفیبول / منیزیم هیدروکسید	$\frac{SWIR8}{SWIR6}$
Amphibole	شاخص آمفیبول	$\frac{SWIR8}{TIR11 \times \frac{SWIR8}{TIR10}}$
Siliceous Rocks	سنگ‌های سیلیس	$\frac{TIR11 \times \frac{SWIR8}{TIR10}}{TIR10 \times \frac{SWIR8}{TIR12}}$

منبع: فرمول‌های شاخص‌ها از پایگاه داده‌ی (Index Database (<https://www.indexdatabase.de>) استخراج شده‌اند.

جدول ۳. شاخص‌های توپوگرافیکی ایجاد شده برای تهیه نقشه CCE

علامت اختصاری	متغیر (واحد)	رابطه / روش محاسبه
Aspect	جهت شیب (درجه)	زاویه از شمال به سمت تصویر افقی بردار عمود بر سطح در یک نقطه‌ی مشخص
Catchment Slope	شیب حوضه (درصد)	اختلاف ارتفاع بین خروجی و دورترین نقطه تقسیم بر فاصله‌ی مربوطه یا روش‌های آماری دیگر.
Channel Network Distance	فاصله از شبکه آبراهه (متر)	فاصله (معمولاً عمودی) هر سلول تا شبکه آبراهه‌ی مبدأ. در SAGA GIS با درون‌یابی تراز پایه شبکه آبراهه محاسبه می‌شود.
Convergence Index (CI)	شاخص همگرایی (درجه یا درصد)	میزان انحراف جهت شیب سلول‌های همسایه از جهتی که به سمت سلول مرکزی است. مقادیر منفی نشان‌دهنده‌ی همگرایی و مثبت نشان‌دهنده‌ی واگرایی جریان است.
Convexity	تحدب (لاپلاسیون)	تحدب لاپلاسیون که با هموارسازی نقشه ارتفاع با یک هسته‌ی گاوسی به دست می‌آید.
Geomorphons	ژئومورفون‌ها	یک روش مبتنی بر شناسایی الگو برای طبقه‌بندی لندفرم‌ها که از یک پنجره‌ی متحرک استفاده می‌کند. ۱۰ نوع اصلی دارد.
Landforms	لندفرم‌ها	معمولاً با استفاده از روش ژئومورفون‌ها، ۴۹۸ الگوی منحصر به فرد در ۱۰ کلاس اصلی شامل: flat, peak, ridge, hollow, footslope, valley, and pit, shoulder, spur, slope طبقه‌بندی می‌شوند.
Local Curvature	انحنای موضعی	به روش‌های مختلفی محاسبه می‌شود؛ برای مثال، مجموع گرادیان‌های (تانژانت شیب) یک سلول نسبت به سلول‌های همسایه آن.
LS Factor	فاکتور طول-شیب	فاکتور توپوگرافیک در معادله‌ی فرسایش RUSLE که از حاصلضرب فاکتور طول شیب (L) و فاکتور درجه‌ی شیب (S) به دست می‌آید. فرمول‌های مختلفی برای آن وجود دارد.
MRRTF	شاخص چندمقیاسی همواری بالایی پشته	شاخصی مکمل برای MRVBF که برای شناسایی مناطق مرتفع و مسطح در مقیاس‌های مختلف طراحی شده است. به طور خودکار در کنار MRVBF محاسبه می‌شود.
MrVBF	شاخص چندمقیاسی همواری کف دره	شاخصی که با ارزیابی صافی و پستی‌بلندی زمین در چندین مقیاس، کف دره‌ها را شناسایی می‌کند.
DEM	مدل رقومی ارتفاعی (متر)	یک نمایش رقومی از سطح توپوگرافی زمین که در قالب یک شبکه‌ی منظم (رستر) ذخیره می‌شود.
Plan Curvature	انحنای افقی (بر ۱ متر)	نرخ تغییر جهت شیب در جهت مماس بر خطوط تراز (عمود بر جهت شیب).
Profile Curvature	انحنای قائم (بر ۱ متر)	نرخ تغییر جهت شیب در جهت شیب (هم‌جهت با جریان).
Relative Slope Position	موقعیت نسبی شیب	موقعیت نسبی یک سلول بین کف دره (مقدار ۰) و قله پشته (مقدار ۱).
Saga Wetness Index	شاخص خیس‌ی توپوگرافیک SAGA	$TWI = \ln(\text{Specific Catchment Area} / \tan(\beta))$
Slope	شیب (درجه یا درصد)	تانژانت زاویه‌ی سطح نسبت به سطح افق
Topographic Position Index	شاخص موقعیت توپوگرافیک (متر)	$TPI = Z_0 - Z_{mean}$
Valley Depth	عمق دره (متر)	اختلاف ارتفاع سلول با تراز برآوردی پشته‌ها (Level Ridge).
Wind Exposure (WE)	در معرض باد بودن (نمای باد)	شاخصی بی‌بعد که میزان قرارگیری یک نقطه در معرض باد را نشان می‌دهد. مقادیر کمتر از ۱ نشان‌دهنده‌ی مناطق سایه‌ی باد و مقادیر بیشتر از ۱ نشان‌دهنده‌ی مناطق در معرض باد است.

مجموعه توابع^۱ (F) مورد استفاده در فرایند تکامل، شامل عملگرهای پایه‌ی ریاضی و منطقی و توابع مقایسه‌ای نظیر کمتر از (LT2A) بود که در جدول ۴ به همراه وزن و تعداد آرگومان^۲ هر یک ارائه شده است.

جدول ۴. مجموعه توابع و ویژگی‌های آن‌ها در الگوریتم GEP

Function	Symbol	Weight	Arity
Addition	+	۴	۲
Subtraction	-	۴	۲
Multiplication	*	۴	۲
Division	/	۱	۲
Exponential	Exp	۱	۱
x to the power of 2	X2	۱	۱
Cube root	3Rt	۱	۱
Minimum of 2 inputs	Min2	۱	۲
Maximum of 2 inputs	Max2	۱	۲
Complement	NOT	۱	۱
Inverse	Inv	۱	۱
Natural logarithm	Ln	۱	۱
Less Than with 2 inputs (A)	LT2A	۱	۲
Greater Than with 2 inputs (A)	GT2A	۱	۲
Equal To with 2 inputs (A)	ET2A	۱	۲
Sine	Sin	۱	۱
Cosine	Cos	۱	۱
Not Equal To with 2 inputs (A)	NET2A	۱	۲

مجموعه پایانه‌ها^۳ (T) نیز شامل هشت متغیر محیطی بود که براساس نتایج الگوریتم حذف بازگشتی ویژگی (RFE) به عنوان تأثیرگذارترین پارامترها انتخاب شدند. این متغیرها عبارت‌اند از: PC3, Brightness, WE, Greenness, MrVBF, DEM, CI, Wetness. همچنین پارامترهای کلیدی الگوریتم شامل تعداد کروموزوم‌ها (۳۰)، تعداد ژن‌ها (۳)، اندازه‌ی سر (۱۵) و اندازه‌ی دم (۱۶) بود که از طریق فرایند آزمون و خطا و با هدف دستیابی به کمترین خطا بهینه‌سازی شدند (جدول ۵). ساختار هر کروموزوم شامل ۳ ژن بوده و نتایج ژن‌های مجزا با استفاده از تابع پیوند جمع^۴ برای تولید خروجی نهایی ترکیب شده‌اند.

جدول ۵. پارامترهای کلیدی الگوریتم GEP و مقادیر تنظیم‌شده برای مدل‌سازی CCE سطحی

پارامترهای مدل	مقدار	پارامترهای مدل	توضیح (در بافت مدل)
Chromosomes	۳۰	کروموزوم‌ها	تعداد کروموزوم‌ها (جمعیت) مورد استفاده در الگوریتم تکاملی
Genes	۳	ژن‌ها	تعداد ژن‌هایی که ساختار هر کروموزوم را تشکیل می‌دهند
Head Size	۱۰	اندازه سر	طول بخش سر ژن (بخشی که حاوی توابع و پایانه‌ها است) در یک کروموزوم. این بخش معمولاً قابلیت تغییرات و تنوع بیشتری دارد
Tail Size	۱۶	اندازه دم	طول بخش دم ژن (بخشی که فقط حاوی پایانه‌ها یا ورودی‌های داده است). اندازه دم معمولاً تابعی از اندازه سر و آرگومان‌های توابع است
Dc Size	۱۶	اندازه Dc (بخشی از دم)	
Linking Function	Addition	تابع پیوند دهنده	تابعی که نتایج ژن‌های مجزا در یک کروموزوم را برای تولید خروجی نهایی مدل با هم ترکیب می‌کند

- Function Set
- Arity
- Terminal Set
- Addition



تابع برازندگی^۱ نیز بر اساس معیار RMSE تعریف شد تا الگوریتم به سمت کمینه سازی خطای پیش بینی هدایت شود. به منظور هدایت فرایند تکامل، عملگرهای ژنتیکی شامل جهش^۲، نو ترکیبی^۳ و درون جادهی^۴ با نرخ های مشخصی به کار گرفته شدند. مقادیر این عملگرها از جمله جهش (۰/۰۰۱۳۸)، تقاطع یکنواخت (۰/۰۰۷۵۵) و باز ترکیب (۰/۰۰۲۷۷) با هدف جلوگیری از همگرایی زودرس و تضمین تنوع جمعیت، تنظیم شدند (جدول ۶).

