



مدیریت جامع نگر و هوشمند خاک و آب

Holistic and Smart Soil and Water Management

دانشکده کشاورزی و منابع طبیعی دانشگاه تهران

College of Agriculture & Natural Resources, University of Tehran



ارزیابی الگوریتم‌های یادگیری ماشین در طبقه‌بندی شوری خاک با استفاده از

سامانه گوگل ارث انجین (مطالعه موردی: دشت آبیک)

آیه جاویدفر^۱، فریدون سرمدیان^۲، محمد گل پرور^۳

۱- دانشجوی کارشناسی ارشد علوم و مهندسی خاک، دانشکده کشاورزی و منابع طبیعی دانشگاه تهران

۲- استاد گروه علوم و مهندسی خاک، دانشکده کشاورزی و منابع طبیعی دانشگاه تهران؛*fsarmad@ut.ac.ir

۳- دانشجوی کارشناسی ارشد علوم و مهندسی خاک، دانشکده کشاورزی و منابع طبیعی دانشگاه تهران

چکیده

افزایش شوری خاک از مهم‌ترین چالش‌های تهدیدکننده پایداری منابع خاک در مناطق خشک است. این پژوهش با هدف طبقه‌بندی شوری خاک با استفاده از الگوریتم‌های Random Forest و CART در بستر Google Earth Engine در دشت آبیک انجام شد. برای این منظور، ۱۷۲ نمونه خاک سطحی برداشت و هدایت الکتریکی آنها اندازه‌گیری شد. پس از تهیه تصاویر Sentinel-2 با پوشش ابر کمتر از ۱۰ درصد و استخراج شاخص‌هایی نظیر SI، Albedo و DEM، دو مدل یادگیری ماشین برای پیش‌بینی مقادیر EC آموزش داده شدند. سپس با آستانه‌گذاری، نقشه‌های طبقه‌بندی شوری خاک در پنج کلاس تولید شد. مدل Random Forest با R^2 بیشتر و RMSE کمتر، عملکرد بهتری نسبت به CART نشان داد و الگوهای مکانی شوری را با دقت بالاتری بازسازی کرد. در حالی که هر دو مدل از نظر آماری قابل قبول بودند، تفکیک مکانی در مدل CART در کلاس‌های شوری کم و متوسط با محدودیت‌هایی همراه بود. این نتایج نشان داد استفاده از الگوریتم‌های پیشرفته یادگیری ماشین در کنار پردازش ابری گوگل ارث انجین، رویکردی کارآمد برای مدیریت و پایش شوری خاک فراهم می‌آورد که می‌تواند در جهت توسعه کشاورزی پایدار و حفاظت از منابع خاک مؤثر باشد.

واژگان کلیدی: جنگل تصادفی، درخت تصمیم، سنجش از دور، شوری خاک، گوگل ارث انجین

مقدمه

پراکندگی و افزایش شوری خاک یکی از چالش‌های اساسی در مدیریت پایدار منابع خاک و تأمین امنیت غذایی در مناطق خشک و نیمه‌خشک است؛ زیرا تجمع یون‌های سدیم و کاهش قابلیت جذب آب توسط گیاهان، باعث کاهش حاصلخیزی خاک و تخریب ساختار آن می‌شود (Shrivastava and Kumar, 2015). افزون بر این، شوری خاک می‌تواند با محدود ساختن نفوذپذیری، افزایش رواناب سطحی و کاهش ظرفیت نگهداری آب در خاک، کارکردهای طبیعی خاک را مختل کرده و منجر به فرسایش، کاهش عملکرد گیاهان و از بین رفتن پوشش گیاهی شود. اثرات این پدیده در مناطق خشک و نیمه‌خشک به دلیل تبخیر و ترقق بالا و بارندگی محدود، شدت بیشتری می‌یابد. در صورت تداوم شرایط شوری و نبود راهکارهای مدیریتی مناسب، این روند می‌تواند منجر به تخریب گسترده اراضی و تهدیدی جدی برای پایداری اکوسیستم‌های کشاورزی محسوب شود (Singh, 2021).

