



Spatial-temporal modeling of soil moisture using optical and thermal remote sensing data and machine learning algorithms

Atefeh Nouraki¹ | Mona Golabi² | Mohammad Albaji³ | Abd Ali Naseri⁴ | Saeid Homayouni⁵

1. Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: atefeh_nouraki@yahoo.com
2. Corresponding Author, Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: m.golabi@scu.ac.ir
3. Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: m.albaji@scu.ac.ir
4. Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran. E-mail: abdalinaseri@scu.ac.ir
5. Centre Eau Terre Environnement, Institut National de la Recherche Scientifique (INRS), Quebec, Canada. E-mail: saeid.homayouni@inrs.ca

Article Info

Article type: Research Article

Article history:

Received: March. 13, 2023

Revised: May. 18, 2023

Accepted: May. 20, 2023

Published online: June. 22, 2023

Keywords:

High Vegetation,
Machine Learning Algorithms,
Remote Sensing,
Soil Moisture,
Soil Physical Properties.

ABSTRACT

Spatiotemporal estimation and monitoring of soil moisture based on remote sensing observations (optical and thermal) is challenging due to its physical nature in high vegetation conditions, necessitating improving and increasing the accuracy of soil moisture estimation in these areas. Therefore, this research aimed to develop a new approach to estimating surface soil moisture in agricultural fields with dense vegetation using machine learning algorithms by incorporating optical and thermal remote sensing data and soil physical properties. For this objective, 16 Landsat-8 satellite images and more than 430 control locations were used during the sugarcane crop's growth period in 2018-2019 at the Hakim Farabi Sugarcane Agro-Industrial company in the Khuzestan province of Iran. A set of 10 scenarios of various unique combinations of the available input variables were developed and then evaluated by five machine learning algorithms, including multiple linear regression (MLR), decision tree-based algorithms (CART and M5P), and ensemble learning-based algorithms (gradient-boosted regression trees (GBRT) and random forest regression (RFR)). According to the results, the highest correlation between input variables and surface soil moisture was observed in Soil Wetness Index (SWI) and Normalized Soil Moisture Index (NSMI) with R values of 0.79 and 0.69, respectively. Also, the highest accuracy of machine learning algorithms based on R², RMSE, and MAE results was obtained in GBRT (0.99, 0.011, and 0.006) and RFR (0.99, 0.014, and 0.007), respectively. In general, the findings of this research show the importance of using variables based on Landsat-8 remote sensing data in combination with ensemble learning algorithms that can be independent of any ground measurements.

Cite this article: Nouraki, A., Golabi, M., Albaji, M., Naseri, A. A., & Homayouni, S. (2023) Spatial-temporal modeling of soil moisture using optical and thermal remote sensing data and machine learning algorithms, *Iranian Journal of Soil and Water Research*, 54 (4), 637-653. <https://doi.org/10.22059/ijswr.2023.356707.669469>

© The Author(s).

Publisher: University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2023.356707.669469>



مدلسازی زمانی-مکانی رطوبت خاک با استفاده از داده های سنجش از دور نوری-حرارتی و مدل های

یادگیری ماشین

عاطفه نورکی^۱ | منا گلابی^۲ | محمد الباجی^۳ | عبدعلی ناصری^۴ | سعید همایونی^۵۱. گروه آبیاری و زهکشی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران، رایانامه: atefeh_nouraki@yahoo.com۲. نویسنده مسئول، گروه آبیاری و زهکشی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران، رایانامه: monagolabi@yahoo.com۳. گروه آبیاری و زهکشی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران، رایانامه: m.albaji@scu.ac.ir۴. گروه آبیاری و زهکشی، دانشکده مهندسی آب و محیط زیست، دانشگاه شهید چمران اهواز، اهواز، ایران، رایانامه: abdalinaseri@scu.ac.ir۵. گروه سنجش از دور محیطی و ژئوماتیک، مرکز آب، زمین و محیط زیست، INRS-کبک، کانادا، رایانامه: saeid.homayouni@inrs.ca

چکیده

اطلاعات مقاله

برآورد و پایش زمانی-مکانی رطوبت خاک سطحی براساس مشاهدات سنجش از دوری (نوری و حرارتی) بدلیل ماهیت فیزیکی در شرایط پوشش گیاهی متراکم چالش برانگیز است که نیاز به بهبود و افزایش دقت تخمین رطوبت خاک در این مناطق را ضروری می سازد. لذا این پژوهش با هدف توسعه یک رویکرد جدید در برآورد رطوبت خاک سطحی در مزارع کشاورزی با شرایط پوشش گیاهی متراکم، براساس ترکیب داده های سنجش از دور نوری-حرارتی و اطلاعات فیزیکی خاک با استفاده از مدل های یادگیری ماشین انجام شد. بدین منظور از ۱۶ تصویر ماهواره لندست-۸ و بیش از ۴۳۰ نقطه کنترل زمینی در طول دوره رشد گیاه نیشکر در سال زراعی ۱۳۹۸-۱۳۹۹ در منطقه کشت و صنعت نیشکر حکیم فارابی خوزستان استفاده گردید. ۱۰ سناریوی مختلف براساس متغیرهای ورودی طراحی شد و سپس توسط پنج الگوریتم یادگیری ماشین شامل رگرسیون خطی چندگانه، مدل های مبتنی بر درخت تصمیم (classification and regression tree و M5-pruned) و مدل های مبتنی بر یادگیری جمعی (رگرسیون درختان توسعه یافته و رگرسیون جنگل تصادفی) مورد ارزیابی قرار گرفتند. مطابق با نتایج، بیشترین همبستگی متغیرها با رطوبت خاک سطحی در شاخس های خبسی خاک و رطوبت خاک نرمال شده با مقادیر ضریب همبستگی برابر ۰/۷۹ و ۰/۶۹ مشاهده شد. همچنین بیشترین دقت مدل های یادگیری ماشین بر اساس آماره های R2، RMSE و MAE به ترتیب در مدل های رگرسیون درختان توسعه یافته (۰/۹۹، ۰/۰۱۱ و ۰/۰۰۱) و رگرسیون جنگل تصادفی (۰/۹۹، ۰/۰۱۴ و ۰/۰۰۷) به دست آمد. به طور کلی یافته های این پژوهش بیان گر اهمیت استفاده از ویژگی های بیوفیزیکی مستخرج از داده های ماهواره لندست-۸ در ترکیب با مدل های یادگیری جمعی است که می تواند مستقل از هرگونه اندازه گیری زمینی باشد.

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۱/۱۲/۲۲

تاریخ بازنگری: ۱۴۰۲/۲/۲۸

تاریخ پذیرش: ۱۴۰۲/۲/۳۰

تاریخ انتشار: ۱۴۰۲/۴/۱

واژه های کلیدی:

پوشش گیاهی متراکم،
مدل های یادگیری ماشین،
سنجش از دور،
رطوبت خاک،
خصوصیات فیزیکی خاک.

استناد: نورکی؛ عاطفه، گلابی؛ منا؛ الباجی؛ محمد، ناصری؛ عبدعلی، همایونی؛ سعید، (۱۴۰۲) مدلسازی زمانی-مکانی رطوبت خاک با استفاده از داده های سنجش از دور نوری-حرارتی و مدل های یادگیری ماشین، مجله تحقیقات آب و خاک ایران، ۵۴ (۴)، ۶۳۷-۶۵۳

<https://doi.org/10.22059/ijswr.2023.356707.669469>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2023.356707.669469>

مقدمه

رطوبت خاک یکی از کلیدی‌ترین متغیرهای چرخه هیدرولوژیکی است که فرآیندهای مختلف سطح زمین، خاک و اتمسفر از جمله تبخیر و تعرق، نفوذ، رواناب و جریان‌های زیرسطحی را کنترل و به هم مرتبط می‌کند (Robinson et al., 2008). به منظور مدیریت پایدار منابع آب کشاورزی و جلوگیری از تنش آبی گیاهان، بررسی و پایش مکانی-زمانی وضعیت رطوبت خاک بسیار ضروری است. در دو دهه گذشته توانایی روش‌های اندازه‌گیری رطوبت خاک به‌طور قابل توجهی در مقیاس‌های نقطه‌ای تا جهانی افزایش یافته است. به عنوان مثال در مقیاس‌های نقطه‌ای و محلی، بازتاب سنجی حوزه زمانی (TDR)^۱، حسگرهای الکترومغناطیسی (FDR)^۲ (Jones et al., 2005)، کاوشگر نوترونی (SMNP)^۳ (Zreda et al., 2008)، و سایر حسگرهای غیرمخرب مانند روش رادار نفوذی به زمین (GPR)^۴ با موفقیت برای اندازه‌گیری رطوبت خاک استفاده شده است. اگرچه اندازه‌گیری‌های متعدد در بسیاری از نقاط یک منطقه می‌تواند مقدار متوسط رطوبت خاک را برای آن منطقه تعیین کند اما زمانی که اندازه‌گیری‌های زیاد و در مقیاس گسترده مورد نیاز باشد، چنین روش‌هایی دشوار، پرهزینه و وقت‌گیر بوده و استفاده از آن‌ها برای دوره‌های طولانی مدت امکان‌پذیر نمی‌باشد. علاوه بر این توزیع آب در خاک به خواص فیزیکی و هیدرولیکی خاک، توپوگرافی و سایر شرایط محیطی بستگی دارد (Korres et al., 2015) و به دلیل تأثیرات متقابل و پیچیده این متغیرها، استفاده از روش‌های نقطه‌ای برای برآورد دقیق تغییرات مکانی-زمانی رطوبت خاک عملاً غیر کاربردی است و استفاده از تکنولوژی و روش‌های جدید در این زمینه را ضروری می‌سازد.

داده‌های سنجش از دور به دلیل قابلیت بروزرسانی و در دسترس بودن اطلاعات، امکان نظارت مستمر، سریع و مقرون به صرفه رطوبت خاک را در مناطق گسترده فراهم می‌سازد (Acharya et al., 2021). مطالعات مختلفی نشان دادند که بازتابش انعکاسی خاک، تشعشع حرارتی و پراکنش امواج مایکروویو به شکل وسیعی با رطوبت خاک همبستگی دارند (Araya et al., 2021; Fatholouloumi et al., 2020). از طرفی توان تفکیک‌پذیری مکانی درشت مقیاس سنسورهای مایکروویو غیرفعال و همچنین اثرات اختلال زبری خاک و پوشش گیاهی بر پراکنش امواج مایکروویو، امکان استفاده از این سنسورها را محدود می‌کند. درحالی‌که نقاط ضعف این روش را می‌توان تا حد زیادی با استفاده از داده‌های سنجش از دور نوری-حرارتی برطرف کرد. تحقیقات زیادی برای پایش رطوبت خاک سطحی با استفاده از روش‌های ترکیب اطلاعات از نوارهای مرئی، مادون قرمز نزدیک و طول موج کوتاه انجام شده است (Babaeian et al., 2018; Kheillouk et al., 2020). با اینحال نتایج Na et al (2021) نشان داد که داده‌های سنجش از دور نوری، تغییرات پوشش گیاهی را با تأخیر نسبت به تغییرات رطوبت خاک نشان می‌دهند که مقدار زمان تأخیر شاخص‌های نوری به مرحله رشد گیاه بستگی دارد.

