



Application of Data Mining Methods in Estimation of Hazelnut Yield in Orchards Equipped with Pressurized Irrigation in Gilan Province

Masoumeh Yeganeh^{*1}, Naeim Adoos², Afshin Ashraf Zadeh³

1. Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
2. Department of Civil Engineering, Faculty of Civil, Islamic Azad University Rasht, Rasht, Iran
3. Department of Water Engineering, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

Received: 07 March 2024/ **Revised:** 25 April 2024/ **Accepted:** 06 June 2024

<https://doi.org/10.22034/arwe.2024.2014876.1015>

Abstract

Kiwi is one of the Iran's agricultural products, which is mainly exported to the world market. Data mining methods are well able to provide manufacturers with the necessary information in the field of product performance modeling. This research has investigated the efficiency of data mining methods of advanced Feedforward backdrop Neural Network, K-Nearest Neighbor, Genetic Programming and Multivariate Linear Regression in estimating kiwi yield in Gilan province using water and soil characteristics. 74 data series were obtained from the field measurement of water and soil information and yield of kiwi fruit in orchards equipped with pressure irrigation system in 2021-2022. Water and soil data including maximum daily evaporation and transpiration, soil electrical conductivity and soil reaction index, clay percentage, silt percentage, soil organic matter percentage, water electrical conductivity and water reaction index, and irrigation volume as model inputs and crop yield, model output selected. The results showed that the leading artificial neural network model has a better performance than the other three models due to the higher explanation coefficient statistics (0.96) and lower root mean square error (0.019). Also, Genetic Programming has correlation coefficient (0.89), root mean square error (0.033) and K-Nearest Neighbor method has correlation coefficient (0.88) root mean square error (0.059) and Multivariate Linear Regression method has coefficient The correlation was (0.58) and the root mean square error (0.093) which indicates the higher accuracy of the Genetic Programming method. Therefore, the advanced artificial neural network model can act as a powerful tool in the estimation of Kiwi performance.

Keywords: Genetic Programming, Multivariate Linear Regression, Feed Forward Backdrop Neural Network, K-Nearest Neighbor.



کاربرد روش‌های داده‌کاوی در تخمین عملکرد کیوی در باغات مجهز به آبیاری تحت فشار در استان گیلان

معصومه یگانه^{۱*}، نعیم آدوس^۲، افشین اشرف‌زاده^۳

۱. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران

۲. گروه عمران، دانشکده عمران، دانشگاه آزاد اسلامی واحد رشت، رشت، ایران

۳. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه گیلان، رشت، ایران

دریافت: ۱۷ اسفند ماه ۱۴۰۲ / اصلاحات: ۰۶ اردیبهشت ماه ۱۴۰۳ / پذیرش: ۱۷ خرداد ماه ۱۴۰۳

<https://doi.org/10.22034/arwe.2024.2014876.1015>

چکیده

کیوی یکی از محصولات کشاورزی ایران است که به صورت عمده در بازار جهانی صادر می‌شود. روش‌های داده‌کاوی به خوبی قادرند تا در زمینه مدل‌سازی عملکرد محصول اطلاعات لازم را در اختیار تولیدکنندگان قرار دهد. این تحقیق به بررسی کارآیی روش‌های داده‌کاوی شبکه عصبی مصنوعی پیشرو، K-نزدیک‌ترین همسایگی، برنامه‌ریزی ژنتیک و رگرسیون خطی چند متغیره در برآورد عملکرد کیوی در استان گیلان با استفاده از خصوصیات آب و خاک پرداخته است. ۷۴ سری داده از اندازه‌گیری میدانی اطلاعات آب و خاک و عملکرد محصول کیوی باغ‌های مجهز به سیستم آبیاری تحت فشار در سال ۱۴۰۰-۱۴۰۱ به دست آمد. داده‌های آب و خاک شامل حداکثر تبخیر و تعرق روزانه، هدایت الکتریکی خاک و شاخص واکنش خاک، درصد رس، درصد سیلت، درصد ماده آلی خاک، هدایت الکتریکی آب و شاخص واکنش آب و حجم آبیاری به عنوان ورودی‌های مدل و میزان عملکرد محصول، خروجی مدل انتخاب گردید. نتایج نشان داد که مدل شبکه عصبی مصنوعی پیشرو به دلیل آماره‌های ضریب تبیین بیشتر (۰/۹۶) و جذر میانگین مربعات خطا کمتر (۰/۱۹) عملکرد بهتری نسبت به سه مدل دیگر دارد. همچنین برنامه‌ریزی ژنتیک دارای ضریب همبستگی (۰/۸۹)، جذر میانگین مربعات خطا (۰/۳۳) و روش K-نزدیک‌ترین همسایگی دارای ضریب همبستگی (۰/۸۸)، جذر میانگین مربعات خطا (۰/۵۹) و روش رگرسیون خطی چند متغیره دارای ضریب همبستگی (۰/۵۸) و جذر میانگین مربعات خطا (۰/۹۳) بوده که حاکی از دقت بالاتر روش برنامه‌ریزی ژنتیک است. بنابراین، مدل شبکه عصبی مصنوعی پیشرو می‌تواند به عنوان یک ابزار قدرتمند در تخمین عملکرد کیوی عمل نماید.