سامانه Google Earth Engine (GEE) به‌واسطه توان محاسباتی ابری و دسترسی به آرشیوهای وسیع ماهواره‌ای همچون Landsat و Sentinel امکان پردازش سریع شاخص‌های شوری خاک و اجرای الگوریتم‌های پیچیده یادگیری ماشین را در زمانی

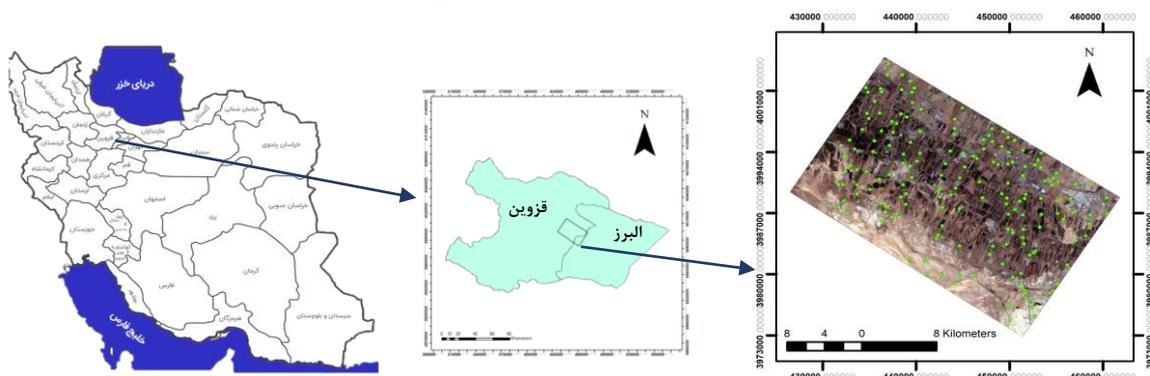
طولانی و مقیاس‌های وسیع فراهم می‌کند. مطالعات متعدد نشان داده‌اند که پیاده‌سازی شاخص‌های شوری و مدل‌های یادگیری ماشین در محیط GEE می‌تواند دقتی هم‌ردیف روش‌های میدانی سنتی ارائه کند و در عین حال هزینه و زمان اجرای پروژه را به طور قابل توجهی کاهش دهد (Aksoy et al., 2022; Zhang et al., 2024). در میان الگوریتم‌های یادگیری ماشین، درخت تصمیم (CART) و جنگل تصادفی (Random Forest) به دلیل سادگی تفسیری، مقاومت در برابر نویز داده‌ها و توانایی مقابله با روابط غیرخطی، کاربرد گسترده‌ای در طبقه‌بندی شوری خاک یافته‌اند. مدل Random Forest با تجمع نتایج چندین درخت CART، قابلیت تعمیم‌پذیری و پایداری بالاتری دارد، اگرچه CART به دلیل ساختار تک‌درختی و ساده‌تر به نمونه‌های آموزشی زیاد نیاز دارد و به بیش‌برازش حساس‌تر است (Aksoy et al., 2022).

بررسی‌ها در زمینه شوری خاک در ایران و جهان، بیانگر کارایی بالای الگوریتم‌های یادگیری ماشین هستند. به عنوان مثال Ma و همکاران (۲۰۲۳) با استفاده از مدل‌های XGBoost، Random Forest و LightGBM و داده‌های Landsat و سامانه Google Earth Engine، تغییرات مکانی-زمانی شوری خاک را در طی ۲۵ سال در چین تحلیل کردند. Kaplan و همکاران (۲۰۲۳) با استفاده از داده‌های Sentinel-2 و مدل‌های یادگیری ماشین شامل M5 Prime، Random Forest، Linear Regression و Instance-Based k، شوری خاک در مناطق بیابانی امارات را پیش‌بینی کردند. جعفری و همکاران (۱۴۰۴) نیز با استفاده از مدل‌های یادگیری ماشین از جمله شبکه عصبی مصنوعی و جنگل تصادفی، نقشه‌های شوری و قلیا بودن خاک را در سه عمق مدیریتی در منطقه آبیگ تهیه کردند.

نوآوری این پژوهش در به‌کارگیری هم‌زمان الگوریتم‌های یادگیری ماشین با پردازش ابری در بستر Google Earth Engine برای طبقه‌بندی شوری خاک در پنج کلاس مشخص نهفته است. در حالی که بسیاری از مطالعات پیشین تنها به پیش‌بینی پیوسته هدایت الکتریکی خاک (EC) پرداخته‌اند، این تحقیق با بهره‌گیری از شاخص‌های منتخب توانست نقشه‌های طبقه‌بندی‌شده و قابل تفسیر شوری خاک را با دقت بالا تولید کند. علاوه بر این، مقایسه عملکرد مکانی و عددی دو الگوریتم و ارزیابی تفصیلی محدودیت‌های مدل CART در تفکیک سطوح مختلف شوری، گامی نوین در تحلیل فضایی و بهینه‌سازی پیش‌بینی شوری خاک محسوب می‌شود. در همین راستا، این پژوهش با هدف ارزیابی و مقایسه عملکرد این الگوریتم‌های یادگیری ماشین در طبقه‌بندی شوری خاک، با بهره‌گیری از قابلیت‌های سامانه گوگل ارث انجین در دشت آبیگ انجام شده است.