روش مثلثی رایج‌ترین روش سنجش از دور حرارتی است که بر پایه رابطه فیزیکی بین نمایه گیاهی و دمای سطح زمین است که به شکل یک نمودار پراکنش نقطه‌ای نمایش داده می‌شود و در آن از مفهوم فضای مثلثی در بازبایی رطوبت خاک سطحی در مناطق با پوشش گیاهی استفاده می‌شود (Potopová et al., 2020). بنابراین در این پژوهش توجه ویژه‌ای به داده‌های رایگان و قابل دسترس نوری-حرارتی ماهواره لندست-۸ در زمینه برآورد رطوبت خاک سطحی شده است. Babaeian et al (2021) در پژوهشی در دانشگاه آریزونا در آمریکا، به بررسی قابلیت شاخص‌های نوری در برآورد رطوبت خاک در منطقه ریشه گیاه گندم پرداختند. نتایج آماری آن‌ها نشان داد که رابطه معنی‌داری بین رابطه تجربی و رطوبت خاک با خطای ۰/۰۲ سانتی متر مکعب بر سانتی متر مکعب وجود دارد. Duethmann et al (2022) در حوضه آبریز Demnitzer Millcreek در آلمان به ارزیابی وضعیت رطوبت خاک با استفاده از داده‌های سنجش از دور پرداختند. نتایج این پژوهش نشان داد بین SWI^۵ و رطوبت اندازه‌گیری شده خاک، همبستگی معنی‌داری وجود دارد (R=۰/۶۴-۰/۱۶). لازم به ذکر است استفاده از رویکردهای تجربی نوری-حرارتی نیاز به بررسی طیف وسیعی از شرایط پوشش گیاهی و رطوبت خاک در یک ناحیه‌ی خاص دارند که همیشه نمی‌توان آن‌ها را برآورد کرد (Barrett & Petropoulos, 2013).

از این رو اخیراً، استفاده از تکنیک‌های مدل‌های یادگیری ماشین برای تخمین رطوبت خاک به دلیل توانایی آن‌ها در مقابله با بسیاری از محدودیت‌ها، توجه بیشتری را به خود جلب کرده است (Fatholouloumi et al., 2020; Han et al., 2018). مزیت اصلی مدل‌های

1 Time-domain reflectometer

2 Frequency domain reflectometry

3 Soil moisture neutron probe

4 Ground penetrating radar

5 Soil water index

یادگیری ماشین توانایی آن‌ها در شناسایی خودکار و کارآمد روابط پیچیده و استخراج ویژگی‌های مکانی-زمانی از انواع مختلف داده‌های خام است که فرصتی برای تخمین دقیق‌تر پروفیل رطوبت خاک فراهم می‌کند. تحقیقاتی زیادی در زمینه برآورد رطوبت خاک با ترکیب روش‌های مختلف سنجش از دور و یادگیری ماشین صورت گرفته است. به طور مثال Fatholouloumi et al (2020) رطوبت خاک را با استفاده از شاخص‌های مشتق شده از داده‌های لندست-۸ و مدل‌های یادگیری ماشین (الگوریتم‌های جنگل تصادفی (RF) و رگرسیون خطی چندگانه

2 (MLR)) پیش‌بینی کردند. نتایج آن‌ها نشان داد که الگوریتم RF بیشترین عملکرد را در مدل‌سازی رطوبت خاک داشته است. Ge et al (2019) با استفاده از شاخص‌های نوری حاصل از تصاویر پهپاد در ترکیب با الگوریتم‌های RF و ماشین یادگیری سریع (ELM) به برآورد رطوبت خاک سطحی در مناطقی از شهر فوکانگ چین پرداختند. نتایج این بررسی حاکی از برتری مدل RF نسبت به مدل ELM بود. همچنین آن‌ها بیان کردند که ترکیب شاخص‌های نوری و الگوریتم RF امکان تخمین رطوبت خاک سطحی را با دقت بالا (ضریب تبیین ۰/۹۱ و درصد انحراف نسبی ۳/۳۹) در مقیاس منطقه‌ای فراهم می‌کند.

از آن‌جا که در غالب مناطق ایران اساساً داده رطوبت خاک اندازه‌گیری نمی‌شود و در صورت استخراج نیز داده‌ها به صورت ایستگاهی و نقطه‌ای خواهد بود لذا یافتن جایگزینی مناسب برای رفع این نقیصه در مدیریت آب کشاورزی بسیار راه‌گشا خواهد بود. با توجه به موارد ذکر شده، اهداف اصلی این مطالعه شامل موارد زیر است:

۱. برآورد قابل اعتماد رطوبت خاک سطحی در مقیاس منطقه‌ای با استفاده از مدل‌های یادگیری ماشین به کمک داده‌های نوری - حرارتی حاصل از تصاویر لندست-۸ و اطلاعات فیزیکی خاک.
۲. بررسی کارایی الگوریتم‌های مختلف یادگیری ماشین در بازیابی رطوبت خاک و معرفی مدل بهینه.
۳. شناسایی مهم‌ترین پارامترهای فیزیکی خاک و متغیرهای سنجش از دور نوری - حرارتی در برآورد رطوبت خاک.

مواد و روش‌ها

منطقه مورد مطالعه

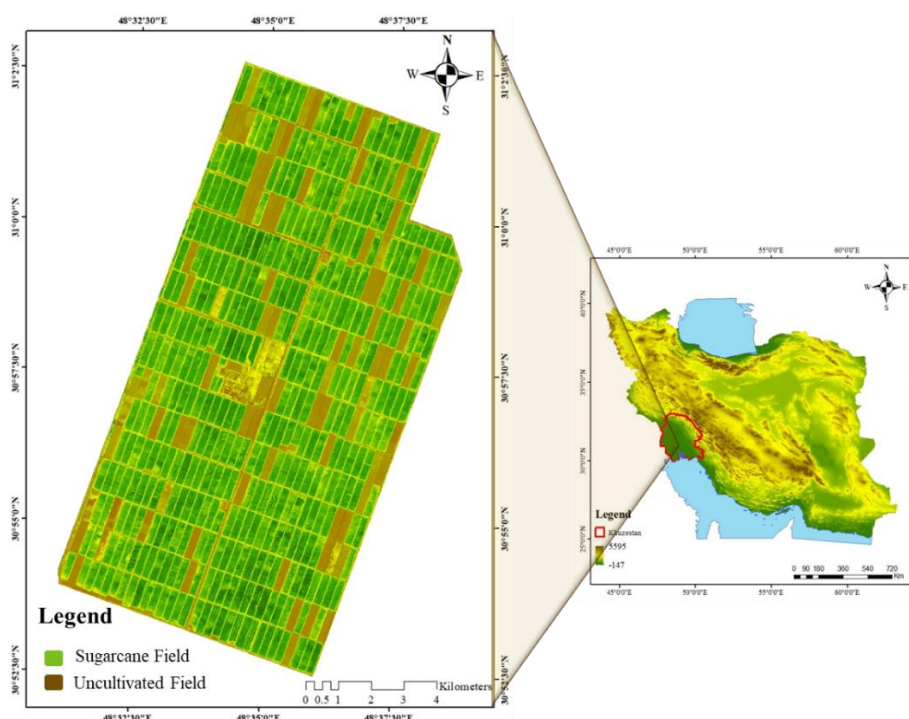
این پژوهش در مزارع نیشکر کشت و صنعت حکیم فارابی در سال زراعی ۱۳۹۹-۱۳۹۸ انجام شد. شرکت کشت و صنعت حکیم فارابی یکی از واحدهای هفت گانه شرکت توسعه نیشکر و صنایع جانبی استان خوزستان است که در ۳۵ کیلومتری جنوب شهرستان اهواز و شرق رودخانه کارون، در عرض جغرافیایی بین ۳۰ و ۳۱ درجه و طول جغرافیایی ۴۸ درجه قرار دارد. مساحت کل کشت و صنعت ۱۴۰۰۰ هکتار و سطح زیر کشت آن ۱۲۰۰۰ هکتار است که دارای ۴۸۰ مزرعه ۲۵ هکتاری (۱۰۰۰×۲۵۰) است. موقعیت جغرافیایی منطقه مورد مطالعه در شکل ۱ آمده است. حداکثر ارتفاع آن از سطح آزاد دریا حدود ۱۲ متر و شیب اراضی این واحد بین ۰/۱ تا ۰/۲ در هزار است. این منطقه دارای آب و هوای خشک و نیمه خشک با میانگین سالانه بارندگی ۲۰۳ میلی‌متر، میانگین سالانه دمای هوا ۲۲/۹۲ درجه سانتی‌گراد و تبخیر سالانه از تشت کلاس A به میزان ۲۲۷۰/۲۴ میلی‌متر می‌باشد.

داده‌های زمینی مورد استفاده

جهت مدل‌سازی دقیق تغییرات رطوبتی در طول دوره مورد مطالعه از داده‌های رطوبت خاک سطحی اندازه‌گیری شده استفاده شد. بدین منظور نمونه‌گیری‌ها در ۲۷ نقطه مختلف در منطقه مورد مطالعه انجام و نمونه‌ها بلافاصله برای توزین به آزمایشگاه منتقل شدند. پس از اندازه‌گیری اولیه، نمونه‌ها به مدت ۲۴ ساعت در دمای ۱۰۸ درجه سانتی‌گراد قرار داده شدند و در نهایت وزن خشک نمونه‌ها اندازه‌گیری و مقدار رطوبت جرمی خاک با استفاده از معادله ۱ استخراج شد. در نهایت با ضرب این مقدار در وزن مخصوص ظاهری، رطوبت حجمی خاک محاسبه گردید.

$$\theta_m = \frac{M_w - M_s}{M_s} \quad \text{رابطه ۱}$$

که در آن θ_m درصد رطوبت جرمی خاک، M_w وزن تر نمونه خاک (گرم) و M_s وزن خشک نمونه خاک (گرم) است.



