



## **Estimating Soil Moisture from Fusion of Soil Physical/Hydraulic Properties and Optical Remote Sensing Observations Using Machine Learning**

**Shokoufeh Shokri<sup>1</sup>, Ahmad Farrokhian Firouzi<sup>2✉</sup>, Ebrahim Babaiean<sup>3</sup>**

1. Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

Email: [shokoufeh.shokri@gmail.com](mailto:shokoufeh.shokri@gmail.com)

2. Corresponding Author, Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran. Email: [a.farrokhian@scu.ac.ir](mailto:a.farrokhian@scu.ac.ir)

3. Department of Environmental Science, University of Arizona, Arizona, USA, Email: [cbabaeian@email.arizona.edu](mailto:cbabaeian@email.arizona.edu)

---

### **Article Info**

**Article type:** Research Article

**Article history:**

**Received:** Apr. 8, 2022

**Revised:** June. 27, 2022

**Accepted:** July. 30, 2022

**Published online:** Sep. 23, 2022

**Keywords:**

Modeling,  
Soil hydraulic parameters,  
Sugarcane,  
Infrared remote sensing.

---

### **ABSTRACT**

Soil moisture content (SM) is a critical state variable that significantly affects both the hydrological cycle and agricultural production. Therefore, accurate estimation of soil moisture is important for agricultural water resources management. Remote sensing observations in the near- and shortwave infrared have large potential for estimating soil moisture. In addition, soil physical and hydraulic properties affect spatial and temporal variability of soil moisture. The objective of this research was to derive different models for soil moisture estimation in Amir Kabir sugarcane agro-industry fields, Kuzestan province using a combination of soil physical/hydraulic properties and remote sensing observations with machine learning algorithms. Consequently, 166 ground control points and 16 Sentinel-2 satellite images were investigated during the growth period of sugarcane in the year 2021. Six machine learning algorithms including decision tree (DT), support vector machine (SVM), Linear regression, Boosted and Bagged trees, and neural network were used for modeling. Seven models were derived from the combination of soil physical/hydrological properties and remote sensing indices in a hierarchical manner to predict soil moisture content at the field scale. The results indicated that the combination of soil physical/hydraulic properties with remote sensing indices enhances the accuracy of soil moisture estimation. It is observed that almost all developed models performed well for estimating soil moisture, with an RMSE of 0.04-0.06 cm<sup>3</sup>cm<sup>-3</sup> and an R<sup>2</sup> of approximately 0.80. The STR parameter was found to be more sensitive to changes in soil water content than NIR reflectance. Therefore, STR was identified as the most important feature in estimating soil moisture content. Moreover, stepwise linear regression with RMSE value of 0.042 cm<sup>3</sup> cm<sup>-3</sup> performed the best in soil moisture estimation. According to the results, the models successfully capture the spatiotemporal dynamics of soil moisture and can be used for irrigation scheduling and precision irrigation management at the field scale.

---

Cite this article: Shokri, Sh., Farrokhian Firouzi, A., & Babaiean, E. (2022) Estimating Soil Moisture from Fusion of Soil Physical/Hydraulic Properties and Optical Remote Sensing Observations Using Machine Learning. *Iranian Journal of Soil and Water Research*, 53 (7), 1575-1591.

© The Author(s).

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/ijswr.2022.341492.669243>



## برآورد رطوبت خاک به کمک تلفیق ویژگی‌های فیزیکی و هیدرولیکی خاک با داده‌های نوری سنجش‌ازدور با استفاده از روش یادگیری ماشین

شکوفه شکری<sup>۱</sup>، احمد فرخیان فیروزی<sup>۲</sup>، ابراهیم بابائیان<sup>۳</sup>

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: [shokoufeh.shokri@gmail.com](mailto:shokoufeh.shokri@gmail.com)

۲. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: [a.farrokhan@scu.ac.ir](mailto:a.farrokhan@scu.ac.ir)

۳. گروه علوم محیط‌زیست، دانشگاه آریزونا، آریزونا، آمریکا، ایمیل: [ebabaeian@email.arizona.edu](mailto:ebabaeian@email.arizona.edu)

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	رطوبت خاک به‌عنوان متغیری پویا در مکان و زمان، یکی از عوامل اصلی اثرگذار در چرخه آب در طبیعت و تولید محصولات کشاورزی محسوب می‌شود؛ بنابراین برآورد دقیق آن برای مدیریت بهینه منابع آب در بخش کشاورزی حائز اهمیت است. داده‌های انعکاس طیفی سنجش‌ازدور در طول موج مادون‌قرمز نزدیک و دور قابلیت زیادی برای برآورد رطوبت خاک دارند و از طرفی ویژگی‌های فیزیکی و هیدرولیکی خاک بر تغییرپذیری مکانی و زمانی رطوبت خاک اثرگذارند. هدف از این پژوهش توسعه و ارزیابی مدل‌های مختلف حاصل از ترکیب متغیرهای سنجش‌ازدور و فیزیکی خاک برای برآورد رطوبت خاک در مزارع کشت‌و‌صنعت امیرکبیر خوزستان با استفاده از روش‌های مختلف یادگیری ماشین بود. بدین منظور ۱۶۶ نقطه کنترل زمینی و ۱۶ تصویر ماهواره سنتینل-۲ در طول دوره رشد گیاه نیشکر در سال ۱۴۰۰ مورد استفاده قرار گرفت. از ترکیب ویژگی‌های فیزیکی / هیدرولیکی و شاخص‌های سنجش‌ازدور، هفت مدل به‌صورت سلسله‌مراتبی به دست آمد که با شش الگوریتم یادگیری ماشین شامل درخت تصمیم‌گیری، ماشین بردار خطی، رگرسیون خطی، درخت توسعه‌یافته، درخت کیسه‌گذاری و شبکه عصبی تلفیق و ارزیابی شدند. نتایج نشان داد ترکیب ویژگی‌های فیزیکی / هیدرولیکی و شاخص‌های سنجش‌ازدور دقت برآورد رطوبت خاک را افزایش می‌دهد. تقریباً همه مدل‌های به‌دست‌آمده با مقدار $RMSE = 0.040-0.060 \text{ cm}^3 \text{ cm}^{-3}$ و $R^2$ حدود $0.80$ برآورد قابل قبولی از مقدار رطوبت خاک ارائه دادند. متغیر STR در مقایسه با NIR به دلیل حساسیت بیشتر به مقدار آب خاک، اهمیت بالاتری در برآورد رطوبت خاک از خود نشان داد. بعلاوه، روش رگرسیون خطی گام‌به‌گام با مقدار $RMSE$ برابر $0.042 \text{ cm}^3 \text{ cm}^{-3}$ در مقایسه با سایر مدل‌های یادگیری ماشین با دقت بالاتری رطوبت خاک را برآورد کرد. نتایج نشان داد که مدل‌های ارائه‌شده قادر به برآورد تغییرات مکانی و زمانی رطوبت خاک هستند، لذا می‌توان از آن‌ها برای برنامه‌ریزی دقیق آبیاری و مدیریت بهینه آب در مقیاس مزرعه استفاده کرد.
واژه‌های کلیدی: مدل‌سازی، پارامترهای هیدرولیکی خاک، کشت‌و‌صنعت امیرکبیر، نیشکر، سنجش‌ازدور مادون‌قرمز.	
تاریخ دریافت: ۱۴۰۱/۱/۱۹	
تاریخ بازنگری: ۱۴۰۱/۴/۶	
تاریخ پذیرش: ۱۴۰۱/۵/۸	
تاریخ انتشار: ۱۴۰۱/۷/۱	

استناد: شکری، شکوفه؛ فرخیان فیروزی، احمد؛ بابائیان، ابراهیم (۱۴۰۱). برآورد رطوبت خاک به کمک تلفیق ویژگی‌های فیزیکی و هیدرولیکی خاک با داده‌های نوری سنجش‌ازدور با استفاده از روش یادگیری ماشین. *مجله تحقیقات آب و خاک ایران*، ۵۳ (۷)، ۱۵۹۱-۱۵۷۵.

DOI: <http://doi.org/10.22059/ijswr.2022.341492.669243>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

## مقدمه

رطوبت خاک تأثیر مستقیمی بر امنیت غذایی، سلامت انسان و عملکرد زیست‌بوم دارد. برآورد دقیق و به‌موقع آن در مدیریت آب در کشاورزی و چرخه آب در طبیعت از اهمیت زیادی برخوردار است (Jung et al., 2010; McColl et al., 2017). از این‌رو، مدل‌سازی و پایش رطوبت خاک مورد توجه پژوهشگران زیادی قرار گرفته است، به طوری که شناخت تغییرات آب‌وهوایی برای توسعه سامانه‌های کشت پایدار و همچنین بهبود کشاورزی دقیق و امنیت غذایی نیازمند آگاهی از تغییرات مکانی و زمانی رطوبت خاک است (Zareie et al., 2016).

در مطالعه حاضر که در کشت‌و صنعت امیرکبیر اجرا شده است، با توجه به این که سطح زیر کشت نیشکر در واحدهای کشت‌و صنعت فعال در استان خوزستان دارای مساحتی بالغ بر ۸۴۰۰۰ هکتار و از مهم‌ترین کشت‌و صنعت‌های موجود در ایران است، سهم زیادی از اراضی کشاورزی استان را به خود اختصاص داده است؛ بنابراین کنترل تک‌تک مزارع و رسیدگی به مسائل و مشکلات آن‌ها امری مشکل است. لذا هر ساله هزینه، وقت و نیروی انسانی زیادی جهت بازرسی و کنترل تعدادی از مزارع در هر کشت‌و صنعت به‌عنوان مزارع شاهد برای کنترل وضعیت این اراضی صرف می‌گردد، در این راستا نیاز مدیران به یک سامانه پایش خودکار و کارآمد برای صرفه‌جویی در زمان و انرژی بیش‌ازپیش آشکار است (Hamzeh and Naseri., 2013).

از آنجایی که نیشکر گیاهی چندساله و دارای دوره رشد طولانی و نیاز آبی زیاد است، در طول دوره رشد به آب فراوان احتیاج داشته و از طرفی نسبت به کم‌آبی حساس است. در عین حال، ریشه آن به غرقاب شدن درازمدت حساس است (New L, 1971). با توجه به خاصیت ذاتی گیاه نیشکر و تراکم بالای گیاه، عملاً حرکت کردن در داخل مزارع آن غیرممکن است؛ بنابراین با استفاده از ابزار سنسور از دور می‌توان نقاط مسئله‌دار در داخل یک قطعه زراعی را در طول فصل رشد از جمله وضعیت رطوبتی خاک، یکنواختی آبیاری، کفایت آبیاری، ماندابی شدن آب در یک نقطه از مزرعه، آفت‌زدگی و وجود بیماری را در گیاه مشاهده کرد (Veysi et al., 2017).