مواد و روش‌ها

منطقه مورد مطالعه با مساحتی حدود ۵۵ هزار هکتار، عمدتاً شامل اراضی مرتعی و زراعی بوده و در دشت آبیگ بین استان‌های البرز و قزوین قرار دارد. کشاورزی در این منطقه بیشتر بر کشت محصولات پاییزه نظیر گندم و جو، و همچنین یونجه و ذرت متمرکز است. بخش‌های شمالی عمدتاً دارای کلاس بافت لوم تا لوم شنی و بخش‌های جنوبی دارای کلاس بافت رسی تا لوم رسی می‌باشند. مارن‌های قرمز همراه با لایه‌های گچ و نمک از مهم‌ترین مواد مادری تشکیل‌دهنده خاک‌های این منطقه به شمار می‌روند (موسوی و همکاران، ۱۴۰۰). بر اساس روش نمونه‌برداری تصادفی تعداد ۱۷۲ پروفیل در منطقه حفر گردید و میزان هدایت الکتریکی آن در خاک سطحی (عمق ۰ تا ۲۵ سانتی متر) به روش استاندارد اندازه‌گیری شدند (Spark, 1996). شکل ۱ موقعیت قرارگیری منطقه مورد مطالعه به همراه نقاط نمونه‌برداری را نشان می‌دهد.



شکل ۱ - موقعیت منطقه مورد مطالعه به همراه نقاط نمونه برداری شده

داده‌های هدایت الکتریکی (EC) در قالب یک فایل اکسل به همراه مختصات آنها مرتب‌سازی و به عنوان ورودی داده‌های زمینی به سامانه Google Earth Engine اضافه شدند. به منظور جداسازی مناطق شهری و صنعتی، از نرم افزار گوگل ارث برای ماسک کردن استفاده گردید. در گام بعد، تصاویر Sentinel-2 مربوط به بازه زمانی سال ۱۴۰۲ تا ۱۴۰۳ با درصد ابر کمتر از ۱۰ درصد استخراج و میانگین‌گیری شدند. دو مدل یادگیری ماشین، جنگل تصادفی (Random Forest) با ۳۰۰ درخت و درخت تصمیم (CART)، برای پیش‌بینی شوری خاک به صورت رگرسیونی آموزش داده شدند. برای آموزش مدل‌ها، علاوه بر داده‌های EC، از شاخص‌های سنجش از دور که شامل شاخص شوری (SI)، ضریب آلبدو (Albedo)، و مدل رقومی ارتفاع (DEM) استفاده شد (جدول ۱). شاخص SI یکی از مهم‌ترین شاخص‌های مرتبط با شوری خاک است. ضریب آلبدو بیانگر نسبت انرژی بازتاب‌شده از سطح زمین به انرژی دریافت‌شده از خورشید است و می‌تواند وجود پوشش گیاهی یا نمک‌های سطحی را نمایان سازد. به‌طور کلی، سطوح دارای پوشش گیاهی متراکم آلبدوی کمتری دارند، در حالی که خاک‌های شور و بدون پوشش گیاهی دارای آلبدوی بالاتری هستند (Liu et al., 2021).

جدول ۱- شاخص‌های استفاده شده به عنوان ورودی مدل‌های یادگیری ماشین

علامت اختصاری	متغیر محیطی (واحد)	معادله	منبع	مرجع
SI	شاخص شوری (-)	$SI = \sqrt{B \times R}$	تصاویر ماهواره‌ای	Allbed and Kumar (2013)
Albedo	ضریب آلبدو (-)	$Albedo = 0.356B + 0.13R + 0.373NIR + 0.085SR1 + 0.072SR2 - 0.0018$	تصاویر ماهواره‌ای	Liu et al., (2021)
DEM	ارتفاع (متر)	-	مدل رقومی ارتفاع	USGS/SRTMGL1_003