جدول ۶. نرخ های عملگرهای ژنتیکی در مدل الگوریتم GEP

Genetic Operators	Optimal Evolution
Strategy	۰/۰۰۲۶۳
Mutation	۰/۰۰۱۳۸
Function Insertion	۰/۰۰۲۰۶
Biased Leaf Mutation	۰/۰۰۵۴۶
Biased Leaf Mutation	۰/۰۰۵۴۶
Conservative Mutation	۰/۰۰۳۶۴
Conservative Function Mutation	۰/۰۰۵۴۶
Permutation	۰/۰۰۵۴۶
Conservative Permutation	۰/۰۰۵۴۶
Biased Mutation	۰/۰۰۵۴۶
Inversion	۰/۰۰۵۴۶
Tail Mutation	۰/۰۰۵۴۶
Tail Inversion	۰/۰۰۵۴۶
IS Transposition	۰/۰۰۵۴۶
RIS Transposition	۰/۰۰۵۴۶
Stumbling Mutation	۰/۰۰۱۴۱
Uniform Recombination	۰/۰۰۷۵۵
Uniform Gene Recombination	۰/۰۰۷۵۵
One-Point Recombination	۰/۰۰۲۷۷
Two-Point Recombination	۰/۰۰۲۷۷
Gene Recombination	۰/۰۰۲۷۷
Gene Transposition	۰/۰۰۲۷۷
Random Chromosomes	۰/۰۰۲۶
Random Cloning	۰/۰۰۱۰۲
Best Cloning	۰/۰۰۲۶
RNC Mutation	۰/۰۰۲۰۶
Constant Fine-Tuning	۰/۰۰۲۰۶
Constant Range Finding	۰/۰۰۰۰۸۵
Constant Insertion	۰/۰۰۱۲۳
Dc Mutation	۰/۰۰۲۰۶
Dc Inversion	۰/۰۰۵۴۶
Dc IS Transposition	۰/۰۰۵۴۶
Dc Permutation	۰/۰۰۵۴۶

تعیین ضوابط توقف^۵ برای اجرای الگوریتم های هوش محاسباتی، یک مرحله حیاتی برای تضمین پایداری و کارایی مدل است. در

1. Fitness Function
2. Mutation
3. Recombination
4. Insertion
5. Stopping Criteria

نرم‌افزار GeneXproTools، فرآیند بهینه‌سازی تا زمانی ادامه می‌یابد که هیچ بهبود قابل توجهی در مقدار، که در این پژوهش با ریشه میانگین مربعات خطا تعریف شده است، مشاهده نگردد؛ این مکانیسم از بیش‌برازش جلوگیری کرده و به مدل اجازه می‌دهد تا در یک تعادل بهینه بین پیچیدگی و دقت همگرا شود. (Lawal et al., 2021)

پس از پیکربندی کامل مدل و استانداردسازی داده‌ها، الگوریتم GEP اجرا شد و خروجی نهایی آن به صورت یک معادله‌ی قطعی^۱ ارائه گردید. این معادله با استفاده از زبان برنامه‌نویسی R بر روی داده‌های رستری منطقه اعمال شد تا نقشه‌های رقومی توزیع مکانی CCE در سطح دشت آبیک تولید شود. این مرحله، امکان تعمیم روابط استخراج‌شده از داده‌های نقطه‌ای به کل منطقه را فراهم آورد و گامی اساسی در جهت مدل‌سازی مکانی ویژگی‌های خاک محسوب می‌شود.

اعتبارسنجی

دقت پیش‌بینی مدل الگوریتم GEP با استفاده از داده‌های مستقل و سه معیار استاندارد و مکمل ارزیابی شد: ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE). این سه معیار به دلیل ارائه اطلاعات مکمل و جامع از عملکرد مدل، استانداردهای پذیرفته‌شده‌ای برای سنجش مدل‌های نقشه‌برداری رقومی خاک محسوب می‌شوند. R^2 نشان‌دهنده نسبت واریانس CCE است که توسط متغیرهای مستقل تبیین می‌شود (رابطه ۱). RMSE میانگین خطای پیش‌بینی را بر حسب واحدهای اصلی CCE اندازه‌گیری می‌کند، اما به دلیل مجذور کردن خطاها، حساسیت بیشتری نسبت به خطاهای بزرگ و داده‌های پرت دارد (رابطه ۲). در مقابل، MAE محاسبه میانگین قدرمطلق خطاها، به هر خطا وزن یکسانی داده و در برابر داده‌های پرت مقاوم‌تر است (رابطه ۳). ارائه همزمان RMSE و MAE امکان ارزیابی جامع‌تری از عملکرد مدل را فراهم می‌آورد، زیرا مقایسه این دو معیار می‌تواند وجود خطاهای بزرگ در پیش‌بینی‌های مدل را آشکار سازد (Hodson, 2022).

$$R^2 = 1 - \frac{\sum (Y_{i \text{ actual}} - Y_{i \text{ predicted}})^2}{\sum (Y_{i \text{ actual}} - \bar{Y}_{\text{actual}})^2} \quad \text{رابطه ۱}$$

$$RMSE = \sqrt{\frac{\sum (Y_{i \text{ actual}} - Y_{i \text{ predicted}})^2}{n}} \quad \text{رابطه ۲}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{i \text{ actual}} - Y_{i \text{ predicted}}| \quad \text{رابطه ۳}$$

که در آن n تعداد مشاهدات، $Y_{i \text{ actual}}$ مقدار واقعی CCE برای مشاهده i ، $Y_{i \text{ predicted}}$ مقدار پیش‌بینی شده CCE برای مشاهده i و $\bar{Y}_{\text{actual}}$ مقدار میانگین مقادیر واقعی CCE است. این معیارها برای ارزیابی دقت مدل‌های نقشه‌برداری رقومی خاک استاندارد هستند.

نتایج و بحث

آماره‌های توصیفی داده‌های CCE

CCE به عنوان شاخصی از مجموع کربنات‌ها نقشی کلیدی در تعیین pH و دسترسی گیاهان به عناصر غذایی ایفا می‌کند. دامنه گسترده مقادیر CCE (از ۰ تا ۳۴/۶۳٪) نشان‌دهنده ناهمگنی مکانی قابل توجه در توزیع آهک در منطقه است که احتمالاً ناشی از تفاوت در فرآیندهای ژئوشیمیایی و هیدرولوژیکی نقاط مختلف است (Mousavi et al., 2021). چولگی مثبت (۱/۰۸) حاکی از توزیع نامتقارن و وجود نقاطی با غلظت‌های بسیار بالای کربنات است. مقدار کشیدگی (۱/۳۸) در محدوده نرمال قرار دارد که نشان‌دهنده تمرکز داده‌ها حول میانگین است. این الگو با یافته‌های مشابه در مناطق خشک و نیمه‌خشک مبنی بر تأثیرپذیری شدید توزیع کربنات‌ها از فرآیندهای تبخیر، آبشویی و توپوگرافی همخوانی دارد (Taghizadeh et al., 2018).

متغیرهای محیطی انتخاب شده

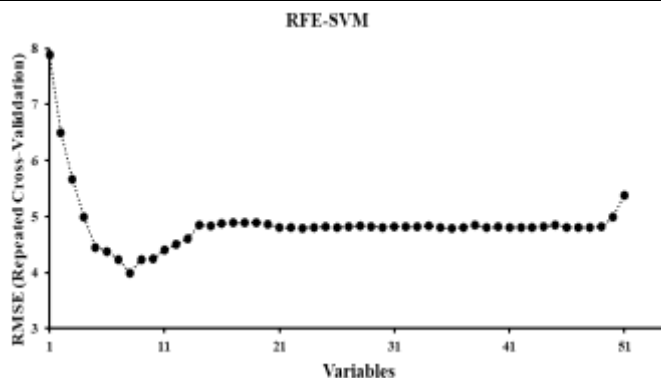
برای غلبه بر چالش‌هایی نظیر هم‌خطی چندگانه و نویز اطلاعاتی، که می‌توانند پایداری و قدرت تعمیم مدل‌های یادگیری ماشین را کاهش دهند، از الگوریتم RFE مبتنی بر ماشین بردار پشتیبان (SVM-RFE) استفاده شد. این روش با رویکرد گام‌به‌گام، متغیرهایی را که کمترین



تأثیر را بر عملکرد مدل دارند، بر اساس میزان کاهش شاخص RMSE حذف می‌کند (Awad & Fraihat, 2023; Guo et al., 2022). تحلیل منحنی بهینه‌سازی نشان داد (شکل ۲) که با حذف حدود ۸۴٪ از متغیرهای ورودی (کاهش از ۵۱ به ۸ متغیر)، مدل به حداقل مقدار خطا (در حدود ۰/۴) دست می‌یابد. افزودن متغیرهای بیشتر فراتر از این آستانه تنها خطا را کاهش نداد بلکه به دلیل ورود متغیرهای همبسته و نویز، منجر به تثبیت یا افزایش مجدد خطا شد (Montesinos López et al., 2022). این یافته‌ها با مطالعات پیشین در زمینه انتخاب متغیر در نقشه‌برداری رقومی خاک همخوانی دارد؛ به‌طوری که Mousavi et al., (2022) نیز با استفاده از رویکردی مشابه، از میان ۵۵ متغیر محیطی، ۱۳ متغیر مؤثر را برای پیش‌بینی کربن آلی خاک در همین منطقه انتخاب کردند. همچنین (Zhao et al., 2023) با استفاده از رویکرد RFE، از میان تعداد زیادی متغیر محیطی، زیرمجموعه‌ای بهینه را برای پیش‌بینی pH خاک انتخاب کردند.