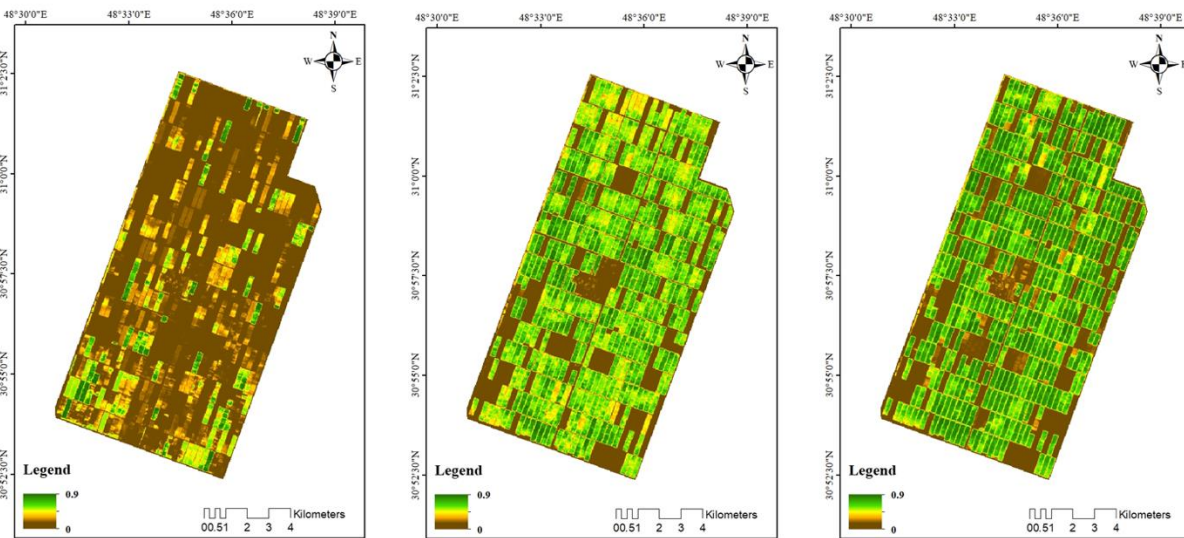
شکل ۱- موقعیت جغرافیایی منطقه مورد مطالعه

به منظور اندازه‌گیری ویژگی‌های فیزیکی خاک، ابتدا نمونه‌های خاک در هوا خشک شدند و پس از کوبیدن از الک (۲ میلی‌متری) عبور داده شدند و سپس فراوانی نسبی ذرات خاک (درصد رس، سیلت و شن) به روش هیدرومتر (Day, 1965) و وزن مخصوص ظاهری^۱ (BD) به روش حجمی (Blake, 1965) محاسبه شد.

تهیه داده‌ها و پردازش تصاویر

در این پژوهش از تصاویر ماهواره لندست-۸ جهت برآورد رطوبت خاک استفاده گردید. ماهواره لندست-۸ در تاریخ ۴ فوریه ۲۰۱۳ از کالیفرنیا، آمریکا به فضا پرتاب شد. در این ماهواره اخذ مستمر داده‌ها توسط دو سنجنده تصویر بردار عملیاتی زمین^۲ (OLI) و مادون قرمز حرارتی^۳ (TIRS) صورت می‌گیرد. سنجنده OLI دارای نه باند طیفی در محدوده مرئی، مادون قرمز نزدیک و دور و سنجنده TIRS دارای دو باند طیفی در محدوده مادون قرمز حرارتی دور می‌باشد. تعداد ۱۶ تصویر بدون ابر از ماهواره لندست-۸ متناظر با اندازه‌گیری‌های زمینی رطوبت خاک، از سایت (USGS^۴) (<https://glovis.usgs.gov/>) دانلود و تصحیحات اتمسفری و رادیومتریک آن‌ها با نرم افزار^۵ Envi 5.3.1 انجام شد. در نهایت پس از محاسبه مقادیر بازتابش باندهای طیفی، شاخص‌های گیاهی^۶ (NDVI)، رطوبتی (NSMI^۷، NTR^۸) و حرارتی (SWI) مورد نظر استخراج شدند (جدول ۱). لازم به ذکر است که دمای سطح زمین^۹ (LST) به روش کانال تکی محاسبه شده است. NDVI یک شاخص سبزی‌نگی است که با نسبت تابش جذب شده توسط فتوسنتز مرتبط است و فعالیت کلروفیل گیاهان را منعکس می‌کند. شکل (۲) نقشه تغییرات مکانی شاخص NDVI را در طول دوره رشد نشان می‌دهد. با بررسی تغییرات مکانی شاخص NDVI مشاهده می‌شود که روند تغییرات آن در طول دوره مورد بررسی به صورت افزایشی بوده است؛ به طوری که در مهرماه مساحت بیشتری از منطقه دارای NDVI بیشتر از ۰/۷ بود.

1 Balk density
 2 Operational land imager
 3 Thermal infrared sensor
 4 United states geological survey
 5 Environment for visualizing images
 6 Normalized difference vegetation index
 7 Normalized soil moisture index
 8 NIR transformed reflectance
 9 Land surface temperature



۱۲ اسفند ۱۳۹۸

۲ تیر ۱۳۹۹

۵ مهر ۱۳۹۹

شکل ۲. تغییرات مکانی شاخص NDVI در طول دوره رشد

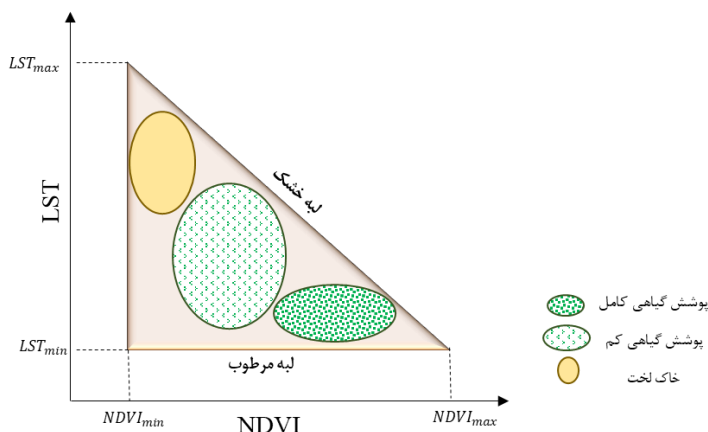
NTR یک شاخص فیزیکی است که از نظریه انتقال تابشی دو شار Kubelka-Munk مشتق شده و تقریباً با محتوای آب سطحی گیاه همبستگی دارد. NSMI یک شاخص رطوبتی است که می‌تواند اطلاعات مناسبی از وضعیت رطوبت خاک سطحی را در شرایط رطوبتی پایین‌تر از ۵۰ درصد حجمی ارائه کند (Lobell & Asner, 2002). شاخص رطوبتی SWI بر پایه رابطه فیزیکی بین شاخص پوشش گیاهی و LST عمل می‌کند و به شکل یک مثلث پراکنش نقطه‌ای نمایش داده می‌شود.

جدول ۱. شاخص‌های استخراج شده از تصاویر لندست-۸

منبع	فرمول	نماد	متغیر
Gates et al. (1965)	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	NDVI	شاخص تفاضلی پوشش گیاهی نرمال شده
Haubrock et al. (2008)	$\frac{\rho_{SWIR1} - \rho_{SWIR2}}{\rho_{SWIR1} + \rho_{SWIR2}}$	NSMI	شاخص رطوبت خاک نرمال شده
Kubelka and Munk (1931)	$\frac{(1 - \rho_{NIR})^2}{2 \rho_{NIR}}$	NTR	بازتاب تبدیل شده NIR
Sandholt et al. (2002)	$\frac{LST_{Max} - LST}{LST_{Max} - LST_{Min}}$	SWI	شاخص خیسگی خاک

که در آن: ρ_{NIR} انعکاس مادون قرمز نزدیک، ρ_{Red} انعکاس باند قرمز، ρ_{SWIR1} : انعکاس مادون قرمز طول موج کوتاه ۱، ρ_{SWIR2} : انعکاس مادون قرمز طول موج کوتاه ۲، LST: دمای سطح زمین، LST_{Max} : لبه خشک و LST_{Min} : لبه مرطوب در فضای مثلثی

این شاخص بسته به شرایط آب و هوایی و پوشش گیاهی متغیر است و برای استخراج رطوبت خاک از مفهوم فضای مثلثی استفاده می‌کند (شکل ۳). بدین منظور با استفاده از روابط ۲ و ۳ لبه‌های خشک (LST_{max}) و مرطوب (LST_{min}) هر تصویر محاسبه می‌شود و سپس براساس فضای مثلثی، شاخص SWI قابل محاسبه می‌شود.



شکل ۳. نمودار مثلثی دما/ پوشش گیاهی (Sandholt et al., 2002)

$$LST_{max} = a + b \text{ NDVI} \quad \text{رابطه ۲}$$

$$LST_{min} = c + d \text{ NDVI} \quad \text{رابطه ۳}$$

که در آن‌ها، NDVI شاخص تفاضلی پوشش گیاهی نرمال شده و a ، b ، c و d ضرایب‌های خط هستند.

مدل‌های یادگیری ماشین

رگرسیون خطی چندگانه

روش MLR، امکان تحلیل هم‌زمان اثر تعدادی متغیر مستقل بر یک متغیر وابسته را فراهم می‌نماید. از روش گام به گام به منظور وارد کردن متغیرها در رگرسیون خطی چندگانه استفاده شد. در این روش تمامی متغیرهای مستقل وارد مدل می‌شوند و اگر نقش معنی‌داری در رگرسیون نداشته باشند از آن حذف می‌شوند (Eshaghi et al., 2010). معادله رگرسیون خطی چندگانه که در این پژوهش استفاده شده، در معادله ۴ نشان داده شده است.

$$Y = a + bX_1 + cX_2 + dX_3 + \dots \quad \text{رابطه ۴}$$

که در آن Y رطوبت خاک، X_1, X_2, X_3, \dots پارامترهای پیش‌بینی کننده و a, b, c, \dots ضرایب معادله می‌باشد.

مدل‌های مبتنی بر درخت تصمیم

درختان تصمیم در کلاس یادگیری ماشین‌های نظارت شده قرار می‌گیرند و به دلیل پیاده‌سازی و تفسیر آسان، می‌توانند با نتایج قابل اعتمادی برای انواع متغیرهای پیوسته و گسسته اعمال شوند. درخت CART یک روش رگرسیونی است که در سال ۱۹۸۴ توسط Breiman et al. (1984) معرفی شد. الگوریتم CART داده‌های ورودی را به صورت بازگشتی دودویی تقسیم می‌کند و همچنین قادر به پردازش متغیرهای عددی و دسته‌ای است. این الگوریتم گره‌های والدین را دقیقاً به دو گروه فرزند منشعب می‌کند و به‌طور بازگشتی منشعب کردن را تا زمانی که انشعاب دیگری نتواند ساخته شود، ادامه می‌دهد. پس از ساخت درخت، عملیات هرس کردن درخت انجام می‌شود. Wang and Witten (1996) الگوریتم M5P را از الگوریتم M5 برای پیش‌بینی داده‌های پیوسته بازسازی و پیشنهاد کردند. مزیت اصلی M5P این است که با کاهش اندازه درخت می‌تواند در مجموعه داده‌های بزرگ بهتر از M5 عمل کند. این مدل برخلاف سایر مدل‌های یادگیری ماشین (ANN^۱ و SVR^۲) به صراحت الگوها و روابط ضمنی داده‌ها را بر اساس چندین اصل، قانون و معادلات رگرسیونی توصیف می‌کند. توسعه درخت M5P شامل سه مرحله اصلی ساخت درخت، هرس درخت و صاف کردن درخت است. برای توضیح بیشتر به پژوهش Wang and Witten (1996) مراجعه شود.