کلمات کلیدی: برنامه‌ریزی ژنتیک، رگرسیون خطی چند متغیره، شبکه عصبی پیشرو، K-نزدیک‌ترین همسایگی.

مقدمه

گیوی به‌عنوان یک محصول تازه تجاری شده، در مقایسه با سایر محصولات کشاورزی که در شرایط آب و هوایی مشابه آن در کشور کشت می‌شود، بازده اقتصادی بالایی دارد. با توجه به شرایط مورد نیاز گیوی، شرایط آب و هوایی و خاکی مناسب برای کشت این محصول، دو استان گیلان و مازندران مراکز تولید گیوی در کشور می‌باشند. مهم‌ترین مسئله برای کشاورزان، کسب درآمد و نیل به حداکثر محصول است. این موضوع، نیازمند درک کامل روابط بین مقدار عملکرد با ویژگی‌های باغات می‌باشد. برآورد عملکرد محصول به‌عنوان یک ویژگی دیرپافت از روی ویژگی‌های زودپافت (مانند ویژگی‌های فیزیکی و شیمیایی خاک و آب) از اهمیت فراوانی برخوردار می‌باشد. یکی از استفاده‌های مدل و مدلسازی در مطالعات ارزیابی علوم کشاورزی در پیش‌بینی مقدار محصول است. از روشهای نوین برای پیش‌بینی عملکرد محصول استفاده از روشهای داده‌کاوی از جمله رگرسیون چند متغیره، روش K-نزدیک‌ترین همسایه، شبکه‌های عصبی مصنوعی، و برنامه‌ریزی ژنتیک می‌باشد. یکی از رایج‌ترین و ساده‌ترین روشها در پیش‌بینی مقدار محصول مدل‌های رگرسیونی است که با استفاده از ویژگی‌های اراضی، تخمین تولید محصول از طریق آنها میسر می‌گردد. در مدل‌های رگرسیونی چند متغیره^۱، تنها ورودی‌ها و خروجی‌ها مورد بررسی قرار می‌گیرند و بدون اینکه روابط ریاضی بین متغیرهای ورودی و خروجی را مورد توجه قرار دهند و تنها مقدار خروجی را برآورد می‌کنند که این موضوع، امکان تفسیر و بررسی میزان تأثیر هر یک از متغیرهای ورودی و خروجی را برای کاربر مشکل می‌سازد (Bahmani, 2020). در برخی پژوهش‌هایی که با روش‌های رگرسیون چند متغیره به تخمین عملکرد پرداخته‌اند، نتایج کارآیی ضعیف این مدل‌ها را نشان می‌دهد (Tashakkori, 2021). در پژوهش ایوبی و همکاران (Ayoubi et al., 2009) رابطه بین عملکرد دانه جو را با ویژگی‌های خاک از طریق به‌کارگیری رگرسیون چند متغیره بررسی نمودند، بیان نمودند روابط رگرسیونی می‌تواند ۸۰ درصد از عملکرد دانه را توجیه کنند. یکی دیگر از روشها برای پیش‌بینی تولید محصول، مدل شبکه عصبی مصنوعی^۲ می‌باشد. این مدل مجموعه‌ای از واحدهای متصل به نام نورون‌های مصنوعی است و هر نورون می‌تواند سیگنالی را به نورون‌های دیگر منتقل کند (Yue et al., 2018). این مدل‌ها بیانگر یک سامانه غیر خطی است که بدون داشتن اطلاعات کامل از نحوه ارتباط پارامترها و قوانین حاکم بین آنها سامانه‌های مختلف را شبیه‌سازی می‌کند. حافظی و همکاران (Hafezi et al., 2021) در پیش‌بینی عملکرد نیشکر از مدل شبکه عصبی استفاده نمودند، نتایج حاکی از عملکرد دقیق این مدل‌ها می‌باشد. جعفری و همکاران (Jafari et al., 2023) نشان دادند که مدل شبکه عصبی بر مبنای خصوصیات هواشناسی قادر به پیش‌بینی عملکرد گندم دیم با دقت مناسب می‌باشد.