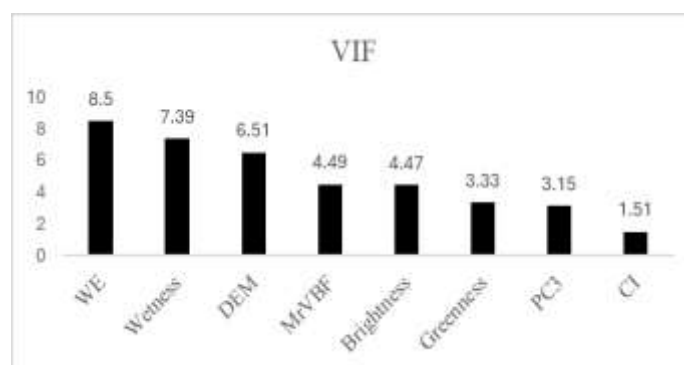
جدول ۷. آمار توصیفی CCE خاک سطحی در محدوده مطالعاتی

آماره	میانگین	میانه	کمترین مقدار	بیشترین مقدار	انحراف معیار	ضریب تغییرات	چولگی	کشیدگی
داده‌های آموزش	۱۰/۸۴	۱۰/۱۰	۰	۳۱	۶/۰۸	۰/۵۶	۱/۳۱	۱/۵۶
داده‌های آزمون	۱۱/۳۷	۱۰	۰/۴۷	۳۴/۶۳	۶/۵۰	۰/۵۷	۰/۹۵	۱/۸۱
کل داده‌ها	۱۱/۳۰	۱۰	۰	۳۴/۶۳	۶/۴۵	۰/۵۷	۱/۰۸	۱/۳۸



شکل ۲. الگوی تغییرات خطای RMSE در طی فرایند حذف بازگشتی ویژگی‌ها با الگوریتم RFE-SVM

پس از شناسایی متغیرهای نهایی، برای اطمینان از عدم وجود هم‌خطی چندگانه میان آن‌ها، از شاخص تورم واریانس^۱ (VIF) استفاده شد (شکل ۳). مطابق با رویه‌ی استاندارد در پژوهش‌های مشابه، آستانه‌ی VIF کمتر از ۱۰ به عنوان معیار عدم وجود هم‌خطی چندگانه در نظر گرفته می‌شود (Roy & Saha, 2019). همان‌طور که در شکل ۳ مشاهده می‌شود، مقدار VIF برای تمامی متغیرهای منتخب به طور قابل توجهی کمتر از ۱۰ است. این امر نشان می‌دهد که متغیرهای انتخاب‌شده از استقلال خطی مناسبی برخوردار بوده و می‌توانند بدون ایجاد ناپایداری در تخمین ضرایب، به‌طور همزمان در مدل‌سازی مورد استفاده قرار گیرند.

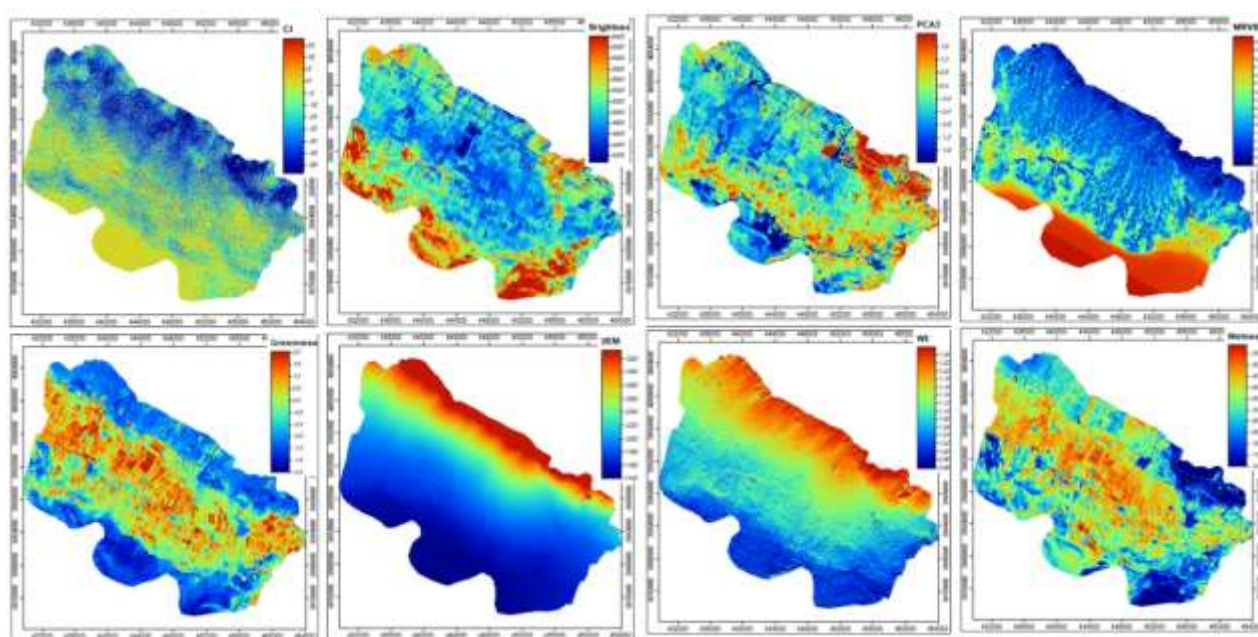


شکل ۳. مقادیر VIF برای متغیرهای محیطی منتخب در مدل‌سازی CCE سطحی

در نهایت، الگوریتم SVM-RFE منجر به انتخاب هشت متغیر محیطی به عنوان تأثیرگذارترین پارامترها بر توزیع مکانی CCE سطحی شد. آمار توصیفی این متغیرها در جدول ۶ ارائه شده است. توزیع مکانی این متغیرها در منطقه‌ی مطالعاتی در شکل ۴ نمایش داده شده است. انتخاب این مجموعه، حاکی از تأثیر توأمان عوامل توپوگرافی، هیدرولوژیکی و طیفی بر توزیع کربنات کلسیم در منطقه است. این یافته با پژوهش (Mousavi et al., 2022) در منطقه‌ای مشابه همخوانی دارد؛ آن‌ها نیز با استفاده از رویکردی مشابه، نقش تعیین‌کننده‌ی ترکیبی از متغیرهای توپوگرافی و طیفی را در مدل‌سازی ویژگی‌های خاک تأیید کردند.

جدول ۶. آمار توصیفی متغیرهای انتخاب شده با روش RFE-SVM

متغیر محیطی	میانگین	میان	حداقل	حداکثر	انحراف معیار	ضریب تغییرات
Brightness	۵۲۷۷/۴۸۶	۵۲۱۹/۱۵۲	۳۷۸۴/۶۲۱	۷۲۳۷/۳۴۰	۵۴۲/۷۵۴	۱۰/۲۸۴
CI	-۲۰/۵۹۷	-۲۰/۱۴۱	-۷۲/۳۲۴	۴۳/۴۱۱	۲۴/۶۷۱	-۱۱۹/۷۷۶
DEM	۱۲۲۳/۳۸۷	۱۲۰۵/۰۰۰	۱۱۴۲/۰۰۰	۱۴۰۲/۰۰۰	۶۴/۵۲۳	۵/۲۷۵
Greenness	-۰/۰۹۳	-۰/۰۹۶	-۱/۶۹۲	۲/۴۱۶	۰/۹۰۰	-۹۷۱/۲۲۷
MrVBF	۲/۴۰۹	۱/۹۹۰	۰/۰۰۰	۶/۸۹۲	۱/۶۷۴	۶۹/۴۹۳
PC3	-۰/۰۶۴	-۰/۰۰۶	-۲/۷۵۶	۲/۷۱۴	۰/۸۸۳	۱۳۸۹/۷۶۷
Wetness	-۶۹۹/۷۵۷	-۶۶۴/۲۹۶	-۱۴۹۵/۶۸۰	-۱۵۶/۵۵۳	۲۵۹/۰۱۱	-۳۷/۰۱۵
WE	۱/۱۲۸	۱/۱۲۸	۰/۹۵۶	۱/۲۷۷	۰/۰۷۳	۶/۴۷۱



شکل ۴. نقشه‌ی توزیع مکانی متغیرهای محیطی منتخب توسط الگوریتم SVM-RFE برای مدل‌سازی CCE سطحی در دشت آبیک

تحلیل روابط همبستگی متغیرهای محیطی با CCE

بررسی ضرایب همبستگی پیرسون (جدول ۷) نشان می‌دهد که Brightness با ضریب ۰/۴۸، قوی‌ترین همبستگی مثبت را با CCE دارد که ناشی از افزایش بازتاب طیفی خاک در محدوده مرئی و مادون قرمز نزدیک در اثر حضور کربنات‌ها و پوسته‌های نمکی-کربناتی است (Sun & Ellwood, 2025; Ben-Dor & Banin, 1994; Makhamreh, 2006). پژوهش‌های متعدد در مناطق خشک و نیمه‌خشک نشان داده‌اند که کربنات‌ها به‌عنوان یکی از مهم‌ترین عوامل کنترل‌کننده‌ی روشنایی خاک، می‌توانند تغییرات طیفی قابل‌توجهی ایجاد کنند که با روش‌های سنجش از دور قابل آشکارسازی است (Lotfollahi et al., 2023). در مقابل، Wetness با ضریب -۰/۴۴، همبستگی منفی معناداری نشان می‌دهد که به تأثیر دوگانه‌ی آب‌سویی در مناطق با زهکشی مناسب و کاهش بازتاب طیفی در اثر رطوبت سطحی مرتبط است (Metternicht & Zinck, 2003; Asgari Hafshejani & Jafari, 2023). همبستگی مثبت Greenness با CCE (۰/۳۵) که در نگاه اول غیرمنتظره به‌نظر می‌رسد، با توجه به پژوهش‌های اخیر، ناشی از بهبود ساختمان خاک و افزایش دسترسی به عناصر غذایی



توسط کربنات‌هاست که شرایط را برای استقرار پوشش گیاهی مساعد می‌سازد (White & Holland, 2018; Zeng et al., 2025). با این وجود، برای تأیید این رابطه و شناخت دقیق‌تر مکانیسم‌های حاکم، انجام پژوهش‌های تکمیلی در مقیاس‌های محلی با تفکیک پوشش‌های گیاهی مختلف توصیه می‌شود.

