

مقاله علمی-پژوهشی

## مدل‌سازی ردپای آب با استفاده از رگرسیون چندکی مبتنی بر یادگیری ماشین و تحلیل عدم قطعیت

مهرناز یحیی زاده<sup>۱</sup>، ام‌البین بذرافشان<sup>۲\*</sup>، آرش ملکیان<sup>۳</sup>، حسین زمانی<sup>۴</sup>

تاریخ دریافت: ۱۴۰۴/۰۴/۲۸ تاریخ پذیرش: ۱۴۰۴/۰۶/۲۸

### چکیده

این پژوهش با هدف مدل‌سازی ردپای آب سبز و آبی گوجه‌فرنگی در استان بوشهر با استفاده از مدل‌های ترکیبی یادگیری ماشین QXGBoost، QRF، و XGBoost و RF و تحلیل عدم قطعیت آن‌ها انجام شد. نتایج نشان داد که ردپای آبی (۸۰٪) سهم غالب در مصرف آب دارد، درحالی‌که ردپای سبز تنها ۲۰٪ را تشکیل می‌دهد که نشان‌دهنده وابستگی شدید به منابع آبی و آسیب‌پذیری سیستم در برابر تغییرات اقلیمی است. تحلیل همبستگی متغیرها مشخص کرد که بارش (P) و بارش مؤثر (Peff) بیشترین تأثیر مثبت را بر ردپای سبز دارند، درحالی‌که دما و تبخیر و تعرق (ETc) اثر کاهنده نشان دادند. برای ردپای آبی، ETc و Tmax بیشترین همبستگی مثبت را داشتند. مدل QXGBoost با ضریب تعیین (Rsq=۰/۹۵) و خطای کم (RMSE=۰/۱۶) به‌عنوان مدل بهینه انتخاب شد. همچنین، ارزیابی عدم قطعیت با روش UNEEC نشان داد که QXGBoost با پوشش پیش‌بینی فاصله‌ای (PICP=۰/۹۳) و فاصله اطمینان متقارن، بهترین عملکرد را دارد. این مطالعه راهکارهای کاهش ردپای آب شامل بهبود آبیاری (زیرسطحی و هوشمند)، کشت ارقام مقاوم به خشکی و مدیریت رطوبت خاک (مالچ‌پاشی) را پیشنهاد می‌کند. یافته‌ها ضرورت تغییر الگوی کشت فعلی به سمت سیستم‌های پایدارتر با تکیه بر مدل‌سازی پیشرفته و پایش مستمر منابع آب را تأکید می‌کنند.

**واژه‌های کلیدی:** ردپای آب، رگرسیون چندکی، گوجه‌فرنگی، مدیریت آب کشاورزی، یادگیری ماشین

### مقدمه

مصرف جهانی آب در قرن گذشته تا شش برابر افزایش یافته است (Mahato et al., 2022). در چهار دهه اخیر مصرف جهانی آب ۴۰ درصد افزایش یافته است، که عمدتاً ناشی از نیازهای آبیاری است که ۷۰ درصد از کل برداشت آب جهانی را تشکیل می‌دهد (Harshdeep et al., 2020). طبق برآوردهای اخیر، مصرف جهانی آب تقریباً به (۴۶۰۰ کیلومتر مکعب) در سال رسیده است (Hindiye et al., 2021; Cheng et al., 2023; Hoosain et al., 2023). این نشان‌دهنده مبنای فعلی، برای درک تقاضای آب شیرین در تمام بخش‌ها است. کمبود آب یک چالش جهانی مهم به ویژه در اقلیم‌های خشک و نیمه خشک است. کمبود در این مناطق ناشی از عوامل طبیعی مانند بارندگی کم، تبخیر زیاد و دماهای بالا است که با رشد جمعیت و فعالیت‌های انسانی مانند برداشت بیش از حد آب‌های زیرزمینی و شیوه‌های آبیاری ناکارآمد بیشتر می‌شود. ایران که با آب و هوای خشک و نیمه خشک مشخص می‌شود، به‌عنوان یکی از کشورهای که در معرض فشار آبی قرار دارند، با منابع محدود آب شیرین و افزایش تقاضای آب دست و پنجه نرم می‌کند (Hofste et al., 2019). کشاورزی، عامل اصلی و قربانی کم‌آبی، مسئول بیش از ۷۰ درصد برداشت‌های آب در ایران و تا ۹۵ درصد در برخی از

آب یک عامل ضروری برای تولید محصولات کشاورزی و بقای انسان است و نقش مهمی در امنیت غذایی دارد. چالش‌های زیست‌محیطی در اکثر کشورهای جهان در نتیجه بهره‌برداری نادرست از منابع زیست‌محیطی توجه بیشتری را به خود جلب کرده‌اند (Fan et al., 2018). استفاده بی‌رویه از منابع آب همراه با تغییرات اقلیمی، بر پتانسیل‌های بازیابی اکوسیستم تأثیر منفی گذاشته است (Chowdhary et al., 2018; Duckett et al., 2016). از سوی دیگر، شتاب رشد اقتصاد جهانی منجر به مشکلات زیست‌محیطی شده است. ارزیابی‌های آماری نشان می‌دهد که

۱- دانشجوی دکتری علوم و مهندسی آبخیزداری، گروه مهندسی منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران

۲- عضو گروه مهندسی منابع طبیعی، دانشگاه هرمزگان، بندرعباس، ایران

۳- عضو گروه احیای مناطق خشک و کوهستانی، دانشگاه تهران، کرج، ایران

۴- عضو گروه آمار، دانشکده علوم پایه، دانشگاه هرمزگان، بندرعباس، ایران

\*- نویسنده مسئول: [Q.bazrafshan@hormozgan.ac.ir](mailto:Q.bazrafshan@hormozgan.ac.ir)

DOI: [10.22034/ijid.2025.535509.2613](https://doi.org/10.22034/ijid.2025.535509.2613)

ضریب همبستگی  $0.72$  و میانگین خطای مطلق کمتر از  $0.48$  (متر مکعب بر تن) است.

(Elbeltagi et al., 2020a) قابلیت‌های دقیق رویکردهای یادگیری ماشینی را از طریق کاربرد شبکه عصبی مصنوعی (ANN) در تخمین رد پای آب ذرت مورد بررسی قرار دادند مطالعه آنها نشان داد که چگونه مدل‌های یادگیری ماشینی می‌توانند در مناطق جغرافیایی مختلف، با حداقل انحراف بین مقادیر واقعی و پیش‌بینی شده، به دقت بالایی پیش‌بینی بالایی کردند که قابلیت اطمینان از مدل‌های یادگیری ماشینی را برای کاربردهای در مقیاس منطقه‌ای برجسته می‌کند. پژوهش دیگر، (Elbeltagi et al., 2020b) بر مدل‌سازی دینامیک تخییر و تعرق محصول (ETc) بلندمدت با استفاده از شبکه‌های عصبی عمیق (DNN) در یک محیط نیمه‌خشک نشان داد که چگونه روش‌های یادگیری ماشینی می‌توانند به طور مؤثر از داده‌های اقلیمی محدود برای تولید پیش‌بینی‌های قوی، به‌ویژه برای برنامه‌ریزی بلندمدت مدیریت آب کشاورزی، استفاده کنند (Tao et al., 2023). با مدل‌سازی موفقیت‌آمیز هر دو مؤلفه رد پای آب آبی و سبز، انعطاف‌پذیری ابزارهای یادگیری ماشینی را بیشتر نشان دادند.

یادگیری ماشین<sup>۲</sup> (ML) به‌عنوان یک جایگزین امیدوارکننده ظهور کرده است و رویکردهای مدل‌سازی متعددی را ارائه می‌دهد که مزایای متمایزی را برای تجزیه و تحلیل کارآمدتر و عملی‌تر جریان آب فراهم می‌کنند. این روش‌شناسی، بینش‌های ارزشمندی را برای محققان و سیاست‌گذارانی که در جهت مدیریت پایدار منابع آب تلاش می‌کنند، فراهم می‌کند. رگرسیون چندکی یک روش ارزشمند برای تخمین توابع شرطی چندکی است که توسط (Koenker, 2005) معرفی شد. درحالی‌که رگرسیون خطی رابطه بین میانگین شرطی متغیر پاسخ و یک یا چند متغیر مستقل را توصیف می‌کند، ممکن است در برخی شرایط در تحلیل داده‌ها عملکرد ضعیفی داشته باشد. رگرسیون چندکی به جای محاسبه میانگین شرطی، از میانه یا چندک‌های شرطی شرطی متغیر پاسخ استفاده می‌کند و نیاز به نرمال بودن داده‌ها ندارد، ضمن این‌که حتی در برابر داده‌های پرت نیز مقاومت خوبی از خود نشان می‌دهد. درواقع، رگرسیون چندکی شکل گسترش یافته‌ای از رگرسیون خطی است که در پردازش پس‌از مدل‌سازی فرایندهای هیدرومتئورولوژیکی و الگوریتم‌های یادگیری ماشینی مورد توجه قرار گرفته است (Papacharalampous et al., 2019)، زیرا می‌تواند به صورت طبیعی و مؤثر در الگوریتم‌های یادگیری ماشینی ادغام شود. از جمله مطالعات مرتبط می‌توان به پژوهش‌های ژو و همکاران (Xu et al., 2015)، عبدالله و همکاران (Patidar et al., 2022)، پاتیدار و همکاران

(Hoekstra & Mekonnen, 2012). آب یک عامل مهم برای توسعه پایدار می‌باشد و یکی از استراتژی‌های کشورهای کم‌آب، کاهش تولید محصولات کشاورزی می‌باشد، که با واردات کالاها آب بر سبب کاهش فشار بر منابع آب داخلی منطقه می‌شود (Lenzen et al., 2013).

مفهوم ردپای آب<sup>۱</sup> (WF) که توسط هوکسترا (Hoekstra, 2003) معرفی شد، به عنوان ابزاری ارزشمند برای تعیین کمیت آب مصرفی در تولید کالاها و خدمات، به طور گسترده شناخته شده است. در بخش کشاورزی، ارزیابی ردپای آب به طور گسترده برای درک توزیع مکانی-زمانی تخصیص آب برای کشت محصولات کشاورزی به کار گرفته شده است (Lovarelli et al., 2016; Luan et al., 2018; Mekonnen & Gerbens-Leenes 2020). در چرخه هیدرولوژی، منابع آب به دو دسته آب آبی و آب سبز تقسیم‌بندی می‌شوند. آب‌های زیرزمینی و آب‌های سطحی، آب آبی را تشکیل می‌دهند، در حالی‌که به رطوبت خاک در مناطق غیر اشباع آب سبز می‌گویند. منشاء آب آبی و آب سبز، بارندگی است. آب باران بعد از نفوذ در خاک و قبل از این‌که به منطقه اشباع برسد، آب سبز را تشکیل می‌دهد، در حالی‌که آب‌های زیرزمینی (منطقه اشباع) و همچنین رواناب حاصل از بارندگی که به رودخانه‌ها و دریاچه‌ها، پشت سد‌ها و تالاب‌ها می‌پیوندد، آب آبی را شامل می‌شوند (Hoekstra, 2002 and Hung). از این نظر، ردپای آب (WF) (Hoekstra, 2008) می‌تواند یکی از ابزارهای مقابله با چالش‌های امنیت آب به ویژه در بخش کشاورزی باشد (Elbeltagi et al., 2020a). در سال‌های اخیر استفاده از هوش مصنوعی در قالب الگوریتم یادگیری ماشینی به‌طور موفقیت‌آمیزی برای پیش‌بینی خطر و ریسک در علوم طبیعی به کار گرفته می‌شود که به دلیل اجرای آسان، امکان آموزش سریع، اعتبارسنجی و رسیدن با دقت بالا از ویژگی‌های یادگیری ماشینی محسوب می‌شود. همچنین به کمک این روش‌ها می‌توان سیستم‌های پیچیده و غیرخطی را با دقت بالا مدل‌سازی نمود (Choubin & Malekian, 2017). در سال‌های اخیر پژوهشگران مختلفی ردپای آب را با استفاده از مدل‌های یادگیری ماشینی مورد بررسی قرار داده‌اند.