امام قلی‌زاده و همکاران (Emamgholizadeh et al., 2015) با مطالعه بر تخمین عملکرد کنگد بیان نمودند که مدل شبکه عصبی کارآیی بیشتری نسبت به رگرسیون چندگانه دارد. طی پژوهشی که دهمرده و همکاران (Dahmardeh et al., 2021) در تخمین عملکرد گندم در سد سیستان با استفاده از برخی خصوصیات خاک در مقیاس مزرعه انجام دادند دریافتند که استفاده از این روش می‌تواند به یک ابزار قدرتمند در زمینه تخمین عملکرد محصول باشد. مدل شبکه عصبی پیشنهادی امامی و همکاران (Emami et al., 2021) برای عملکرد محصول شلیل مغان برآزش خوبی با داده‌های واقعی داشت. احمدی و همکاران (Ahmadi et al., 2023) در شبیه‌سازی عملکرد گندم دیم با استفاده از شاخص‌های خشکسالی و مدل شبکه عصبی نشان داد نتایج آنها از کارآیی بهتر، ضریب دقت بیشتر و خطای کمتری نسبت به سایر مدل‌ها است. تحقیقات تشکری و همکاران (Tashakkori et al., 2021) نشان داد که مدل شبکه عصبی بر پایه ویژگی‌های خاک قادر به پیش‌بینی عملکرد گل زعفران با کمترین خطا می‌باشد. یکی از پرکاربردترین فرآیند آموزشی شبکه عصبی مصنوعی، روش پیشرو با پس‌انتشار خطا^۳ است، در این روش یادگیری الگو در طول آموزش روی می‌دهد که موجب افزایش دقت مدل و کاهش خطا می‌شود. (Sivaneasan et al., 2017) تابش خورشیدی را با پیش‌پردازش منطق فازی با روش پیشرو با پس‌انتشار خطا مدلسازی کرده‌اند، نتایج نشان داد که میزان خطا با استفاده از روش پس‌انتشار کاهش می‌یابد (Moodi et al., 2022). در تخمین مقاومت فشاری با شبکه عصبی پیشرو با پس‌انتشار خطا دریافتند شبکه پس‌انتشار خطا باعث کاهش خطا نسبت به مدل‌های تحلیلی می‌شود. از دیگر روشهای متداول برای پیش‌بینی تولید محصول روش K-نزدیک‌ترین همسایه^۴ می‌باشد. این

1. Multiple Linear Regression (MLR)
 2. Artificial neural network (ANN)
 3. Feed Forward Backprop (FFB)
 4. K-Nearest Neighbors (KNN)

روش مزایای دارد از جمله: سادگی، کارایی، شهودی و دارای عملکرد طبقه‌بندی رقابتی، دقت مناسب (Zheng, 2017). در پیش‌بینی عملکرد گندم توسط انصاری (Ansari, 2023) با روش‌های فرآکوشی در استان البرز با استفاده از داده‌های سطح زیر کشت و تولید سالانه، روش K-نزدیکترین همسایه نسبت به روشهای برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان بهترین دقت را نشان داد. سلاجقه و همکاران (Salajegheh et al., 2010) با روش K-نزدیکترین همسایه مقادیر دبی را در رودخانه کرج را پیش‌بینی کردند. نتایج نشان داد که این عملکرد مناسبی در تخمین دبی دارد. اسلامی و همکاران (Eslami et al., 2021) در بررسی احتمال وقوع آتش‌سوزی در جنگل‌های حوزه بابلرود با روشهای K-نزدیکترین همسایه و مدل شبکه عصبی نشان داد که روش شبکه عصبی مصنوعی از دقت بالاتری برخوردار است. ستاری و همکاران (Satari et al., 2015) با استفاده از روش نزدیکترین همسایگی به تخمین میزان هدایت الکتریکی و کل جامدات محلول پرداخته و کارایی بالای این روش را اعلام کردند. مدرسی و همکاران (Modaresi et al., 2022) با مقایسه شبیه‌سازی جریان ماهانه ورودی با خصوصیات دما و بارش در سد کرخه نشان دادند که روش K-نزدیکترین همسایه نسبت به مدل شبکه عصبی دقت پایین‌تری دارد اما در تعیین میانگین سالانه بارش توسط انارکی و همکاران (Anaraki et al., 2020) مدل شبکه عصبی نسبت به روش K-نزدیکترین همسایه دارای دقت بیشتری بود.