جدول ۷. ضرایب همبستگی پیرسون بین متغیرهای ورودی مدل و CCE

	CI	DEM	MrVBF	PC3	WE	Brightness	Greenness	Wetness	CCE
CI	۱/۰۰								
DEM	-۰/۴۹	۱/۰۰							
MrVBF	۰/۴۰	-۰/۷۱	۱/۰۰						
PC3	-۰/۰۳	۰/۱۷	-۰/۲۶	۱/۰۰					
WE	-۰/۵۱	۰/۸۹	-۰/۸۷	۰/۲۰	۱/۰۰				
Brightness	-۰/۰۳	-۰/۲۱	۰/۴۸	-۰/۰۶	-۰/۳۳	۱/۰۰			
Greenness	-۰/۲۳	۰/۰۲	-۰/۱۳	-۰/۱۱	۰/۱۰	۰/۲۲	۱/۰۰		
Wetness	۰/۲۴	-۰/۰۷	-۰/۱۵	۰/۰۵	۰/۰۱	-۰/۷۹	-۰/۵۹	۱/۰۰	
CCE	-۰/۰۴	-۰/۱۲	۰/۱۶	-۰/۰۵	-۰/۲۰	۰/۴۸	۰/۳۵	-۰/۴۴	۱/۰۰

عوامل توپوگرافی، هرچند همبستگی خطی ضعیف‌تری نشان می‌دهند، نقش کلیدی در کنترل هیدرولوژی و رسوب‌گذاری دارند. DEM با ضریب $-0/12$ ، الگوی کاهش CCE با افزایش ارتفاع را به دلیل زهکشی بهتر و آبشویی طبیعی در ارتفاعات شمالی نشان می‌دهد که با یافته‌های پیشین در منطقه همخوانی دارد. MrVBF با ضریب $0/16$ ، اراضی مسطح و آبرفتی جنوبی را شناسایی می‌کند که محل تجمع رسوبات کربناتی و تشکیل کلسی‌کرت‌ها هستند (Mousavi et al., 2021; Rasaei et al., 2025; Gallant & Dowling, 2003). CI با وجود همبستگی خطی ناچیز ($-0/04$)، از طریق شناسایی نقاط فرورفته و مستعد تجمع آب، نقش غیرخطی خود را در تشکیل کربنات‌های سطحی نشان می‌دهد (Alijani & Sarmadian, 2015). در مقابل، شاخص WE با ضریب $-0/20$ ، علاوه بر نشان‌دادن نقش فرسایش بادی (Sherrod et al., 2015)، به خوبی توانایی جداسازی عوارض ژئومورفولوژیکی در مقیاس منطقه‌ای را دارد. بررسی نقشه‌ی WE (شکل ۴) نشان می‌دهد که این شاخص به‌وضوح پهنه‌های هموار آبرفتی نظیر مخروط‌افکنه‌ها، کف دره‌ها و دشت‌های جنوبی را از ارتفاعات و خط‌الرأس‌های شمالی تفکیک می‌کند. بنابراین، همبستگی منفی WE با CCE را می‌توان هم به فرسایش بادی نسبت داد و هم به این نکته که اراضی پست جنوبی که خود محل تجمع CCE هستند، ذاتاً توسط توپوگرافی اطراف محافظت شده و WE پایینی دارند.

در مجموع، درخشندگی و خیس‌ی به‌عنوان قوی‌ترین متغیرهای کنترل‌کننده‌ی CCE شناسایی شدند که حاکی از تأثیر متقابل فرآیندهای تبخیر (در اراضی پست جنوبی) و آبشویی (در ارتفاعات شمالی) است (Mahmoudabadi et al., 2017; Tejedor et al., 2003). شاخص‌های توپوگرافی و پوشش گیاهی نیز به‌صورت غیرخطی و از طریق تعامل با یکدیگر بر توزیع کربنات‌ها اثر می‌گذارند که ضرورت به‌کارگیری مدل‌های داده‌کاوی را برای آشکارسازی این روابط پیچیده توجیه می‌کند (Lotfollahi et al., 2023). متغیرهای منتخب این مرحله، به‌عنوان ورودی مدل GEP در بخش بعدی استفاده خواهند شد.

ساختار و تفسیر مدل نهایی GEP

پس از تکمیل فرآیند تکاملی و دستیابی به نقطه توقف بهینه، الگوریتم GEP معادله‌ای تحلیلی و تفسیرپذیر (رابطه ۴) استخراج نمود که ترکیبی از متغیرهای محیطی (شاخص‌های گیاهی و پارامترهای توپوگرافی)، ضرایب عددی، عملگرهای ریاضی و توابع شرطی سفارشی شده (مانند gepGT2E و gepGT2D) را در خود جای داده است. ساختار نهایی این معادله از پیوند سه ژن (زیرمدل) مجزا با استفاده از تابع پیونددهنده جمع حاصل شده است که ضمن مدل‌سازی همزمان روابط غیرخطی، تعاملی و آستانه‌ای، امکان تفسیر مستقل سهم هر ژن در پیش‌بینی مقدار CCE را فراهم می‌کند. عبارت ریاضی صریح به‌دست‌آمده، قابلیت کاربرد مستقیم برای پیش‌بینی مکانی CCE در مناطق فاقد داده را دارد.

$$CCE = \text{gepGT2E} \left(\max \left(\text{gepGT2D} \left(7.54, DEM \right) \times \frac{DEM + MrVBF}{2}, 7.30 + PC3 \right), PC3^2 \right) + \text{gepLT2E} \left(\text{gepGT2E} \left(3.04, (Greenness - PC3) \times MrVBF \right) - PC3, Brightness \right) + \left(1.0 - \left(-7.16 + \frac{1.0 - \min(Brightness \times 6.16, WE)}{2.0} + Wetness \right) \right) \quad (\text{رابطه ۴})$$

که در آن توابع شرطی و ریاضی به صورت زیر تعریف می‌شوند:

$$\text{gepGT2D}(x, y) = \begin{cases} x \times y & x > y \\ x/y & x \leq y \end{cases}$$

$$\text{gepLT2E}(x, y) = \begin{cases} x + y & x < y \\ x \times y & x \geq y \end{cases}$$

$$\text{gepGT2E}(x, y) = \begin{cases} x + y & x > y \\ x \times y & x \leq y \end{cases}$$

مدل طراحی شده با بهره‌گیری از منطق درخت‌های بیان چندگانه، ساختاری سلسله‌مراتبی دارد که قابلیت تحلیل عملکرد مدل در مواجهه با تغییرات متغیرها را فراهم می‌سازد. برخلاف مدل‌هایی مانند شبکه‌های عصبی، در این مدل، روابط ریاضی به صورت صریح بین متغیرها تعریف شده‌اند که تحلیل رفتار مدل را در سناریوهای مختلف ممکن می‌سازد (Ferreira et al., 2001).

الف) ژن اول: مدل سازی وابستگی‌های توپوگرافی و تجزیه مؤلفه اصلی اولین جزء مدل، یک ساختار شرطی پیچیده را نشان می‌دهد (رابطه ۵) که عمدتاً بر اساس مدل ارتفاع رقومی، واریانس شیب متوسط (MrVBF) و مؤلفه اصلی سوم (PC3) استوار است. این ژن از تابع شرطی gepGT2E استفاده می‌کند، که نشان‌دهنده یک نقطه آستانه‌ای در رابطه بین ورودی‌ها است:

$$CCE_{Gene1} = \text{gepGT2E} \left(\max \left(\text{gepGT2D}(7.54, DEM) \times \frac{DEM + MrVBF}{2}, 7.30 + PC3 \right), PC3^2 \right) \quad (\text{رابطه ۵})$$

در این ساختار، مدل رقومی ارتفاع و MrVBF به صورت تعاملی و شرطی (توسط gepGT2D) ترکیب می‌شوند. استفاده از تابع max نشان می‌دهد که در این ژن، برآورد مقدار CCE توسط عاملی که تأثیر بیشتری دارد (ترکیب توپوگرافی یا PC3 خطی)، کنترل می‌شود. نهایتاً، این ترکیب با مربع PC3 توسط gepGT2E مقایسه شده تا تأثیر مؤلفه‌های اصلی بر اساس یک آستانه (شرط gepGT2E) تنظیم شود.

ب) ژن دوم: تلفیق پوشش گیاهی و توپوگرافی با ساختار شرطی دوگانه ژن دوم بر ترکیب تأثیر شاخص پوشش گیاهی (Greenness)، PCA3 و MrVBF تمرکز دارد (رابطه ۶). این ژن یک ساختار شرطی تو در تو را به کار می‌برد تا وابستگی مقدار CCE به پویایی‌های محیطی را آشکار سازد:

$$CCE_{Gene2} = \text{gepLT2E}(\text{gepGT2E}(3.04, (\text{Greenness} - PC3) \times MrVBF) - PCA3, \text{Brightness}) \quad (\text{رابطه ۶})$$

در این ژن، ابتدا تأثیر متغیر Greenness، با کسر PC3، نرمالایز شده و سپس در MrVBF ضرب می‌شود. این حاصل توسط gepGT2E با یک مقدار ثابت (۳/۰۴) مقایسه شده تا یک شرط داخلی ایجاد گردد. نتیجه این عبارت داخلی با Brightness توسط gepLT2E مرتبط می‌شود. این ساختار پیچیده نشان می‌دهد که رابطه مقدار CCE و پوشش گیاهی/توپوگرافی، به شدت غیرخطی و وابسته به مقادیر آستانه‌ای است که در آن، تغییر در پوشش گیاهی، بسته به شرایط توپوگرافی (از طریق MrVBF)، می‌تواند تأثیر متفاوتی بر مقدار CCE بگذارد.