بادروچ و بذرافشان (۱۴۰۱) با هدف تخمین و پیش‌بینی ردپای آب سبز و آبی محصول زراعی گندم استان فارس با استفاده از مدل‌های یادگیری ماشینی شامل مدل شبکه عصبی، دو کرنل لوجستیک و تانژانت هائیربولیک، مدل جنگل تصادفی، رگرسیون بردار پشتیبان (تابع کرنل سیگموئید) طی سال‌های (۱۳۸۴-۱۳۹۶) مورد بررسی قرار دادند. نتایج این پژوهش نشان داد بهترین مدل برای برآورد ردپای آب گندم مدل شبکه عصبی با تابع لوگ لجستیک با

۱۳۸۳). برداشت بی‌رویه از منابع آب زیرزمینی نه‌تنها موجب تشدید شوری آب شده، بلکه کاهش چشمگیر عملکرد محصولات کشاورزی را در پی داشته است. ارزیابی دقیق ردپای آب در تولید گوجه‌فرنگی بوشهر به عنوان راهکاری اساسی برای مدیریت پایدار منابع آبی این منطقه ضروری به نظر می‌رسد. با توجه به شرایط خاص اقلیمی و اهمیت کشت خارج از فصل گوجه‌فرنگی در تأمین امنیت غذایی، تدوین استراتژی‌های کارآمد برای بهینه‌سازی مصرف آب اجتناب‌ناپذیر است.

این تحقیق با به‌کارگیری مدل‌های ترکیبی یادگیری ماشین مانند XGBoost و Random Forest، و مدل رگرسیون چندکی آنها (QRF و QXGBoost) به مدل‌سازی ردپای آب سبز و آبی گوجه‌فرنگی در شرایط کشت خارج از فصل در استان بوشهر می‌پردازد و از روش UNEEC برای تخمین عدم قطعیت و کمی‌سازی خطاهای پیش‌بینی استفاده می‌کند. کمتر تحقیقی به کمی‌سازی عدم قطعیت این مدل‌ها پرداخته است. همچنین، با ادغام داده‌های اقلیمی، مدیریتی و کشاورزی، مدل‌سازی پویا و واقع‌بینانه‌ای ارائه می‌دهد و در نهایت، راهکارهای مدیریتی مبتنی بر سناریوهای کمی برای کاهش فشار بر منابع آب زیرزمینی پیشنهاد می‌کند. این مطالعه با ترکیب یادگیری ماشین و روش‌های کمی‌سازی عدم قطعیت، چارچوبی علمی برای مدیریت هوشمند آب در کشت گوجه‌فرنگی ایجاد می‌کند که می‌تواند به سیاست‌گذاران و کشاورزان در تخصیص بهینه منابع آب و جلوگیری از بحران‌های آبی کمک کند.

## مواد و روش

### مواد و منطقه مورد مطالعه

استان بوشهر با اقلیم گرم و خشک، به دلیل تابش آفتاب فراوان (۳۰۰ روز در سال)، دمای مطلوب زمستانه (۱۵-۲۵°C) و خاک مناسب، شرایط ایده‌آلی برای کشت خارج از فصل گوجه‌فرنگی دارد (Bazrafshan et al., 2022). این منطقه با کاهش تنش‌های دمایی (عدم یخ‌زدگی) و کنترل آفات در فصل خنک، امکان تولید محصول در زمانی که عرضه کم است (پاییز و زمستان) را فراهم می‌کند، که ارزش اقتصادی و صادراتی بالایی دارد. به‌طور کلی، بوشهر پتانسیل بالایی برای کشت پرسود گوجه‌فرنگی خارج از فصل دارد. در این پژوهش اطلاعات مربوط به سطح زیر کشت، عملکرد در واحد سطح، میزان تولید، میزان کود، راندمان آبیاری، تاریخ کشت و تیپ خاک از سال‌نامه سازمان جهاد کشاورزی استخراج گردید. همچنین داده‌های اقلیمی از سازمان هواشناسی هر استان (IRIMO, 2022) در منطقه مورد نظر تهیه گردید که داده‌ها شامل سری زمانی بارش، درصد رطوبت نسبی، ساعات آفتابی، حداقل و حداکثر درجه حرارت هوا و سرعت باد طی دوره آماری مورد مطالعه می‌باشد.

2023) و کسرابی و همکاران (Kasraei et al., 2021) اشاره کرد. انتخاب مدل‌های یادگیری ماشین در این مطالعه بر اساس قابلیت‌های اثبات‌شده آن‌ها در مدیریت پیچیدگی‌های پیش‌بینی خشکسالی، به‌ویژه در تنظیمات دو متغیره که شامل روابط غیرخطی و داده‌های با ابعاد بالا هستند، صورت گرفته است. به‌طور خاص، مدل‌های جنگل تصادفی (RF) و جنگل تصادفی چندکی (QRF) به دلیل توانایی درک وابستگی‌های غیرخطی و قابلیت تفسیرپذیری انتخاب شده‌اند و گزینه‌های قابل اعتمادی برای مجموعه‌داده‌های محیطی با ویژگی‌های نویز و مقادیر گم‌شده محسوب می‌شوند (Breiman, 2001, Meinshausen and Ridgeway, 2006). مدل تقویت‌گرایان شدید<sup>۱</sup> (XGB) و تقویات‌گرایان شدید چندکی<sup>۲</sup> (QXGB) به دلیل عملکرد برتر در وظایف پیش‌بینی که نیاز به دقت بالا دارند و همچنین اثربخشی در پردازش مجموعه‌داده‌های بزرگ با تعاملات پیچیده، در این مطالعه گنجانده شده‌اند. قابل ذکر است که XGB به دلیل سرعت، مقیاس‌پذیری و توانایی کاهش بیش‌برازش از طریق تکنیک‌های تنظیم پارامترهای شناخته‌شده، مورد توجه قرار گرفته است (Chen and Guestrin 2016; Amini and Li, 2019).

تخمین عدم قطعیت مبتنی بر خطاهای محلی و خوشه‌بندی (UNEEC) به عنوان یک چارچوب مؤثر برای کمی‌سازی عدم قطعیت در مدل‌های یادگیری ماشین ظهور کرده است. این روش که بر پایه تحلیل خطاهای محلی و تکنیک‌های خوشه‌بندی استوار است، عملکردی قوی در پیش‌بینی سیل (Dogulu et al., 2014)، مدل‌سازی غلظت نیترات (Rahmati et al., 2020) و پیش‌بینی خشکسالی (Zamani et al., 2025) نشان داده است. این پیشرفت‌ها، UNEEC را به عنوان یک روش بنیادین برای مقابله با عدم قطعیت در سیستم‌های پیچیده محیطی تثبیت کرده است. تحقیقات نشان داده به عدم قطعیت مدل‌های ML در شبیه‌سازی هیدرولوژیکی کمتر توجه شده است.

استان بوشهر با بیش از ۱۵ هزار هکتار سطح زیر کشت گوجه‌فرنگی (معادل ۲۶ درصد اراضی آبی استان)، به عنوان یکی از قطب‌های اصلی تولید این محصول در کشور شناخته می‌شود. تولید عمده گوجه‌فرنگی در این منطقه به صورت خارج از فصل انجام می‌گیرد که به دلیل درآمدزایی بالا، نقش حیاتی در معیشت کشاورزان منطقه ایفا می‌کند. با متوسط عملکرد بیش از ۵۰ تن در هکتار، این محصول جایگاه ویژه‌ای در اقتصاد کشاورزی استان دارد. اما قرارگیری بوشهر در منطقه خشک و نیمه خشک و مواجهه با چالش‌های جدی مانند کمبود منابع آبی، خشکسالی‌های مکرر و افت کیفیت آب، ادامه روند فعلی کشت را با تهدید مواجهه ساخته است (پوزش شیرازی،

1- Extreme Gradient Boosting

2- Quantile XGBoost

اصلی (Hoekstra and Chapagain., 2008) طی دوره آماری ۱۳۷۸ تا ۱۴۰۳ برآورد گردید. جدول (۱) متوسط متغیرهای گیاهی و اقلیمی مورد استفاده را نشان می‌دهد.

به‌منظور انجام این پژوهش داده‌های مشاهداتی ردپای آب آبی و سبز (به‌عنوان متغیرهای وابسته) و دیگر داده‌های مربوط به کشت و داده‌های اقلیمی (به‌عنوان متغیرهای مستقل) ارزیابی و استفاده شد. در این پژوهش ردپای آب آبی در استان بوشهر، با به‌کارگیری چارچوب

جدول ۱- متوسط متغیرهای گیاهی و اقلیمی منطقه مورد استفاده

| مقدار         | متغیرهای گیاهی و اقلیمی       |
|---------------|-------------------------------|
| ۱۷۹           | دوره رشد (روز)                |
| آبان- فروردین | دوره رشد (ماه شروع تا پایان)  |
| ۴۴/۲۸         | عملکرد (ton/ha)               |
| ۳۷/۳۵         | دمای حداکثر (Tmax) (°C)       |
| ۱۵/۶۴         | دمای حداقل (Tmin) (°C)        |
| ۲۸/۴۴         | دمای میانگین (Tmean) (°C)     |
| ۱۹۳/۸۲        | بارش (mm)                     |
| ۵۳/۹۷         | رطوبت نسبی (RH) (%)           |
| ۲۵۱/۶۷        | سرعت باد (VW) (hr)            |
| ۱۲۵/۹۸        | بارش موثر (Peff) (mm)         |
| ۳۴۱/۳۹        | تبخیر و تعرق گیاهی (ETc) (mm) |
| ۱۵۲/۶۵        | نیاز آبیاری (mm)              |
| ۰/۴۵-۰/۵۵     | ضریب گیاهی (Kc I)             |
| ۱/۱-۰/۲       | ضریب گیاهی (Kc II)            |
| ۰/۷-۰/۸       | ضریب گیاهی (Kc III)           |

بریمین مدل Bagging را به RF ارتقا داد. مدل جنگل تصادفی از دو پارامتر استفاده می‌کند: تعداد درختان رگرسیون برای رشد، و تعداد ویژگی‌هایی که به‌طور تصادفی در هر تقسیم نمونه‌برداری شده‌اند (Breiman., 2001; Dong et al., 2020). در نتیجه این ویژگی، هنگامی که همه درختان آموزش داده شدند، بدون اعمال (هرس)، هر مورد جدید توسط همه درختان تجزیه و تحلیل می‌شود و پیش‌بینی نهایی با میانگین نتایج به‌دست می‌آید. آموزش الگوریتم پایه در مدل RF مانند مدل Bagging موازی است. مدل RF پیش‌بینی را با دقت بالا، بدون برآزش بیش از حد، برای مدل‌سازی ویژگی‌های هیدرولوژیکی ارائه می‌دهد (Breiman, 2001; Ruiz-Álvarez et al., 2021; Wang et al., 2017). هنگامی که مدل آموزش داده شد، پیش‌بینی را می‌توان به‌صورت زیر به‌دست آورد.

$$Y = \frac{1}{M} \sum_{m=1}^M T_m(f_t) \quad (3)$$

که در آن M تعداد درخت‌ها،  $T_m$  درخت تصمیم منفرد و  $f_t$  بردار پیش‌بینی کننده است (Ruiz-Álvarez et al., 2021).