مقدار رطوبت خاک بر انعکاس طیفی خاک تأثیر می‌گذارد (Kaleita et al., 2005). آب در محدوده مادون قرمز نزدیک، انعکاسی ندارد و به راحتی از دیگر پدیده‌ها تفکیک می‌شود. خاک خشک در تمامی باندها، بالاترین بازتاب را داشته و روشن‌تر به نظر می‌رسد و خاک مرطوب دارای کمترین بازتاب بوده و تیره‌تر دیده می‌شود. بنابراین داده‌های محدوده نوری سنسوردها منابع مطلوبی برای نقشه‌برداری رطوبت خاک هستند (Scott Christopher et al., 2003). از نظر تئوری، تغییرات در انعکاس طیفی خاک زمانی رخ می‌دهد که منافذ خاک به طور نسبی با ملکولهای آب پر شده و تغییر در انعکاس طیف الکترومغناطیسی اتفاق افتد (Irons et al., 1992). در نتیجه داده‌های نوری سنسور از دور در گستره مادون قرمز نزدیک ۱ (NIR) و مادون قرمز طول موج کوتاه ۲ (SWIR) قابلیت زیادی در برآورد رطوبت خاک از خود نشان داده است (Sadeghi et al., 2015). طیف مرئی ۳- مادون قرمز نزدیک (Vis-NIR) نیز نتایج موفقیت‌آمیزی در برآورد رطوبت خاک سطحی داشته است (Kaleita et al., 2005). بسیاری از مطالعات نشان می‌دهند که برآورد رطوبت خاک سطحی (عمق ۷/۶-۰ سانتی‌متر) با استفاده از انعکاس طول موج‌های مرئی و مادون قرمز طول موج کوتاه امکان‌پذیر است (Kaleita et al., 2005).

شاخص تفاضلی پوشش گیاهی (NDVI) که با استفاده از طول موج مادون قرمز نزدیک و قرمز در بخش طیف مرئی به دست می‌آید (رابطه ۲) با نسبت تابش جذب شده فتوسنتزی ارتباط دارد و فعالیت کلروفیل گیاهان را انعکاس می‌دهد. در یک پیکسل مشخص افزایش NDVI نشان‌دهنده افزایش پوشش گیاهی سبز است. از آنجایی که پوشش گیاهی به‌طور غیرمستقیم بر پویایی رطوبت خاک تأثیر می‌گذارد بنابراین در این مطالعه از این شاخص به‌عنوان یکی از داده‌های نوری سنسور از دور به همراه با سایر ویژگی‌های فیزیکی و هیدرولیکی خاک در ایجاد مدل‌های ترکیبی پیشنهادی برای برآورد رطوبت خاک استفاده شده است.

از شناخته‌شده‌ترین روش‌ها در حوزه مرئی / مادون قرمز نزدیک برای تعیین وضعیت رطوبت خاک تشکیل فضای مثلثی است که ارتباط بین رطوبت خاک و شاخص پوشش گیاهی (VI) و دمای سطح زمین (LST) را بیان می‌کند. Babaeian et al., (2021) و Hosseini Chamani et al., (2019) بیان داشتند که انعکاس مادون قرمز طول موج کوتاه (SWIR) با شرایط محیطی به‌طور معنی‌داری تغییر نمی‌کند و همبستگی خطی بالایی با رطوبت خاک دارد، بنابراین ایده جایگزینی LST<sup>۵</sup> با STR<sup>۶</sup> را مطرح کردند که از انعکاس تبدیل یافته مادون قرمز طول موج کوتاه (SWIR) برای اندازه‌گیری رطوبت خاک در حوزه نوری استفاده می‌کند و از آن به‌عنوان مدل دوزنقه‌ی

1 near-infrared

2 short wave- infrared

3 Visible spectrum

4 Normalized difference vegetation index

5 Land surface temperature

6 Shortwave infrared Transformed Reflectance



نوری از آن یاد می‌شود. پارامتر STR به‌عنوان انعکاس تبدیل‌شده SWIR (رابطه ۳) عملکرد موفقیت‌آمیزی درزمینه‌ی برآورد رطوبت خاک دارد (Sadeghi *et al.*, 2015) و به‌عنوان دومین ورودی داده‌ی نوری سنجنش‌ازدور در این مطالعه برای ایجاد مدل‌های ترکیبی پیشنهادی تعریف‌شده است.

در مطالعات بسیاری روش‌های رگرسیونی به دلیل روش‌شناسی ساده و سابقه طولانی که دارند به‌منظور ارزیابی فرآیندها موردتوجه بوده‌اند. بااین‌حال در تحلیل رگرسیون‌های سنتی برخی مفروضات آماری موردنیاز مانند داده‌های پرت، غیرخطی بودن و ناهمسانی ممکن است منجر به استفاده محدود از آن‌ها شوند. روش‌های یادگیری ماشین مانند شبکه عصبی، جنگل تصادفی و درخت‌های تصمیم برای غلبه بر مشکلات فوق‌پدید آمده‌اند. یادگیری ماشین به جست‌وجو و واکاوی خودکار یا نیمه‌خودکار مجموعه داده‌های بزرگ به‌منظور کشف همبستگی‌ها، الگوها و قوانین معنی‌دار در میان داده‌ها می‌پردازد (Yeh and Lien., 2009). برتری اصلی یادگیری ماشین توانایی آن در شناسایی خودکار و کارآمد روابط پیچیده و استخراج ویژگی‌های مکانی-زمانی از انواع مختلف داده‌های زودیافت است.

تاکنون مدل‌های ترکیبی زیادی در زمینه‌ی برآورد رطوبت خاک با داده‌های سنجنش‌ازدور با استفاده از روش‌های یادگیری ماشین مورد ارزیابی قرار گرفته است. (Chakrabarti *et al.*, 2018) به‌منظور کاهش مقیاس رطوبت خاک از الگوریتم جدیدی بر اساس درختان رگرسیون کیسه‌گذاری (BRT) باهدف ایجاد همبستگی بین محصولات سنجنش‌ازدور با وضوح بالا و مشاهدات رطوبت خاک استفاده کردند. (Ge *et al.*, 2019) برای برآورد رطوبت خاک از تصاویر فراترینی و الگوریتم‌های جنگل تصادفی (RF) و ماشین یادگیری افراطی یا سریع (ELM) توسط شش روش پیش‌آزمایشی به همراه چهار شاخص طیفی استفاده کردند. دقت اعتبار روش برآورد شده در مقایسه با مدل‌های خطی بالا بود. مقایسه دو الگوریتم یادگیری ماشین در این پژوهش نشان داد که مدل‌های RF نسبت به مدل‌های ELM برتری دارند. در نتیجه نقشه رطوبت خاک پیش‌بینی‌شده بسیار شبیه به نقشه رطوبت خاک اندازه‌گیری شده بود. این پژوهشگران دریافتند الگوریتم‌های یادگیری ماشین، ارزیابی برآورد رطوبت خاک را در روش ترکیب شاخص‌های طیفی با دقت بالا ( $R^2 = 0.907$ ) پیش‌بینی می‌کنند. (Cai *et al.*, 2019) مدل پیش‌بینی رطوبت خاک را با روش یادگیری عمیق با استفاده از الگوریتم رگرسیون شبکه عصبی عمیق<sup>۱</sup> (DNNR) در ترکیب با سری‌های زمانی ارائه دادند. این مدل توانست با استفاده از داده‌های هواشناسی انتخاب‌شده پیش‌بینی دقیقی از رطوبت داشته باشد و مبنای مؤثری برای آبیاری و صرفه‌جویی در مصرف آب فراهم کند. در پژوهشی دیگر (Adab *et al.*, 2020) قابلیت روش‌های یادگیری ماشین برای برآورد رطوبت خاک سطحی با استفاده از الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان<sup>۲</sup> (SVM)، شبکه عصبی مصنوعی<sup>۳</sup> (ANN) و رگرسیون شبکه الاستیک<sup>۴</sup> (EN) با باندهای نوری و حرارتی لندست ۸ موردبررسی قرار دادند. با ترکیب انعکاس طیفی سطحی و داده‌های مکانی رطوبت خاک نتایج قابل‌قبولی به دست آوردند. مقایسه‌های آماری آن‌ها نشان داد که روش RF با بالاترین مقدار بازده نش-ساتکلایف<sup>۵</sup> ( $0.73$ ) برای برآورد رطوبت خاک برای پوشش انواع کاربری‌های مختلف زمین مناسب است. در مطالعه‌ای دیگر (Babaeian *et al.*, 2021) به‌منظور توسعه روشی جدید برای برآورد رطوبت ناحیه ریشه، با استفاده از روش یادگیری ماشین خودکار<sup>۶</sup> دوازده مدل را از ترکیب داده‌های انعکاس نوری با وضوح مکانی بالا و اطلاعات فیزیکی و هیدرولیکی خاک به دست آوردند. مدل‌های استخراج‌شده برآوردی دقیق از مقدار رطوبت ناحیه ریشه با مقادیر RMSE<sup>۷</sup> کمتر از  $0.02$  سانتی‌متر مکعب بر سانتی‌متر مکعب و مقادیر همبستگی بالای  $0.90$  داشتند. (Acharya *et al.*, 2021) در مطالعه‌ای به‌منظور ارزیابی برآورد رطوبت خاک با استفاده از داده‌های خاک، پوشش گیاهی و هواشناسی از روش‌های یادگیری ماشین هم‌چون درخت‌های طبقه‌بندی و رگرسیون<sup>۸</sup> (CART)، رگرسیون

1 Random Forest  
2 Extreme learning machine  
3 Deep Neural Network Regression  
4 Support Vector Machines  
5 Artificial Neural Networks  
6 Elastic Net Regression  
7 Nash-Sutcliffe  
8 Automated Machine Learning  
9 Root mean square error  
10 Classification and regression trees

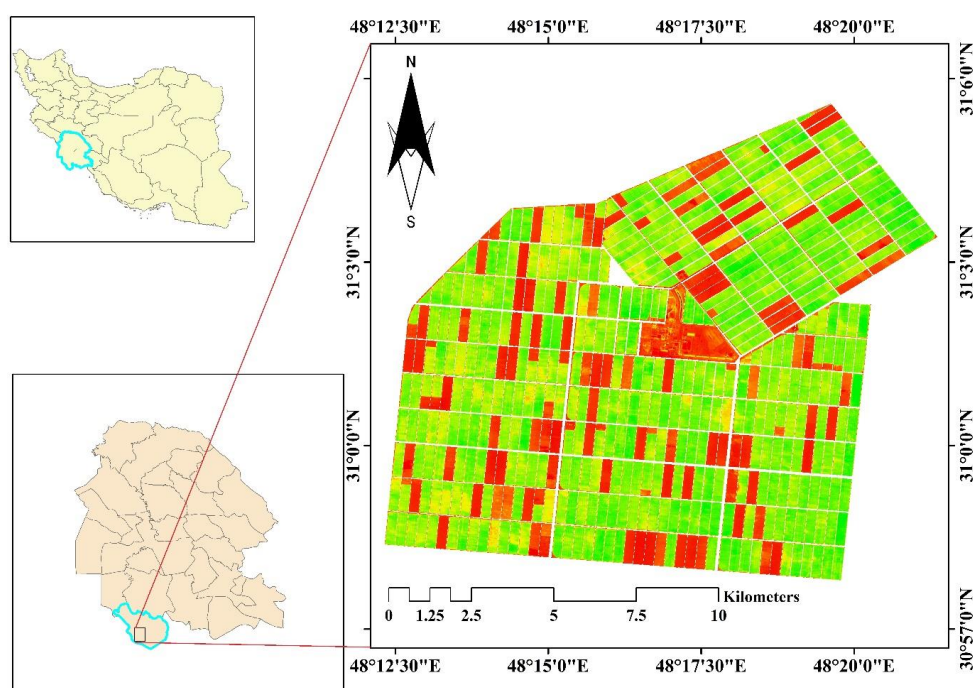
جنگل تصادفی<sup>۱</sup> (RFR)، رگرسیون درختان تقویت‌شده<sup>۲</sup> (BRT)، رگرسیون خطی چندگانه<sup>۳</sup> (MLR)، رگرسیون بردار پشتیبان<sup>۴</sup> (SVR) و شبکه‌های عصبی مصنوعی<sup>۵</sup> (ANN) استفاده کردند. نتایج این پژوهش نشان داد که الگوریتم‌های RFR و BRT با میانگین خطای مطلق  $MAE < 0.040 \text{ m}^3 \text{ m}^{-3}$  بهترین عملکرد را در پیش‌بینی رطوبت خاک دارند.