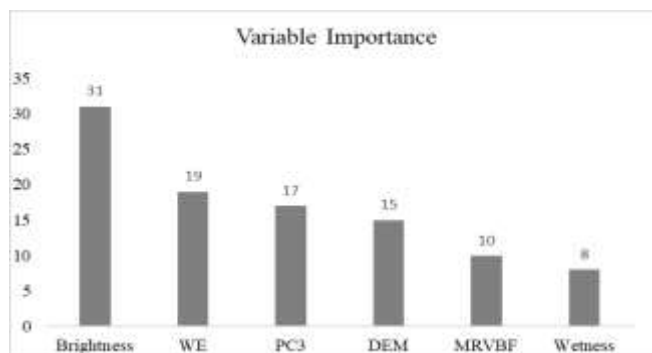
ج) ژن سوم: تأثیر خطی و محدود شده متغیرهای اصلی و پارامترهای زمین ژن سوم ساختاری ساده‌تر و عمدتاً خطی را به نمایش می‌گذارد (رابطه ۷)، که بیشتر بر تنظیم ضرایب و تأثیر متغیرهای اصلی (پارامترهای PC3، Greenness، MrVBF) متمرکز است:

$$CCE_{Gene3} = \left(1.0 - \left(-7.16 + \frac{1.0 - \min(\text{Brightness} \times 6.16, WE)}{2.0} + Wetness \right) \right) \quad (\text{رابطه ۷})$$

در این ژن، عملیات min نشان‌دهنده محدودسازی یا کنترل تأثیر یکی از متغیرها بر دیگری است. مقدار ۷/۱۶ و ۱/۰ به عنوان ضرایب ثابت عمل می‌کنند که توسط الگوریتم GEP به صورت بهینه یاد گرفته شده‌اند تا بایاس مدل را تنظیم کنند. این ژن، عملکرد جبرانی دارد و معمولاً مسئول جذب باقیمانده‌های خطی و تنظیم نهایی مدل است و کمترین وابستگی را به توابع شرطی پیچیده نشان می‌دهد.

اهمیت نسبی متغیرهای محیطی در مدل

بررسی اهمیت نسبی متغیرها در مدل GEP (شکل ۵) نشان می‌دهد که شاخص Brightness با سهم ۳۱٪، مؤثرترین متغیر در پیش‌بینی CCE است که ناشی از افزایش بازتاب طیفی خاک در حضور کربنات‌ها و پوسته‌های نمکی-کربناتی می‌باشد (Ben-Dor & Banin, 1994; Mulder et al., 2011). متغیرهای WE و PC3 به ترتیب با سهم ۱۹٪ و ۱۷٪ در رتبه‌های بعدی قرار گرفتند که حاکی از نقش توپوگرافی و نوسانات طیفی مرتبط با رطوبت یا کانی‌شناسی سطحی در توزیع کربنات‌هاست (Sherrod et al., 2015; Mousavi et al., 2022). عوامل توپوگرافی شامل DEM (۱۵٪) و MrVBF (۱۰٪) با وجود همبستگی خطی ضعیف، در ساختار غیرخطی GEP سهم قابل توجهی در پیش‌بینی داشتند که نقش غیرمستقیم آن‌ها را از طریق کنترل هیدرولوژی و رسوب‌گذاری نشان می‌دهد (Gallant & Dowling, 2003). در مقابل، شاخص Wetness با سهم ۸٪ کمترین تأثیر را داشتند که به رفتار غیرخطی و آستانه‌ای آن‌ها و نیز غلبه‌ی فرآیند تبخیر در شرایط خشک منطقه نسبت داده می‌شود (Metternicht & Zinck, 2003). به‌طور کلی، الگوی اهمیت متغیرها نشان می‌دهد که توزیع مکانی CCE عمدتاً توسط ویژگی‌های طیفی سطح خاک، فرآیندهای انتقال مواد (توپوگرافی) و نوسانات ترکیبی طیفی کنترل می‌شود (Emamgolizadeh et al., 2023; Rasaei et al., 2025).

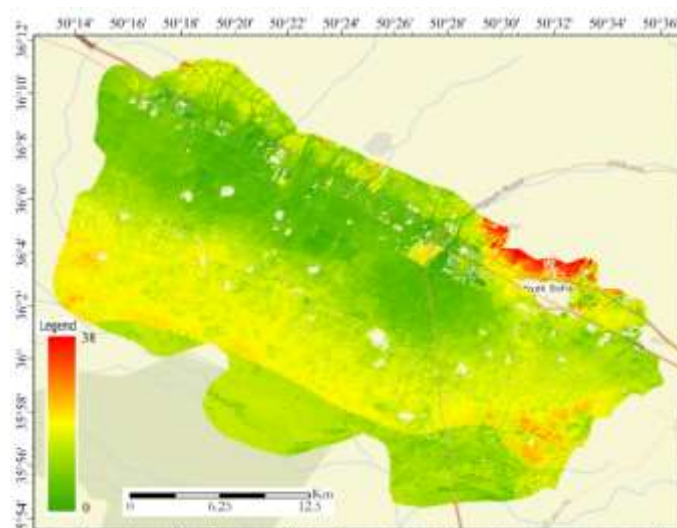


شکل ۵. نمودار اهمیت نسبی پارامترهای ورودی به الگوریتم GEP به درصد

توزیع مکانی CCE بر اساس مدل GEP

پس از توسعه‌ی مدل GEP با استفاده از داده‌های جدولی، گام بعدی شامل بازنویسی کد مدل به زبان R برای اعمال بر روی کل منطقه و تولید نقشه‌ی رستری بود. این بازنویسی به دلیل تفاوت اساسی در ساختار داده‌ها نیازمند تغییرات در کد بود، از جمله جایگزینی حلقه‌های مبتنی بر رکورد با عملیات‌های برداری برای پردازش هم‌زمان آرایه‌های بزرگ و تطبیق توابع شرطی با ساختارهای قابل اجرا در R برای داده‌های پیکسلی. پس از این تغییرات، کد نهایی بر روی لایه‌های رستری استانداردشده‌ی متغیرهای محیطی اعمال شد و نقشه‌ی پیوسته‌ی CCE تولید گردید.

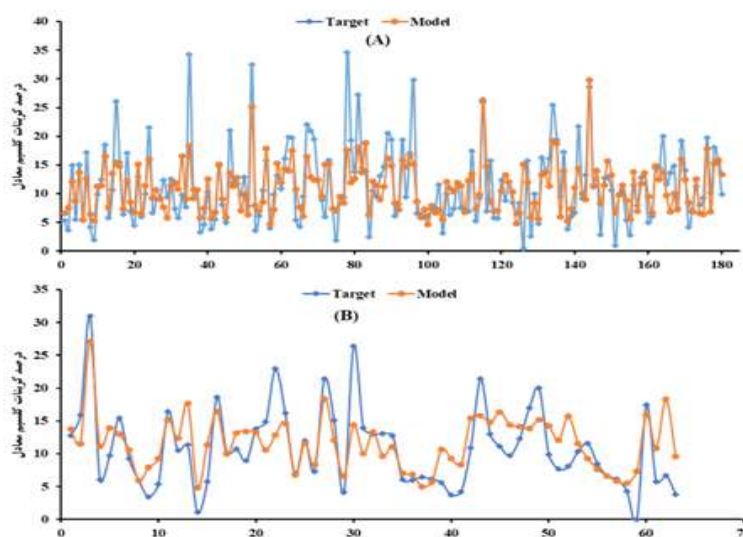
نقشه‌ی پیش‌بینی‌شده‌ی CCE با استفاده از الگوریتم GEP (شکل ۶) الگوی مکانی مشخصی را در منطقه‌ی مورد مطالعه نشان می‌دهد که از شمال به جنوب دارای تغییرات معنی‌داری است. بر اساس تحلیل منطقه‌ای انجام‌شده بر روی سه کلاس کاربری اراضی، میانگین CCE در اراضی مرتعی شمالی ۱۲۸/۰۸ درصد و انحراف معیار ۶/۱۱ برآورد گردید و در مقایسه با سایر کلاس‌ها، دارای بیشترین تنوع مکانی (حداکثر ۵۹/۲۷ درصد) است. این الگو نشان‌دهنده‌ی تأثیر توپوگرافی و زهکشی بر توزیع کربنات‌ها در ارتفاعات بالاتر و مناطق با زهکشی مناسب می‌باشد. در بخش مرکزی، اراضی کشاورزی با میانگین CCE معادل ۱۱/۳۶ درصد و انحراف معیار ۳/۸۸ کمترین تنوع درون کلاسی را نشان می‌دهند. در بخش‌های جنوبی منطقه که عمدتاً با کاربری مرتع کم‌پوشش مشخص می‌شوند، میانگین CCE به ۱۱/۰۲ درصد و انحراف معیار ۴/۵۴ می‌رسد که با ارتفاع کم، سطح آب زیرزمینی بالا و تجمع املاح و شوری همراه است (Lu et al., 2025). این یافته‌ها با پژوهش‌های پیشین در زمینه‌ی نقش توپوگرافی و زهکشی در انباشت یا شست‌شوی کربنات‌های خاک همخوانی کامل دارد (Mousavi et al., 2021; Teimouri Bardyani & Sarmadian, 2024). به‌طور کلی، نقشه‌ی پیش‌بینی‌شده توانایی مدل GEP در شناسایی مناطق بحرانی از نظر کربنات کلسیم را به‌وضوح نشان می‌دهد و می‌تواند به‌عنوان ابزاری کارآمد در مدیریت اراضی در مناطق نیمه‌خشک مورد استفاده قرار گیرد.