نیاز آبی، نیاز آبیاری و بارش موثر با استفاده از مدل AGWAT (IRIMO., 2001) محاسبه شد. میزان تبخیر و تعرق گیاهی و نیاز آبیاری با استفاده از روش FAO-Penman- Monteith در شرایط استاندارد و غیراستاندارد محاسبه شد (Allen et al., 1998).

در روابط زیر،  $WF_{Green}$  ردپای آب سبز،  $WF_{Blue}$  ردپای آب آبی بر حسب متر مکعب در هر کیلوگرم می‌باشد. همچنین  $P_e$  مجموع بارندگی موثر (با استفاده از روش USDA S.C. Method) در طول دوره رشد گیاه (میلی‌متر)،  $ET_c$  تبخیر و تعرق هر گیاه (میلی‌متر در روز)،  $Y$  عملکرد محصول (تن در هکتار) است (Chapagain et al., 2006).

$$WF_{Green} = \frac{(P_e) * 10}{Y} \quad (1)$$

$$WF_{Blue} = \frac{(ET_c - P_e) * 10}{Y} \quad (2)$$

### مدل جنگل تصادفی<sup>۱</sup>

مدل جنگل تصادفی گونه‌ای از روش‌های دسته‌بندی و رگرسیون مبتنی بر ترکیب تعداد زیادی درخت تصمیم مستقل و هم‌توزیع است که اولین بار توسط بریمین در سال ۲۰۰۱ معرفی شد.

1 Random Forest (RF)

یادگیرنده اول، ابتدا به کل فضا داده‌های ورودی برازش داده می‌شود و سپس مدل برای رفع اشکالات یک یادگیرنده ضعیف برازش داده می‌شود. این فرایند برازش چند بار تکرار می‌شود تا زمانی که معیار توقف برآورد شود. پیش‌بینی نهایی مدل از مجموع پیش‌بینی هر یادگیرنده به دست می‌آید (Chen and Guestrin, 2016). معادلات آن به صورت زیر می‌باشد (Fan et al., 2018).

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad (6)$$

که در آن،  $f_t(x_i)$  یادگیرنده در مرحله  $t$  است،  $f_i^{(t)}$  و  $f_i^{(t-1)}$  پیش‌بینی‌های مراحل  $t$  و  $(t-1)$  هستند و  $x_i$  متغیر ورودی است.

$$obj^{(t)} = \sum_{k=1}^n \ell(\bar{y}y) + \sum_{k=1}^t \Omega(f_i) \quad (7)$$

که در آن،  $\ell$  تابع از دست دادن است،  $n$  تعداد مشاهدات استفاده شده و  $\Omega$  عبارت منظم‌سازی است و به صورت زیر تعریف می‌شود.

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \|W\|^2 \quad (8)$$

که در آن،  $W$  بردار امتیازها در برگ‌ها،  $\lambda$  پارامتر منظم‌سازی و  $\gamma$  حداقل زیان مورد نیاز برای تقسیم بیش‌تر گره برگ است.

### رگرسیون چندکی تقویت گرادیان شدید

رگرسیون چندکی (QXGB) ترکیبی از الگوریتم گرادیان تقویت‌شده (XGBoost) و رگرسیون چندکی است که برای مدل‌سازی رابطه بین متغیرهای پیش‌بین و کوانتایل‌های شرطی متغیر پاسخ (به جای میانگین شرطی) استفاده می‌شود. این روش در مسائلی که توزیع داده‌ها ناهموار است یا نیاز به برآورد عدم قطعیت وجود دارد، بسیار کارآمد عمل می‌کند. فرمول ریاضی تابع هزینه (Loss Function) برای کوانتایل  $\tau$  به صورت زیر است:

$$L_{\tau}(y, \hat{y}) = \sum_{i: y_i \geq \hat{y}_i} \tau |y_i - \hat{y}_i| + (1 - \tau) |y_i - \hat{y}_i| \quad (9)$$

$\tau$ : چندک هدف (مثلاً ۰.۵ برای میانه)؛  $y$ : مقادیر واقعی،  $\hat{y}$ : مقادیر پیش‌بینی شده (Zamani et al., 2025).

### معیار ارزیابی مدل‌ها

در این تحقیق برای ارزیابی کارایی مدل‌ها و روش‌های استفاده شده برای برآورد ردپای آب، از رابطه ۱۰ تا ۱۲ معیارهای سنجش خطای ضریب همبستگی (CC)، ریشه میانگین مربعات خطا

### جنگل رگرسیون چندکی<sup>۱</sup>

جنگل رگرسیون چندکی مدل توسعه‌یافته جنگل تصادفی است که ارزش تمام مشاهدات را ارزیابی می‌کند و یک توزیع مشروط از اطلاعات فراهم می‌کند. بنابراین QRF روشی دقیق برای تخمین چندک‌های شرطی متغیرهای پیش‌بینی کننده بالا با ابعاد بالا ارائه می‌دهد. از مهم‌ترین ویژگی‌های این روش عملکرد بالای آن در اندازه‌گیری اهمیت متغیرها، به منظور مشخص کردن نقش هر متغیر در پیش‌بینی پاسخ می‌باشد.

توزیع شرطی  $X|Y$  را با استفاده از وزن‌های  $\omega_i(x)$  که از جنگل رگرسیون چندکی؛  $X$  برای یک نقطه داده به دست می‌آیند، تخمین می‌زند.

$$F = (y|X = x) = \sum_{i=1}^n \omega_i(x) . I(Y_i \leq y) \quad (4)$$

$\omega_i(x)$  وزن‌های مربوط به مشاهدات آموزشی  $(X_i, Y_i)$  که از طریق فاصله درختی تعیین می‌شوند.

$I(Y_i \leq y)$  تابع نشانگر که اگر  $y \leq Y_i$ ، یک و در غیر این صورت صفر است.

CDF چندک  $\tau$  ام ( $0 < \tau < 1$ ) به عنوان معکوس تابع توزیع تجمعی تخمین زده می‌شود.

$$Q_{\tau}(X) = \inf \{y : F(y|X = x) \geq \tau\} \quad (5)$$

در عمل، این مقدار با مرتب‌سازی مقادیر  $Y_i$  در هر برگ و محاسبه چندک نمونه‌ای از توزیع وزن‌دار به دست می‌آید (Meinshausen and Ridgeway, 2006; Vaysse & Lagacherie, 2017; Dharumarajan et al., 2020).

### تقویت گرادیان شدید<sup>۲</sup>

مدل XGB توسط چن و همکاران اولین بار در سال ۲۰۱۶ به عنوان روش  $k$ -طبقه بندی و درخت‌های رگرسیونی برای اجرای گرادیان ماشین بوستینگ پیشنهاد گردید (Chen et al., 2016). مدل XGBoosting بر اساس مفهوم "افزایشی" است که همه گروه‌های پیش‌بینی یادگیرنده ضعیف را به منظور ساختن یک یادگیرنده قوی با استفاده از تکنیک تجمیع استفاده می‌شود. بعلاوه مدل XGBoosting سعی می‌کند که از بیش برازش و محاسبات سنگین جلوگیری کند. در این امر با در نظر گرفتن تکنیک بهینه‌سازی جهت بدست آوردن بهترین ترکیب کارا از متغیرهای پیشگوی نرمالیزه شده بدست می‌آید. علاوه بر این، محاسبات همزمان به طور خودکار برای عملکردهای مدل XGBoosting در طول آموزش انجام می‌شود.

1- Quantile Regression Forests  
2- Extreme Gradient Boosting (XGB)

بیشتر و میزان RMSE و MAE کمتر باشد، مدل از کارایی بیشتری برای ارائه بهترین جواب برخوردار است (بادرود و بذرافشان، ۱۴۰۱).  
 (RMSE) و میانگین خطای مطلق (MAE) در مرحله آموزش و آزمایش استفاده شد (جدول ۲). بر این اساس هر چه میزان CC

۲-جدول معیارهای سنجش خطا

| مقدار بهینه | رابطه  | معیارهای سنجش خطا            |
|-------------|--|------------------------------|
| ۱           | $CC = \frac{N \sum HF - (H)(\sum F)}{\sqrt{\sum(H^2) - (\sum H)^2} \sqrt{N \sum(F^2) - (\sum F)^2}}$ | (۱۰) Correlation Coefficient |
| صفر         | $RMSE = \sqrt{\frac{1}{N} \sum_i (H - F)^2}$   | (۱۱) Root Mean Square Error  |
| صفر         | $MAE = \frac{1}{N} \sum_i  H - F $   | Mean Absolute Error<br>(۱۲)  |

H: داده های مشاهده شده، F: داده های برآورد شده، N: تعداد داده ها

( $\sigma_j$ ) برای هر خوشه است که به صورت زیر محاسبه می شوند:

که در آن:

$$\bar{e} = \frac{1}{n_j} \sum_{i \in C_j} e_i \quad (14)$$

$$\sigma_j = \sqrt{\frac{1}{n_j} \sum_{i \in C_j} (e_i - \bar{e})^2} \quad (15)$$

$n_j$ : تعداد نمونه های موجود در خوشه  $C_j$

#### مرحله ۴: ساخت فاصله اطمینان پیش بینی

برای هر خوشه، فاصله اطمینان  $(1-\alpha)\%$  به صورت زیر

ساخته می شود:

$$[L_i, U_i] = [Y + q_{\alpha/2}^{(j)}, Y_i + q_{1-\alpha/2}^{(j)}] \quad (16)$$

که در آن که در آن  $q_{\alpha/2}^{(j)}$  و  $q_{1-\alpha/2}^{(j)}$  به ترتیب نشان دهنده کوانتیل های  $\alpha/2$  و  $1-\alpha/2$  از توزیع خطا در خوشه  $C_j$  هستند. PICP به صورت زیر محاسبه می شود:

$$PICP = \frac{1}{n} \sum_{i=1}^n I(Y_i \in [L_i, U_i]) \times 100\% \quad (17)$$

$n$ : تعداد کل نمونه ها،  $Y_i$ : مقدار مشاهده شده واقعی برای نمونه

$i$ :  $[L_i, U_i]$ : بازه پیش بینی برای نمونه  $i$  (به ترتیب کرانه های

پایین و بالا)،  $I(0)$ : تابع نشانگر (۱) اگر  $Y_i$  در بازه قرار گیرد، در

غیر این صورت (۰) (Dogulu et al., 2015).