با توجه به اصل پویایی مکانی و زمانی رطوبت خاک و ارتباط آن با ویژگی‌های فیزیکی و هیدرولیکی خاک، برآورد رطوبت خاک با استفاده از ترکیب داده‌های فیزیکی/هیدرولیکی خاک با داده‌های سنجش‌ازدور در مقیاس مزرعه ضروری است. اما تاکنون مدل‌های ترکیبی و ارزیابی آن‌ها با استفاده از روش‌های مختلف یادگیری ماشین در زمینه برآورد رطوبت خاک در مقیاس مزرعه ارائه نشده است. بنابراین هدف اصلی این پژوهش (۱) توسعه روشی سریع، کم‌هزینه و قابل‌اعتماد برای برآورد رطوبت خاک با استفاده از ترکیب داده‌های فیزیکی خاک و مشاهدات سنجش‌ازدور نوری به کمک روش‌های مختلف یادگیری ماشین و (۲) شناسایی مهم‌ترین پارامترهای فیزیکی خاک و متغیرهای سنجش‌ازدور نوری برای برآورد رطوبت خاک در مقیاس مزرعه برای گیاه نیشکر بود.

## مواد و روش‌ها

### منطقه مورد مطالعه

در کشت‌و صنعت نیشکر امیرکبیر واقع در جنوب استان خوزستان، یکی از هفت شرکت توسعه‌ی کشت نیشکر و صنایع جانبی، در سال زراعی ۱۴۰۰ به‌منظور امکان نظارت و مدیریت بر فرآیند پایش و مدیریت مصرف آب انتخاب شد. مساحت کل کشت و صنعت ۱۴۰۰۰ هکتار و سطح زیر کشت آن ۱۲۰۰۰ هکتار است که دارای ۴۸۰ مزرعه ۲۵ هکتاری است (شکل ۱).



شکل ۱: موقعیت منطقه مطالعاتی (کشت و صنعت امیرکبیر)

### اندازه‌گیری ویژگی‌های فیزیکی و هیدرولیکی خاک

با استفاده از رطوبت حجمی ۱۶۶ نمونه خاک در تاریخ‌های ۲۸ تیر، ۱۷ مرداد و ۱۶ شهریورماه از لایه سطحی خاک (۱۰-۰ سانتی متر) به‌صورت مستقیم و به روش وزنی انجام شد و از حاصل ضرب آن در وزن مخصوص ظاهری، رطوبت حجمی خاک محاسبه گردید.

- 1 Random forest regression
- 2 Boosted regression trees
- 3 Multiple linear regression
- 4 Support vector regression
- 5 Artificial neural networks

فراوانی نسبی ذرات خاک (درصد رس، سیلت و شن) به روش هیدرومتر، وزن مخصوص ظاهری (BD) به روش حجمی (ASTM, 1999)، تخلخل با استفاده از وزن مخصوص ظاهری با فرض وزن مخصوص حقیقی  $2/65$  گرم در سانتی متر مکعب محاسبه شد (Flint and Flint., 2002). مقدار آب در ظرفیت مزرعه<sup>۱</sup> ( $\theta_{FC}$ ) و نقطه پژمردگی دائمی<sup>۲</sup> ( $\theta_{PWP}$ )، دو پارامتر مهم برای مدیریت آبیاری کشاورزی هستند هستند که با استفاده از منحنی رطوبتی آب در خاک<sup>۳</sup> (SWC) با توابع انتقالی محاسبه شد. توابع انتقالی برای محاسبه خواص هیدرولیکی خاک از ویژگی‌های پایه‌ای خاک مانند فراوانی نسبی ذرات خاک (درصد رس، سیلت و شن) و وزن مخصوص ظاهری خاک استفاده می‌کند. در این مطالعه از توابع انتقالی<sup>۴</sup> ROSETTA برای تخمین پارامترهای مدل ونگنوختن<sup>۵</sup> ( $\theta_r, \theta_s, \alpha, n$ ) و در نتیجه برای تعیین  $\theta_{PWP}$  و  $\theta_{FC}$  استفاده شد (Schaap et al. 2001; Babaeian et al., 2021).

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha h)^n]^{1-1/n}} \quad \text{رابطه (۱)}$$

که در آن  $\theta$  مقدار رطوبت خاک ( $\text{cm}^{-3} \text{ cm}^{-3}$ )،  $\theta_r$ : مقدار رطوبت باقی مانده ( $\text{cm}^{-3} \text{ cm}^{-3}$ )،  $\theta_s$ : مقدار رطوبت اشباع ( $\text{cm}^{-3} \text{ cm}^{-3}$ )،  $\alpha$  و  $n$ : پارامترهای شکل هستند.

داده‌های ماهواره‌ای سنتینل-۲ با عنوان تصاویر سطح 2B از سایت ماهواره سنتینل به آدرس <https://scihub.copernicus.eu/dhus/#/home> از ۱۰ اردیبهشت تا ۱۶ شهریورماه در طول فصل رشد گیاه نیشکر دانلود و با نرم‌افزار ArcMap 10.5 و QGIS 2.14 تصحیحات اتمسفری، رادیومتریک و هندسی تصاویر بررسی شد. درجات روشنایی هر تصویر خام به مقادیر انعکاس طیفی تبدیل و انعکاس طیفی خاک و پوشش گیاهی در باندهای قرمز و مادون قرمز نزدیک برای محاسبه شاخص تفاضلی نرمال شده گیاهی (NDVI) و (STR) انعکاس تبدیل شده مادون قرمز طول موج کوتاه (SWIR) به شرح زیر استفاده شد (شکل ۲ و روابط ۲ و ۳). با استفاده از نرم‌افزار Matlab 2021 مجموعه داده ماهواره‌ای به ۷۰٪ داده‌های آموزشی و ۳۰٪ اعتبار سنجی تقسیم و الگوریتم‌های موردنظر اعمال شد.

$$NDVI = \frac{R_{NIR} - R_{red}}{R_{NIR} + R_{red}} \quad \text{رابطه (۲)}$$

$$STR = \frac{(1 - R_{SWIR})^2}{2R_{SWIR}} \quad \text{رابطه (۳)}$$

که در آن NDVI: شاخص تفاضلی پوشش گیاهی،  $R_{NIR}$ : انعکاس مادون قرمز نزدیک،  $R_{red}$ : انعکاس باند قرمز و SWIR: انعکاس مادون قرمز طول موج کوتاه است. چارچوب پردازش مجموعه داده‌ها با الگوریتم‌های موجود و مشخصات مدل‌ها و پارامترهای ورودی به آن برای برآورد رطوبت خاک در شکل ۳ و جدول ۱ نشان داده شده است. روش‌های یادگیری ماشین در این مطالعه عبارت‌اند از الگوریتم‌های رگرسیون خطی گام‌به‌گام<sup>۶</sup>، ماشین بردار پشتیبان، شبکه عصبی، درخت تصمیم‌گیری<sup>۹</sup> و یادگیری جمعی<sup>۱۰</sup> شامل درختان توسعه‌یافته<sup>۱۱</sup> و درختان کیسه‌گذاری<sup>۱۲</sup>.

### رگرسیون گام‌به‌گام

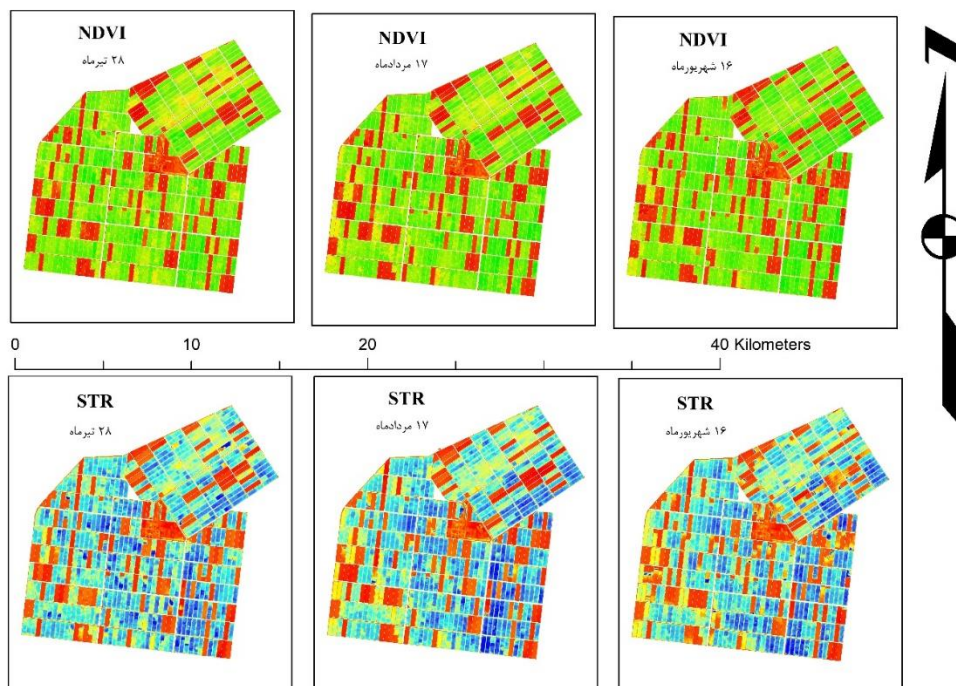
یکی از انواع رگرسیون خطی چند متغیره است. در این روش‌ها می‌توان هم‌زمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. این روش برای برازش مدل‌های رگرسیونی با انتخاب خودکار متغیرهای پیش‌بینی استفاده می‌شود. در این روش رابطه بین متغیرها به صورت تابعی از متغیر وابسته است که متغیرهای مستقل را پیش‌بینی می‌کند (رابطه ۴):

- 1 Field capacity
- 2 Permanent wilting point
- 3 Soil Water Characteristic
- 4 pedotransfer functions
- 5 van Genuchten SWC model
- 6  $\theta_r$  and  $\theta_s$  are the residual and saturated volumetric moisture contents ( $\text{cm}^3 \text{ cm}^{-3}$ ),
- 7  $\alpha$  ( $\text{cm}^{-1}$ ) and  $n$  ( ) are scaling and shape parameters of the van Genuchten model, respectively.
- 8 Stepwise Regression
- 9 Decision tree (DT)
- 10 Ensembling learning
- 11 Boosted trees
- 12 Bagged trees