شکل ۶. نقشه مقادیر CCE خاک سطحی با استفاده از الگوریتم GEP

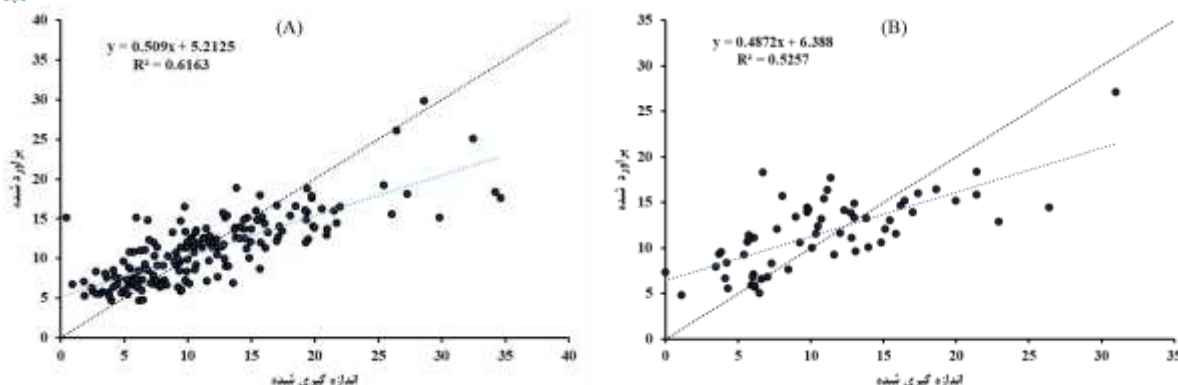
ارزیابی عملکرد مدل GEP

بررسی هم‌زمان نمودارهای توالی نمونه‌ها (شکل ۸) و نمودارهای پراکندگی رگرسیونی (شکل ۹) نشان می‌دهد که الگوریتم GEP توانسته است روند کلی تغییرات CCE را با دقتی متوسط مدل‌سازی کند. در فاز آموزش، ضریب تعیین 0.62 به دست آمد که از عملکرد نسبتاً خوب مدل در یادگیری الگوهای غالب حکایت دارد (Adeli et al., 2024). با این حال، خط رگرسیون این فاز $(y = 0.509x + 5.21)$ دارای شیبی به مراتب کمتر از واحد است که نشان‌دهنده انقباض شدید مدل به سمت میانگین می‌باشد (Kumar et al., 2017). به عبارت علمی‌تر، مدل دامنه‌ی تغییرات واقعی را فشرده می‌کند: مقادیر پایین CCE را بیش‌برآورد و مقادیر بالا را کم‌برآورد می‌نماید. این رفتار در نمودار توالی نمونه‌های آموزش (شکل ۸) نیز تأیید می‌شود؛ جایی که در نمونه‌های دارای مقادیر بالای CCE، انحراف نقاط پیش‌بینی شده از مشاهده‌شده به وضوح قابل مشاهده است.



شکل ۸. عملکرد مدل الگوریتم GEP در پیش‌بینی مقادیر CCE در برابر داده‌های واقعی برای هر نمونه (A: داده‌های آموزش، B: داده‌های آزمون)

در فاز آزمون، عملکرد مدل با کاهش محسوسی مواجه شد؛ ضریب تعیین به 0.52 (کاهش ۱۶ درصدی) تنزل یافت و خطاهای مطلق افزایش جزئی داشتند (RMSE از $4/13$ به $4/25$ و MAE از $2/98$ به $3/34$). خط رگرسیون فاز آزمون $(y = 0.4872x + 6.388)$ شیبی حتی کمتر از فاز آموزش را نشان می‌دهد که بیانگر تشدید پدیده انقباض میانگین در مواجهه با داده‌های جدید است (Omondigbe et al., 2024). این افت عملکرد در داده‌های آزمون شکل ۸ به وضوح قابل درک است؛ پراکندگی نقاط پیش‌بینی شده از نقاط مشاهده‌شده افزایش یافته و مدل در بازتولید مقادیر حدی (به‌ویژه CCE بسیار کم یا زیاد) با چالش جدی مواجه شده است.



شکل ۹. نمودار خروجی الگوریتم GEP در پیش‌بینی مقادیر CCE (A: داده‌های آموزش، B: داده‌های آزمون)

دلایل اصلی این عملکرد ضعیف در دامنه‌های حدی را می‌توان در توزیع نامتوازن داده‌های آموزشی و کمبود نمونه در بازه‌های افراطی جست‌وجو کرد که یادگیری الگوهای نادر را برای مدل دشوار می‌سازد (Ahangar & Piri, 2026). علاوه بر این، تغییرپذیری ذاتی خاک و عدم دسترسی به متغیرهای کلیدی مانند عمق مؤثر خاک یا نوع ماده‌ی مادری که در تعاملات غیرخطی با CCE هستند، می‌تواند به این انحرافات دامن بزند (Phillips, 2009; Nenkam et al., 2024). با وجود این محدودیت‌ها، توزیع متقارن باقیمانده‌ها حول صفر در هر دو فاز، حاکی از عدم وجود سوگیری سیستماتیک در مدل است که نقطه‌ی قوتی محسوب می‌شود (Gholizadeh et al., 2018). با این حال، شیب‌های بسیار پایین (حدود ۰/۵) و کاهش R^2 در فاز آزمون، قابلیت تعمیم‌پذیری مدل را به شدت محدود می‌کند و نشان می‌دهد که مدل GEP در شکل فعلی خود، برای پیش‌بینی دقیق CCE در مناطق با مقادیر حدی قابل اعتماد نیست و برای کاربرد عملی در این مناطق، نیازمند نمونه‌برداری هدفمند و افزودن متغیرهای مؤثرتر است.

جدول ۸. شاخص‌های ارزیابی کمی عملکرد الگوریتم GEP

شاخصه‌های آماری	Statistical indicators	آزمون	آموزش
ریشه میانگین مربعات خطا	RMSE	۴/۲۵	۴/۱۳
خطای میانگین مطلق	MAE	۳/۳۴	۲/۹۸
ضریب تعیین	R-square	۰/۵۲	۰/۶۲

نتیجه‌گیری

در این پژوهش، ترکیب الگوریتم RFE-SVM با GEP برای پیش‌بینی مکانی CCE خاک در منطقه‌ای نیمه‌خشک در ایران مورد ارزیابی قرار گرفت. رویکرد RFE-SVM با کاهش ۸۴ درصدی متغیرها (از ۵۱ متغیر اولیه به ۸ متغیر کلیدی)، ضمن کاهش پیچیدگی مدل، پایداری و قابلیت تعمیم آن را بهبود بخشید. متغیرهای منتخب شامل درخشندگی (با اهمیت نسبی ۰/۴۸) و رطوبت (با اهمیت نسبی ۰/۴۴) - به خوبی توانستند الگوی فضایی CCE را در امتداد گرادیان شمالی-جنوبی منطقه تبیین کنند، به گونه‌ای که مقادیر بالاتر CCE در بخش‌های جنوبی و پست با ارتفاع کم و تجمع املاح و مقادیر پایین‌تر در ارتفاعات شمالی با زهکشی مناسب همراه بود.

الگوریتم GEP نیز با ترکیب سه ژن مستقل، معادله‌ای ریاضی و تفسیرپذیر ارائه داد که در آن ژن اول بر اساس DEM و MrVBF و PC2، ژن دوم با تلفیق پوشش گیاهی و توپوگرافی، و ژن سوم با تنظیم خطی متغیرهای اصلی، ساختاری جامع برای پیش‌بینی CCE ایجاد کردند.

ارزیابی عملکرد مدل در فاز آموزش ($R^2=0.62$, $RMSE=4.13$) و فاز آزمون ($R^2=0.52$, $RMSE=4.25$) نشان‌دهنده توانایی نسبی مدل در یادگیری الگوها و تعمیم به داده‌های جدید است، هرچند شیب خطوط رگرسیون (۰/۵۰۹ در آموزش و ۰/۴۸۷ در آزمون) که به مراتب کمتر از واحد هستند، بیانگر انقباض شدید مدل به سمت میانگین و فشرده‌شدگی دامنه‌ی پیش‌بینی‌هاست که عمدتاً ناشی از توزیع نامتوازن داده‌های آموزشی و کمبود نمونه در دامنه‌های حدی (مقادیر بسیار پایین یا بالای CCE) می‌باشد. با وجود این محدودیت‌ها، عدم وجود سوگیری سیستماتیک (توزیع متقارن باقیمانده‌ها حول صفر) و عملکرد نسبتاً پایدار مدل در پیش‌بینی مقادیر متوسط CCE، پتانسیل بالای رویکرد ترکیبی RFE-SVM و GEP را برای نقشه‌برداری رقومی خاک در مناطق نیمه‌خشک تأیید می‌کند. با این حال، تغییرپذیری

ذاتی خاک، عدم دسترسی به متغیرهای کلیدی مانند عمق مؤثر خاک، نوع ماده‌ی مادری و الگوی زهکشی و همچنین تعاملات غیرخطی پیچیده، از عوامل مؤثر در بروز انحرافات پیش‌بینی در دامنه‌های حدی شناسایی شدند که لزوم توسعه مدل‌های ترکیبی یا استفاده از رویکردهای یادگیری عمیق را آشکار می‌سازد.