از دیگر روشهای در پیش‌بینی تولید محصول، روش برنامه‌ریزی ژنتیک^۱ است. برنامه‌ریزی ژنتیک شاخه‌ای از الگوریتم‌های تکاملی است که توانایی مدلسازی فرآیندهای کاملاً غیرخطی و پویا را دارد. در برنامه‌ریزی ژنتیک، افراد همانند شاخه‌های مجزا هستند. همچنین بر ساختار درختی مجموعه‌ها تأکید می‌شود. این روش به دلیل دارا بودن دقت کافی، از کاربرد بیشتری برخوردار است. احمدی و همکاران (Ahmadi et al., 2015) در تخمین تبخیر تعرق ماهانه با مدل برنامه‌ریزی ژنتیک نشان دادند این مدل دقت بالا و خطای کمی دارد. پرویز و همکاران (Parviz et al., 2022) در برآورد توزیع مکانی شاخص واکش خاک نشان دادند برنامه‌ریزی ژنتیک و رگرسیون خطی به ترتیب، کمترین و بیشترین مقدار خطا را دارند. باباعلی و دهقانی (Baba Ali and Deghani, 2012) کارایی دو روش برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی را در برآورد بارش ماهانه برای حوضه کاکارضا بررسی کردند و اظهارداشتند که مدل ژنتیک، توانایی بالایی در تخمین بارش ماهانه دارد (Zahiri and Azamathulla, 2014) برای پیش‌بینی دبی جریان در مقاطع مرکب از روش برنامه‌ریزی ژنتیک خطی استفاده کردند. نتایج این تحقیق نشان داد دقت روش برنامه‌ریزی ژنتیک خطی بالاتر از مدل درختی می‌باشد.

با توجه به اهمیت موضوع تخمین عملکرد محصول، این تحقیق با هدف بررسی عملکرد روشهای شبکه عصبی مصنوعی پیشرو، برنامه‌ریزی ژنتیک روش K-نزدیک‌ترین همسایگی و مدل‌های رگرسیونی چند متغیره برای باغات کیوی در استان گیلان مجهز به سیستم آبیاری قطره‌ای انجام می‌شود.

مواد و روشها

منطقه مطالعاتی و داده‌های مورد استفاده

در این پژوهش باغات کیوی در آستانه‌اشرفیه، املش، انزلی، تالش، رشت، رضوانشهر، رودسر، فومن، لنگرود و ماسال (مناطق اصلی تولیدکننده کیوی استان گیلان) انتخاب گردید. موقعیت جغرافیایی مناطق مورد مطالعه در شکل ۱ نشان داده شده است. اطلاعات آب و خاک و عملکرد کیوی در باغات تحت مدیریت بهره‌برداران (تعداد ۷۴ باغ) بصورت میدانی برداشت شده‌اند. خاکهای تحت کشت کیوی در شمال ایران با زهکشی مناسب و بافت لومی شنی هستند که کودپذیری بالایی دارند. وجود چنین شرایطی اجازه توسعه عمقی و جانبی مناسبی را به ریشه‌های کیوی در عمق ۳۰ سانتی‌متری خاک گسترش می‌یابد. با توجه به عمق ریشه درخت کیوی و به ازای یک هکتار باغ، سه نمونه خاک از عمق ۰-۳۰ سانتی‌متری برداشت شد. خصوصیات فیزیکی و شیمیایی خاک شامل درصد ماده آلی (%OC)، درصد سیلت (%Silt)، درصد رس (%Clay)، شاخص واکنش خاک (pHsoil)، هدایت الکتریکی خاک (ECe) اندازه‌گیری شد. نمونه‌های خاک پس از انتقال به آزمایشگاه و خشک شدن در مجاورت هوا با چکش چوبی کوبیده و از الک ۲ میلیمتری عبور داده شدند.