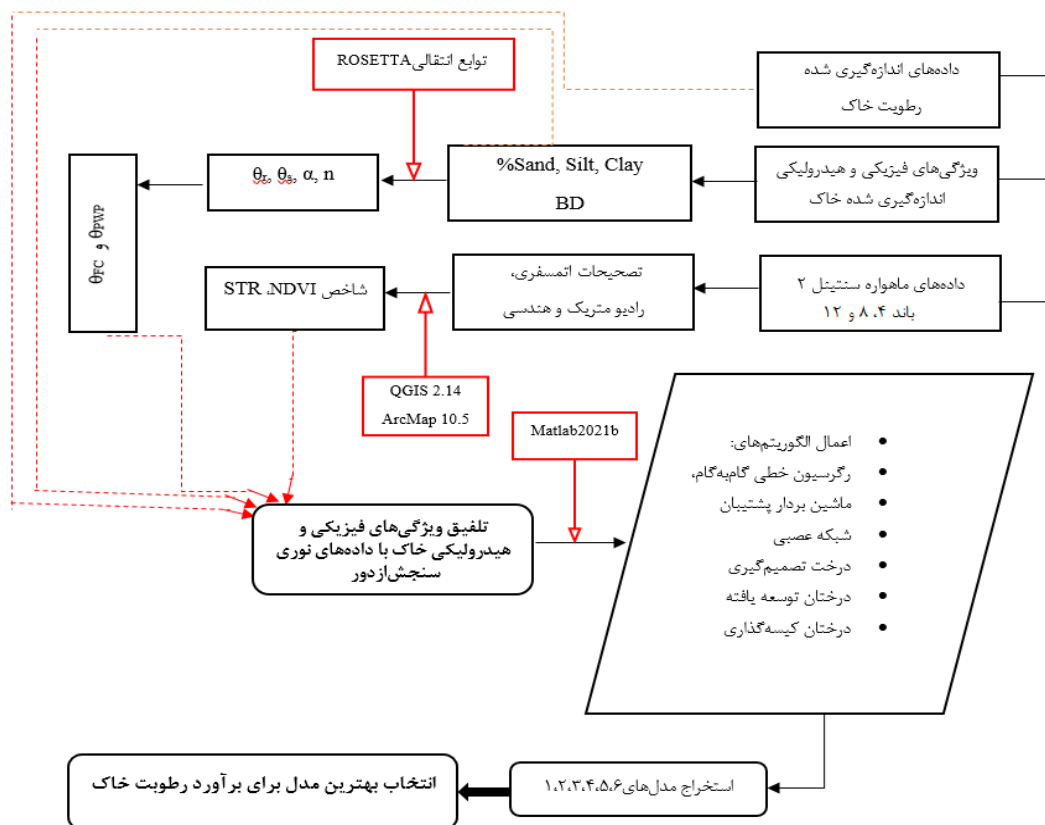
$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

رابطه ۴)

که در این معادله Y متغیر پیش‌بینی شونده (رطوبت خاک)،  $x_1$  تا  $x_n$  متغیرهای پیش‌بینی کننده و a و b ضرایب ثابت معادله هستند. در رگرسیون گام‌به‌گام تمامی متغیرهای مستقل وارد مدل می‌شوند و آن متغیر مستقلی که تأثیر چندانی بر متغیر وابسته نداشته باشد از مدل حذف می‌شود (Gorthi, 2011).



شکل ۲: شاخص NDVI و STR در طول دوره رشد گیاه نیشکر



شکل ۳: فلوچارت برآورد رطوبت خاک سطحی با استفاده از روش یادگیری ماشین

### ماشین بردار پشتیبان (SVM)

یک روش طبقه‌بندی نظارت‌شده است که یک ابر صفحه یا مرز بین دو کلاس داده را پیدا می‌کند که حاشیه بین دو کلاس را به حداکثر رساند. صفحات زیادی وجود دارند که می‌توانند کلاس را از هم جدا کنند، اما فقط یک صفحه می‌تواند حاشیه یا فاصله بین کلاس‌ها را به حداکثر برساند. ماشین بردار پشتیبان شامل استفاده از زیرمجموعه‌ای از نقاط داده است که بر اساس یک حاشیه خطی از پیش تعریف‌شده برای جا دادن یک مدل رگرسیون بین متغیر وابسته و متغیرهای مستقل تعریف می‌شود. به آن دسته از نقاط داده، بردارهای پشتیبانی می‌گویند. فرض کنید، نمونه‌های  $X = \{x_1, x_2, \dots, x_i\}$  و مقادیر هدف مربوطه  $Y = \{y_1, y_2, \dots, y_i\}$ ،  $y_i \in R$  باشد، آنگاه هدف SVR یافتن تابع  $f(x)$  است که حداکثر انحراف استاندارد از هدف به دست آمده  $y_i$  برای همه داده‌های آموزشی را داشته باشد. (Yang et al., 2009)

### شبکه عصبی (ANN)

معمولاً از لایه‌های ورودی، پنهان و خروجی تشکیل شده است که به صورت مجزا مرتب شده‌اند (Priddy et al., 2008). لایه‌های پنهان اطلاعات مفیدی را از لایه ورودی استخراج می‌کنند که برای تخمین خروجی استفاده می‌شود. شبکه‌های عصبی مصنوعی ابزار محاسباتی قدرتمندی هستند که از طریق بسیاری از عناصر به هم پیوسته ساده به نام نورون‌ها با قابلیت منحصر به فرد در تشخیص روابط زیربنایی با رویدادهای ورودی و خروجی ساخته شده‌اند. مجموعه‌ای از نورون‌ها که در آرایه بعدی چیده شده‌اند یک لایه نامیده می‌شود که در آن هر لایه شامل یک یا چند گره مجزا است. تعداد متغیرهای ورودی لازم برای پیش‌بینی متغیرهای خروجی مورد نظر، تعداد گره‌های ورودی را تعیین می‌کند. پیچیدگی مدل‌سازی به تعداد بهینه گره‌های پنهان و لایه‌های پنهان بستگی دارد (تعداد بیشتر لایه‌های پنهان منجر به مدل بزرگ‌تر و پیچیده‌تر می‌شود) (Grimes et al., 2003). مدل‌های ANN توانایی یادگیری انسان را با یادگیری از مجموعه داده‌های آموزشی تقلید می‌کنند. آن‌ها ابزار قدرتمندی برای روابط غیرخطی چند متغیره هستند (Twarakavi et al., 2006). این مدل‌ها برای همبستگی‌های ورودی-خروجی فرآیندهای غیرخطی در منابع آب و هیدرولوژی (Ahmad et al., 2003) استفاده می‌شوند.

### درخت تصمیم (DT)

یک نوع طبقه‌بندی کننده نظارت‌شده سلسله مراتبی است که جزء روش‌های دانش‌پایه به شمار می‌آید؛ به عبارت دیگر در این روش با داشتن اطلاعات در مورد پدیده‌های موضوعی و با گرفتن تصمیمات مناسب از بالا به پایین کلاس‌بندی انجام می‌شود (Breiman et al., 1996). در این روش مجموعه‌ای از شرط‌های منطقی، به صورت الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود. ایجاد درخت تصمیم دو مرحله دارد: مرحله اول، ایجاد و رشد درخت است که این مرحله شامل پیوند و انشعاب است. مرحله دوم، مرحله توقف و هرس است. هدف مرحله اخیر، به حداقل رساندن خطای پیش‌بینی است (Taghizadeh-Mehrjardi et al., 2012).

### یادگیری جمعی

روشی است که در آن از الگوریتم‌های یادگیری چندگانه به طور هم‌زمان استفاده می‌شود. به طور کلی می‌توان یادگیری جمعی را به دودسته روش درختان توسعه‌یافته و کیسه‌گذاری تقسیم کرد. Friedman (2002) درختان رگرسیونی توسعه‌یافته (BRT) را به عنوان یک مدل درخت تصمیم تعریف کرد که توسط الگوریتم تقویت گرادیان بهبود می‌یابد و ترکیبی از دو روش آماری بوستینگ<sup>۱</sup> و رگرسیون درختی است. بوستینگ یک روش مرحله‌ای روبه‌جلو است که در آن، مدل‌های درختی به صورت تکرارپذیر با زیرمجموعه‌ای از داده‌های آموزشی برازش داده می‌شوند. در برازش رگرسیون درختی توسعه‌یافته باید دو پارامتر، نرخ یادگیری و پیچیدگی درخت مشخص گردد. مقدار یادگیری، سهم هر درخت متوالی را در مدل نهایی تعیین می‌کند. پیچیدگی درخت، اثرات اصلی یا اثرات متقابل بین متغیرها را نشان می‌دهد (Elith et al., 2008; Natekin et al., 2013). مدل‌های BRT به دلیل توانایی عملکرد بهتر با داده‌های پیچیده، در بین دانشمندان داده محبوب‌تر هستند (Friedman et al., 2001). BRT به تقسیم‌بندی تأثیرات متغیرهای مستقل (پیش‌بینی کننده) بر روی متغیر وابسته

1 Boosted Regression Trees

2 Boosting



(رطوبت خاک برای این مطالعه) کمک می‌کند.

کیسه گذاری<sup>۱</sup> نیز روشی است که توسط (Breiman et al., 1996) توسعه یافته است. ثبات و دقت پیش‌بینی را می‌توان با استفاده از این روش بهبود بخشید. نسخه‌های متعدد مدل‌ها با بوت استرپ مجموعه داده‌ها به‌عنوان مجموعه‌های آموزشی برای ساخت مدل‌های مختلف شکل می‌گیرند. طبقه‌بندی در این روش بر اساس بیشترین رأی برای پیش‌بینی یک متغیر در نظر گرفته می‌شود درحالی‌که در رگرسیون، میانگین تمام نسخه‌ها را به‌عنوان پیش‌بینی یک متغیر در نظر می‌گیرد. در این مطالعه به‌منظور منظم کردن برآوردهای رطوبت خاک از روش یادگیری جمعی غیرخطی و غیر پارامتری کیسه گذاری شده استفاده شد (Gorthi 2011).

جدول ۱: پارامترهای ورودی مورداستفاده برای ۷ مدل پیشنهادی

متغیرهای ورودی	مدل
STR, NDVI	مدل ۱
STR, NDVI, $\theta_{FC}$ , $\theta_{pwp}$	مدل ۲
STR, NDVI, Text, BD	مدل ۳
STR, NDVI, Text, BD, $\theta_{FC}$ , $\theta_{pwp}$	مدل ۴
STR, Text, BD, $\theta_{FC}$ , $\theta_{pwp}$	مدل ۵
NDVI, Text, BD, $\theta_{FC}$ , $\theta_{pwp}$	مدل ۶
STR, NDVI, Text, BD, $\theta_{FC}$ , $\theta_{pwp}$ , $\Phi$	مدل ۷

STR: انعکاس تبدیل‌شده طول‌موج مادون‌قرمز کوتاه (SWIR)، NDVI: شاخص تفاضلی نرمال شده گیاهی، Text: بافت خاک (درصد توزیع اندازه ذرات خاک)، BD: وزن مخصوص ظاهری خاک (گرم بر سانتی‌متر مکعب)،  $\theta_{FC}$ : مقدار رطوبت در ظرفیت مزرعه (سانتی‌متر مکعب بر سانتی‌متر مکعب)،  $\theta_{pwp}$ : مقدار رطوبت در نقطه پژمردگی دائم (سانتی‌متر مکعب بر سانتی‌متر مکعب)؛  $\Phi$  مقدار تخلخل خاک.

#### ارزیابی دقت مدل‌ها

با استفاده از آماره‌های ضریب تعیین<sup>۲</sup> ( $R^2$ )، جذر میانگین مربعات خطا<sup>۳</sup> (RMSE) و میانگین خطای مطلق<sup>۴</sup> (MAE) انجام شد.

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

که در اینجا  $\theta_{obs}$  مقدار رطوبت خاک اندازه‌گیری شده و  $\theta_{est}$  مقدار رطوبت خاک برآورد شده،  $\bar{\theta}_{obs}$  و  $\bar{\theta}_{est}$  به ترتیب میانگین رطوبت خاک اندازه‌گیری شده و برآورد شده است. ضریب تعیین نشان می‌دهد که چند درصد تغییرات متغیر وابسته توسط متغیر مستقل تعیین می‌شوند.

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

مقدار RMSE نشان‌دهنده این است که مقادیر پیش‌بینی شده تا چه حد اندازه‌گیری را بیشتر یا کمتر تخمین زده‌اند. بهترین حالت برای RMSE این است که مقادیر پیش‌بینی شده و اندازه‌گیری شده باهم برابر باشند یعنی برابر با صفر باشد.

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

MAE یکی از توابع خطا است که از فاصله بین مقدار پیش‌بینی و مقدار واقعی به‌عنوان معیار استفاده می‌کند، به عبارتی میانگین قدر مطلق تفاضل بین مقدار پیش‌بینی و واقعی را محاسبه می‌کند.