بر اساس یافته‌های این پژوهش، برای بهبود دقت و کاربردپذیری مدل در تحقیقات آتی، انجام نمونه‌برداری هدفمند در مناطق با CCE بسیار بالا و بسیار پایین و استفاده از رویکردهای ترکیبی یا یادگیری عمیق توصیه می‌شود. به‌طور کلی، توزیع مکانی CCE در منطقه‌ی مورد مطالعه، الگویی پیچیده اما قابل تبیین را منعکس می‌کند که با بهبود کیفیت و توزیع داده‌های آموزشی و به‌کارگیری تحلیل‌های پیشرفته‌تر، می‌تواند به دقت و کاربرد عملی بالاتری در مدیریت اراضی و برنامه‌ریزی پایدار کشاورزی در مناطق نیمه‌خشک دست یابد.

ملاحظات اخلاقی

حامی مالی

حمایت مالی از این پژوهش از طرف دانشگاه تهران، دانشکده مهندسی و فناوری کشاورزی در قالب پژوهانه پایان‌نامه دانشجویی نویسنده اول انجام شده است.

مشارکت نویسندگان

طراحی پژوهش: فریدون سرمیدیان، نظارت بر مراحل انجام پژوهش: فریدون سرمیدیان، بررسی و کنترل نتایج: فریدون سرمیدیان، جمع‌آوری داده‌ها: فریدون سرمیدیان، تهیه گزارش پژوهش: محسن دویرند، سجاد تیموری بردیانی؛ انجام محاسبات و تحلیل داده‌ها: محسن دویرند نویسندگان به‌طور مساوی در کلیه مراحل طراحی و انجام پژوهش، گردآوری داده‌ها، تجزیه و تحلیل آماری داده‌ها، تحلیل و تفسیر اطلاعات و نتایج، تهیه پیش‌نویس مقاله، بررسی و کنترل نتایج، اصلاح، بازبینی و نهایی‌سازی مقاله مشارکت داشتند.

اعلامیه هوش مصنوعی مولد و فناوری‌های مبتنی بر هوش مصنوعی در فرایند نگارش

نویسندگان از ابزارهای هوش مصنوعی برای نگارش این مقاله استفاده ننموده‌اند.

بیانیه دسترسی به داده‌ها

داده‌های پژوهش حاضر از طریق درخواست از نویسندگان قابل دسترسی است.

سپاسگزاری

از معاونت محترم پژوهشی دانشگاه تهران به خاطر حمایت مالی/ حمایت معنوی/ همکاری در اجرای پژوهش حاضر سپاسگزاری می‌شود.

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

منابع

- رحمانی سیالرز؛ شهلا، کشاورزی؛ علی، سرمیدیان؛ فریدون، و فرحبخش؛ محسن. (۱۴۰۳). ارزیابی شاخص‌های زیست‌محیطی آلودگی خاک به عنصر سرب در بخشی از اراضی شهری استان تهران. *تحقیقات آب و خاک ایران*: 55(9), 1485-1503. doi: 10.22059/ijswr.2024.378270.669735
- تقی زاده؛ روح الله، امیریان چکان؛ علیرضا، سرمیدیان؛ فریدون، محمدی؛ جهانگرد. (۱۳۹۶). بررسی توزیع جانبی و عمودی کربنات کلسیم در خاک با استفاده از زمین آمار و توابع اسپلاین. *تحقیقات کاربردی خاک*, 15(2), 1-15.
- تیموری بردیانی؛ سجاد و سرمیدیان؛ فریدون. (۱۴۰۳). نقشه‌برداری رقومی کربنات کلسیم معادل و درصد رس خاک با استفاده از تصاویر ماهواره‌های لندست ۸ و پریسما توسط الگوریتم جنگل تصادفی. *تحقیقات آب و خاک ایران*, ۵۵(۳), ۳۸۱-۳۹۹. doi: 10.22059/ijswr.2024.363941.669558
- موسوی؛ سید روح اله، سرمیدیان؛ فریدون، امید؛ محمود و بوگارت؛ پاتریک. (۱۴۰۰). مدل‌سازی تغییرات عمقی کربنات کلسیم معادل خاک با استفاده از الگوریتم‌های یادگیری ماشین در دشت قزوین. آب و خاک, ۳۵(۵), ۷۱۹-۷۳۴. doi: 10.22067/jswr.2021.71748.1076



- سرمدیان، فریدون و قوامی، محمد سجاد . (۱۳۹۸). ارزیابی تناسب اراضی با استفاده از روش TOPSIS و مقایسه آن با روش‌های پارامتریک برای محصول ذرت در بخشی از اراضی قزوین. *تحقیقات آب و خاک ایران*, 50(9), 2275-2287. doi: 10.22059/ijswr.2019.273788.668095
- کاربخش راوری، الهه و سرمدیان، فریدون . (۱۴۰۴). ارزیابی تناسب اراضی با روش مارکوس و مقایسه آن با روش پارامتریک. *تحقیقات آب و خاک ایران*, 56(7), 1821-1838. doi: 10.22059/ijswr.2025.394553.669933
- رسائی، زهرا، سرمدیان، فریدون و جعفری، اعظم . (۱۴۰۴). انبوهش‌زدایی واحدهای نقشه خاک با استفاده از مدل دسمارت: ترکیب مدل‌های مبتنی بر سیستم درختی و داده‌های جدید خاک‌رخی. *تحقیقات آب و خاک ایران*, ۵۶(۶), ۱۶۰۹-۱۶۲۹. doi: 10.22059/ijswr.2025.391334.669894

REFERENCES

- AbdelRahman, M. A., Metwaly, M. M., Yossif, T. M., & Moursy, A. R. (2025). Using hyperspectral data to estimate and map surface and subsurface soil salinity, pH, and calcium carbonates in arid region. *Environmental Sustainability*, 8(2), 367-387.
- Adeli, H., et al. (2024). Application of machine learning models for soil property prediction: A systematic review and meta-analysis. *Geoderma*, 443, 115049. (This is a comprehensive and recent review supporting the use of ML in soil science).
- Ahangar, A. G., & Piri, J. (2026). Evaluation of standard, black-box, and bayesian RSM-SVR models in the semi-arid area of south-eastern Iran for predicting soil chemical properties. *Scientific Reports*, 16(1), 11183.
- Alijani, Z., & Sarmadian, F. (2014). The role of topography in changing of soil carbonate content. *India J Sci Res*, 6, 263-271.
- Asgari Hafshejani, N., & Jafari, S. (2023). The study of particle size distribution of calcium carbonate and its effects on some soil properties in khuzestan province. *Iran Agricultural Research*, 36(2), 71-80.
- Awad, M., & Fraihat, S. (2023). Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems. *Journal of Sensor and Actuator Networks*, 12(5), 67.
- Ben-Dor, E., & Banin, A. (1994). Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment*, 48(3), 261-274.
- Bolan, N., Srivastava, P., Rao, C. S., Satyanaraya, P. V., Anderson, G. C., Bolan, S., ... & Kirkham, M. B. (2023). Distribution, characteristics and management of calcareous soils. *Advances in agronomy*, 182, 81-130. <https://doi.org/10.1016/bs.agron.2023.06.002>
- Emamgolizadeh, S., Mousavi, S. R., & Taghizadeh-Mehrjardi, R. (2023). Predicting soil organic carbon using gene expression programming in a semi-arid region of Iran. *Journal of Soils and Sediments*, 23(4), 1567–1580.
- Fayyaz, M., Davatgar, N., Mousavi, S. H., & Emamgolizadeh, S. (2024). Gene Expression Programming (GEP) for predicting soil properties: A review. *Journal of Environmental Management*, 350, 119565.
- Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. arXiv preprint cs/0102027.
- Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water resources research*, 39(12).
- Gholami, A., Fathabadi, A., & Taghizadeh-Mehrjardi, R. (2022). Machine learning approaches for predicting soil properties in semi-arid regions. *Environmental Earth Sciences*, 81(18), 456.
- Gholizadeh, H., Gamon, J. A., Zygielbaum, A. I., Wang, R., Schweiger, A. K., & Cavender-Bares, J. (2018). Remote sensing of biodiversity: Soil correction and data dimension reduction methods improve assessment of α -diversity (species richness) in prairie ecosystems. *Remote Sensing of Environment*, 206, 240-253.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2023.113567>
- Guo, J., Wang, K., & Jin, S. (2022). Mapping of Soil pH Based on SVM-RFE Feature Selection Algorithm. *Agronomy*, 12(11), 2742. <https://doi.org/10.3390/agronomy12112742>
- Hartemink, A. E., & Barrow, N. J. (2023). The effects of pH on nutrient availability depend on both soils and plants. *Plant and Soil*, 487(1), 21-37.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... &

- Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2), e0169748.
- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10.
- Karbakhsh-ravari, E. and Sarmadian, F. (2025). Land suitability assessment using the Marcos method and its comparison with the parametric method. *Iranian Journal of Soil and Water Research*, 56(7), 1821-1838. doi: 10.22059/ijswr.2025.394553.669933.in Persian
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- Kumar, S. (2017). A survey of deep learning methods for relation extraction. arXiv preprint arXiv:1705.03645.
- Lawal, A. I., Kwon, S., Hammed, O. S., & Idris, M. A. (2021). Blast-induced ground vibration prediction in granite quarries: An application of gene expression programming, ANFIS, and sine cosine algorithm optimized ANN. *International Journal of Mining Science and Technology*, 31(2), 265-277.
- Lotfollahi, L., Delavar, M. A., Biswas, A., Jamshidi, M., & Taghizadeh-Mehrjardi, R. (2023). Modeling the spatial variation of calcium carbonate equivalent to depth using machine learning techniques. *Environmental monitoring and assessment*, 195(5), 607. <https://doi.org/10.1007/s10661-023-11126-8>
- Lu, T., Luo, P., Wang, J., Lu, Y., Huo, A., & Liu, L. (2025). Soil salinity accumulation and groundwater degradation due to overexploitation over recent 40-year period in Yaoba Oasis, China. *Soil and Tillage Research*, 248, 106398.
- Mahmoudabadi, E., Karimi, A., Haghnia, G. H., & Sepehr, A. (2017). Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. *Environmental monitoring and assessment*, 189(10), 500.
- Makhamreh, Z. (2006). Analysis of Spectral Reflectance for Estimation of Soil Quality along a Climatic Gradient in the Eastern Mediterranean Region.
- Metternicht, G. I., & Zinck, J. A. (2003). Remote sensing of soil salinity: potentials and constraints. *Remote sensing of environment*, 85(1), 1-20.
- Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 109-139). Cham: Springer International Publishing.
- Mousavi, S., Sarmadian, F., Omid, M. and Bogaert, P. (2021). Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Water and Soil*, 35(5), 719-734. <https://doi.org/10.22067/jsw.2021.71748.1076>. in Persian.
- 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.
- Mousavi, S. R., Taghizadeh-Mehrjardi, R., & Rostaminia, M. (2023). Digital mapping of soil organic carbon stocks in a semi-arid region using machine learning and remote sensing. *Catena*, 224, Article 106985. <https://doi.org/10.1016/j.catena.2023.106985>
- Mulder, V. L., De Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1-2), 1-19.
- Nawar, S., & Mouazen, A. M. (2021). Machine learning approaches for soil classification and prediction of soil properties using proximal sensing. *Soil and Tillage Research*, 209, Article 104966. <https://doi.org/10.1016/j.still.2021.104966>
- Nenkam, A. M., Wadoux, A. M. C., Minasny, B., Silatsa, F. B., Yemefack, M., Ugbaje, S. U., ... & McBratney, A. B. (2024). Applications and challenges of digital soil mapping in Africa. *Geoderma*, 449, 117007.
- Nguyen, T. T., Shoukry, A. E., & Saraji, S. (2025). Investigating reactive transport and precipitation patterns of calcium carbonate in fractured porous media. *Journal of Colloid and Interface Science*, 679, 467-480.
- Omondigbe, O. P., Roudier, P., Lilburne, L., Ma, Y., & McNeill, S. (2024). Quantifying uncertainty in the prediction of soil properties using mid-infrared spectra. *Geoderma*, 448, 116954.
- Pereira, P., Bogunovic, I., Muñoz-Rojas, M., & Brevik, E. C. (2018). Soil ecosystem services, sustainability, valuation and management. *Current Opinion in Environmental Science & Health*, 5, 7-13.
- Phillips, J. D. (2009). Soils as extended composite phenotypes. *Geoderma*, 149(1-2), 143-151.
- Rahmani Siyalarz, S., Keshavarzi, A., Sarmadian, F. and Farahbakhsh, M. (2024). Assessment of environmental indices for soil lead contamination in a part of Shahr-e-Ray, Tehran Province. *Iranian Journal of Soil and Water Research*, 55(9), 1485-1503. doi: 10.22059/ijswr.2024.378270.669735. in Persian.



- Ramezan, C. A. (2022). Transferability of recursive feature elimination (RFE)-derived feature sets for support vector machine land cover classification. *Remote Sensing*, 14(24), 6218.
- Rasaei, Z., Sarmadian, F., & Jafari, A. (2025). Improving the Disaggregation of Soil Map Units Using the DSMART Method: Integrating Tree-Based Models and New Soil Profile Data. *Iranian Journal of Soil and Water Research*, 56(6), 1609-1629. in Persian.
- Rillig, M. C., Van der Heijden, M. G., Berdugo, M., Liu, Y. R., Riedo, J., Sanz-Lazaro, C., ... & Delgado-Baquerizo, M. (2023). Increasing the number of stressors reduces soil ecosystem services worldwide. *Nature Climate Change*, 13(5), 478-483 .
- Sarmadian, F. and GHavami, M. J. (2020). Land Suitability Evaluation Using TOPSIS Method and Its Comparison with Parametric Methods for Maize Production in Part of Qazvin. *Iranian Journal of Soil and Water Research*, 50(9), 2275-2287. doi: 10.22059/ijswr.2019.273788.668095. in Persian.
- Shahabi, A., Nabiollahi, K., Davari, M., Zeraatpisheh, M., Heung, B., Scholten, T., & Taghizadeh-Mehrjardi, R. (2022). Spatial prediction of soil properties through hybridized random forest model and combination of reflectance spectroscopy and environmental covariates. *Geocarto International*, 37(27), 18172-18195.
- Sherrod, L. A., Erskine, R. H., & Green, T. R. (2015). Spatial Patterns and Cross-Correlations of Temporal Changes in Soil Carbonates and Surface Elevation in a Winter Wheat-Fallow Cropping System. *Soil Science Society of America Journal*, 79(2), 417-427.
- Shirazi, F. R. A., Shabbazi, F., Rezaei, H., & Biswas, A. (2024). Multi-property digital soil mapping at 30-m spatial resolution down to 1 m using extreme gradient boosting tree model and environmental covariates. *Remote Sensing Applications: Society and Environment*, 33, 101123.
- Soil Survey Staff. (2022). Kellogg Soil Survey Laboratory Methods Manual. USDA Natural Resources Conservation Service.
- Sun, Y. and Ellwood, M.J. (2025) Calcium Carbonate Cycling in the Southern Ocean: Insights from Dissolved Calcium and Potential Alkalinity Tracers. *Limnology and Oceanography Letters*.
- Tadić, J. M., et al. (2023). Satellite imagery and digital elevation models for soil mapping: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 308-325.
- Taghizadeh, R. Amirian Chekan, A. Sarmadian. And Mohammadi, J.(2018). Study of lateral and vertical distribution of soil calcium carbonate using geostatistics and spline functions. *Applied Soil Research*, 5(2), 1-15. in Persian.
- Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., & Zeraatpisheh, M. (2021). Digital soil mapping: A review of challenges and opportunities in Iran. *Geoderma Regional*, 25, Article e00390.
- Taghizadeh-Mehrjardi, R., Zeraatpisheh, M., Amirian-Chakan, A., & Scholten, T. (2024). A brief review of digital soil mapping in Iran. *Remote Sensing of Soil and Land Surface Processes*, 217-228.
- Teimouri Bardyani, S., & Sarmadian, F. (2024). Digital mapping of soil properties (Calcium Carbonate and soil clay percentage) using landsat 8 and Prisma satellite images by the random forest algorithm. *Iranian Journal of Soil and Water Research*, 55(3), 381-399. in Persian.
- Tejedor, M., Jiménez, C. C., & Díaz, F. (2003). Use of volcanic mulch to rehabilitate saline-sodic soils. *Soil Science Society of America Journal*, 67(6), 1856-1861.
- UNCCD. (2017). Global Land Outlook, first edition. United Nations Convention to Combat Desertification.
- Wadoux, A. M. C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science*, 71(2), 133-136.
- Waschkowski, F., Li, H., Deshmukh, A., Grenga, T., Zhao, Y., Pitsch, H., ... & Sandberg, R. D. (2025). Gradient information and regularization for gene expression programming to develop data-driven physics closure models. *Flow, Turbulence and Combustion*, 114(1), 145-175
- White, P. J., & Holland, J. E. (2018). Calcium in plant physiology and its availability from the soil. In White, PJ & Holland, JE Proceedings of the International Fertiliser Society 827: Calcium in Plant Physiology and its Availability from the Soil. (pp. 1-32).
- Wilford, J., De Caritat, P., & Bui, E. (2015). Modelling the abundance of soil calcium carbonate across Australia using geochemical survey data and environmental predictors. *Geoderma*, 259, 81-92.
- Yu, Z., Kuang, L., Jiang, Y., Li, W., Zhang, J., Zhou, Y., ... & Ye, Y. (2024). Soil attributes are more important than others in shaping the diversity of cultivated land quality types, southern China. *Ecological Indicators*, 166, 112472.
- Zeng, S., Liu, Z., Jiang, Y., Goldscheider, N., Yang, Y., Zhao, M., ... & Shi, L. (2025). A greening Earth has reversed the trend of decreasing carbonate weathering under a warming climate. *Nature*

Communications, 16(1), 2583.

- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445-452.
- Zhou, H., Wang, X., Wu, Y., & Zhang, X. (2021). Mechanical properties and micro-mechanisms of marine soft soil stabilized by different calcium content precursors based geopolymers. *Construction and Building Materials*, 305, 124722. <https://doi.org/10.1016/j.conbuildmat.2021.124722>
- Zhou, J., Tian, Q., Nazar, S., & Huang, J. (2024). Hyper-tuning gene expression programming to develop interpretable prediction models for the strength of corncob ash-modified geopolymer concrete. *Materials Today Communications*, 38, 107885.
- Zou, J., Wei, Y., Zhang, Y., Liu, Z., Gai, Y., Chen, H., ... & Song, Q. (2024). Remote sensing inversion of soil organic matter in cropland combining topographic factors with spectral parameters. *Frontiers in Environmental Science*, 12, 1420557.