مدل‌های مبتنی بر یادگیری جمعی

در سال‌های اخیر، با توسعه فناوری کامپیوتر و تئوری یادگیری ماشین، الگوریتم‌های یادگیری جمعی به طور فزاینده‌ای در زمینه‌های مختلف، به‌ویژه در تحقیقات کشاورزی مورد استفاده قرار گرفته‌اند (Babaeian et al., 2021; Nguyen et al., 2022). یادگیری جمعی یا گروهی عمدتاً برای ترکیب چند یادگیرنده به منظور به دست آوردن یک مدل بهتر و جامع با نظارت قوی استفاده می‌شود. الگوریتم‌های رایج یادگیری جمعی شامل الگوریتم‌های تقویتی (Boosting)، کیسه‌گذاری (Bagging) و انباشته‌سازی (Stacking) هستند. روش تقویتی یک روش مرحله‌ای روبه‌جلو است که در آن، مدل‌های درختی به صورت تکرارپذیر با زیرمجموعه‌ای از داده‌های آموزشی برازش داده می‌شوند. GBRT توسط Friedman (2001) به عنوان یک تکنیک یادگیری جمعی با الگوریتم تقویتی معرفی شد که در قالب مجموعه‌ای از مدل‌های درختی به صورت متوالی آموزش داده می‌شود. روش کیسه‌گذاری یا تجمیع بوت استرپ، یک روش مجموعه‌ای است که شامل آموزش چندین بار الگوریتم با استفاده از زیر مجموعه‌های مختلف نمونه برداری شده از داده‌های آموزشی است. RFR توسط Friedman (2001) به عنوان یک تکنیک یادگیری جمعی با الگوریتم کیسه‌گذاری معرفی شد.

در این پژوهش به منظور مدل‌سازی، مجموعه داده‌ها به ۷۰ درصد داده‌های آموزش و ۳۰ درصد داده‌های آزمایش تقسیم شدند. جهت پیش‌بینی مدل MLR از نرم افزار SPSS^۱ استفاده شد و مدل‌سازی سایر الگوریتم‌های یادگیری ماشین در محیط Spyder با زبان برنامه نویسی Python انجام شد. همچنین از روش اعتبار سنجی متقابل 10-folds به منظور بهینه‌سازی و تنظیم فرآیندهای مدل‌های یادگیری ماشین استفاده شد. جدول ۲ مشخصات سناریوها و متغیرهای ورودی در برآورد رطوبت خاک را نشان می‌دهد. مراحل انجام این پژوهش در شکل ۴ ارائه شده است.

جدول ۲. سناریوهای تعریف شده در این پژوهش

متغیرهای ورودی	سناریو
B, G, R, NIR, SWIR1, SWIR2	۱
B, G, R, NIR, SWIR1, SWIR2, LST	۲
B, G, R, NIR, SWIR1, SWIR2, LST, SWI	۳
B, G, R, NIR, SWIR1, SWIR2, NDVI, NSMI, NTR	۴
B, G, R, NIR, SWIR1, SWIR2, LST, SWI, NDVI, NSMI, NTR,	۵
B, G, R, NIR, SWIR1, SWIR2, LST, NDVI, NSMI, NTR, SWI, Sand, Silt, Clay, BD	۶
B, G, R, NIR, SWIR1, SWIR2, LST, Sand, Silt, Clay, BD	۷
NDVI, NSMI, NTR, SWI, Sand, Silt, Clay, BD	۸
LST, NIR, SWIR2, NDVI, NSMI, NTR, SWI	۹
SWIR2, LST, SWI	۱۰

که در آن: B: انعکاس باند آبی، G: انعکاس باند سبز، R: انعکاس باند قرمز، NIR: انعکاس مادون قرمز نزدیک، SWIR1: انعکاس مادون قرمز طول موج کوتاه ۱، SWIR2: انعکاس مادون قرمز طول موج کوتاه ۲، LST: دمای سطح زمین، SWI: شاخص خیسگی خاک، NDVI: شاخص تفاضلی پوشش گیاهی نرمال شده، NSMI: شاخص رطوبت خاک نرمال شده، NTR: بازتاب تبدیل شده NIR، Sand: شن، Silt: ماسه، Clay: رس، BD: چگالی ظاهری

ارزیابی دقت مدل‌ها

به منظور ارزیابی دقت و میزان عملکرد مدل‌های بررسی شده در این پژوهش و مقایسه آن‌ها با رطوبت خاک اندازه‌گیری شده از معیارهای ضریب تبیین (R^2)، جذر میانگین مربعات خطا ($RMSE^2$) و میانگین قدر مطلق خطا (MAE^3) براساس روابط زیر استفاده شد.

$$R^2 = \frac{[\sum_{i=1}^N (\theta_{obs} - \bar{\theta}_{obs})(\theta_{est} - \bar{\theta}_{est})]^2}{\sum_{i=1}^N (\theta_{obs} - \bar{\theta}_{obs})^2 \sum_{i=1}^N (\theta_{est} - \bar{\theta}_{est})^2} \quad (\text{رابطه ۵})$$

1 Statistical package for the social sciences

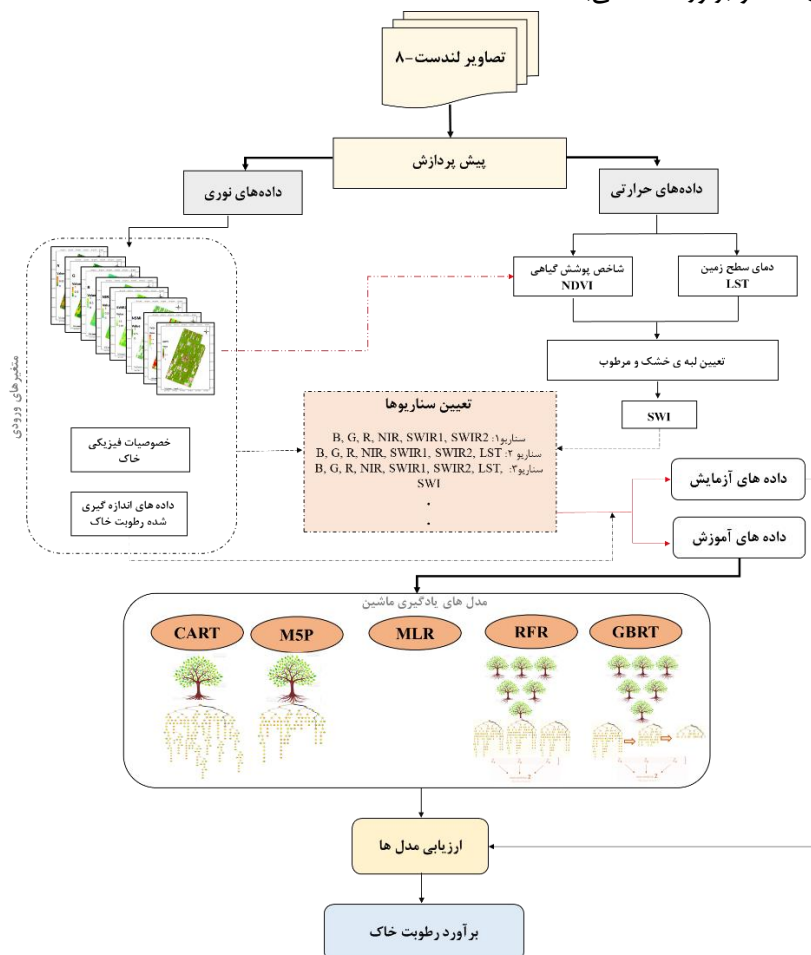
2 Root mean square error

3 Mean absolute error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_{est} - \theta_{obs})^2} \quad \text{رابطه ۶}$$

$$MAE = \frac{\sum_{i=1}^N |\theta_{est} - \theta_{obs}|}{N} \quad \text{رابطه ۷}$$

که در اینجا θ_{obs} مقدار رطوبت خاک اندازه گیری شده و θ_{est} مقدار رطوبت خاک برآورد شده، $\bar{\theta}_{est}$ و $\bar{\theta}_{obs}$ به ترتیب میانگین رطوبت خاک اندازه گیری شده و برآورد شده می باشند.