#### تحلیل عدم قطعیت

##### روش UNEEC تخمین عدم قطعیت با استفاده از خوشه بندی

##### خطا

یک روش ناپارامتریک برای کمی سازی عدم قطعیت در مدل های پیش بین است. این روش از داده های تاریخی بدون نیاز به فرضیات پارامتریک درباره توزیع خطاها برای تخمین عدم قطعیت استفاده می کند. در ادامه روش شناسی کامل این رویکرد به همراه معیار PICP ارائه شده است (Rahmati et al., 2020).

**مرحله ۱:** آموزش مدل و محاسبه خطاها: ابتدا مدل پیش بین بر روی داده های تاریخی آموزش داده می شود. برای هر نمونه آموزشی  $i$ ، پیش بینی نقطه ای ( $\hat{Y}_i$ ) محاسبه می شود. سپس خطاهای پیش بینی با استفاده از معادله (۱) محاسبه می گردند:

$$e_i = Y_i - \hat{Y}_i \quad (13)$$

که در آن:

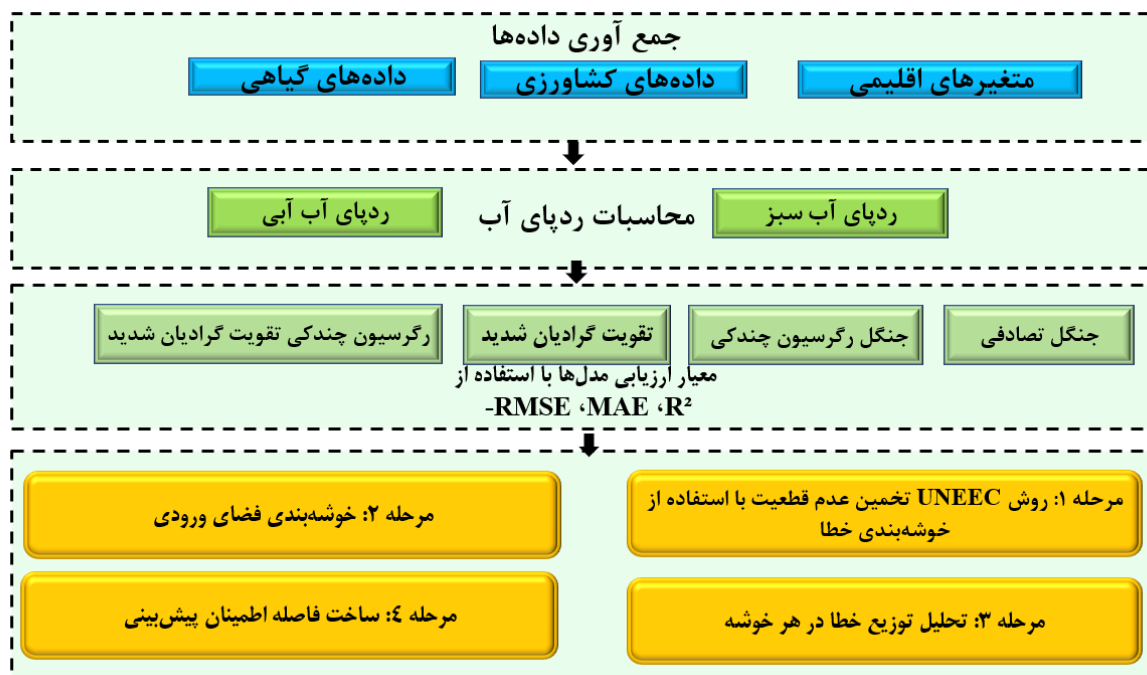
$Y_i$ : مقدار واقعی متغیر مورد مطالعه،  $Y_i \hat{Y}_i$ : مقدار پیش بینی شده توسط مدل.

##### مرحله ۲: خوشه بندی فضای ورودی

فضای ورودی  $X$  به  $K$  خوشه  $C_1$  تا  $C_k$  تقسیم می شود. این خوشه بندی می تواند با روش هایی مانند k-means انجام شود. هدف از این مرحله، گروه بندی نمونه های مشابه بر اساس شرایط ورودی است.

##### مرحله ۳: تحلیل توزیع خطا در هر خوشه

برای هر خوشه  $C_j$ ، توزیع خطاهای مربوطه تحلیل می شود. این تحلیل شامل محاسبه میانگین خطا ( $\bar{e}_j$ ) و انحراف معیار خطاها



شکل ۱- مراحل انجام تحقیق

## نتایج و بحث

### متغیرهای موثر بر ردپای آبی و سبز

تحلیل نمودار حرارتی ردپای سبز (شکل ۲- الف) نشان می‌دهد که متغیرهای اقلیمی و مدیریتی تأثیر متفاوتی بر ردپای آب سبز گوجه‌فرنگی در بوشهر دارند. بارش (P) و بارش مؤثر (Peff) بیشترین همبستگی مثبت را با ردپای آب سبز نشان می‌دهند، که نشان‌دهنده نقش کلیدی این عوامل در تأمین رطوبت خاک برای کشت دیم است. در مقابل، دماهای حداکثر (Tmax) و حداقل (Tmin) و تبخیر و تعرق گیاهی (ETc) همبستگی منفی دارند، که احتمالاً به دلیل افزایش تبخیر و کاهش رطوبت قابل دسترس برای گیاه است. رطوبت نسبی (RHmean, RHmax, RHmin) نیز تأثیر مثبت ولی کم‌تری دارد، در حالی که سرعت باد (SH) همبستگی منفی نشان داد. عملکرد محصول (CropYield) ارتباط ضعیفی با ردپای آب سبز داشت، که نشان می‌دهد عوامل دیگری مانند مدیریت زراعی و آبیاری نیز در تولید نقش مهمی ایفا می‌کنند. به‌طور کلی، این تحلیل تأکید می‌کند که بارش و مدیریت بهینه رطوبت خاک عوامل کلیدی در افزایش ردپای آب سبز و پایداری کشت گوجه‌فرنگی در شرایط خشک بوشهر هستند.

(شکل ۲- ب)، نمودار حرارتی ارتباط بین متغیرهای مختلف و ردپای آب آبی (WFBlue) در کشت گوجه‌فرنگی بوشهر را نشان

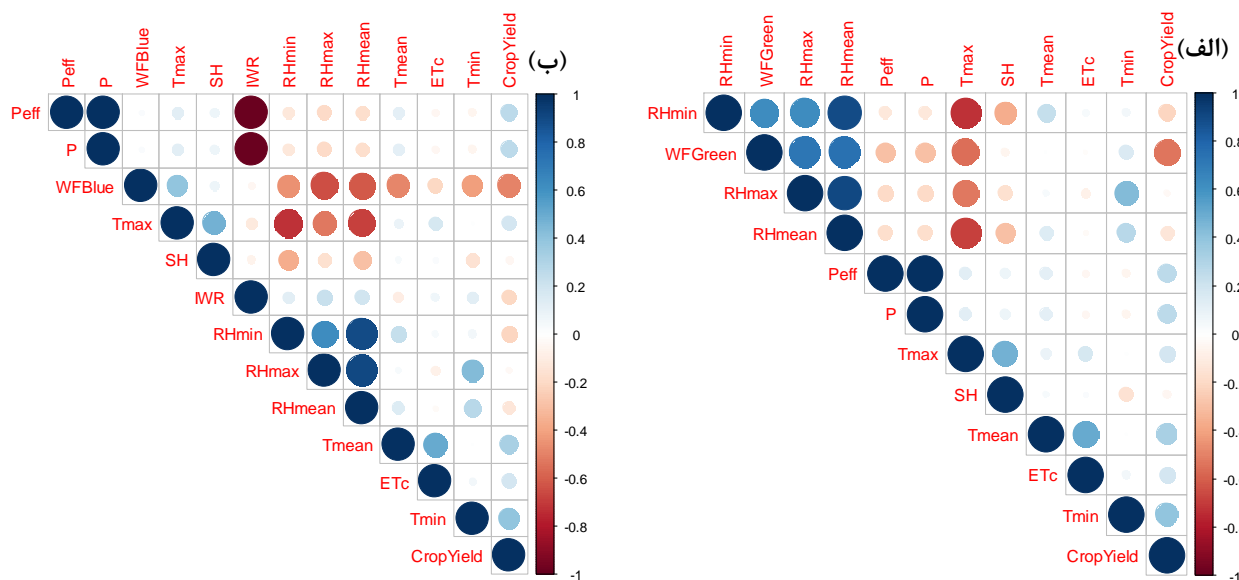
می‌دهد. متغیرهای ETc (تبخیر و تعرق) و Tmax (حداکثر دما) با رنگ‌های گرم (قرمز) بیشترین تأثیر مثبت را بر WFBlue دارند، چرا که افزایش آنها منجر به مصرف آب بیشتر می‌شود. در مقابل، RHmean (رطوبت نسبی متوسط) با رنگ‌های سرد (آبی) تأثیر کاهنده نشان می‌دهد، زیرا رطوبت بالاتر نیاز آبی را کاهش می‌دهد. ساعات آفتابی (SH) و عملکرد محصول (CropYield) تأثیر محسوسی ندارند، در حالی که بارندگی (P) و حداقل دما (Tmin) نیز نقش چندانی در این تحلیل ایفا نکرده‌اند. این نتایج نشان می‌دهد که مدیریت تبخیر و تعرق و کنترل دما می‌تواند راهکار مؤثری برای کاهش ردپای آب آبی در این منطقه باشد.

نتایج مطالعه حاضر نشان می‌دهد که بارش (P) و بارش مؤثر (Peff) بیشترین همبستگی مثبت را با ردپای آب سبز (WFGreen) گوجه‌فرنگی در بوشهر دارند، یافته‌ای که با پژوهش‌های جهانی (Mekonnen & Hoekstra, 2011; Ridoutt et al., 2010) همسو است و تأثیر حیاتی بارش را در تأمین آب سبز در مناطق خشک تأیید می‌کند. از سوی دیگر، دما و تبخیر و تعرق (ETc) همبستگی منفی با WFGreen نشان دادند که این نتیجه با یافته‌های (Chukalla et al., 2015) مطابقت دارد و نشانگر کاهش کارایی مصرف آب سبز با افزایش دماست. همچنین، اگرچه رطوبت نسبی (RH) تأثیر مثبتی بر WFGreen داشت، اما این تأثیر نسبتاً ضعیف بود که با نتایج لوواری و همکاران (Lovarelli et al., 2018) همخوانی دارد و بیانگر آن

بارش (P) بر WFBblue با یافته‌های هرات و همکاران (Herath et al., 2011) در استرالیا همخوانی دارد که بیان می‌کند در سیستم‌های آبیاری محور، بارش نقش محدودی در کاهش مصرف آب آبی ایفا می‌کند. برای بهبود مدیریت آب در این منطقه، راهکارهای عملی مانند استفاده از روش‌های حفظ رطوبت خاک (مالچ‌بندی و کشت حفاظتی) برای افزایش سهم آب سبز و بهینه‌سازی زمان آبیاری با پایش ETC واقعی طبق توصیه Allen و همکاران، ۱۹۹۸ پیشنهاد می‌شود. همچنین، به کارگیری سیستم‌های آبیاری زیرسطحی می‌تواند تبخیر را کاهش داده و کارایی مصرف آب آبی را بهبود بخشد. این راهکارها نه تنها با چارچوب نظری Hoekstra et al., (2006) در تمایز بین منابع آب سبز و آبی هماهنگ است، بلکه می‌تواند الگوی کشت پایدارتری را در مناطق خشک مانند بوشهر ایجاد کند.