شکل ۱. موقعیت مناطق مورد مطالعه
 Fig. 1. The location of the studied areas

در نمونه‌های خاک مقادیر در عصاره گل اشباع (pHsoil و ECe)، درصد ذرات معدنی خاک (رس و سیلت) به روش هیدرومتری با قرائت چهار زمانه اندازه‌گیری شدند. جهت تعیین خصوصیات کیفی آب بعد از روشن شدن پمپ چاه‌ها در زمان آبیاری باغات به مدت ۱۵ دقیقه، از هر باغ ۳ نمونه به حجم یک لیتر در فاصله زمانی یک ساعت برداشت گردید و در کمتر از ۱۲ ساعت به آزمایشگاه کیفیت آب برای تعیین خصوصیات کیفی آب انتقال داده شد. خصوصیات آب شامل شاخص واکنش آب (pHwater)، هدایت الکتریکی (ECw) اندازه‌گیری شد. همچنین با روشن کردن سیستم آبیاری برای یک دور آبیاری، حجم آب مصرفی (V) و زمان آبیاری (H) نیز اندازه‌گیری شد. حداکثر تبخیر و تعرق روزانه (ET) از داده‌های هواشناسی هر شهرستان تعیین شد. مقدار عملکرد کیوی در هکتار برحسب کیلوگرم در پایان فصل برداشت سال ۱۴۰۰-۱۴۰۱ تعیین شد. از ۷۴ باغ کیوی، ۷۴ سری داده از پارامترهای آب و خاک و عملکرد باغات با هماهنگی سازمان جهاد کشاورزی استان گیلان، برای باغات کیوی مجهز به سیستم آبیاری تحت فشار اطلاعات به دست آمد.

نرمال‌سازی داده‌ها

برای جلوگیری از تأثیر نامطلوب مقیاس متفاوت متغیرها بر مدل‌ها، ضروری است داده‌ها از طریق پیش‌پردازش تصحیح شوند. هدف از نرمال‌سازی آن است که داده‌ها به اعدادی بین صفر و یک تبدیل شوند. برای ورودی‌های نزدیک به صفر و یک، تغییرات وزن نورونها حداقل خواهد بود؛ زیرا در این اعداد عناصر پردازشگر (نورونها) به دلیل شکل توابع آستانه نظیر تابع سیگموئیدی کند عمل می‌کنند ولی برای مقادیر ورودی‌های نزدیک به ۰/۵ پاسخ نورونها به سیگنال ورودی سریعتر خواهد بود (Kumar et al., 2002). نرمال‌سازی داده‌ها به نحوی صورت گرفت که میانگین سری داده‌ها برابر ۰/۵ گردد. بدین منظور از رابطه ۱ برای نرمال‌سازی استفاده شد (Hafezi et al., 2022).

$$X_s = \left(\frac{X_i - X_{mean}}{X_{max} - X_{min}} \right) \times 0.5 + 0.5 \quad (1)$$

در رابطه ۱، XS: مقدار استاندارد شده، Xi: مقدار واقعی، Xmean: متوسط مقادیر، Xmax: مقدار حداکثر، Xmin: مقدار حداقل می‌باشد.

مدل برنامه‌ریزی ژنتیک

برنامه‌ریزی ژنتیک (GP) تعمیم‌یافته الگوریتم ژنتیک می‌باشد. به این ترتیب که جمعیتی در جهت تکامل به صورت انتخابی، جمعیت نامناسب را رها کرده و فرزندان اصلاح‌شده ایجاد می‌کنند. برنامه‌ریزی ژنتیک یک روش برنامه‌ریزی خودکار می‌باشد که راه‌حل مسئله را با استفاده از برنامه کامپیوتری ارائه می‌کند. این روش جدیدترین شیوه از بین روش‌های الگوریتم تکاملی می‌باشد که به دلیل دارا دقت بودن کافی، از کاربرد بیشتری برخوردار است. در این روش در ابتدای فرآیند هیچ‌گونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مؤلفه‌های آن می‌باشد. برنامه‌ریزی ژنتیک برخلاف الگوریتم ژنتیک روی ساختار درختی فرمول‌ها به جای سلسله ارقام دودویی عمل می‌کند.

ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورداستفاده در روابط ریاضی) و متغیرهای مسئله و اعداد ثابت ایجاد می‌شوند. قبل از مراحل اجرایی برنامه‌ریزی ژنتیک گام‌های مقدماتی زیر تعیین شوند (Soltani et al., 2011).

۱- یک جمعیت اولیه از توابع مرکب نشان‌دهنده الگوهای پیش‌بینی، به صورت تصادفی در نظر گرفته می‌شود.

۲- معرفی جمعیت اولیه به رایانه و ارزیابی هر یک از افراد جمعیت مذکور با استفاده از توابع برازش.

۳- انتخاب ژن‌های مؤثر به منظور تکثیر، جهش و تولیدمثل افراد جدید با صفات اصلاح‌شده.

۴- اعمال فرآیند توسعه‌ای تکراری بر روی فرزندان در هر تولید. این گام به تعداد معین و تا حصول بهترین پاسخ تکرار خواهد شد.