## نتایج و بحث

### تغییرپذیری مکانی ویژگی‌های فیزیکی و هیدرولیکی خاک در مزرعه

آمار پایه شامل مقادیر حداکثر، حداقل و میانگین خصوصیات فیزیکی خاک اندازه‌گیری شده شامل فراوانی نسبی ذرات خاک (درصد رس، سیلت و شن)، وزن مخصوص ظاهری، تخلخل خاک، رطوبت‌های ظرفیت زراعی و نقطه پژمردگی در جدول ۲ ذکر شده است. نتایج نشان می‌دهد که بافت خاک سراسر مزرعه بسیار متغیر است (silty clay loam, silty clay loam, sandy clay loam, loam, sandy loam, clay, clay loam, silt loam). مقدار وزن مخصوص ظاهری خاک از ۱/۴۵ تا ۱/۹۰ گرم بر سانتی‌متر مکعب با مقدار متوسط ۱/۶۷ گرم بر سانتی‌متر مکعب متغیر است.

جدول ۲: آمار پایه ویژگی‌های فیزیکی و هیدرولیکی اندازه‌گیری شده نمونه‌های خاک به‌عنوان متغیرهای ورودی ۷ مدل پیشنهادی

میانگین	حداکثر	حداقل	خصوصیات خاک
۲۶/۵۶	۴۸	۱۳/۲۵	درصد رس
۳۵/۲۳	۵۱/۷۵	۲۱/۵۰	درصد سیلت
۳۸/۲۱	۶۳	۹/۵۰	درصد شن
۱/۶۷	۱/۹۰	۱/۴۵	چگالی ظاهری ( $\text{gr}/\text{cm}^3$ )
۰/۲۳	۰/۴۶	۰/۱۶	رطوبت در ظرفیت زراعی ( $\text{cm}^3/\text{cm}^3$ )
۰/۱۳	۰/۵۸	۰/۰۷	رطوبت در نقطه پژمردگی ( $\text{cm}^3/\text{cm}^3$ )
۰/۰۹	۰/۵۲	۰/۰۲	رطوبت حجمی اندازه‌گیری شده
۳۷	۴۵	۲۸	درصد تخلخل

جدول ۳: میانگین پارامترهای منحنی رطوبتی مدل ونگنوختن (۱۹۸۰) با استفاده از روش توابع انتقالی (Schaap et al. 2001) ROSETTA

مقدار	پارامتر هیدرولیکی
۰/۰۸	$\theta_r$ ( $\text{cm}^3/\text{cm}^3$ )
۰/۳۶	$\theta_s$ ( $\text{cm}^3/\text{cm}^3$ )
۰/۰۲	$\alpha$ ( $\text{cm}^{-1}$ )
۱/۳۲	$n$ (-)

$\theta_r$ : مقدار رطوبت باقی‌مانده،  $\theta_s$ : مقدار رطوبت اشباع،  $\alpha$  و  $n$ : پارامترهای شکل منحنی رطوبتی هستند.

جدول ۳ مقادیر میانگین پارامترهای مدل مشخصه رطوبت خاک ونگنوختن (معادله ۱) را نشان می‌دهد که از توابع انتقالی ROSETTA (Schaap et al., 2001) با استفاده از درصد رس، سیلت و شن و وزن مخصوص ظاهری خاک به‌عنوان پارامترهای ورودی به‌دست آمده است و سپس برای تعیین مقادیر رطوبت‌های ظرفیت زراعی و نقطه پژمردگی از آن‌ها استفاده شده است (جدول ۲).  $\theta_{FC}$  و  $\theta_{PWP}$  که عمدتاً تحت تأثیر بافت خاک بوده (Sadeghi et al., 2018) به ترتیب دارای مقادیر میانگین  $(\text{cm}^{-3} \text{ cm}^{-3})$  ۰/۲۳ و ۰/۱۳ به دست آمد. در مدل Acharya et al., (2021) مقادیر شن، سیلت و رس به ترتیب ۲۹، ۳۰ و ۳۴ درصد در افزایش RMSE نقش داشتند که نشان‌دهنده اهمیت نسبی آن‌ها است.

### ارزیابی عملکرد مدل‌ها در برآورد رطوبت خاک

با شاخص‌های اندازه‌گیری شامل ضریب تعیین ( $R^2$ )، جذر میانگین مربعات خطا (RMSE) و میانگین خطای مطلق (MAE) انجام شد. دقت برآورد مدل‌ها در جداول ۴ تا ۶ نشان می‌دهد که مدل‌های ۴ و ۶ به ترتیب به‌عنوان قوی‌ترین و ضعیف‌ترین مدل برآورد رطوبت خاک عمل کرده‌اند (شکل ۳ و ۴). همان‌طور که مشاهده می‌شود بهترین روش ارزیابی با بالاترین مقدار ضریب تعیین و کمترین مقدار RMSE و MAE مربوط به روش رگرسیون خطی گام‌به‌گام در ۷ مدل پیشنهادی است. همه الگوریتم‌های یادگیری ماشین مورد مطالعه به‌غیر از شبکه عصبی مصنوعی از نظر مقادیر RMSE،  $R^2$  و MAE به‌خوبی عمل کردند. بیشترین مقدار RMSE برابر با ۰/۱۱، ۰/۰۷۰، ۰/۰۶۵ و ۰/۰۶۰ در این الگوریتم و به ترتیب مربوط به مدل‌های ۶، ۴، ۷، ۵ است. نتایج مطالعات Prakash. et al., (2019) که با استفاده از رگرسیون خطی چندگانه، ماشین بردار پشتیبان و شبکه عصبی، سه مجموعه داده مختلف برای پیش‌بینی رطوبت خاک استفاده کردند

نشان داد که رگرسیون خطی چندگانه نسبت به ماشین بردار پشتیبان و شبکه عصبی برتری دارد. همچنین، بیشترین مقادیر  $R^2$  (۰/۸۷) مربوط به رگرسیون خطی گام‌به‌گام در مدل ۴، ماشین بردار پشتیبان در مدل ۱ (۰/۸۶)، درختان توسعه‌یافته در مدل ۳ (۰/۸۵)، شبکه عصبی در مدل ۱، ۲ (۰/۸۴)، درخت تصمیم‌گیری در مدل ۴ و ۵ (۰/۸۳) و درختان کیسه‌گذاری در مدل ۲ (۰/۸۱) است. کمترین مقادیر خطا (MAE) نیز به ترتیب مربوط به رگرسیون خطی گام‌به‌گام در مدل ۴ (۰/۰۳۴)، ماشین بردار پشتیبان، شبکه عصبی و درختان کیسه‌گذاری در مدل ۱ (۰/۰۳۶ و ۰/۰۳۷ و ۰/۰۳۸)، درختان توسعه‌یافته در مدل ۱، ۳ و ۴ (۰/۰۳۸)، درخت تصمیم‌گیری در مدل ۱، ۴ و ۵ (۰/۰۴۰) است. طبق نتایج Adab et al., (2020) مدل‌های شبکه عصبی در مقایسه با ماشین بردار پشتیبان در شرایط مرطوب پیش‌بینی رطوبت خاک را بهتر نشان می‌دهند. از آنجایی که میانگین رطوبت خاک در ۱۶۶ نقطه اندازه‌گیری شده برابر با ۰/۰۹ سانتی‌متر مکعب بر سانتی‌متر مکعب است پس با نتایج این پژوهش سازگاری دارد. نتایج Acharya et al., (2021) نشان داد که مدل‌های یادگیری ماشینی مانند جنگل تصادفی، درختان توسعه‌یافته و ماشین بردار پشتیبان عملکرد بهتری نسبت به مدل‌های شبکه عصبی، رگرسیون خطی چندگانه و درختان طبقه‌بندی و رگرسیون (CART) دارند و این موضوع احتمالاً به دلیل الگوریتم‌های بهینه‌سازی پارامترهای انتخاب‌شده با دقت مختلف برای آموزش مدل‌ها است. برتری جنگل تصادفی و ماشین بردار پشتیبان نسبت به مدل‌های دیگر نیز در مطالعات مختلف گزارش شده است (Achieng, 2019؛ Kalra and Ahmad, 2009). در نتایج Araya et al., (2021) دو الگوریتم درختان توسعه‌یافته و جنگل تصادفی با مقدار MAE کمتر از ۰/۴ رطوبت خاک، عملکرد بهتری داشتند که با مطالعات دیگر مطابقت دارد و نشان می‌دهند مدل‌های رگرسیون مبتنی بر درخت تصمیم بهتر از سایر الگوریتم‌های یادگیری ماشینی عمل می‌کنند (Caruana and Niculescu-Mizil, 2006).

جدول ۴: مقایسه مقادیر  $R^2$  در شش الگوریتم یادگیری ماشینی برای هفت مدل پیشنهادی

مدل‌ها	الگوریتم‌ها	رگرسیون خطی گام‌به‌گام	ماشین بردار پشتیبان	درختان توسعه‌یافته	درخت تصمیم‌گیری	درختان کیسه‌گذاری	شبکه عصبی
مدل ۱	۰/۸۶	۰/۸۶	۰/۸۶	۰/۸۱	۰/۸۲	۰/۸۰	۰/۸۴
مدل ۲	۰/۸۷	۰/۸۷	۰/۸۵	۰/۸۳	۰/۸۲	۰/۸۱	۰/۸۴
مدل ۳	۰/۸۶	۰/۸۶	۰/۸۵	۰/۸۵	۰/۸۱	۰/۷۶	۰/۷۷
مدل ۴	۰/۸۷	۰/۸۷	۰/۸۳	۰/۸۳	۰/۸۳	۰/۷۹	۰/۶۶
مدل ۵	۰/۸۵	۰/۸۵	۰/۸۴	۰/۸۲	۰/۸۳	۰/۸۰	۰/۷۴
مدل ۶	۰/۵۴	۰/۵۴	۰/۵۲	۰/۴۶	۰/۴۸	۰/۵۳	۰/۱۱
مدل ۷	۰/۸۵	۰/۸۵	۰/۸۳	۰/۸۴	۰/۸۱	۰/۷۶	۰/۷۱

جدول ۵: مقایسه مقادیر RMSE در شش الگوریتم یادگیری ماشینی برای هفت مدل پیشنهادی

مدل‌ها	الگوریتم‌ها	رگرسیون خطی گام‌به‌گام	ماشین بردار پشتیبان	درختان توسعه‌یافته	درخت تصمیم‌گیری	درختان کیسه‌گذاری	شبکه عصبی
مدل ۱	۰/۰۴۴	۰/۰۴۴	۰/۰۴۵	۰/۰۵۱	۰/۰۵۰	۰/۰۵۳	۰/۰۴۷
مدل ۲	۰/۰۴۴	۰/۰۴۴	۰/۰۴۷	۰/۰۴۹	۰/۰۵۱	۰/۰۵۳	۰/۰۵۴
مدل ۳	۰/۰۴۵	۰/۰۴۵	۰/۰۴۶	۰/۰۴۷	۰/۰۵۲	۰/۰۵۹	۰/۰۵۷
مدل ۴	۰/۰۴۲	۰/۰۴۲	۰/۰۴۹	۰/۰۴۹	۰/۰۴۹	۰/۰۵۵	۰/۰۷۰
مدل ۵	۰/۰۴۶	۰/۰۴۶	۰/۰۴۸	۰/۰۵۰	۰/۰۵۰	۰/۰۵۳	۰/۰۶۰
مدل ۶	۰/۰۸۱	۰/۰۸۱	۰/۰۸۲	۰/۰۸۵	۰/۰۸۷	۰/۰۸۱	۰/۱۱
مدل ۷	۰/۰۴۶	۰/۰۴۶	۰/۰۵۰	۰/۰۴۸	۰/۰۵۲	۰/۰۵۹	۰/۰۶۵

بیشترین دقت تخمین (RMSE) در شش روش یادگیری ماشینی ذکر شده به ترتیب مربوط به رگرسیون خطی گام‌به‌گام در مدل ۴ (۰/۰۴۲)، ماشین بردار پشتیبان و شبکه عصبی در مدل ۱ (۰/۰۴۵ و ۰/۰۴۷)، درختان توسعه‌یافته در مدل ۳ (۰/۰۴۷)، درخت تصمیم‌گیری در مدل ۴ (۰/۰۴۹)، درختان کیسه‌گذاری در مدل ۱، ۲، ۳ (۰/۰۵۳) است. Chakrabarti et al., (2018) که از الگوریتم درختان کیسه‌گذاری برای کاهش مقیاس رطوبت خاک استفاده کردند، پیشنهاد استفاده از این الگوریتم با ویژگی‌هایی مانند شیب، خواص خاک ارائه

کردند.