است که رطوبت نسبی تنها در صورت همراهی با بارش می‌تواند تأثیر معناداری بر افزایش ردپای سبز داشته باشد. این یافته‌ها به‌طور کلی بر اهمیت مدیریت بهینه منابع آب سبز در شرایط اقلیمی خشک و نیمه‌خشک تأکید می‌کنند.

نتایج این مطالعه نشان می‌دهد که ردپای آب آبی (WFBblue) گوجه‌فرنگی در بوشهر به شدت تحت تأثیر تبخیر و تعرق (ETc) و دمای حداکثر (Tmax) قرار دارد که با یافته‌های پژوهش‌های بین‌المللی مانند مطالعه Aldaya و همکاران (۲۰۱۲) در اسپانیا و Sun و همکاران (۲۰۱۳) در چین همسو است. این مطالعات تأکید می‌کنند که افزایش دما و ETc می‌تواند نیاز آبی محصول را تا ۱۵٪ افزایش دهد. از سوی دیگر، نقش کاهنده رطوبت نسبی (RHmean) بر WFBblue نیز مشابه نتایج Pellicer-Martínez & Martínez-Paz (2016) در منطقه مورسیا است که نشان دادند افزایش رطوبت نسبی می‌تواند ردپای آب آبی را تا ۱۰٪ کاهش دهد. در مقابل، تأثیر ناچیز



شکل ۲- نمودار حرارتی همبستگی بین متغیرهای مؤثر بر ردپای سبز (الف) و (ب) ردپای آبی در محصول گوجه فرنگی بوشهر

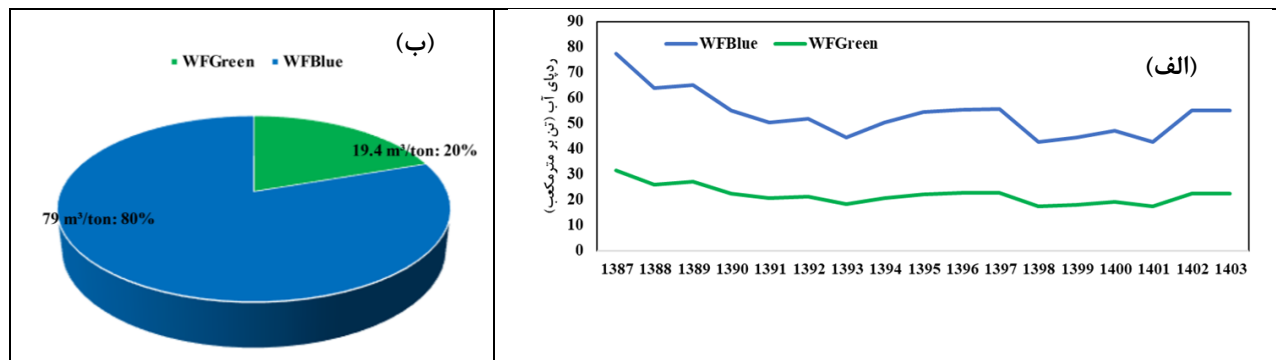
مدیریت پایدار آب از طریق بهبود روش‌های آبیاری، تغییر الگوی کشت و پایش مستمر منابع آب در منطقه می‌باشد. تحلیل داده‌های ردپای آب گوجه‌فرنگی در استان بوشهر نشان می‌دهد (شکل ب) که سهم آب آبی با ۷۹ مترمکعب بر تن (۸۰ درصد) بسیار بیشتر از سهم آب سبز با ۱۹/۴ مترمکعب بر تن (۲۰ درصد) است. این وضعیت بیانگر وابستگی شدید تولید گوجه‌فرنگی به منابع آب شیرین (آب‌های زیرزمینی و سطحی) در منطقه‌ای با اقلیم خشک و گرم است که به‌طور طبیعی با محدودیت منابع آبی مواجه می‌باشد. چنین الگوی مصرفی نه تنها فشار مضاعفی بر ذخایر

### تحلیل ردپای آب در محصول گوجه فرنگی

شکل (۳) تغییرات ردپای آبی و سبز را طی سال‌های ۱۳۸۷ تا ۱۴۰۳ نمایش می‌دهد. براساس نتایج، ردپای آبی (مصرف آب از منابع سطحی و زیرزمینی) گوجه‌فرنگی در بوشهر طی سال‌های اخیر روندی صعودی داشته، به‌ویژه بین سال‌های ۱۳۹۲ تا ۱۳۹۵ که احتمالاً ناشی از کاهش بارندگی و افزایش وابستگی به آبیاری است، در حالی که ردپای سبز (استفاده از آب باران) از ۱۳۹۴ به بعد کاهش چشمگیری یافته که نشان‌دهنده تأثیر تغییرات اقلیمی و کاهش بارش‌های مؤثر است. این روندها حاکی از افزایش فشار بر منابع آبی زیرزمینی و لزوم

آب سبز (۲۰ درصد) نشان می‌دهد سیستم کشت موجود از ظرفیت بارش‌های طبیعی و رطوبت خاک به نحو مطلوبی بهره‌برداری نمی‌کند.

استراتژیک آب منطقه وارد می‌کند، بلکه در بلندمدت می‌تواند منجر به افت سطح سفره‌های آب زیرزمینی، شور شدن منابع آب باقیمانده و تخریب اکوسیستم‌های آبی شود. این در حالی است که سهم ناچیز



شکل ۳- تغییرات ردپای سبز و آبی طی دوره آماری (الف) و (ب) سهم اجزای ردپای آب در تولید گوجه‌فرنگی در بوشهر

براساس نتایج، ۸۰ درصد سهم ردپای آب مربوطه به آب آبی است. مقایسه این شاخص‌ها با استانداردهای جهانی حاکی از ناپایداری الگوی فعلی مصرف آب در کشت گوجه‌فرنگی بوشهر است. برای دستیابی به تولید پایدار، لازم است راهکارهای جامعی شامل توسعه سیستم‌های آبیاری کارآمد (مانند آبیاری قطره‌ای زیرسطحی)، استفاده از ارقام مقاوم به خشکی، بهینه‌سازی زمان کشت متناسب با دوره‌های بارشی، و اجرای برنامه‌های مدیریت یکپارچه منابع آب در دستور کار قرار گیرد. این اقدامات می‌تواند ضمن کاهش فشار بر منابع آب زیرزمینی، سهم آب سبز را افزایش داده و تاب‌آوری بخش کشاورزی منطقه را در برابر تغییرات اقلیمی بهبود بخشد. در مقایسه با مطالعات مشابه در سایر مناطق، الگوی مصرف آب گوجه‌فرنگی در بوشهر نگران‌کننده به نظر می‌رسد. تحقیقات نشان می‌دهد در مناطق خشک مشابه مانند اسپانیا سهم آب سبز در کشت گوجه‌فرنگی معمولاً بین ۳۰ تا ۵۰ درصد است (Chahpagian and Orr., 2009). براساس نتایج، ردپای آبی با سهم ۸۰ درصدی، بخش غالب مصرف آب در تولید گوجه‌فرنگی استان بوشهر را تشکیل می‌دهد، در حالی که سهم ردپای سبز تنها ۲۰ درصد بود. این وابستگی شدید به منابع آب آبی عمدتاً ناشی از شرایط اقلیمی خشک منطقه با بارش کم، تبخیر و تعرق بالا و الگوی کشت خارج از فصل است که باعث کاهش دسترسی به آب سبز (ناشی از بارش) و افزایش اتکا به آبیاری تکمیلی از منابع زیرزمینی و سطحی می‌شود (Dehghanpir et al., 2025). این وضعیت، آسیب‌پذیری سیستم تولید گوجه‌فرنگی را در برابر تغییرات اقلیمی و افت منابع آب زیرزمینی آشکار می‌سازد.

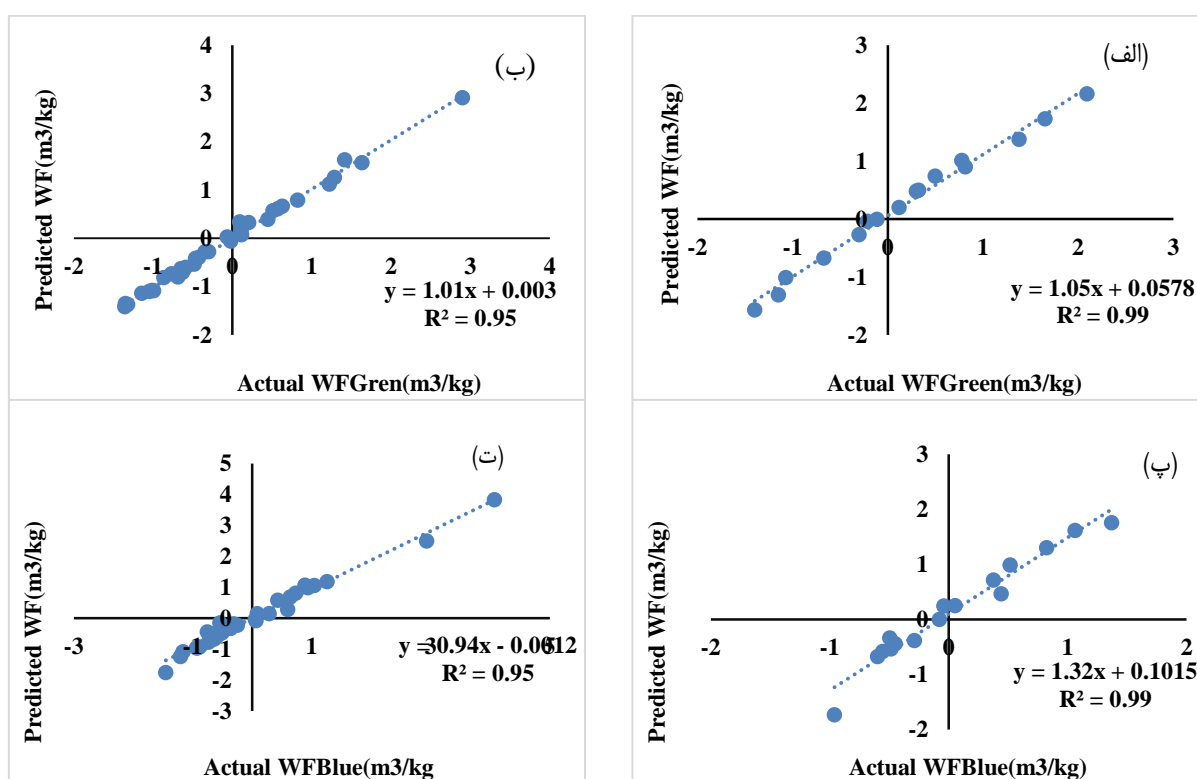
مدل بهینه QXGBoost نشان می‌دهد. مطالعات متعددی نشان داده‌اند که ترکیب روش‌های مبتنی بر

#### بررسی کارایی مدل‌های مورد بررسی

نتایج جدول (۳) نشان داد که تمام مدل‌های مورد استفاده شامل

بالا (مانند ردپای آب) عملکرد بهتری دارند. علاوه بر این، پژوهش‌های اخیر (Zhang et al., 2020) تأیید کرده‌اند که ادغام مکانیسم‌های بهینه‌سازی گرادیان در XGBoost با چارچوب رگرسیون چندکی، به مدل اجازه می‌دهد تا نه تنها پیش‌بینی‌های دقیق‌تری ارائه دهد، بلکه تخمین‌های معتبری از دامنه عدم قطعیت نیز به دست آورد. این ویژگی به طور خاص در کاربردهای مدیریت منابع آب که نیازمند ارزیابی ریسک هستند، بسیار ارزشمند است بنابراین، برتری QXGBoost در این مطالعه نه تنها با مبانی نظری همخوانی دارد، بلکه توسط شواهد تجربی متعددی در حوزه‌های مشابه نیز تأیید شده است.