۵- معیار پایان و ارائه نتایج مانند تعداد تولید جمعیت جدید، تعیین یک مقدار مشخص برای برازش فرمول‌ها

در تحقیق حاضر جهت مدلسازی الگوریتم برنامه‌ریزی ژنتیک از Eureqa pro 1.24.0 استفاده شده است.

روش K-نزدیکترین همسایه

بر خلاف توابع انتقالی کلاسیک، روش نزدیکترین همسایگی (KNN) از هیچ تابع ریاضی از پیش تعریف شده‌ای برای تخمین متغیرهای مختلف استفاده نمی‌کند. این روش یک طبقه‌بندی یادگیری با ناظر می‌باشد که نمونه‌ها را بر اساس شباهت و فاصله از نمونه‌های آموزشی طبقه‌بندی می‌کند که در طبقه‌بندی و پیش‌بینی استفاده می‌شود. بر اساس این روش می‌توان نمونه جدیدی را بر اساس بیشترین گروه که نزدیکترین همسایگی را با آن نمونه داشته باشند، تقسیم کرد. می‌توان گفت که این روش، k تعداد از نمونه مشابه را پیدا می‌کند و بر اساس آنها ارزش نمونه مطالعه‌شده را پیش‌بینی می‌کند. نخستین گام در استفاده از این مدل، یافتن رابطه‌ای برای محاسبه فاصله بین داده‌های مورد آزمایش و داده‌های تعلیم است. برای تعیین این فاصله از فاصله اقلیدسی استفاده می‌شود:

$$d(X, Y) = \sqrt{\sum_{i=1}^N (X_i - Y_i)^2} \quad (2)$$

که در آن X نماینده داده‌های تعلیم با پارامترهای مشخص (Xi تا Xn) و y نماینده داده‌های آموزش با همان تعداد پارامترهای مشخص (y1 تا yn) است (Jagtap et al., 2004).

$$X = (x_1, x_2, \dots, x_n) \quad (3)$$

$$Y = (y_1, y_2, \dots, y_n) \quad (4)$$

پس از تعیین فاصله اقلیدسی بین داده‌ها، نمونه‌های بانک داده به ترتیب صعودی از کمترین فاصله (حداکثر تشابه) تا بیشترین فاصله از نمونه طبقه‌بندی و ارزش‌گذاری می‌شوند. سپس تعداد نقاطی (k) از داده‌های آزمایش برای تخمین ویژگی‌های بانک داده موردنظر یافت می‌شود. تعیین تعداد همسایه‌ها (k)، مهم‌ترین مرحله به‌شمار می‌آید (Xindung and Kumar, 2009). برای مدلسازی با روش K-نزدیکترین همسایه از نرم‌افزار Statistica 13.5.0.17 استفاده گردید.

مدل شبکه عصبی پیشرو

شبکه عصبی مصنوعی دارای لایه‌های ورودی، پنهان (میانی) و خروجی است. (Brown et al., 2008). در این پژوهش از شبکه عصبی شبکه عصبی پیشرو جهت مدلسازی استفاده گردید. پیشرو بودن به این معناست که نورون‌های عصبی در لایه‌های متوالی قرار گرفته‌اند و خروجی خود را به جلو می‌فرستند (Lokus 2015). اتصال نورونها در این شبکه به این صورت است که تمام نورون‌های لایه ورودی به

نورون‌های لایه پنهان و تمام نورون‌های لایه پنهان به نورون‌های لایه خروجی متصل است. برای یادگیری شبکه از روش پس‌انتشار خطا استفاده می‌شود که طی این روش نورون‌ها وزن خود را برای به‌دست آوردن دانش جدید انطباق می‌دهند (Badde et al., 2013). یادگیری در شبکه عصبی پیشرو با پس‌انتشار خطا در طول مرحله آموزش روی می‌دهد که در آن هر الگوی ورودی از مجموعه آموزش یادگیری در لایه ورودی اعمال می‌شود و بعد به جلو انتشار می‌یابد یک شبکه پس‌انتشار با دارا بودن انحراف^۱، یک لایه سیگموئید و یک لایه خروجی خطی توانایی تخمین زدن هر تابعی با تعداد نقاط ناپیوستگی محدود را دارا است.