شکل ۴. روند کلی تحقیق

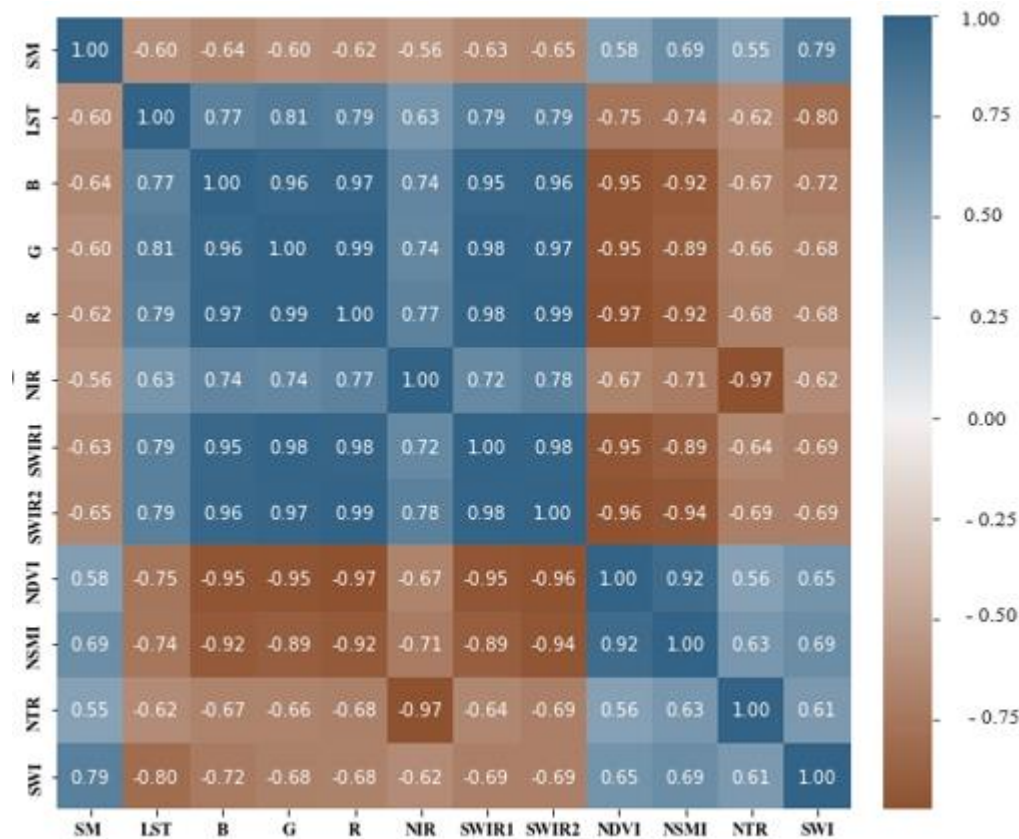
نتایج و بحث

همبستگی رطوبت خاک اندازه گیری شده با داده های سنجنش از دور

در گام نخست به کمک آزمون همبستگی پیرسون به بررسی ارتباط بین رطوبت خاک اندازه گیری شده و متغیرهای کمکی مشتق شده از تصاویر لندست-۸ (باندها و شاخص های نوری) پرداخته شد (شکل ۵). نتایج این آزمون همبستگی معنی داری را بین رطوبت خاک و ۱۱ متغیر مبتنی بر داده های سنجنش از دور نشان می دهد ($p < 0.05$). بالاترین مقادیر همبستگی با رطوبت خاک سطحی توسط متغیرهای SWI، NSMI، SWIR2 و B مشاهده شد؛ در حالی که باندهای R، G، NIR، SWIR1 و LST کمترین میزان همبستگی را نشان دادند (شکل ۵). همان طور که Nolet et al (2014) بر اساس رفتار هیدرولیکی آب در شرایط غیراشباع بیان کردند، همبستگی ضعیف باندهای نوری احتمالاً بدلیل رفتار غیرخطی و کاهش بین رطوبت خاک و بازتاب طیفی حاصل از آن می باشد. از طرفی بازتاب باند NIR در خاک لخت و همراه با پوشش گیاهی کاملاً متفاوت است که منجر به پیچیده تر شدن استفاده از رابطه بازتاب NIR در تخمین رطوبت سطحی در مکان هایی با مخلوطی از خاک و پوشش گیاهی می گردد (Khellouk et al., 2020). در همین راستا Adab et al (2020) در مطالعه ای به بررسی روابط بین رطوبت خاک سطحی و باندهای نوری - حرارتی پرداختند. نتایج آن ها حاکی از رفتار متفاوت باندها در شرایط

مختلف پوشش سطح خاک بود؛ به طوری که باند SWIR1 و SWIR2 به ترتیب بیشترین همبستگی را با رطوبت خاک سطحی لخت و همراه با پوشش گیاهی داشتند.

به طور کلی شاخص‌های نوری با توانایی کاهش خطای پس زمینه جهت استخراج اطلاعات جدید و ساده‌سازی فرآیندهای تفسیر و پردازش تصاویر ماهواره‌ای، معمولاً بهتر از باندهای نوری منفرد می‌توانند تغییرات رطوبت خاک را نشان دهند که به طور خاص این ویژگی در نتایج Wang et al (2020) مشاهده می‌شود. Babaeian et al (2021) و Acharya et al (2021) نشان دادند که بازتاب طیفی خاک علاوه بر محتوای آب می‌تواند تحت تأثیر عوامل ذاتی خاک مانند مقدار ماده آلی، توزیع اندازه ذرات، ترکیب معدنی، زبری سطح و رنگ عناصر خاک قرار گیرد. بنابراین بررسی خواص فیزیکی خاک به عنوان یک متغیر کمکی در رویکردهای طیفی برآورد رطوبت خاک سطحی می‌تواند تأثیر بسزایی داشته باشد.



شکل ۵- ضریب همبستگی پیرسون (R) بین داده‌های سنجش از دور حاصل از لندست-۸ و رطوبت خاک اندازه‌گیری شده

ارزیابی عملکرد مدل‌های یادگیری ماشین

جدول ۳ نتایج به کارگیری ۱۰ سناریو مختلف (جدول ۲) در پیش‌بینی رطوبت خاک با استفاده از مدل‌های یادگیری ماشین را نشان می‌دهد. از نظر شاخص‌های اندازه‌گیری (R^2 , RMSE و MAE) تقریباً همه الگوریتم‌های یادگیری ماشین مورد مطالعه عملکرد قابل قبولی داشتند. برآوردهای رطوبت خاک در تمام مدل‌های یادگیری ماشین روندی تقریباً یکسان در سناریوهای ۱ تا ۱۰ داشت (جدول ۳). عملکرد سناریو اول با ترکیب داده‌های سنجش از دور نوری (شامل باندهای الکترومغناطیسی RGB, NIR و SWIR) در مدل‌های یادگیری ماشین مورد مطالعه منجر به کمترین مقدار R^2 (۰/۹۸ - ۰/۴۴) و بیشترین مقدار RMSE (۰/۱۱ - ۰/۰۲ $\text{cm}^3\text{cm}^{-3}$) و MAE (۰/۰۶ - ۰/۰۱ $\text{cm}^3\text{cm}^{-3}$) شد. در حالی که با اضافه شدن باند حرارتی و شاخص SWI به داده‌های ورودی (سناریوهای ۲ و ۳) عملکرد مدل‌های یادگیری ماشین به طور چشم‌گیری بهبود یافت. همچنین سناریو سوم در تمام مدل‌های یادگیری ماشین تخمین‌های دقیق‌تری از سناریوی دوم ارائه داد (جدول ۳).

جدول ۳. مقایسه مقادیر ضریب تبیین (R²)، جذر میانگین مربعات خطا (RMSE) و میانگین مربعات خطا (MAE) در پیش بینی رطوبت خاک سطحی با استفاده از پنج الگوریتم یادگیری ماشین برای ۱۰ سناریو پیشنهادی در مرحله آموزش

سناریو	R ²					(cm ³ cm ⁻³) RMSE					(cm ³ cm ⁻³) MAE				
	RFR	GBRT	M5P	CART	MLR	RFR	GBRT	M5P	CART	MLR	RFR	GBRT	M5P	CART	MLR
۱	۰/۸	۰/۹۸	۰/۴۸	۰/۵	۰/۴۴	۰/۰۷۴	۰/۰۲۳	۰/۱۱	۰/۱۱	۰/۰۹۶	۰/۰۵۹	۰/۰۰۷	۰/۰۹۹	۰/۰۸۹	۰/۰۷۸
۲	۰/۹۱	۰/۹۸	۰/۷۵	۰/۵۸	۰/۴۴	۰/۰۵	۰/۰۲۴	۰/۰۷۹	۰/۱۰	۰/۰۹۶	۰/۰۳۹	۰/۰۰۷۳	۰/۰۶۵	۰/۰۸۱	۰/۰۷۸
۳	۰/۹۹	۰/۹۹	۰/۸۷	۰/۷۶	۰/۶۷	۰/۰۱۸	۰/۰۱۶	۰/۰۵۵	۰/۰۷۵	۰/۰۷۴	۰/۰۱	۰/۰۰۴	۰/۰۴۲	۰/۰۵۴	۰/۰۶۰
۴	۰/۹۲	۰/۹۸	۰/۸۱	۰/۶۰	۰/۴۴	۰/۰۴۸	۰/۰۱۸	۰/۰۶۸	۰/۰۹۸	۰/۰۹۶	۰/۰۳۸	۰/۰۰۶	۰/۰۵۴	۰/۰۷۹	۰/۰۷۸
۵	۰/۹۹	۰/۹۹	۰/۸۹	۰/۸۳	۰/۷۰	۰/۰۱۵	۰/۰۱	۰/۰۵۱	۰/۰۶۴	۰/۰۷۱	۰/۰۰۸	۰/۰۰۱۱	۰/۰۴	۰/۰۴۶	۰/۰۵۷
۶	۰/۹۴	۰/۹۹	۰/۸۸	۰/۷۳	۰/۷۰	۰/۰۴	۰/۰۱۶	۰/۰۵۴	۰/۰۸	۰/۰۷۱	۰/۰۲۹	۰/۰۰۴	۰/۰۴۱	۰/۰۵۸	۰/۰۵۶
۷	۰/۹۲	۰/۹۱	۰/۷۷	۰/۶۸	۰/۴۷	۰/۰۴۷	۰/۰۵۴	۰/۰۷۶	۰/۰۸۸	۰/۰۹۴	۰/۰۳۷	۰/۰۲۷	۰/۰۶۲	۰/۰۶۴	۰/۰۷۷
۸	۰/۹۳	۰/۹۵	۰/۸۶	۰/۸۴	۰/۶۹	۰/۰۴۲	۰/۰۴	۰/۰۵۹	۰/۰۶	۰/۰۷۲	۰/۰۳۱	۰/۰۲۰	۰/۰۴۳	۰/۰۴۷	۰/۰۵۷
۹	۰/۹۹	۰/۹۹	۰/۹۰	۰/۸۷	۰/۶۹	۰/۰۱۴	۰/۰۱۱	۰/۰۵۴	۰/۰۶	۰/۰۷۱	۰/۰۰۷۶	۰/۰۰۱	۰/۰۴۲	۰/۰۴۶	۰/۰۵۷
۱۰	۰/۹۸	۰/۹۶	۰/۸۶	۰/۸۴	۰/۶۷	۰/۰۳۴	۰/۰۳۲	۰/۰۶	۰/۰۶۹	۰/۰۷۵	۰/۰۱۶	۰/۰۱	۰/۰۴۷	۰/۰۵۲	۰/۰۶۰

همان‌طور که در جدول ۳ مشاهده می‌شود، دقت برآورد رطوبت خاک در سناریو چهارم با افزودن شاخص‌های NSMI، NDVI و NTR به عنوان متغیرهای ورودی در ترکیب با باندهای نوری حاصل از ماهواره لندست-۸ در مقایسه با سناریو سوم کاهش یافته است. این نتیجه برتری استفاده از شاخص SWI نسبت به شاخص‌های NSMI، NDVI و NTR به عنوان متغیر کمکی در برآورد رطوبت خاک را نشان می‌دهد. لازم به ذکر است که نتایج استفاده از شاخص‌های گیاهی در شرایط کمبود رطوبت خاک به دلیل مکانیسم‌های ذاتی گیاهان در مقابله با تنش آبی قابل اتکا نخواهد بود. از طرفی در شرایط کمبود رطوبت، رسانایی پوشش گیاهی به بخار آب کاهش می‌یابد و این رخداد بر بازتابندگی پرتوهای برخوردی اثر می‌گذارد، که پیامد آن افزایش مصرف انرژی در شارهای گرمای محسوس و انباشت در برگ خواهد بود، از این‌رو دمای سطح تغییر خواهد کرد (Goetz, 1997) و می‌توان با استفاده از باند و شاخص‌های حرارتی (مانند SWI) رطوبت موجود در خاک را بهتر شناسایی کرد. در همین راستا نتایج مطالعات Potopová et al (2020) نشان داد که شاخص SWI بهتر از شاخص‌های EVI و SPEI، تغییرات مکانی-زمانی خشکسالی‌های کشاورزی را نمایش می‌دهد.