مدل‌های ۲ و ۳ که هر کدام دارای چند پارامتر فیزیکی و هیدرولیکی (مدل ۲ شامل رطوبت‌های ظرفیت زراعی و نقطه پژمردگی و مدل ۳ شامل بافت و جرم مخصوص ظاهری خاک) به‌عنوان ورودی هستند، همبستگی مشابهی از خود نشان دادند. در مدل ۴ که ترکیبی از پارامترهای ذکر شده با داده‌های سنجش‌ازدور (STR و NDVI) است، به دلیل افزودن متوالی اطلاعات ورودی فیزیکی و هیدرولیکی خاک از جمله فراوانی نسبی ذرات خاک، وزن مخصوص ظاهری و تخلخل خاک، رطوبت خاک ظرفیت زراعی و نقطه پژمردگی با افزایش همبستگی و کاهش مقدار خطا همراه است. در رگرسیون خطی گام‌به‌گام مقادیر خطای RMSE و MAE به ترتیب برابر با ۰/۰۴۲ و ۰/۰۳۴ دارای کمترین مقدار است و این موضوع اهمیت خواص فیزیکی و هیدرولیکی خاک را به‌عنوان پارامترهای مهم ورودی در روش رگرسیون گام‌به‌گام نشان می‌دهد، اما در شبکه عصبی دقت تا حدودی کاهش یافته و مقدار RMSE و MAE به ترتیب برابر با ۰/۰۷۰ و ۰/۰۵۲ است؛ بنابراین در روش شبکه عصبی با افزایش پارامترهای ورودی به دلیل عدم قطعیت ذاتی همراه با هر پارامتر، افزایش مقدار خطا محتمل‌تر است؛ بنابراین در این پژوهش روش رگرسیون گام‌به‌گام در تلفیق ویژگی‌های فیزیکی و هیدرولیکی خاک برای برآورد غیرمستقیم رطوبت خاک موفق عمل کرده است.

در مدل‌های ۲، ۴ و ۷ که رطوبت‌های ظرفیت زراعی و نقطه پژمردگی به‌عنوان پارامترهای ورودی همراه با متغیرهای سنجش‌ازدور (STR و NDVI) استفاده قرار گرفته‌اند، با افزودن برخی پارامترهایی فیزیکی و هیدرولیکی خاک به‌عنوان متغیر ورودی، مقدار ضریب همبستگی در برخی مدل‌های یادگیری ماشین اندکی کاهش یافته که می‌تواند به دلیل اثر خطای ورودی همراه با هر ویژگی خاک و نیز همبستگی درونی بین رطوبت خاک ظرفیت زراعی و نقطه پژمردگی با توزیع اندازه ذرات و تخلخل خاک باشد. از این رو مدل ۲ تنها با در نظر گرفتن متغیرهای رطوبت‌های ظرفیت زراعی و نقطه پژمردگی نتایج نسبتاً بهتری در اکثر مدل‌های یادگیری ماشین ارائه نموده است. مطالعات نشان داده است که در روش‌های یادگیری ماشین افزایش تعداد پارامترهای ورودی می‌تواند تا حدودی منجر به کاهش دقت تخمین‌ها شود (Whittaker et al., 2010; Oymak and Soltanolkotabi, 2020). از طرفی، پارامترهای فراوانی نسبی ذرات خاک، وزن مخصوص ظاهری و مقادیر رطوبت خاک ظرفیت زراعی و نقطه پژمردگی در مقایسه با تخلخل خاک تخمین‌های بهتری را ارائه کردند. افزودن فراوانی نسبی ذرات خاک (درصد رس، سیلت و شن) به داده‌های ورودی (مثلاً مدل ۳) منجر به افزایش ضریب تعیین ( $R^2$ ) و کاهش مقدار خطای RMSE در الگوریتم درختان توسعه یافته شد.

با توجه به نتایج Acharya et al., (2021) مقادیر  $R^2$ ، RMSE و MAE در برآورد رطوبت خاک با شش روش مختلف یادگیری ماشین نشان دادند که جنگل تصادفی و درختان توسعه یافته دارای RMSE کمتر از ۰/۰۵۰ مترمکعب در مترمکعب نتایج رضایت بخشی دارند در حالی که مدل شبکه عصبی عملکرد ضعیف‌تری ( $R^2=0.53$ ,  $RMSE=0.085$ ) را برای داده‌های آزمایشی نشان داد. مقادیر  $R^2$  به ترتیب برای الگوریتم‌های درختان توسعه یافته و ماشین بردار پشتیبان ۰/۶۷ و ۰/۶۵ بود. در مطالعه آن‌ها مدل‌های ماشین بردار پشتیبان ( $RMSE=0.050$  و  $MAE=0.039$ ) و درختان توسعه یافته ( $RMSE=0.048$  و  $MAE=0.037$ ) قادر به ثبت رابطه بین رطوبت خاک در هر مزرعه با مقدار رطوبت حجمی ثبت شده در هر ایستگاه هواشناسی مربوطه، بارندگی، مقدار تبخیر و تعرق، پوشش گیاهی و عوامل خاک با مقادیر خطای پایین‌تر (RMSE و MAE) بود.

در مدل ۷ که بیشترین تعداد ورودی‌های مدل را دارد، مقدار دقت در مقایسه با مدل ۶ به‌طور زیادی بالا است. تفاوت مدل ۶ و ۷ در افزودن متغیر STR به‌عنوان ورودی مدل است که باعث افزایش دقت مدل ۷ در مقایسه با مدل ۶ در تمامی مدل‌های یادگیری ماشین شده است. بیشترین مقدار همبستگی در مدل ۷ مربوط به روش رگرسیون خطی گام‌به‌گام با  $R^2=0.85$  است. این یافته نقش قابل توجه متغیر SRT را در تخمین رطوبت خاک و نقش ویژگی‌های فیزیکی و هیدرولیکی خاک را در توصیف پویایی زمانی و مکانی رطوبت خاک به‌خوبی در مقیاس مزرعه نشان می‌دهد.

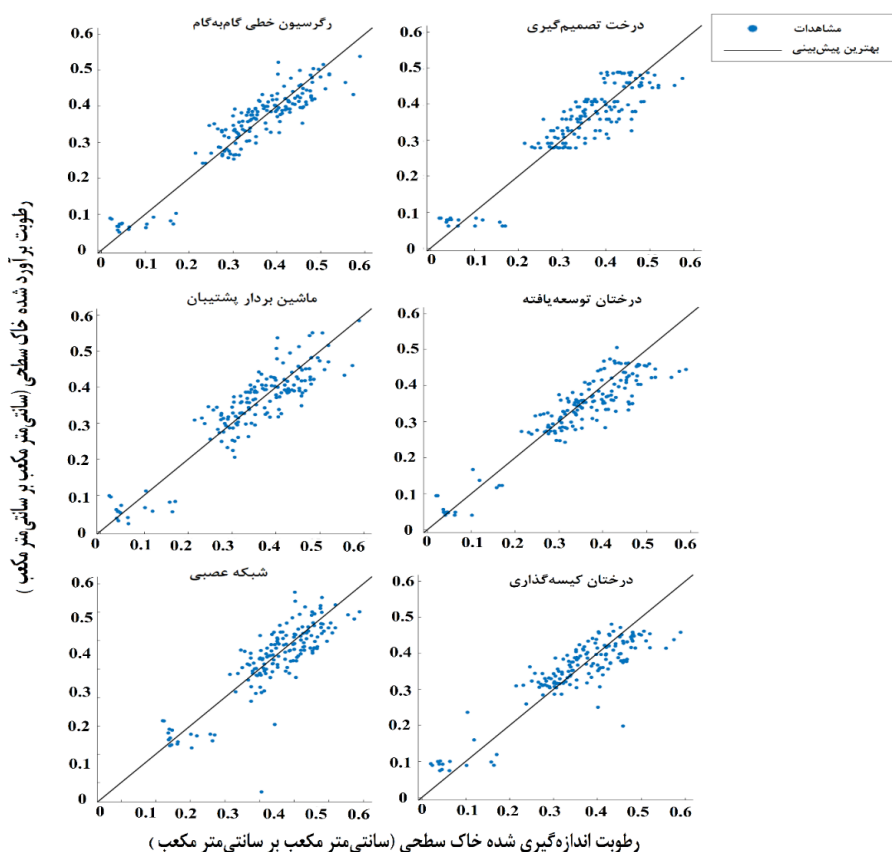
Babaeian et al (2021) نشان دادند به‌طور کلی، مقادیر خطای (RMSE) برای مدل‌هایی که از NDVI به‌عنوان پارامتر ورودی برای تخمین رطوبت خاک استفاده نمی‌کنند بالاترین (۰/۰۵۰-۰/۰۶۰ سانتی‌متر مکعب بر سانتی‌متر مکعب) است. هم‌چنین مقادیر RMSE برای مدل‌هایی که از پارامترهای فیزیکی و هیدرولیکی خاک به‌عنوان ورودی استفاده می‌کند، کمتر (۰/۰۴۰-۰/۰۴۰ سانتی‌متر مکعب بر سانتی‌متر مکعب) است. با این حال، آن‌ها برای تخمین رطوبت خاک نزدیک به سطح و منطقه ریشه بر اساس مدل‌هایی که در آن از اطلاعات فیزیکی خاک به‌عنوان ورودی مدل استفاده شد، عملکرد بهتری به دست آوردند. آن‌ها افزایش همبستگی را به افزودن متوالی اطلاعات ورودی فیزیکی و هیدرولیکی خاک از جمله پارامترهای کلیدی فراوانی نسبی ذرات خاک، چگالی ظاهری و مقادیر رطوبت‌های ظرفیت زراعی و

نقطه پژمردگی دانستند.