روش پس‌انتشار استاندارد یک روش با کاهش شیب^۲ می‌باشد که در آن وزن‌های شبکه در جهت خلاف شیب تابع کارایی^۳ حرکت می‌کنند. روش گرادیان توأم^۴ از الگوریتم‌های سریع پس‌انتشار از تکنیک‌های استاندارد بهینه‌سازی عددی می‌باشد. در این الگوریتم اندازه هر گام به ازای اندازه هر تکرار تنظیم می‌شود به گونه‌ای یک عملیات جست‌وجو بین تمامی شیب‌های توأم انجام می‌شود تا مناسب‌ترین جهت در راستای حداقل‌سازی تابع کارایی تعیین شود. این الگوریتم‌ها با جست‌وجو در جهت بیشترین شیب در اولین تکرار آغاز می‌شوند. پس یک جست‌وجوی خطی برای تعیین فاصله بهینه با جست‌وجوی خطی انجام می‌گیرد. جست‌وجوی جدید تعیین شده با جهت قبلی توأم می‌شود. رویه عمومی برای تعیین جهت جست‌وجوی جدید ترکیبی از جهت جست‌وجوی جدید و جهت جست‌وجوی قبلی است. این الگوریتم‌ها بسیار سریع هستند. جهت مدل‌سازی با شبکه عصبی پیشرو با پس‌انتشار خطا (FFB) از نرم افزار MathworkMatlab2023 استفاده گردید.

مدلسازی رگرسیونی

هدف کلی مدل‌سازی رگرسیون خطی چند متغیره (MLR)، پیدا کردن رابطه بین چند متغیر مستقل و یک متغیر وابسته است. به منظور انجام (MLR) به روش گام به گام، نرم‌افزار IBM SPSS Statistica26 مورد استفاده قرار گرفت

ارزیابی مدل‌ها

در این پژوهش از ۷۴ سری داده جمع‌آوری شده اطلاعات آب و خاک و عملکرد محصول از باغات کیوی، ۵۹ سری داده به منظور واسنجی مدل (۸۰ درصد داده‌ها) و ۱۴ سری داده (۲۰ درصد داده‌ها) برای صحت‌سنجی مدل استفاده شد. برای انجام مدل‌سازی، یک متغیر شامل عملکرد محصول به عنوان متغیر وابسته در خروجی مدل‌ها و نه متغیر شامل متغیرهای درصد ماده آلی، درصد سیلت، درصد رس، شاخص واکنش خاک، هدایت الکتریکی خاک، شاخص واکنش آب، هدایت الکتریکی آب، حجم آب مصرفی، زمان آبیاری و حداکثر تبخیر و تعرق به عنوان متغیرهای مستقل در ورودی مدل‌ها مورد استفاده قرار گرفت. برای ارزیابی و مقایسه کارایی مدل‌های به کار رفته در این تحقیق از دو شاخص جذر میانگین مربعات خطا (RMSE) و ضریب تبیین (R^2) استفاده شد که به ترتیب در روابط ۵ و ۶ ارائه شده‌اند:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (5)$$

$$R^2 = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad (6)$$

در روابط فوق n تعداد داده‌ها، O_i و P_i ، به ترتیب مقادیر اندازه‌گیری شده و پیش‌بینی شده عملکرد هستند. هر چه شاخص RMSE به صفر نزدیکتر باشد، نشان‌دهنده خطای کمتر و کارایی بهتر مدل می‌باشد. مقدار شاخص ضریب تبیین در دامنه منفی یک تا یک تغییر می‌کند. مقدار یک و منفی یک به ترتیب نشان‌دهنده همبستگی کامل و معکوس مقادیر مشاهداتی و شبیه‌سازی شده می‌باشد. هر چه مقدار این شاخص به یک نزدیکتر باشد، کارایی مدل بهتر است.

نتایج و بحث

توصیف آماری داده‌ها

خلاصه آماری ویژگی‌های آب و خاک و عملکرد محصول کیوی در جدول ۱ آورده شده است. با توجه به نتایج ضریب همبستگی پیرسون مشاهده می‌گردد متغیرهای حداکثر تبخیر و تعرق روزانه، هدایت الکتریکی خاک، شاخص واکنش خاک، درصد سیلت، درصد رس، شاخص واکنش آب و حجم آب موردنیاز همبستگی مثبت با عملکرد محصول نشان دادند. این رابطه بدین مفهوم است که با افزایش مقدار هریک از این متغیرها عملکرد محصول افزایش می‌یابد. هرچه ضریب همبستگی در بازه ۰ تا ۱ به عدد ۱ نزدیک‌تر باشد همبستگی قوی‌تر است. شاخص واکنش خاک بیشترین ضریب همبستگی مثبت را دارد. درصد ماده آلی خاک، هدایت الکتریکی آب، زمان آبیاری همبستگی منفی با عملکرد محصول نشان دادند بدین مفهوم است که با کاهش مقدار هریک از این متغیرها عملکرد محصول کاهش می‌یابد. هرچه ضریب همبستگی به عدد -۱ نزدیک‌تر باشد همبستگی قوی‌تر است. بنابراین هدایت الکتریکی آب بیشترین ضریب همبستگی منفی را دارد.