رگرسیون چندکی با الگوریتم‌های پیشرفته یادگیری ماشین مانند XGBoost می‌تواند منجر به بهبود قابل توجهی در دقت پیش‌بینی‌ها شود (Meinshausen & Ridgeway 2006). برتری مدل QXGBoost در این مطالعه را می‌توان به چند عامل کلیدی نسبت داد به طوری که، این مدل با ترکیب مزایای XGBoost در پردازش روابط غیرخطی و پیچیده بین متغیرها و توانایی رگرسیون چندکی در مدیریت داده‌های پرت و عدم قطعیت‌ها (Koenker & Hallock., 2001)، قادر است توزیع شرطی متغیر هدف را با دقت بالاتری مدل‌سازی کند. علاوه بر این، مطالعاتی مانند (Bromberg et al, 2019) نشان داده‌اند که رویکردهای چندکی در مقایسه با روش‌های نقطه‌ای سنتی، به ویژه در مدل‌سازی پدیده‌های طبیعی با تغییرپذیری



شکل ۴- نمودار پراکنش مقادیر مشاهده‌ای و برآورد شده توسط مدل QXGBoost در ردپای آب سبز در مرحله آموزش (الف) مرحله آزمایش (ب)، ردپای آبی در مرحله آموزش (پ) و آزمایش (ت)

قطعیت دارد. این مدل در مقایسه با QRF که PICP بالایی (۰/۷۱) در آموزش و (۰/۹۲ در تست) نشان می‌دهد، فاصله اطمینان باریک‌تری ارائه کرده که نشانگر دقت بالاتر است. مدل‌های RF و XGBoost عملکرد ضعیف‌تری در پوشش عدم قطعیت دارند، به‌ویژه RF با PICP=۰/۴۰ در آموزش و (۰/۲۵ در تست) که نشان‌دهنده تخمین نادرست محدوده عدم قطعیت است. برای ردپای آبی (WFBlue)، QXGBoost مجدداً با PICP=۰/۷۸ در آموزش و (۰/۷۹ در تست و

#### تحلیل عدم قطعیت مدل‌ها

ارزیابی عدم قطعیت مدل‌ها با روش UNEEC برآورد شد و نتایج در جدول (۴) نشان داده شده است. نتایج ارزیابی عدم قطعیت مدل‌ها برای ردپای سبز (WFGreen) نشان داد که QXGBoost با داشتن فاصله اطمینان متقارن‌تر (LB=-0.34 و UB=0.37) و پوشش پیش‌بینی‌های فاصله‌ای (PICP=0.80) در مرحله آموزش و همچنین PICP=0.93 در مرحله تست، بهترین عملکرد را در مدیریت عدم

استفاده از روش‌های ترکیبی مانند QXGBoost که توانایی بالایی در تخمین همزمان پیش‌بینی‌های نقطه‌ای و فاصله‌ای دارند، ضروری است.

شکل (۵) اختلاف باندها و میزان PICP را نشان می‌دهد. در این شکل بالا بودن مقدار خطا (UA) در برآورد مدل‌ها بخوبی قابل تشخیص است. لذا نتایج این شکل نشان‌دهنده چالش‌های قابل توجه در مدل‌سازی ردپای آبی (WFBlue) نسبت به ردپای سبز (WFGreen) است. در مدل‌های مختلف، مقادیر UB و LB برای ردپای آبی نوسانات بیشتری دارند که نشان‌دهنده عدم قطعیت بالاتر در پیش‌بینی‌ها است. به‌ویژه در مدل RF برای داده‌های تست، حدود فاصله اطمینان ( $UB=0/56$  و  $LB=4/34$ ) و اختلاف بسیار زیادی دارند که نشان‌دهنده خطای سیستماتیک در تخمین است. همچنین، عرض فاصله اطمینان (UA) برای ردپای آبی در تمام مدل‌ها به‌طور محسوسی بزرگ‌تر از ردپای سبز است.

فاصله اطمینان متعادل ( $UB=0/32$  و  $LB=-0/32$ ) بهترین نتایج را نشان می‌دهد، در حالی که RF با اختلاف زیاد بین حدود بالا و پایین ( $UB=0/69$  و  $LB=0/70$ ) و PICP پایین ( $0/37$ ) ضعیف‌ترین عملکرد را دارد.

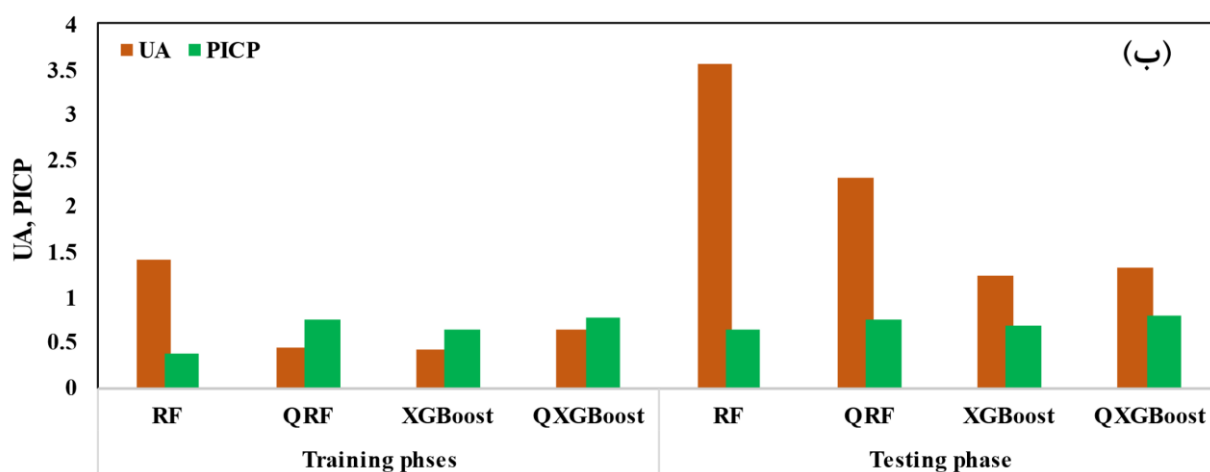
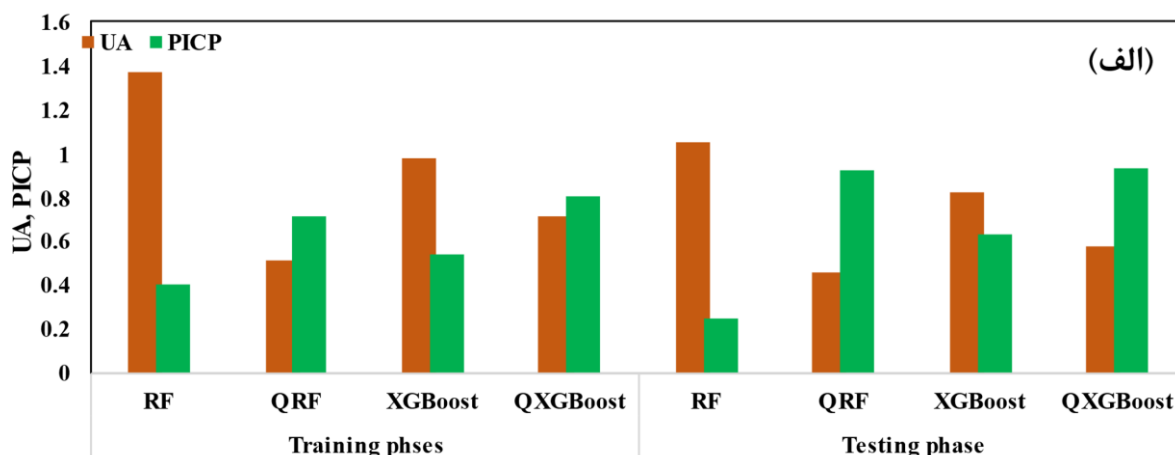
در مقایسه کلی، مدل‌های مبتنی بر رگرسیون چندکی (QRF و QXGBoost) به‌ویژه QXGBoost، در هر دو شاخص ردپای سبز و آبی، توانایی بهتری در تخمین محدوده‌های عدم قطعیت با پوشش مناسب و فاصله اطمینان متقارن‌تر نشان داده‌اند. این برتری احتمالاً ناشی از ماهیت ذاتی رگرسیون کوانتیل در مدلسازی توزیع شرطی داده‌ها است. در مقابل، مدل‌های استاندارد (RF و XGBoost) در تخمین محدوده عدم قطعیت عملکرد نامناسبی داشته‌اند، به‌طوری‌که یا فاصله اطمینان بسیار گسترده (مثل RF در WFBlue یا  $UB=4/34$  و  $LB=-0/32$ ) یا پوشش ناکافی (مثل XGBoost در WFGreen با  $PICP=0/54$ ) ارائه کرده‌اند. این نتایج تأکید می‌کند که برای مسائل حساس به عدم قطعیت مانند مدل‌سازی ردپای آب،

جدول ۳- نتایج شبیه‌سازی اجزای ردپای آب با مدل‌های یادگیری ماشین

| آزمایش       |      | آموزش |       | مدل  |      |          |
|--------------|------|-------|-------|------|------|----------|
| ردپای آب سبز |      |       |       |      |      |          |
| MAE          | RMSE | Rsqr  | MAE   | RMSE | Rsqr |          |
| 0/28         | 0/23 | 0/91  | 0/02  | 0/05 | 0/99 | RF       |
| 0/24         | 0/38 | 0/93  | 0/01  | 0/04 | 0/99 | QRF      |
| 0/18         | 0/27 | 0/90  | 0/02  | 0/03 | 0/99 | XGBoost  |
| 0/12         | 0/13 | 0/95  | 0/06  | 0/01 | 0/99 | QXGBoost |
| ردپای آب آبی |      |       |       |      |      |          |
| 0/17         | 0/35 | 0/92  | 0/028 | 0/04 | 0/96 | RF       |
| 0/27         | 0/29 | 0/84  | 0/016 | 0/04 | 0/96 | QRF      |
| 0/18         | 0/23 | 0/91  | 0/002 | 0/03 | 0/97 | XGBoost  |
| 0/14         | 0/16 | 0/95  | 0/06  | 0/01 | 0/99 | QXGBoost |