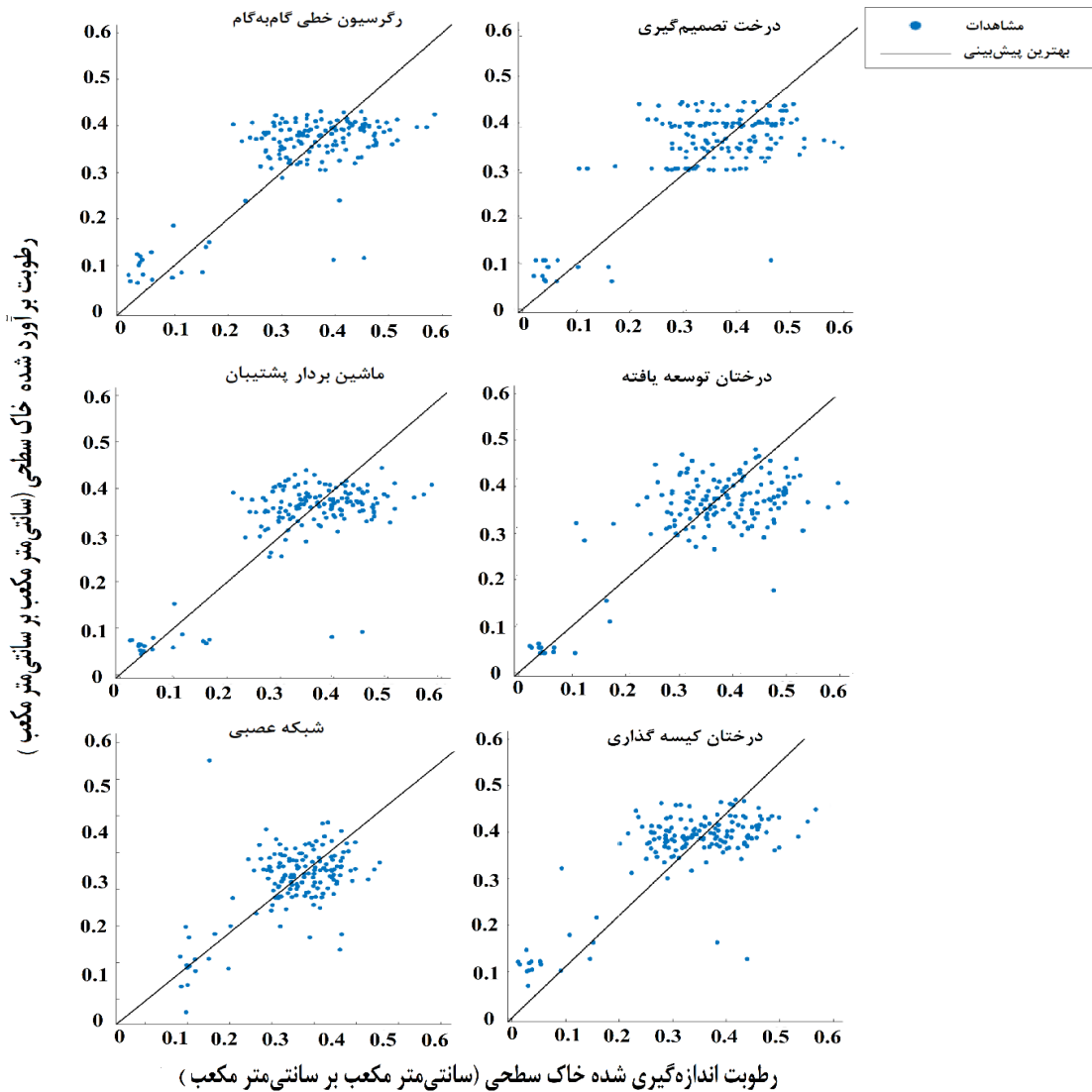
همچنین Araya et al., (2021) دریافتند باندهای تشکیل دهنده پوشش گیاهی NDVI (باندهای قرمز و مادون قرمز نزدیک) در پیش بینی رطوبت خاک مهم و مؤثر هستند و در نتیجه بایستی مدل‌هایی که NDVI دارند از اهمیت بالایی برخوردار باشند؛ اما دریافتند که دو شاخص پوشش گیاهی مورد آزمایش آن‌ها، یعنی NDVI و TTVI، اهمیت چندانی در پیش‌بینی نداشتند بلکه در پیش‌بینی باند قرمز و مادون قرمز نزدیک به‌تنهایی از اهمیت داشتند. همچنین باند مادون قرمز نزدیک از اهمیت کمتری نسبت باند قرمز در پیش‌بینی رطوبت سطحی خاک برخوردار بود، با توجه به حساسیت بیشتر باند مادون قرمز نزدیک به تنش رطوبتی گیاه، این واقعیت استنباط شد که منطقه مورد مطالعه آن‌ها به‌طور کامل دارای پوشش گیاهی است. شایان ذکر است هنگام در نظر گرفتن اهمیت متغیرها، ممکن است متغیرهای درجات مختلفی از اهمیت داشته باشند که بستگی به مرطوب یا خشک بودن خاک دارد. Western et al (1999) دریافتند که جریان تجمعی بهترین پیش‌بینی کننده برای توزیع رطوبت خاک در شرایط مرطوب است، درحالی‌که شاخص تابش خورشیدی قابلیت پیش‌بینی بهتری در شرایط خشک دارد.

جدول ۶ مقایسه مقادیر MAE به دست آمده برای هفت مدل پیشنهادی و الگوریتم‌های مختلف یادگیری ماشین

مدل‌ها	الگوریتم‌ها	رگرسیون خطی گام به گام	ماشین بردار پشتیبان	درختان توسعه یافته	درخت تصمیم‌گیری	درختان کیسه گذاری	شبکه عصبی
مدل ۱		۰/۰۳۵	۰/۰۳۶	۰/۰۳۸	۰/۰۴۰	۰/۰۳۸	۰/۰۳۷
مدل ۲		۰/۰۳۵	۰/۰۳۷	۰/۰۳۹	۰/۰۴۱	۰/۰۴۱	۰/۰۴۱
مدل ۳		۰/۰۳۵	۰/۰۳۷	۰/۰۳۸	۰/۰۴۲	۰/۰۴۵	۰/۰۴۶
مدل ۴		۰/۰۳۴	۰/۰۴۰	۰/۰۳۸	۰/۰۴۰	۰/۰۴۳	۰/۰۵۲
مدل ۵		۰/۰۳۷	۰/۰۳۸	۰/۰۳۹	۰/۰۴۰	۰/۰۴۱	۰/۰۴۶
مدل ۶		۰/۰۶۲	۰/۰۶۳	۰/۰۶۵	۰/۰۶۷	۰/۰۶۴	۰/۰۸
مدل ۷		۰/۰۳۶	۰/۰۳۸	۰/۰۳۹	۰/۰۴۱	۰/۰۴۵	۰/۰۴۸



شکل ۴: همبستگی بین مقادیر اندازه‌گیری و برآورد شده رطوبت خاک حاصل از مدل ۴ (بهترین برآورد)



شکل ۳: همبستگی بین مقادیر اندازه‌گیری و برآورد شده رطوبت خاک حاصل از مدل ۶ (ضعیف‌ترین برآورد)

به‌طور کلی، مقادیر خطا (RMSE) برای مدل ۶ که از STR به‌عنوان متغیر ورودی برای تخمین رطوبت خاک استفاده نمی‌کند بالاترین مقدار به ترتیب برای الگوریتم‌های شبکه عصبی، درخت تصمیم‌گیری، درختان توسعه‌یافته، ماشین بردار پشتیبان، درختان کیسه‌گذاری و رگرسیون خطی گام‌به‌گام (۰/۰۸۱، ۰/۰۸۲، ۰/۱۱، ۰/۰۸۷، ۰/۰۸۱ و ۰/۰۸۵ سانتی‌متر مکعب بر سانتی‌متر مکعب) داشت؛ بنابراین پارامتر STR به‌عنوان مهم‌ترین ویژگی در برآورد رطوبت خاک در این مطالعه شناسایی شد.

بین رطوبت خاک و میزان آب برگ‌های پوشش گیاهی، رابطه خاصی وجود دارد. رطوبت خاک بالا و پایین، محتوای آب برگ‌ها را تحت تأثیر قرار داده و در نهایت منجر به تغییر در مشخصات طیفی می‌شود (Fernández-Novales *et al.*, 2018)؛ بنابراین اهمیت شاخص طیفی پوشش گیاهی به دو دلیل آشکار می‌شود، حساسیت به پارامترهای هدف و عدم حساسیت به عوامل مزاحم و در نتیجه با وجود این شاخص دقت تخمین برای اهداف خاص بهبود می‌یابد زیرا اثرات عوامل مزاحم کاهش می‌یابد (Liang *et al.*, 2015). مطالعات قبلی نشان داده‌اند که انعکاس SWIR نسبت به انعکاس NIR به تغییرات در محتوای آب و خاک حساس‌تر است؛ اما در مدل ۶ که فقط از شاخص تفاضلی پوشش گیاهی (NDVI) به‌عنوان ورودی سنجنش‌ازدور استفاده شده برآوردها کمترین مقدار ممکن را نشان می‌دهند (Sadeghi *et al.*, 2017; Babaeian *et al.*, 2018). این موضوع نشان می‌دهد که پارامترهای ورودی فیزیکی و هیدرولیکی تعریف‌شده به مدل‌ها برای تخمین در شرایط مرطوب مهم‌تر هستند، در حالی که ویژگی‌های پوشش گیاهی (به‌عنوان مثال، NDVI) در شرایط خشک مهم‌تر هستند (Tuller and Or, 2005).

مطالعات قبلی همبستگی خوبی بین رطوبت خاک و مقدار آب گیاه را نشان داده‌اند (Rutter and Sands, 1958). عملکرد نسبتاً

ضعیف برخی مدل‌ها را به حذف متغیر ورودی بیوفیزیکی NDVI نسبت داده که به‌طور غیرمستقیم بر وضعیت رطوبت خاک بستگی دارد. از نظر فیزیکی، این بدان معناست که دسترسی گیاه به آب در سبزی گیاه منعکس می‌شود ( Babaeian et al., 2017; Sadeghi et al., 2018). کمبود آب خاک بر پتانسیل آب برگ تأثیر می‌گذارد که به‌نوبه خود بر محتوای نسبی آب برگ‌های گیاه تأثیر می‌گذارد NDVI. نشان‌دهنده اثر تراکم پوشش گیاهی بر پویایی رطوبت خاک مکانی در طول رشد گیاه است.

## نتیجه‌گیری

این پژوهش به سؤالات زیر پاسخ داد: چگونه می‌توان از روش‌های یادگیری ماشینی برای برآورد رطوبت خاک از ترکیب داده‌های سنجش‌ازدور و ویژگی‌های فیزیکی و هیدرولیکی خاک استفاده کرد؟ مهم‌ترین عوامل پیش‌بینی کننده رطوبت خاک در منطقه مورد مطالعه کدام است؟ در این مطالعه با استفاده از تصاویر ماهواره‌ای در محدوده باندهای قرمز، مادون قرمز نزدیک و مادون قرمز طول موج کوتاه، شاخص پوشش گیاهی NDVI و انعکاس تبدیل یافته مادون قرمز طول موج کوتاه (STR) استخراج گردید. هفت مدل از ترکیب متغیرهای سنجش‌ازدور و خصوصیات فیزیکی و هیدرولیکی خاک توسعه داده شد که با الگوریتم‌های خطی و غیرخطی یادگیری ماشینی ارزیابی گردید تا اهمیت متغیرهای مختلف درک و ماهیت رابطه بین متغیرها و رطوبت خاک روشن شود. نتایج نشان داد در پیش‌بینی رطوبت خاک بر اساس مقادیر  $R^2$ ، RMSE و MAE، رگرسیون خطی گام‌به‌گام، ماشین بردار پشتیبان، درختان توسعه‌یافته از سایر مدل‌ها عملکرد بهتری داشتند. بالاترین مقدار  $R^2$  (0/87) مربوط به رگرسیون خطی گام‌به‌گام در مدل ۴ و کمترین مقدار  $R^2$  (0/11) مربوط به شبکه عصبی در مدل ۶ بود. همچنین مدل ۶ نشان داد پارامتر STR به‌عنوان مهم‌ترین ویژگی در پیش‌بینی رطوبت خاک در مزارع مورد مطالعه است. بر اساس نتایج بدست آمده، مدل ۷ با بیشترین تعداد ورودی‌ها، کامل‌ترین مدل در برآورد رطوبت خاک بود و نقش ویژگی‌های فیزیکی و هیدرولیکی خاک را در توصیف پویایی زمانی و مکانی رطوبت خاک در مقیاس مزرعه به‌خوبی نشان داد. علاوه بر این، متغیرهای NDVI و STR همراه با اطلاعات فیزیکی و هیدرولیکی خاک به‌عنوان ورودی مدل‌ها، در اکثر روش‌های یادگیری ماشینی به‌خوبی مقدار رطوبت خاک را برآورد کردند.