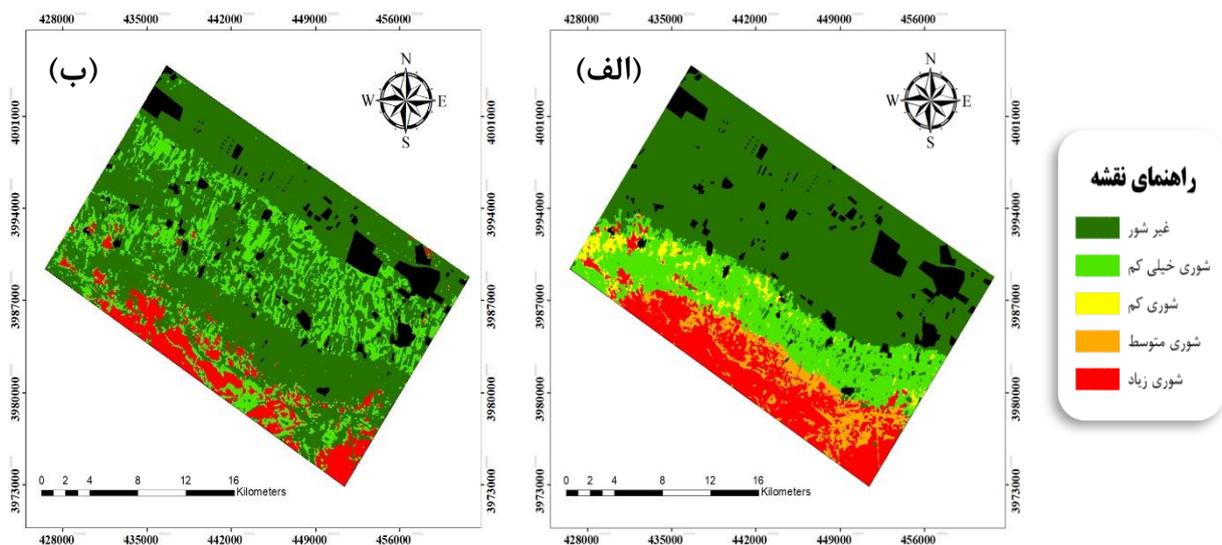
داده‌ها سپس به صورت تصادفی به دو مجموعه آموزشی (۷۰ درصد) و آزمایشی (۳۰ درصد) تقسیم گردیدند. برای ارزیابی عملکرد مدل‌ها، معیارهای ضریب تعیین (R^2) و ریشه میانگین مربعات خطا (RMSE) بر روی داده‌های آزمایشی محاسبه شد. خروجی مدل‌های رگرسیونی، بر اساس آستانه‌های تعریف‌شده در جدول ۲، به ۵ کلاس (غیر شور، شوری خیلی کم، شوری کم، شوری متوسط، شوری زیاد) تقسیم شدند و نقشه‌های نهایی طبقه‌بندی شوری تولید گردیدند. در نهایت برای حذف نویزهای طبقه‌بندی و یکنواخت‌سازی نقشه نهایی، فیلتر (Majority) با شعاع ۲ پیکسل اعمال شد.

جدول ۲- طبقه‌بندی شوری خاک بر اساس میزان هدایت الکتریکی (Brown et al., 1954)

هدایت الکتریکی (dS/m)	کلاس شوری
کمتر از ۲	غیر شور
۲-۴	شوری خیلی کم
۴-۸	شوری کم
۸-۱۶	شوری متوسط
بیشتر از ۱۶	شوری زیاد

نتایج و بحث

نقشه‌های خروجی حاصل از دو مدل Random Forest و CART که در شکل ۲ ارائه شده‌اند، تفاوت قابل توجهی در کیفیت مکانی طبقه‌بندی شوری خاک نشان می‌دهند. همان‌گونه که از این شکل پیداست، مدل Random Forest توانسته الگوهای مکانی شوری را با انسجام و همبستگی بالاتری نسبت به داده‌های زمینی بازسازی کند. این مدل نه تنها مرز بین کلاس‌های مختلف شوری را با دقت بیشتری تفکیک کرده، بلکه در مقایسه با CART، مناطق با شوری متوسط تا بالا را نیز با دقت بهتری شناسایی کرده است.