جدول ۴- عدم قطعیت مدل‌های مورد بررسی در مدل‌سازی ردپای آب در دو مرحله آموزش و آزمایش

| آموزش        |      | آزمایش |      | مدل  |      |       |      |          |
|--------------|------|--------|------|------|------|-------|------|----------|
| ردپای آب سبز |      |        |      |      |      |       |      |          |
| PICP         | UA   | LB     | UB   | PICP | UA   | LB    | UB   |          |
| 0/25         | 1/05 | -0/53  | 0/53 | 0/40 | 1/37 | -0/89 | 0/48 | RF       |
| 0/92         | 0/46 | 0/05   | 0/51 | 0/71 | 0/51 | -0/27 | 0/24 | QRF      |
| 0/63         | 0/82 | -0/47  | 0/35 | 0/54 | 0/98 | -0/30 | 0/69 | XGBoost  |
| 0/93         | 0/58 | -0/34  | 0/24 | 0/80 | 0/71 | -0/34 | 0/37 | QXGBoost |
| ردپای آب آبی |      |        |      |      |      |       |      |          |
| 0/65         | 3/56 | 4/34   | 0/56 | 0/37 | 1/40 | 0/70  | 0/69 | RF       |
| 0/75         | 2/31 | -0/42  | 1/89 | 0/74 | 0/45 | -0/21 | 0/24 | QRF      |
| 0/68         | 1/24 | -0/19  | 1/05 | 0/64 | 0/41 | -0/33 | 0/07 | XGBoost  |
| 0/79         | 1/31 | -0/38  | 0/93 | 0/78 | 0/64 | -0/32 | 0/32 | QXGBoost |



شکل ۵- اختلاف باندها و میزان PICP در مدل‌های مورد استفاده در ردپای سبز (الف) و ردپای آبی (ب)

مدل‌سازی وابستگی‌های پیچیده و چندمتغیره بدون نیاز به فرضیات محدودکننده در مورد توزیع خطاها، برتری محسوس نسبت به روش‌های متعارف دارد. این رویکرد قادر است به طور همزمان ناهمسانی واریانس را شناسایی کرده و مجموعه‌ای از سناریوهای ممکن را تولید کند که به طور واقع‌بینانه‌تری عدم قطعیت ذاتی در پیش‌بینی‌های مبتنی بر داده را منعکس می‌کنند (Dogulu et al., 2015).

### نتیجه‌گیری

نتایج این مطالعه نشان می‌دهد که کشت گوجه‌فرنگی در بوشهر با وابستگی شدید به منابع آبی (۸۰٪) و سهم ناچیز آب سبز (۲۰٪) همراه است که الگویی ناپایدار را در مدیریت منابع آب منطقه رقم زده است. افزایش ردپای آبی و کاهش ردپای سبز طی سال‌های اخیر، حاکی از تشدید فشار بر منابع آب زیرزمینی و سطحی است که

عدم قطعیت در مدل‌سازی ردپای آبی (WFBlue) به طور معناداری بیشتر از ردپای سبز (WFGreen) است که این امر عمدتاً ناشی از سه عامل کلیدی است: اولاً، ردپای آبی وابستگی شدیدی به فعالیت‌های انسانی و مدیریت آبیاری دارد که ذاتاً متغیر و غیرقطعی هستند (Hoekstra & Mekonnen, 2012). ثانیاً، منابع آب آبی (آب‌های سطحی و زیرزمینی) در مقایسه با آب سبز (ناشی از بارش) بیشتر تحت تأثیر تغییرات اقلیمی و برداشت‌های بی‌رویه قرار می‌گیرند که منجر به نوسانات شدید در داده‌ها می‌شود (Fazeli Farsani et al., 2019). ثالثاً، پارامترهای کیفی آب مانند شوری و آلودگی که در مدل‌ها اغلب نادیده گرفته می‌شوند، بر دقت پیش‌بینی‌های WFBlue تأثیر منفی می‌گذارند (Chukalla et al., 2015). این عوامل در کنار پیچیدگی روابط هیدرولوژیکی در سیستم‌های آبیاری، منجر به افزایش خطای سیستماتیک و گسترده‌تری بیشتر فاصله اطمینان در مدل‌سازی ردپای آبی شده‌اند. روش UNEEC به دلیل ماهیت غیرپارامتریک و توانایی در

مؤثر (مانند بارش و تبخیر و تعرق) و ارائه راهکارهای عملیاتی مانند آبیاری زیرسطحی، کشت ارقام مقاوم به خشکی و مدیریت رطوبت خاک، از دیگر دستاوردهای این پژوهش است که می‌تواند به سیاست‌گذاران در مدیریت پایدار منابع آب کمک کند.

مهم‌ترین محدودیت این مطالعه، عدم قطعیت بالای مدل‌ها در پیش‌بینی ردپای آبی ناشی از پیچیدگی عوامل انسانی (مانند سیاست‌های آبیاری و برداشت غیرمجاز آب) و تغییرپذیری منابع آب زیرزمینی بود. همچنین، کمبود داده‌های دقیق مدیریتی (مانند نوع آبیاری و کیفیت آب) و وابستگی شدید مدل‌ها به داده‌های اقلیمی، دقت پیش‌بینی‌ها را تحت تأثیر قرار داد. این محدودیت‌ها نیازمند جمع‌آوری داده‌های دقیق‌تر و ادغام عوامل مدیریتی در مدل‌سازی آینده است. علاوه بر این، عدم امکان مقایسه جامع با مطالعات مشابه در سایر مناطق خشک جهان، ارزیابی کامل‌تر نتایج را محدود کرد.

### منابع

باردوج، ف. و بذرافشان، ا. ۱۴۰۱. مدل‌سازی ردپای آب گندم با استفاده از مدل‌های یادگیری ماشین در استان فارس. آکو هیدرولوژی. ۹(۳): ۶۸۹-۶۷۵.

پوزش شیرازی، م. ۱۳۸۳. بررسی تأثیر مقادیر مختلف پتاسیم بر کارایی مصرف آب و تحمل به خشکی گیاه گوجه فرنگی در استان بوشهر. مجله علوم و کشاورزی ایران. ۳۶(۶): ۱۵۴۸-۱۵۳۹.

Abdallah, M., Mohammadi, B., Zaroug, M.A., Omer, A., Cheraghalizadeh, M., Eldow, M.E. and Duan, Z. 2022. Reference evapotranspiration estimation in hyper-arid regions via D-vine copula based-quantile regression and comparison with empirical approaches and machine learning models. *Journal of Hydrology: Regional Studies*. 44:101259.

Amini A, Li H. 2019. Quantile regression for tree-based methods. *Ann Stat* 47(4):2039-2068.

Bromberg, C.L., Gazen, C., Hickey, J.J., Burge, J., Barrington, L. and Agrawal, S., 2019, December. Machine learning for precipitation nowcasting from radar images. In *Proceedings of the Machine Learning and the Physical Sciences Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, BC, Canada (pp. 1-4).

Breiman, L. 2001. Random forests. *Machine learning*. 45(1): pp.5-32. <https://doi.org/10.1023/A:1010933404324>.

Chen, T. and Guestrin, C. 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Bazarfshan, O., Yahyazadeh, M., Jamshidi, S. and Zamani, H., 2022. Spatial prioritization of tomato cultivation based on water footprint, land

در بلندمدت می‌تواند منجر به بحران‌های جدی مانند افت سفره‌های آب زیرزمینی، شور شدن آب‌های باقیمانده و تخریب اکوسیستم‌ها شود. از سوی دیگر، تأثیرپذیری ردپای سبز از متغیرهای اقلیمی مانند بارش و دما، نشان‌دهنده آسیب‌پذیری این سیستم در برابر تغییرات اقلیمی است. این یافته‌ها ضرورت بازنگری در الگوی کشت فعلی و به‌کارگیری راهکارهای یکپارچه مدیریت آب را بیش از پیش آشکار می‌سازد.

کشت خارج‌ازفصل گوجه‌فرنگی در بوشهر با وجود مزیت‌های اقتصادی، به دلیل وابستگی شدید به منابع آب آبی (۸۰٪ ردپای آب)، الگویی ناپایدار است. برای کاهش ردپای آب و حفظ پایداری این سیستم، ترکیب راهکارهایی مانند مالچ‌پاشی) کاهش ۳۰ درصد تبخیر خاک، کشت ارقام مقاوم به کم‌آبی (مانند ارقام هیبریدهای کم‌آب‌بر)، و آبیاری بهینه (زیرسطحی و هوشمند) ضروری است. بدون این اقدامات، فشار بر منابع آب زیرزمینی تشدید شده و امنیت غذایی منطقه را تهدید خواهد کرد. بنابراین، سیاست‌گذاری برای ترویج این روش‌ها میان کشاورزان و توسعه تحقیقات ارقام سازگار با کم‌آبی، گام کلیدی در جهت کشت پایدار گوجه‌فرنگی در بوشهر محسوب می‌شود. مدل‌سازی با QXGBoost به‌عنوان ابزاری کارآمد، امکان پیش‌بینی ردپای آب و تحلیل عدم قطعیت‌ها را فراهم کرده است. با این حال، تفاوت معنادار در عدم قطعیت مدل‌های ردپای آبی و سبز، لزوم توجه به عوامل انسانی و مدیریتی (مانند الگوی آبیاری و سیاست‌های برداشت آب) را در کنار متغیرهای اقلیمی برجسته می‌سازد. این نتایج تأکید می‌کنند که مدیریت پایدار منابع آب در بوشهر مستلزم کاهش وابستگی به آب آبی از طریق بهبود کارایی آبیاری، افزایش سهم آب سبز با روش‌های حفظ رطوبت خاک، و پایش مستمر منابع آب با استفاده از مدل‌های پیشرفته است.

استفاده از سیستم‌های آبیاری زیرسطحی و هوشمند برای کاهش تبخیر و تعرق، افزایش سهم آب سبز از طریق اجرای روش‌های کشت حفاظتی (مانند مالچ‌پاشی و کشت مستقیم) برای بهره‌برداری بهینه از بارش، تغییر الگوی کشت از طریق معرفی ارقام مقاوم به خشکی و تنظیم تقویم کشت متناسب با دوره‌های بارشی می‌تواند اقدامی مفید در جهت کاهش ردپای آب باشد.