از آنجایی که خصوصیات فیزیکی و هیدرولیکی استفاده‌شده در این مطالعه همچون؛ بافت خاک، تخلخل، وزن مخصوص ظاهری، رطوبت‌های ظرفیت زراعی و نقطه پژمردگی تقریباً در طول فصل رشد ثابت و بدون تغییر هستند، با در دست داشتن این اطلاعات نیاز به اندازه‌گیری مستقیم آن‌ها نیست؛ بنابراین تنها با داشتن پارامترهای سنجش‌ازدور (STR و NDVI) در طول زمان و مکان می‌توان مقادیر آن‌ها را جایگذاری و برای مطالعات آینده معادلات به‌روز تولید کرد. واضح است که روش ارائه‌شده پویایی مکانی و زمانی رطوبت را با دقتی مناسب برای مدیریت دقیق آبیاری در مقیاس مزرعه به تصویر می‌کشد. در نتیجه راه‌های جدیدی را برای بهبود بهره‌وری مصرف آب در بخش کشاورزی به‌منظور حفظ منابع آب در مناطق با محدودیت آب باز می‌کند. پیشنهاد می‌شود در مطالعات آینده از تصاویر سامانه‌های هوایی بدون سرنشین (UAS) برای بالا بردن قدرت تفکیک مکانی در مقیاس مزرعه استفاده شود. سپس نقشه‌های رطوبت خاک با استفاده از مدل‌های توسعه‌یافته در مراحل مختلف رشد گیاه تهیه شود.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

## REFERENCES

- Acharya, U., Daigh, A. and Oduor, P. (2021). Machine Learning for Predicting Field Soil Moisture Using Soil, Crop, and Nearby Weather Station Data in the Red River Valley of the North. *Journal of soilsystems*, 5, 57.
- Achieng, K.O. (2019). Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs. support vector regression models. *Journal of Computers and Geosciences*. 2019, 133, 104320.
- Adab, M., Morbidelli, R., Saltalippi, C., Moradian, M. and Fallah Ghalhari, G-A. (2020). Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Journal of Water*, 12, 3223.
- Ahmad, S. and Simonovic, S.P. (2005). An artificial neural network model for generating hydrograph from hydro-meteorological parameters. *Journal of Hydrology*, 315, 236–251.
- Araya1, S. N., Fryjoff-HungA., Anderson, A., Viers, J. H. and Ghezzehei, Teamrat A. (2021). Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. *Journal of Hydrology and Earth System Sciences*. 25, 2739–2758,
- Babaeian, E., Sadeghi, M., Franz, T.E., Jones, S., and Tuller, M., (2018). Mapping soil moisture with the



- Optical TRapezoid model (OPTRAM) based on long-term MODIS observations. *Journal of Remote Sensing of Environment*, 211, 425–440.
- Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. and Tuller, M. (2021). Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Journal of Remote Sensing of Environment*, 260, 1-13
- Breiman, L. (2001) Random Forest. *Mach. Learn.*, 45, 5–32.
- Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.*, 29, 1189–1232.
- Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L. and Xue, X. (2019). Research on soil moisture prediction model based on deep learning. *Journal of PLoS ONE*, 0214508.
- Caruana, R., and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms, in: Proceedings of the 23rd international conference on Machine learning –. *Journal of ICML '06, ACM Press, New York, USA*, 161–168.
- Chakrabarti, S., Judge, J., Bongiovanni, T., Rangarajan, A. and Ranka, S. (2018). Spatial Scaling Using Temporal Correlations and Ensemble Learning to Obtain High-Resolution Soil Moisture. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, VOL. 56, NO. 3.
- Elith, J., Leathwick, J.R. and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813.
- Engman, E.T. (2000). In Remote Sensing in Hydrology and Water Management; Schultz, G.A. Engman, E.T. Eds. Springer: Berlin/Heidelberg, Germany, pp. 197–216.
- Fernández-Navales, J., Tardaguila, J., Gutiérrez, S., Marañón, M. and Diago, MP. (2018). In field quantification and discrimination of different vineyard water regimes by on-the-go NIR spectroscopy. *Biosystems Engineering* 165:47–58.
- Friedman, J.H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 367–378.
- Flint, A.L. Flint, L.E. Available water. In: Dane, J.H. Topp, G.C. (Eds.), *Methods of Soil Analysis, Part 4, Physical Methods. Soil Science Society of America, Madison*, pp. 229–233.
- Ge, X., Wang, J., Ding, J., Cao, X., Zhang, Z., Liu, J. and Li, X. (2019). Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *Journal of peerj*, 7e:6926.
- Gorthi, S. (2011). *Prediction Models for Estimation of Soil Moisture Content*. All Graduate Theses and Dissertations. Utah State University. Logan, Utah.
- Grimes, D.I.F., Coppola, E., Verdecchia, M. and Visconti, G. (2003). A neural network approach to real-time rainfall estimation for Africa using satellite data. *Journal of Hydrometeorol*, 4, 1119–1133.
- Hamzeh, S., Naseri, A. A., Alavipanah, S. K., Mojaradi, B., Bartholomeus, H. M. and Clevers, J. G. (2013). Estimating salinity stress in sugarcane farms with spaceborne hyperspectral vegetation indices. *International Journal of Applied Earth Observation and Geoinformation*, 21, 282–290.
- Hosseini Chamani, F., Farrokhian Firouzi, A. and Amerikhah, H. (2019). Pedotransfer Function (PTF) for Estimation Soil moisture using NDVI, land surface temperature (LST) and normalized moisture (NDMI) indices. *Journal of Water and Soil Conservation*, 26 (4), 239-254.
- Irons, J.R., Campbell, G.S., Norman, J.M., Graham, D.W. and Kovalick, W.M. (1992). Prediction and measurement of soil bidirectional reflectance. *IEEE Transactions On Geoscience and Remote Sensing*, 30, 249–260.
- James, G., Witten, D. and Hastie, T. (2013). Tibshirani, R. An Introduction to Statistical Learning; Springer: New York, NY, USA.
- Joshi, C. and Mohanty, B.P. (2010). Physical controls of near-surface soil moisture across varying spatial scales in an agricultural landscape during SMEX02. *Journal of Water Resources Research*, 46, 1-21.
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J. and Goulden, M.L. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Journal of Nature*, 467, 951–954.
- Kaleita, A.L. Tian, L.F. Hirschi, M.C. (2005). Relationship between soil moisture content and soil surface reflectance. *Trans. ASAE*, 48, 1979-1986.
- Kalra, A. and Ahmad, S. (2009). Using oceanic-atmospheric oscillations for long lead time streamflow forecasting. *Journal of Water Resources Research*, 45, W03413.
- Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S. and Lin, H. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Journal of Remote Sensing of Environment* 165:123–134.
- Li, T., Hao, X.M. and Kang, S.Z. (2014). Spatiotemporal variability of soil moisture as affected by soil



- properties during irrigation cycles. *Soil Science Society of America, Madison*. 78, 598–608.
- Manns, H.R., Berg, A.A., Bullock, P.R. and McNairn, H. (2014) Impact of soil surface characteristics on soil water content variability in agricultural fields. *Journal of Hydrology. Process*. 28, 4340–4351.
- McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S. and Entekhabi, D. (2017). The global distribution and dynamics of surface soil moisture. *Journal of Nature Geoscience*, 10, 100–104.
- Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Front. Neurorobot*, 7, 21.
- New L. (1971). Influence of alternate furrow irrigation and time of application on grain sorghum production. *Texas Agricultural State Program Report No 2953*.
- Oymak, S. and Soltanolkotabi, M., 2020. Toward moderate Overparameterization: global convergence guarantees for training shallow neural networks. *IEEE Journal of Select. Areas Inform. Theor.*, 1 (1), 84–105.
- Prakash., S., Sharma., A., & Sahu., S. S., (2019). Soil Moisture Prediction Using Machine Learning. *Published in: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*.
- Rutter, A.J. and Sands, K. (1958). The relation of leaf water deficit to soil moisture tension in *Pinus Sylvestris* L. *Journal of New Phytologist*, 57 (1), 50–65.
- Sadeghi, M., Jones, B.S. and Philpot, W.D. (2015). A linear physically-based model for remote sensing of soil moisture using shortwave infrared bands. *Journal of Remote Sensing. Environmental*, 164, 66–76.
- Sadeghi, M., Babaeian, E., Tuller, M. and Jones, S. (2017). The optical Trapezoid model: a novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat- 8 observations. *Journal of Remote Sensing. Environmental*, 198, 52–68.
- Sadeghi, M., Babaeian, E., Arthur, E., Jones, S.B. and Tuller, M. (2018). Soil physical properties and processes. In: Kutz, M. (Ed.), *Handbook of Environmental Engineering*. Wiley, Hoboken, NJ, pp. 137–207. ISBN-10: 1118712943.
- Schaap, M.G., Leij, F.J. and van Genuchten, M.Th. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251, 163–176.
- Scott, C. A., Bastiaanssen, W.G.M. and Ahmad, M.D. (2003). Mapping Root Zone Soil Moisture Using Remotely Sensed Optical Imagery. *Journal of Irrigation and Drainage Engineering*, 129(5), 326-335.
- Sharma, S.K., Mohanty, B.P. and Zhu, J. (2006). Including topography and vegetation attributes for developing pedotransfer functions. *Soil Science Society of America Journal*, 70, 1430–1440.
- Taghizadeh-Mehrjardi, R. Nabiollahi, K. Minasny, B. and Triantafilis, J. (2015). Comparing data mining classifiers to predict spatial distribution of USDAfamily soil groups in Baneh region, Iran. *Geoderma*, 253, 67-77.
- Tuller, M., and Or, D. (2005). Water retention and characteristic curve. *Journal of Hydrology Environmental Science*, 4 (1), 278–289.
- Twarakavi, N.K., Misra, D. and Bandopadhyay, S. (2006). Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data. *Journal of Hydrology Natural Resources Research*, 15, 15–26.
- Veysi, Sh., Naseri, A.A., Hamzeh, S. and Bartholomeus, H. (2017). A satellite based crop water stress index for irrigation scheduling insugarcane fields. *Journal of Hydrology Agricultural Water Management*, 189 70–86.
- Western, A. W., Grayson, R. B., Blöschl, G., Willgoose, G. R., and McMahon, T. A. (1999). Observed spatial organization of soil moisture and its relation to terrain indices. *Journal of Water Resources Research*, 35, 797–810.
- Whittaker, G., Confesor, R., Di Luzio, M., and Arnold, J.G. (2010). Detection of Overparameterization and Overfitting in an automatic calibration of SWAT. *Journal of Trans. ASABE*, 53 (5), 1487–1499.
- Yang, H., Huang, K., King, I. and Lyu, M.R. (2009). Localized support vector regression for time series prediction. *Neurocomputing*, 72, 2659–2669.
- Yeh, I.C. and Lien, C.H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*. Volume 36, Issue 2, Part 1, 36, 2473–2480.
- Zareie, A., Amin, M.S.R. and Amador-Jiménez, L.E. (2016). Thornthwaite moisture index modeling to estimate the implication of climate change on pavement deterioration. *Journal of Transportation Engineering*, 142 (4), 04016007.