شکل ۲ - نقشه طبقه‌بندی شوری خاک با الگوریتم الف (Random Forest) و ب (CART). مناطق شهری به رنگ سیاه نمایش داده شدند

نتایج کمی ارزیابی مدل‌ها نیز که در جدول ۳ آمده‌اند، مؤید همین موضوع‌اند. مدل Random Forest با ضریب تعیین (R^2) بسیار بالا و مقدار RMSE کمتر، بالاترین دقت را در برآورد مقادیر EC نشان داد. در حالی که مدل CART با وجود دستیابی به R^2 نسبتاً بالا، دارای خطای بصری بالاتری بود. در حالی که هر دو مدل از منظر عددی قابل قبول هستند، توان تفکیک مکانی به‌ویژه در کلاس‌های شوری کم و متوسط، در مدل CART محدودتر بوده است و منجر به طبقه‌بندی نادرست شده است.

جدول ۳- مقایسه صحت کارایی مدل‌ها بر اساس معیارهای ضریب تعیین و ریشه میانگین مربعات خطا

الگوریتم یادگیری ماشین	ضریب تعیین (R^2)	ریشه میانگین مربعات خطا (RMSE)
Random Forest	۰/۹۹۶	۰/۸۴
CART	۰/۹۹۰	۱/۳۵

این یافته‌ها با نتایج پژوهش Aksoy و همکاران (۲۰۲۲) مطابقت دارد. در آن مطالعه نیز مدل Random Forest نسبت به CART و سایر الگوریتم‌ها، عملکرد دقیق‌تری در بازنمایی مکانی شوری خاک داشت؛ حتی در شرایطی که اختلاف چندانی از نظر شاخص R^2 مشاهده نمی‌شد. همان‌طور که در آن پژوهش نیز تأکید شده، دقت عددی تنها معیار کافی برای ارزیابی مدل‌های شوری خاک نیست، بلکه سنجش تطابق مکانی نقشه‌های خروجی با واقعیت میدانی نیز اهمیت ویژه‌ای دارد.

نتیجه‌گیری

این مطالعه نشان داد که الگوریتم Random Forest در طبقه‌بندی شوری خاک، نسبت به CART، قابلیت تفکیک مکانی و بازنمایی دقیق‌تر الگوهای شوری خاک را دارد، به‌گونه‌ای که می‌تواند داده‌های سنجش از دور را با واقعیت میدانی همخوانی بیشتری دهد. این برتری، علی‌رغم نزدیکی نسبی معیارهای صحت‌سنجی دو مدل، اهمیت کیفیت مکانی نقشه‌های شوری را در کاربردهای عملی برجسته می‌کند. علاوه بر الگوریتم‌ها، به‌کارگیری Google Earth Engine به‌عنوان سامانه‌ای قدرتمند، زمینه اجرای تحلیل‌های پیشرفته و پردازش حجم بالای داده‌های سنجش از دور را به شکل بهینه فراهم نمود. این بستر محاسباتی، امکان تسریع روند تحلیل‌ها و تولید نقشه‌های طبقه‌بندی با دقت مکانی قابل قبول را بدون نیاز به زیرساخت‌های سخت‌افزاری پیچیده فراهم کرد و بهره‌وری پژوهش را ارتقا داد.

بنابراین، نتایج این مطالعه بر ضرورت ترکیب الگوریتم‌های یادگیری ماشین قوی با ابزارهای پردازش ابری تاکید دارد تا بتواند در حوزه پایش و مدیریت شوری خاک، راهکارهای کارآمد و عملیاتی ارائه دهد. این رویکرد می‌تواند زمینه‌ساز توسعه پایدار کشاورزی و حفاظت از منابع خاک در مناطق تحت تأثیر شوری باشد.