با کاربرد داده‌های سنجش از دور نوری-حرارتی حاصل از تصاویر لندست-۸ در سناریو پنجم مقدار آماره‌های خطا RMSE و MAE کاهش یافت و نتایج مشابه نتایج سناریوی سوم رقم خورد. با اضافه شدن خصوصیات فیزیکی خاک به متغیرهای سنجش از دور در مجموعه داده‌های ورودی (سناریو ششم) دقت عملکرد الگوریتم‌های CART، M5P، GBRT و RFR در تخمین رطوبت خاک کاهش یافت. حذف شاخص‌های گیاهی و رطوبتی از متغیرهای ورودی سناریوی ششم (سناریو هفتم) منجر به کاهش محسوس دقت برآورد در تمام مدل‌های یادگیری ماشین شد. از طرفی عملکرد بهتر سناریوی هشتم نسبت به سناریوهای ششم و هفتم را می‌توان به حذف باندهای نوری و افزودن شاخص‌های گیاهی و رطوبتی (SWI، NTR، NSMI، NDVI) در ترکیب با خصوصیات فیزیکی خاک نسبت داد.

بهترین عملکرد مدل‌های یادگیری ماشین (بالاترین R² و کمترین RMSE و MAE) در سناریو نهم با ترکیب متغیرهای ورودی SWI، NTR، NSMI، NDVI، SWIR2 و NIR مشاهده شد (جدول ۳). نوع رابطه رطوبت خاک با پوشش گیاهی و LST از نظر فیزیکی متفاوت است. به طوری که تغییرات رطوبت خاک با تاخیر توسط NDVI قابل مشاهده است، این در حالی است که LST بلافاصله در پاسخ به تنش آبی تغییر می‌کند. همین ویژگی باعث عملکرد متفاوت در شاخص‌های SWI، NDVI، NTR و باندهای LST و NIR می‌شود که با نتایج Potopová et al (2020) هم‌خوانی دارد. همچنین Fernández-Novales et al (2018) در مطالعه‌ای بیان کردند که تغییرات رطوبت خاک، محتوای آب برگ‌ها را تحت تاثیر قرار می‌دهد و در نهایت منجر به تغییر مشخصات طیفی منعکس شده از پوشش گیاهی می‌شود. از طرفی Araya et al (2021) نشان دادند که باندهای Red و NIR نسبت به شاخص‌های NDVI و TTVI¹ در برآورد رطوبت خاک سطحی از اهمیت بالاتری برخوردار هستند. علاوه بر باندهای NIR و LST، باند SWIR2 نیز با تغییرات رطوبت خاک ارتباط موثری دارد؛ چرا که انواع خاک در این محدوده طیفی بازتاب‌های قابل توجهی دارند. این ارتباط در نتایج Adab et al (2020) قابل مشاهده می‌باشد. از طرفی دقت برآوردها در سناریوی دهم با حذف متغیرهای NIR، NDVI، NSMI و NTR از مجموعه داده‌های

1 Thiam's transformed vegetation index

ورودی قبلی نسبت به حالت حضورشان کاهش یافت. بنابراین اهمیت ویژه‌ی شاخص‌های گیاهی و رطوبتی در برآورد رطوبت خاک به دلیل حساسیت موثر آن‌ها به رطوبت خاک و همچنین عدم حساسیت به عوامل مزاحم (خطاهای ناشی از پس‌زمینه) بیش از پیش نمایان می‌شود. ضعیف‌ترین عملکرد برآورد رطوبت خاک در بین الگوریتم‌های مورد بررسی در مدل MLR مشاهده شد (جدول ۳). جدول ۴ روابط رگرسیونی خطی چندگانه بین رطوبت خاک سطحی و متغیرهای ورودی در هر سناریو را نشان می‌دهد. همچنین در این الگوریتم کم‌ترین دقت پیش‌بینی در سناریوهای اول، دوم و چهارم ($R^2=0/45$ ، $RMSE=0/096$ و $MAE=0/078$) ثبت شد (جدول ۴). با بررسی روابط این سه سناریو (جدول ۴) مشاهده می‌شود که طی رگرسیون گام به گام تمامی متغیرهای ورودی به جز باند SWIR2 از فرآیند مدل‌سازی نهایی حذف شده‌اند که همین امر منجر به عملکرد ضعیف آن‌ها شده است. همچنین بیشترین مقدار R^2 ($0/70$) و کم‌ترین مقدار $RMSE$ ($0/07$) و MAE ($0/05$) در مدل MLR در سناریوی ششم به دست آمد. مطابق با جدول ۴، متغیرهای B, G, R, NIR, SWIR1, SWIR2, NDVI, Sand, Silt و BD در سناریو ششم اثر معنی‌داری در پیش‌بینی رطوبت خاک نداشته‌اند ($P < 0/05$) و از رابطه نهایی مدل حذف شدند. از طرفی با توجه به ضریب استاندارد شده بتا، متغیر SWI ($\beta=0/578$) موثرترین متغیر در تخمین رطوبت خاک در سناریو ششم شناخته شد. شکل ۵، نمودار پراکنش مقادیر پیش‌بینی شده توسط سناریوی برتر مدل MLR (سناریوی ۶) در مرحله آزمایش و مقادیر مشاهداتی را نشان می‌دهد. بررسی این نمودار نشان می‌دهد که مدل MLR در بهترین حالت خود نیز قادر به ثبت عملکردی مناسب و رضایت بخش در مرحله آزمایش نبوده است ($R^2=0/65$). باقری و همکاران (۱۳۹۸) در پژوهشی به بررسی رطوبت خاک سطحی در دو ماه (اسفند و اردیبهشت) در مزارع کشاورزی جنوب استان تهران پرداختند. آن‌ها رطوبت خاک سطحی را با استفاده از داده‌های ماهواره لندست-۸ و رگرسیون خطی چندگانه به ترتیب با ضریب تبیین $0/72$ و $0/81$ برای ماه‌های اسفند و اردیبهشت به دست آوردند. همچنین Prakash et al (2019) نشان دادند که مدل MLR نسبت به مدل‌های ماشین بردار پشتیبان و شبکه عصبی عملکرد بهتری در پیش‌بینی رطوبت خاک دارد. با این حال با توجه به نتایج نه چندان رضایت‌بخش مدل MLR در این پژوهش، به نظر می‌رسد بین پارامترهای ورودی و رطوبت خاک روابط غیر خطی وجود دارد.

مطابق با جدول ۳، ضرایب همبستگی و آماره‌های خطا حاصل از برآورد رطوبت خاک توسط مدل CART در سناریو ۱ تا ۱۰ از مقادیر R^2 بین $0/5$ تا $0/88$ ، $RMSE$ بین $0/11$ تا $0/056$ و MAE بین $0/089$ تا $0/044$ متغیر بود. همچنین بالاترین دقت برآورد رطوبت خاک توسط این مدل مربوط به سناریوی نهم می‌باشد ($R^2=0/88$ ، $RMSE=0/057$ و $MAE=0/046$). شکل ۶ پراکنش مقادیر رطوبت خاک اندازه‌گیری شده و برآورد شده در مرحله آزمایش با مدل CART در سناریو ۹ را نشان می‌دهد. پراکنش مقادیر برآورد شده مدل CART در رطوبت‌های بالای $0/35$ سانتی متر مکعب در سانتی متر مکعب نسبت به مقدار واقعی آن‌ها نشان دهنده عملکرد ضعیف مدل در این بازه است (شکل ۶). مدل CART با تشکیل یک درخت دودویی با استفاده از تقسیم‌بندی بازگشتی دودویی که حداکثر کاهش را در تغییرپذیری متغیر به همراه دارد (Acharya et al., 2021)، بهتر از MLR عمل کرد. همچنین برآورد قابل قبول رطوبت خاک با استفاده از مدل CART توسط Han et al (2018) در چین گزارش شده است.

همان‌طور که در جدول ۳ مشاهده می‌شود، عملکرد مدل M5P در مرحله آموزش در تمامی سناریوها (بجز سناریوی اول) بهتر از مدل‌های MLR و CART بود. بیش‌ترین دقت برآورد رطوبت خاک در این مدل به ترتیب در سناریو نهم ($R^2=0/90$ ، $MAE=0/042$)، پنجم ($R^2=0/89$ ، $MAE=0/04$) و سوم ($R^2=0/87$ ، $MAE=0/042$) مشاهده شد. بررسی میزان پراکنش مقادیر پیش‌بینی شده رطوبت خاک سطحی از مقادیر مشاهداتی در سه سناریوی برتر مدل M5P (سناریوی ۳، ۵ و ۹)، نشان از عملکرد تقریباً مشابه آن‌ها داشت (شکل ۷). مدل M5P (۹۱ رابطه رگرسیونی چندگانه) برخلاف مدل CART (۵۱۹ رابطه رگرسیونی چندگانه) توانست با کاربرد تعداد روابط کمتر دقت برآورد بالاتری را از رطوبت خاک سطحی در مرحله آموزش نشان دهد.

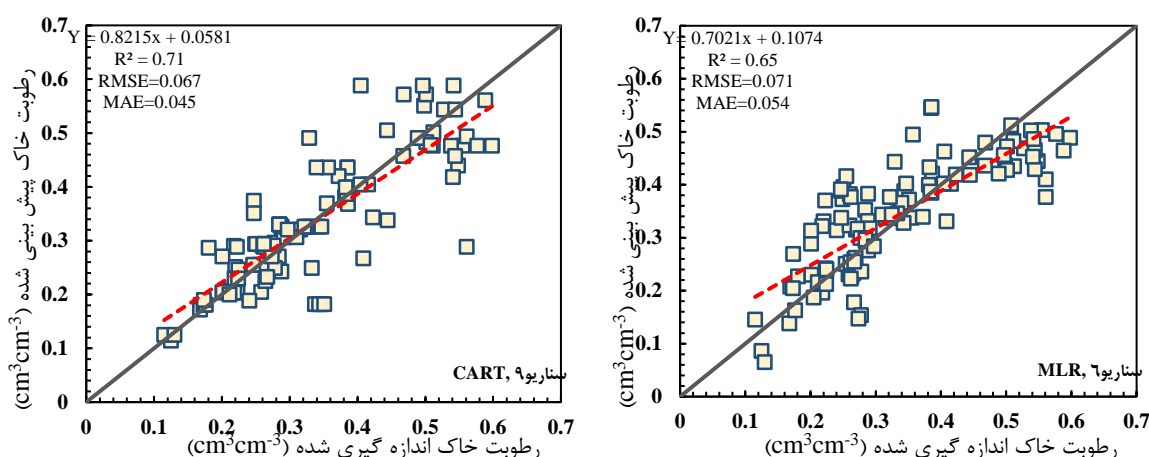
نتایج Acharya et al (2021) نشان داد که مدل‌های یادگیری ماشینی مانند جنگل تصادفی، درختان توسعه یافته و ماشین بردار پشتیبان عملکرد بهتری نسبت به مدل‌های رگرسیون خطی چندگانه و درختان طبقه‌بندی و رگرسیون (CART) دارند.