این مطالعه با به‌کارگیری مدل‌های ترکیبی یادگیری ماشین و تحلیل عدم قطعیت با روش UNEEC، موفق به مدل‌سازی دقیق ردپای آب سبز و آبی گوجه‌فرنگی در استان بوشهر شد. مهم‌ترین دستاورد تحقیق، توسعه مدل QXGBoost به‌عنوان مدل بهینه با قابلیت پیش‌بینی با دقت بالا و عدم قطعیت پایین بود. این مدل نشان داد که ردپای آبی با سهم ۸۰ درصدی، وابستگی شدید سیستم کشت به منابع آب زیرزمینی و سطحی را نشان می‌دهد، در حالی که ردپای سبز تنها ۲۰ درصد سهم دارد. همچنین، شناسایی متغیرهای کلیدی

- Maroufpoor, S., 2020b. Modeling long-term dynamics of crop evapotranspiration using deep learning in a semi-arid environment. *Agricultural Water Management*. 241:106334.
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X. and Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and forest meteorology*. 263: 225-241.
- Freeman, E.A. and Moisen, G.G., 2015. An application of quantile random forests for predictive mapping of forest attributes. In In: Stanton, Sharon M.; Christensen, Glenn A., comps. 2015. *Pushing boundaries: new directions in inventory techniques and applications: Forest Inventory and Analysis (FIA) symposium 2015*. 2015 December 8–10; Portland, Oregon. Gen. Tech. Rep. PNW-GTR-931. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 362. (Vol. 931).
- Fazeli Farsani, I., Farzaneh, M.R., Besalatpour, A.A., Salehi, M.H. and Faramarzi, M., 2019. Assessment of the impact of climate change on spatiotemporal variability of blue and green water resources under CMIP3 and CMIP5 models in a highly mountainous watershed. *Theoretical and Applied Climatology*. 136(1): 169-184.
- Hofste, R.W., Kuzma, S., Walker, S., Sutanudjaja, E.H., Bierkens, M.F., Kuijper, M., Sanchez, M.F., Van Beek, R., Wada, Y., Rodríguez, S.G. and Reig, P., 2019. *Aqueduct 3.0: Updated decision-relevant global water risk indicators*. World Resources Institute. 784.
- Hoekstra, A.Y. and Mekonnen, M.M. 2012. The water footprint of humanity. *Proceedings of the national academy of sciences*. 109(9): 3232-3237.
- Hoekstra, A.Y. and Chapagain, A.K. 2011. *Globalization of water: Sharing the planet's freshwater resources*. John Wiley & Sons.
- Hoekstra, A.Y. 2003. Virtual water trade: A quantification of virtual water flows between nations in relation to international crop trade. In *Proc. the International Expert Meeting on Virtual Water Trade 12*, Delft, 2003.
- Harshdeep, N. R. and Young, W. 2020. Disruptive technologies for improving water security in large river basins. *Water*. 12(10): 2783.
- Hindiyeh, M., Albatayneh, A., Altarawneh, R., Jaradat, M., Al-Omary, M., Abdelal, Q., ... and Jeguirim, M. 2021. Sea level rise mitigation by global sea water desalination using renewable-energy-powered plants. *Sustainability*. 13(17): 9552.
- Hoosain, M. S., Paul, B. S., Doorsamy, W. and Ramakrishna, S. 2023. The influence of circular economy and 4IR technologies on the climate–water–energy–food Nexus and the SDGs. *Water*, 15(4), 787.
- productivity, and economic indices. *Irrigation and Drainage*. 71(5): pp.1363-1378.
- Chapagain, A.K. and Orr, S. 2009. An improved water footprint methodology linking global consumption to local water resources: A case of Spanish tomatoes. *Journal of environmental management*, 90(2), pp.1219-1228.
- Choubin, B. and Malekian, A. 2017. Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. *Environ Earth Sci* 76 (15): 538 [online]
- Chowdhary, P., Raj, A. and Bharagava, R.N., 2018. Environmental pollution and health hazards from distillery wastewater and treatment approaches to combat the environmental threats: a review. *Chemosphere*. 194: 229-246.
- Chapagain, A.K., Hoekstra, A.Y., Savenije, H.H. and Gautam, R. 2006. The water footprint of cotton consumption: An assessment of the impact of worldwide consumption of cotton products on the water resources in the cotton producing countries. *Ecological economics*. 60(1): 186-203.
- Cheng, C. L. and Kawamura, S. 2023. Toward near-zero water consumption grade labelling and evaluation benchmarks for residential buildings. *Journal of Asian Architecture and Building Engineering*. 22(4): 2365-2375.
- Dogulu, N., López López, P., Solomatine, D.P., Weerts, A.H. and Shrestha, D.L., 2014. Estimation of predictive hydrologic uncertainty using quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrology & Earth System Sciences Discussions*. 11(9).
- Dogulu, N., López López, P., Solomatine, D.P., Weerts, A.H. and Shrestha, D.L. 2015. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrology and Earth System Sciences*. 19(7):3181-3201.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. 2020. A survey on ensemble learning. *Frontiers of Computer Science*. 14(2): 241-258. <https://10.1007/s11704-019-8208-z>.
- Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Hegde, R., Singh, S.K. and Lagacherie, P., 2020. Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. *Geoderma Regional*, 20, p.e00250.
- Duckett, D., Feliciano, D., Martin-Ortega, J. and Munoz-Rojas, J. 2016. Tackling wicked environmental problems: The discourse and its influence on praxis in Scotland. *Landscape and Urban Planning*. 154: pp.44-56.
- Elbeltagi, A., Deng, J., Wang, K. and Hong, Y., 2020a. Crop Water footprint estimation and modeling using an artificial neural network approach in the Nile Delta, Egypt. *Agricultural Water Management*, 235, p.106080.
- Elbeltagi, A., Deng, J., Wang, K., Malik, A. and

- Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms. *Water*. 11(10): 2126.
- Patidar, V.K., Wadhvani, R., Shukla, S., Gupta, M. and Gyanchandani, M., 2023, February. Quantile regression comprehensive in machine learning: a review. In 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-6). IEEE.
- Rahmati, O., Falah, F., Dayal, K.S., Deo, R.C., Mohammadi, F., Biggs, T., Moghaddam, D.D., Naghibi, S.A. and Bui, D.T., 2020. Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of Queensland Australia. *Science of the total environment*. 699:134230.
- Ruiz-Álvarez, M., Gomariz-Castillo, F. and Alonso-Sarria, F., 2021. Evapotranspiration response to climate change in semi-arid areas: Using random forest as multi-model ensemble method. *Water*. 13(2): 222.
- Tao, M., Zhang, T., Xie, X. and Liang, X., 2023. Water footprint modeling and forecasting of cassava based on different artificial intelligence algorithms in Guangxi, China. *Journal of Cleaner Production*. 382: 135238.
- Van Loon, A.F. and Van Lanen, H.A. 2013. Making the distinction between water scarcity and drought using an observation-modeling framework. *Water Resources Research*. 49(3): 1483-1502.
- Vaysse, K. and Lagacherie, P. 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma*. 291: 55-64.
- Wang, L., Kisi, O., Hu, B., Bilal, M., Zounemat-Kermani, M. and Li, H., 2017. Evaporation modelling using different machine learning techniques. *International Journal of Climatology*. 37: 1076-1092.
- Xu, K., Yang, D., Xu, X. and Lei, H., 2015. Copula based drought frequency analysis considering the spatio-temporal variability in Southwest China. *Journal of Hydrology*. 527: 630-640.
- Zamani, H., Pakdaman, Z., Shakari, M., Bazrafshan, O. and Jamshidi, S. 2025. Enhancing drought monitoring with a multivariate hydrometeorological index and machine learning-based prediction in the south of Iran. *Environmental Science and Pollution Research*. 32(9): 5605-5627.
- Zhang, S., Wang, Y., Zhang, Y., Wang, D. and Zhang, N. 2020. Load probability density forecasting by transforming and combining quantile forecasts. *Applied energy*. 277:115600.
- Koenker, R. and Hallock, K.F., 2001. Quantile regression. *Journal of economic perspectives*. 15(4): 143-156.
- Hofste, R.W., Kuzma, S., Walker, S., Sutanudjaja, E.H., Bierkens, M.F., Kuijper, M., Sanchez, M.F., Van Beek, R., Wada, Y., Rodríguez, S.G. and Reig, P. 2019. *Aqueduct 3.0: Updated decision-relevant global water risk indicators*. World Resources Institute. 784.
- Hoekstra, A.Y., 2008. *Water neutral: reducing and offsetting the impacts of water footprints*, Value of Water Research Report Series No. 28. Delft, Netherlands: UNESCO-IHE. Recuperado em, 10.
- Koenker, R. 2005. *Quantile regression (Vol. 38)*. Cambridge university press.
- Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E. and Bethel, W. 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. *Environmental Modelling & Software*. 144: 105139.
- Lenzen, M., Moran, D., Bhaduri, A., Kanemoto, K., Bekchanov, M., Geschke, A. and Foran, B., 2013. International trade of scarce water. *Ecological Economics*. 94: 78-85.
- Lovarelli, D., Bacenetti, J. and Fiala, M. 2016. Water Footprint of crop productions: A review. *Science of the total environment*. 548: 236-251.
- Luan, X.B., Yin, Y.L., Wu, P.T., Sun, S.K., Wang, Y.B., Gao, X.R. and Liu, J., 2018. An improved method for calculating the regional crop water footprint based on a hydrological process analysis. *Hydrology and Earth System Sciences*. 22(10):5111-5123.
- Mekonnen, M.M. and Gerbens-Leenes, W. 2020. The water footprint of global food production. *Water*. 12(10): 2696.
- Meinshausen, N. and Ridgeway, G., 2006. Quantile regression forests. *Journal of machine learning research*, 7(6). Mahato, A., Upadhyay, S., & Sharma, D. (2022). Global water scarcity due to climate change and its conservation strategies with special reference to India: A review. *Plant Archives (09725210)*, 22(1).
- Pellicer-Martínez, F. and Martínez-Paz, J.M. 2016. The Water Footprint as an indicator of environmental sustainability in water use at the river basin level. *Science of the Total Environment*. 571: 561-574.
- Papacharalampous, G., Tyralis, H., Langousis, A., Jayawardena, A.W., Sivakumar, B., Mamassis, N., Montanari, A. and Koutsoyiannis, D., 2019.

## Water Footprint Modeling Using Machine Learning-Based Quantile Regression with Uncertainty Analysis

M. Yahyazade<sup>1</sup>, O. Bazrafshan<sup>2\*</sup>, Arash Malekian<sup>3</sup>, Hossein Zamani<sup>4</sup>

Received: Jul.19, 2025

Accepted: Sep.19, 2025

### Abstract

This study aimed to model the green and blue water footprint of tomato cultivation in Bushehr province using hybrid machine learning models (QXGBoost, QRF, XGBoost, and RF) with uncertainty analysis. Results revealed that blue water footprint accounted for the majority (80%) of water consumption, while green water footprint constituted only 20%, indicating heavy reliance on water resources and climate vulnerability. Correlation analysis showed precipitation (P) and effective precipitation (Peff) had the strongest positive impact on green water footprint, whereas temperature and evapotranspiration (ETc) showed negative effects. For blue water footprint, ETc and Tmax exhibited the highest positive correlation. The QXGBoost model demonstrated optimal performance with a determination coefficient (Rs<sub>q</sub>=0.95) and low error (RMSE=0.16). Uncertainty evaluation using UNEEC method confirmed QXGBoost's superiority with prediction interval coverage (PICP=0.93) and symmetric confidence intervals. The study proposes water footprint reduction strategies including improved irrigation (subsurface/smart systems), drought-resistant cultivars, and soil moisture management (mulching). Findings emphasize transitioning to sustainable farming systems through advanced modeling and continuous water resource monitoring.

**Keywords:** Agricultural water management, Machine learning, Tomato Quantile regression, Water footprint

1 -Ph. D Student, Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandarabbas, Iran

2- Department of Natural Resources Engineering, Faculty of Agricultural and Natural Resources Engineering, University of Hormozgan, Bandarabbas, Iran

3- Department of Reclamation of Arid and Mountainous Regions, University of Tehran, Karaj, Iran

4- Department of Mathematics and Statistics, Faculty of Science, University of Hormozgan, Bandarabbas, Iran

(\*-Corresponding Author Email: o.bazrafshan@hormozgan.ac.ir)