فهرست منابع

۱. جعفری، اعظم، سرمیدیان، فریدون و رسائی، زهرا. (۱۴۰۴). عدم قطعیت و نقشه‌برداری مکانی شوری و قلیا بودن خاک با استفاده از روش‌های یادگیری ماشین در سه عمق مدیریتی مختلف در منطقه آبیک. *تحقیقات آب و خاک ایران*, 56(3), 607-629. doi: 10.22059/ijswr.2025.382783.669797
۲. موسوی، سیدروح اله، سرمیدیان، فریدون، امید، محمود و بوگارت، پاتریک. (۱۴۰۰). کاربرد مدل‌های یادگیری ماشین در برآورد مکانی فسفر و پتاسیم خاک در بخشی از اراضی دشت آبیک. پژوهش‌های خاک، ۳۵(۴)، ۳۹۷-۴۱۱. doi: 10.22092/ijsr.2022.355198.618
3. Allbed, A., & Kumar, L. (2013). Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. *Advances in remote sensing*, 2(4), 373-385.
4. Aksoy, S., Yildirim, A., Gorji, T., Hamzehpour, N., Tanik, A., & Sertel, E. (2022). Assessing the performance of machine learning algorithms for soil salinity mapping in Google Earth Engine platform using Sentinel-2A and Landsat-8 OLI data. *Advances in Space Research*, 69(2), 1072-1086.
5. Brown, J.W., Hayward, H.E., Richards, A., Bernstein, L., Hatcher, J.T., Reeve, R.C., Richards, L.A., 1954. *Diagnosis and improvement of saline and alkali soils*. United States Department of Agriculture, Agriculture handbook.
6. Liu, J., Zhang, L., Dong, T., Wang, J., Fan, Y., Wu, H., Geng, Q., Yang, Q., & Zhang, Z. (2021). The Applicability of Remote Sensing Models of Soil Salinization Based on Feature Space. *Sustainability*, 13(24), 13711. <https://doi.org/10.3390/su132413711>
7. Ma, S., He, B., Xie, B., Ge, X., & Han, L. (2023). Investigation of the spatial and temporal variation of soil salinity using Google Earth Engine: a case study at Werigan–Kuqa Oasis, West China. *Scientific Reports*, 13(1), 2754.
8. Kaplan, G., Gašparović, M., Alqasemi, A. S., Aldhaheri, A., Abuelgasim, A., & Ibrahim, M. (2023). Soil salinity prediction using machine learning and sentinel-2 remote sensing data in hyper-arid areas. *Physics and Chemistry of the Earth, Parts a/b/c*, 130, 103400.
9. Shrivastava, P., & Kumar, R. (2014). Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi journal of biological sciences*, 22(2), 123.
10. Singh, A. (2022). Soil salinity: A global threat to sustainable development. *Soil Use and Management*, 38(1), 39-67.
11. Sparks, D.L. (1996). *Methods of Soil Analysis. Part. 3: Chemical Methods*; Soil Science Society of America, American Society of Agronomy: Madison, WI, USA, ISBN 978-0-89118-825-4.
12. Zhang, Y., Wu, H., Kang, Y., Fan, Y., Wang, S., Liu, Z., & He, F. (2024). Mapping the Soil Salinity Distribution and Analyzing Its Spatial and Temporal Changes in Bachu County, Xinjiang, Based on Google Earth Engine and Machine Learning. *Agriculture*, 14(4), 630.

Evaluation of Machine Learning Algorithms for Soil Salinity Classification Using Google Earth Engine (Case Study: The Abyek Plain)

Ayeh Javidfar, Fereydoon Sarmadian, Mohammad Golparvar

MSc student of Soil Science and Engineering, College of Agriculture & Natural Resources, University of Tehran

Professor, Department of Soil Science and Engineering, College of Agriculture & Natural Resources, University of Tehran

MSc student of Soil Science and Engineering, College of Agriculture & Natural Resources, University of Tehran

Abstract

Increasing soil salinity is one of the most critical challenges threatening the sustainability of soil resources in arid regions. This study aimed to classify soil salinity using the Random Forest and Classification and Regression Tree (CART) algorithms on the Google Earth Engine platform in the Abyek Plain. To this end, 172 surface soil samples were collected and their electrical conductivity (EC) measured. After acquiring Sentinel-2 imagery with cloud cover less than 10 % and extracting indices such as the Salinity Index (SI), albedo, and digital elevation model (DEM), two machine learning models were trained to predict EC values. The model outputs were then thresholded to generate five-class soil salinity maps. The Random Forest model demonstrated superior performance compared to CART, achieving a higher coefficient of determination (R^2) and lower root-mean-square error (RMSE), and reconstructed spatial salinity patterns with greater accuracy. Although both models were statistically acceptable, CART exhibited limitations in spatial delineation of low- and medium-salinity classes. These results indicate that employing advanced machine learning algorithms alongside Google Earth Engine's cloud processing presents an efficient approach for soil salinity management and monitoring, which can contribute to sustainable agricultural development and soil resource conservation.

Keywords: Random Forest, Classification and Regression Tree (CART), Remote Sensing, Soil Salinity, Google Earth Engine