نتایج بدست آمده از هر دو مدل یادگیری جمعی (GBRT و RFR) نشان از برتری آن‌ها (بالاترین مقدار R^2 و کم‌ترین مقادیر $RMSE$ و MAE) در تخمین رطوبت خاک در مرحله آموزش نسبت به سایر الگوریتم‌های مورد بررسی دارد (جدول ۳). مشابه مدل M5P بالاترین دقت تخمین رطوبت سطحی در سناریوهای سوم، پنجم و نهم در هر دو مدل‌های یادگیری جمعی با بیشترین مقدار ضریب تبیین ($R^2=0/99$) و کم‌ترین مقدار $RMSE$ ($0/16 - 0/11$ ؛ $RMSE_{GBRT}=0/11 - 0/16$ ؛ $RMSE_{RFR}=0/14 - 0/18$) مشاهده شد. با توجه به نمودارهای همبستگی ۱:۱ بین رطوبت خاک معیار و تخمین زده شده با مدل‌های GBRT و RFR در سناریوهای برتر، مشخص شد که بیش‌ترین

پراکندگی در برآوردهای رطوبت خاک سطحی در مرحله آزمایش بین بازه‌ی مقادیر ۰/۳۵ تا ۰/۴۵ سانتی‌متر مکعب در سانتی‌متر مکعب رخ داده است (شکل ۷). با اینحال هر سه سناریو در دو مدل GBRT و RFR در مرحله آزمایش به خوبی رطوبت خاک را برآورد کرده‌اند. شایان ذکر است که تعداد هاپیرپارامترهای قابل تنظیم در مدل RFR نسبت به الگوریتم GBRT کمتر است که همین امر باعث کاهش پیچیدگی و کاربرد سریع‌تر آن می‌گردد. همچنین با توجه به آماره‌های خطا (RMSE و MAE) (شکل ۷)، می‌توان گفت مدل RFR در برآورد رطوبت خاک در مرحله آزمایش بهترین عملکرد را بین تمام مدل‌های مورد بررسی ارائه کرده است، هرچند که نتایج الگوریتم GBRT نیز کاملاً رضایت بخش بود. نتایج پژوهش حاضر همسو با یافته‌های Acharya et al (2021) بود. آن‌ها از شش روش مختلف یادگیری ماشین (DT^۱, RFR, BRT, MLR, SVR و ANN) برای برآورد رطوبت خاک در امتداد سواحل رودخانه قرمز واقع در شمال آمریکا استفاده کردند. نتایج آن‌ها نشان داد که الگوریتم‌های RFR و BRT بهترین عملکرد را با MAE کمتر از ۰/۰۴ و RMSE به ترتیب ۰/۰۴۵ و ۰/۰۴۸ مترمکعب در مترمکعب داشته است. در حالی که مدل‌های ANN و MLR با مقادیر R² به ترتیب ۰/۵۲ و ۰/۵۳ عملکردهای ضعیفی را ارائه کردند. همچنین Araya et al (2021) به ارزیابی عملکرد پنج الگوریتم یادگیری ماشین (ANN, SVR, RVR, RFR و BRT^۳) در برآورد رطوبت خاک با استفاده از داده‌های سنجش از دور و خصوصیات خاک پرداختند. نتایج آن‌ها نشان داد که مدل‌های رگرسیون مبتنی بر درخت تصمیم (الگوریتم‌های BRT و RFR) توانایی بالاتری در برآورد رطوبت خاک دارند (RMSE کمتر از ۴ درصد) که مطابق با یافته‌های پژوهش حاضر می‌باشد.

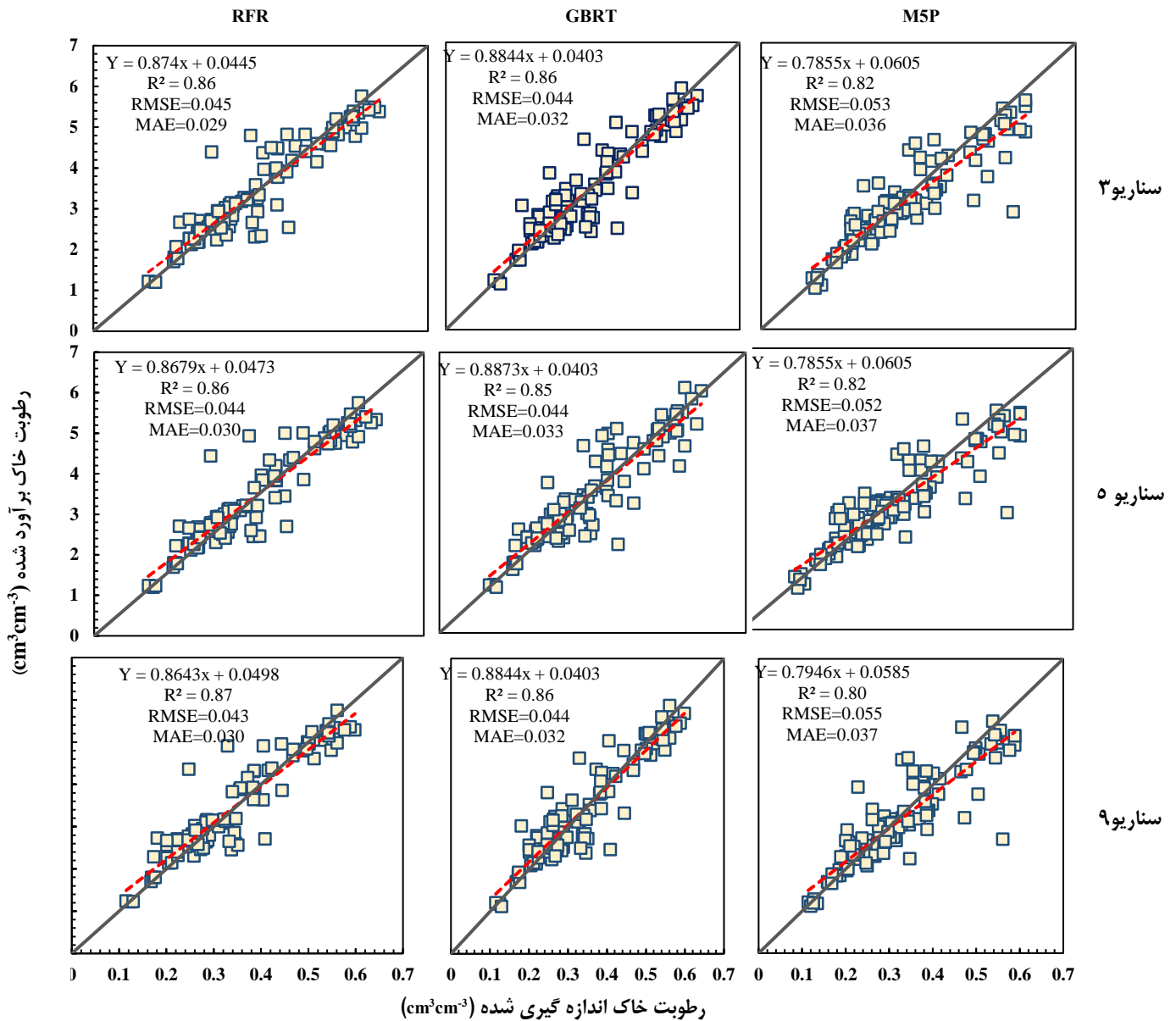
جدول ۴. روابط رگرسیون خطی چندگانه بین رطوبت خاک و متغیرهای ورودی در سناریوهای مختلف

سناریو	مدل‌های رگرسیون خطی چندگانه
۱	SM= 52.455 - 116.143SWIR2
۲	SM= 52.455 - 116.143SWIR2
۳	SM= 21.195 + 36.861SWI - 129.612SWIR2 + 100.803R
۴	SM= 52.455 - 116.143SWIR2
۵	SM= -78.857 + 39.049SWI + 97.958NSMI + 7.509NTR + 0.317LST - 36.910NDVI - 60.829SWIR1
۶	SM= -66.073 + 36.578SWI + 64.091NSMI - 0.514Clay + 13.465NTR + 0.246LST
۷	SM= 74.549 - 115.123SWIR2 - 0.623Clay
۸	SM=14.449 + 37.583SWI + 55.771NSMI - 0.473Clay +5.707NTR
۹	SM= -97.986 + 38.555SWI + 99.413NSMI + 12.359NTR + 0.292LST - 16.590NDVI
۱۰	SM=22.676 + 35.806SWI - 59.135SWIR2



شکل ۶. نمودار پراکندگی مقایسه مقادیر رطوبت خاک برآورد شده از MLR و CART در مرحله آزمایش با رطوبت خاک اندازه گیری شده در سناریوهای برتر (سناریو ۹ و ۶)

1 Decision tree
 2 Relevance vector regression
 3 Boosted regression trees



شکل ۷. نمودار پراکندگی مقایسه مقادیر رطوبت خاک برآورد شده از RFR، GBRT، MSP در مرحله آزمایش با رطوبت خاک اندازه گیری شده در سناریوهای برتر (S₃, S₅, S₉)

نتیجه گیری

در این پژوهش، یک رویکرد جدید، سریع و کم هزینه به منظور برآورد رطوبت خاک سطحی در مقیاس مکانی و زمانی با هدف پایش خشکسالی در مناطق گرم و خشک پیشنهاد شد. پنج مدل یادگیری ماشین مختلف (RFR، GBRT، MSP، CART، MLR) در ترکیب با داده‌های حاصل از تصاویر ماهواره لندست-۸ و برخی خصوصیات فیزیکی خاک برای پیش‌بینی رطوبت خاک سطحی استفاده شدند. همچنین به منظور بررسی رابطه و اهمیت بین متغیرهای ورودی و رطوبت خاک، ۱۰ سناریو مختلف از ترکیب متغیرهای سنجش از دور و خصوصیات فیزیکی خاک توسعه داده شد و بر اساس شاخص‌های آماری بهترین سناریوها از هر مدل معرفی شدند. نتایج نشان داد که مدل‌های یادگیری ماشین براساس مقدار R² به ترتیب کاهشی در اکثر سناریوها به صورت $RFR \geq GBRT > MSP > CART > MLR$ رتبه‌بندی می‌شوند. همچنین بررسی مقادیر خطای (RMSE) سناریوهای برتر مدل‌ها در مرحله آزمایش نشان داد که این مقدار در الگوریتم‌های RFR و GBRT، CART (به ترتیب ۰/۰۶۷، ۰/۰۳۲ و ۰/۰۳۰) در سناریوی نهم و در الگوریتم‌های MSP و MLR (۰/۰۵۲)

و (۰/۰۷۱) به ترتیب در سناریوی پنجم و ششم بیشینه است. نتایج نشان داد که استفاده از خصوصیات فیزیکی خاک در ترکیب با باندهای نوری و شاخص‌های گیاهی، رطوبتی و دمایی در مدل‌های GBRT، M5P، CART، RFR و منجر به پیش‌بینی‌های بهینه از رطوبت خاک سطحی نمی‌شود. در حالی که در مدل MLR استفاده از درصد رس، علاوه بر کاربرد اطلاعات سنجش از دوری تأثیر بسزایی در مدل تخمین رطوبت خاک سطحی داشته است اما عملکرد کلی این الگوریتم رضایت بخش نبود ($R^2=0/68$). عملکرد بهینه در اکثر مدل‌ها زمانی رخ داد که شاخص خسی خاک (SWI) (حاصل از روش مثلثی) به ترکیب شاخص‌های گیاهی و رطوبتی (مانند NDVI، NSMI، NTR) اضافه شد. به طور کلی مطابق با نتایج بدست آمده، مدل‌های یادگیری جمعی (GBRT و RFR) عملکرد بهتری نسبت به سایر مدل‌های مورد مطالعه داشتند. این دو مدل با استفاده از متغیرهای NTR، NSMI، NDVI، SWIR2، NIR، LST و SWI به خوبی تغییرات رطوبت خاک در مقیاس میدانی را بدون نیاز به اطلاعات زمینی نشان دادند. بنابراین استفاده از نتایج پژوهش حاضر می‌تواند در زمینه رفع چالش پایش بلادرنگ رطوبت خاک و گیاه راه‌گشا و کاربردی باشد. برای تأیید هرچه بیشتر نتایج این تحقیق، پیشنهاد می‌شود در مطالعات آینده ارتباط بین رطوبت خاک و داده‌های سنجش از دور در شرایط محیطی و مدیریتی متفاوت بررسی شود.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

منابع

باقری، کیوان؛ باقری، میلاد؛ حسین زاده، علی اصغر و پروین، منصور (۱۳۹۸). برآورد رطوبت خاک با استفاده از سنجش از دور نوری، حرارتی و راداری (مطالعه موردی: اراضی جنوب تهران). *مجله علوم ومهندسی آبخیزداری ایران*، ۱۳ (۴۷)، ۶۳-۷۴

REFERENCES

- Acharya, U., Daigh, A. L., & Oduor, P. G. (2021). Machine learning for predicting field soil moisture using soil, crop, and nearby weather station data in the Red River Valley of the North. *Soil Systems*, 5(4), 57 .
- Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., & Ghalhari, G. A. F. (2020). Machine learning to estimate surface soil moisture from remote sensing data. *Water*, 12(11), 3223 .
- Araya, S. N., Fryjoff-Hung, A., Anderson, A., Viers, J. H., & Ghezzehei, T. A. (2021). Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. *Hydrology and Earth System Sciences*, 25(5), 2739-2758 .
- Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. K., & Tuller, M. (2021). Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Remote sensing of environment*, 260, 112434 .
- Babaeian, E., Sadeghi, M., Franz, T. E., Jones, S., & Tuller, M. (2018). Mapping soil moisture with the Optical TRapezoid Model (OPTRAM) based on long-term MODIS observations. *Remote sensing of environment*, 211, 425-440 .
- Barrett, B., & Petropoulos, G. P. (2013). Satellite remote sensing of surface soil moisture. *Remote sensing of energy fluxes and soil moisture content*, 85 .
- Blake, G. (1965). Bulk density. *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling*, 9, 374-390 .
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks. In: Cole Advanced Books and Software.
- Day, P. R. (1965). Particle fractionation and particle-size analysis. *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling*, 9, 545-567 .
- Duethmann, D., Smith, A., Soulsby, C., Kleine, L., Wagner, W., Hahn, S., & Tetzlaff, D. (2022). Evaluating satellite-derived soil moisture data for improving the internal consistency of process-based ecohydrological modelling. *Journal of hydrology*, 614, 128462 .
- Eshaghi, A., Motamedvaziri, B., & Feiznia, S. (2010). Landslides Hazard Zonation Using Logistic Regression Method (Case Study: Safaroud Watershed). *Territory*, ۶(۲۴)، ۶۷-۷۷ .
- Fatholouloumi, S., Vaezi, A. R., Alavipanah, S. K., Ghorbani, A., & Biswas, A. (2020). Comparison of spectral and spatial-based approaches for mapping the local variation of soil moisture in a semi-arid mountainous area. *Science of the Total Environment*, 724, 138319 .
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232 .
- Gates, D. M., Keegan, H. J., Schleiter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied*



optics, 4(1), 11-20 .

- Ge, L., Hang, R., & Liu, Q. (2019). Retrieving soil moisture over continental us via multi-view multi-task learning. *IEEE Geoscience and Remote Sensing Letters*, 16(12), 1954-1958 .
- Goetz, S. (1997). Multi-sensor analysis of NDVI ,surface temperature and biophysical variables at a mixed grassland site. *International journal of remote sensing*, 18(1), 71-94 .
- Han, J., Mao, K., Xu, T., Guo, J., Zuo, Z., & Gao, C. (2018). A soil moisture estimation framework based on the CART algorithm and its application in China. *Journal of hydrology*, 563, 65-75 .
- Haubrock, S. N., Chabrillat, S., Lemmnitz, C., & Kaufmann, H. (2008). Surface soil moisture quantification models from reflectance data under field conditions. *International journal of remote sensing*, 29(1), 3-29 .
- Jones, S. B., Blonquist, J., Robinson, D. A., Rasmussen, V. P., & Or, D. (2005). Standardizing Characterization of Electromagnetic Water Content Sensors Part 1. Methodology. *Vadose Zone Journal*, 4(4), 1048-1058 .
- Khellouk, R., Barakat, A., Boudhar, A., Hadria, R., Lionboui, H., El Jazouli, A., Rais, J., El Baghdadi, M., & Benabdellouahab, T. (2020). Spatiotemporal monitoring of surface soil moisture using optical remote sensing data: a case study in a semi-arid area. *Journal of Spatial Science*, 65(3), 481-499 .
- Korres, W., Reichenau, T., Fiener, P., Koyama, C., Bogena, H. R., Cornelissen, T., Baatz, R., Herbst, M., Diekkrüger, B., & Vereecken, H. (2015). Spatio-temporal soil moisture patterns—A meta-analysis using plot to catchment scale data. *Journal of hydrology*, 520, 326-341 .
- Kubelka, P., & Munk, F. (1931). An article on optics of paint layers. *Z. Tech. Phys*, 12(593-601), 259-274 .
- Lobell, D. B., & Asner, G. P. (2002). Moisture effects on soil reflectance. *Soil Science Society of America Journal*, 66(3), 722-727 .
- Na, L., Na, R., Bao, Y., & Zhang, J. (2021). Time-lagged correlation between soil moisture and intra-annual dynamics of vegetation on the Mongolian plateau. *Remote Sensing*, 13(8), 1527 .
- Nguyen, T. T., Pham, T. D., Nguyen, C. T., Delfos, J., Archibald, R., Dang, K. B., Hoang, N. B., Guo, W., & Ngo, H. H. (2022). A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Science of the Total Environment*, 804, 150187 .
- Nolet, C., Poortinga, A., Roosjen, P., Bartholomeus, H., & Ruessink, G. (2014). Measuring and modeling the effect of surface moisture on the spectral reflectance of coastal beach sand. *PLoS One*, 9(11), e112151 .
- Potopová, V., Trnka, M., Hamouz, P., Soukup, J., & Castravet, T. (2020). Statistical modelling of drought-related yield losses using soil moisture-vegetation remote sensing and multispectral indices in the south-eastern Europe. *Agricultural Water Management* . ۱۰۶۱۶۸ , ۲۳۶ ,
- Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., & Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review . *Vadose Zone Journal*, 7(1), 358-389 .
- Sandholt, I., Rasmussen, K., & Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote sensing of environment*, 79(2-3), 213-224 .
- Wang, J., Ding, J., Yu, D., Teng, D., He, B., Chen, X., Ge, X., Zhang, Z., Wang, Y., & Yang, X. (2020). Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. *Science of the Total Environment*, 707, 136092 .
- Wang, Y., & Witten, I. H. (1996). Induction of model trees for predicting continuous classes .
- Zreda, M., Desilets, D., Ferré, T., & Scott, R. L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical research letters*, 35(21)

Spatial-temporal modeling of soil moisture using optical and thermal remote sensing data and machine learning algorithms

EXTENDED ABSTRACT

Introduction:

Spatiotemporal estimation and monitoring of soil moisture based on remote sensing observations are essential for managing water resources, improving agricultural land productivity, increasing water use efficiency, and assessing crop drought conditions. In this regard, methods based on optical and thermal remote sensing data have successfully estimated surface soil moisture at different scales. However, the physical nature of these data has limited and challenged their application in dense vegetation conditions, necessitating improving and increasing the estimation accuracy in these areas.

Objective:

This research aims to develop a new approach to estimating surface soil moisture in agricultural fields with dense vegetation conditions, such as sugarcane fields, using machine learning algorithms by incorporating optical and thermal remote sensing data and soil physical properties.

Materials and methods:

This study used 16 Landsat-8 images during the sugarcane crop's growth period in 2018-2019 at Hakim Farabi Sugarcane Agro-Industrial company in the Khuzestan province of Iran. Soil moisture measurements were collected simultaneously as the satellite passed through the study area at more than 430 control locations during the period. A set of 10 scenarios of various unique combinations of the available input variables were developed. Five popular machine learning algorithms evaluated the scenarios, including multiple linear regression (MLR), decision tree-based algorithms (Classification and Regression Trees (CART) and M5-pruned (M5P)), and ensemble learning-based algorithms (gradient-boosted regression trees (GBRT) and random forest regression (RFR)).

Results and discussion:

According to the results, the highest correlation between input variables and surface soil moisture was observed in Soil Wetness Index (SWI) and Normalized Soil Moisture Index (NSMI) with R values of 0.79 and 0.69, respectively. While the NIR band with an R-value of 0.56 showed the lowest correlation. The obtained results showed the high ability of machine learning algorithms to estimate surface soil moisture in the area. The highest accuracy of machine learning algorithms based on R^2 , RMSE, and MAE results was obtained in GBRT (0.99, 0.011, and 0.006) in scenario 9, RFR (0.99, 0.014, and 0.007) in scenario 9, M5P (0.90, 0.054, and 0.042) in scenario 9, CART (0.87, 0.058, and 0.046) and MLR (0.70, 0.07, and 0.056) in scenario 6, respectively. The importance of incorporating soil physical properties, especially clay percentage, with remote sensing data was observed only in the MLR algorithm. While in CART, M5P, GBRT, and RFR algorithms, the use of soil physical properties in combination with optical bands and different vegetation, humidity, and temperature indices did not lead to proper surface soil moisture predictions.

Conclusion:

In general, the findings of this research show the importance of using variables based on Landsat-8 remote sensing data (NTR, NSMI, NDVI, SWIR2, NIR, LST, and SWI) in combination with ensemble learning algorithms (RFR and GBRT) that can be independent of any ground measurements. The proposed method provides valuable results for estimating and monitoring surface soil moisture in high-vegetation areas.

Keywords: High Vegetation, Machine Learning Algorithms, Remote Sensing, Soil Moisture, Soil Physical Properties.