جدول ۱. خلاصه‌ای از وضعیت آماری ویژگی‌های مورد بررسی در منطقه مطالعاتی

Table 1. A summary of the statistical status of the studied features in the study area

Y (ton/ha)	V (m3)	H (hr)	pHwater	ECw (dsm-1)	OC %	Clay %	Silt %	pHsoil	ECe (dsm-1)	ET (mm/day)	ویژگی
21.65	6031	2.58	7.14	0.6	1.69	19.9	28.3	6.76	0.59	3.84	میانگین
-0.13	1	0.42	-2.04	0.19	-0.25	-0.19	-2.29	-0.54	0.79	0.44	چولگی
-0.92	-1	-0.7	9.03	0.43	1.01	-1.51	5.71	-1.2	1.54	-1.2	کشیدگی
19	5300	1	5.84	0.14	0.98	6.9	7	5.12	0.29	2.9	حداقل
24	7300	4	7.5	0.98	2.42	31	37	7.54	1.07	4.72	حداکثر
11	13	24	10	32	16	40	24	12	26	16	ضریب تغییرات
	0.11	-0.08	0.03	-0.15	-0.02	0.11	0.13	0.21	0.37	0.18	همبستگی پیرسون

تمامی ویژگی‌هایی که دارای ضریب چولگی بین -۱ تا ۱ باشند، از توزیع نرمال تبعیت می‌کنند. براساس ضریب چولگی به‌دست آمده، تمام ویژگی‌های محصول کیوی به جز هدایت الکتریکی آب و درصد سیلت از توزیع نرمال برخوردار می‌باشند. نمونه‌های خاک برداشت شده دارای شاخص واکنش خاک متوسط ۶/۷۶ می‌باشند. هدایت الکتریکی خاکها به طور متوسط در حدود ۰/۵۹ دسی زیمنس بر متر می‌باشد که حاکی از پائین بودن مقادیر شوری آنها می‌باشد.

میانگین درصد رس، درصد سیلت و ماده آلی خاک به ترتیب برابر ۲۸/۳۴ درصد، ۱۹/۹۴ درصد و ۱/۶۴ درصد می‌باشد. هم‌چنین برای نمونه‌های آب برداشت شده شاخص واکنش آب متوسط ۷/۱۴ می‌باشند. هدایت الکتریکی آب به طور متوسط در حدود ۰/۶ دسی زیمنس بر متر می‌باشد.

میانگین حجم آب موردنیاز، ساعت آبیاری و حداکثر تبخیر و تعرق روزانه در باغات مورد بررسی به ترتیب برابر ۶۰۳۱ مترمکعب در هکتار، ۲/۵۸ ساعت و ۳/۸۴ میلی‌متر بر روز می‌باشد. کمترین ضریب تغییرات مربوط به ویژگی هدایت الکتریکی آب خاک برابر ۱۰ درصد می‌باشد و بیشترین ضریب تغییرات مربوط به ویژگی درصد رس خاک برابر با ۴۰ درصد می‌باشد.

در پژوهش گمرکچی و همکاران (Gomrokchi et al., 2022) حجم آب مصرفی بیشترین همبستگی با عملکرد گندم را داشت. نتایج تحقیق دهمرده و همکاران (Dehmardeh et al., 2021) نشان داد رس، ماده آلی و سیلت خاک دارای بیشترین و واکنش خاک کمترین اثر بر عملکرد گندم را دارند. ککس و همکاران (Cox et al., 2006) نیز در مورد شاخص واکنش خاک، کمترین ضریب تغییرات را در بین ویژگی‌های خاک به دست آورده‌اند. بهمنی و همکاران (Bahmani et al., 2022) بیشترین ضریب تغییرات را برای هدایت الکتریکی خاک به‌دست آوردند. مقدار ضریب تغییرات برای تمامی ویژگی‌های به جز درصد رس کمتر از ۳۵ درصد و دارای تغییرپذیری کم می‌باشند.

میانگین عملکرد محصول کیوی برابر ۲۱/۴۳ تن در هکتار و دارای ضریب تغییرات ۱۱ درصد می‌باشد که این مقدار اندکی متأثر از تغییرپذیری زیاد خصوصیات فیزیکی و شیمیایی آب و خاک می‌باشد.

نتایج مدل‌سازی

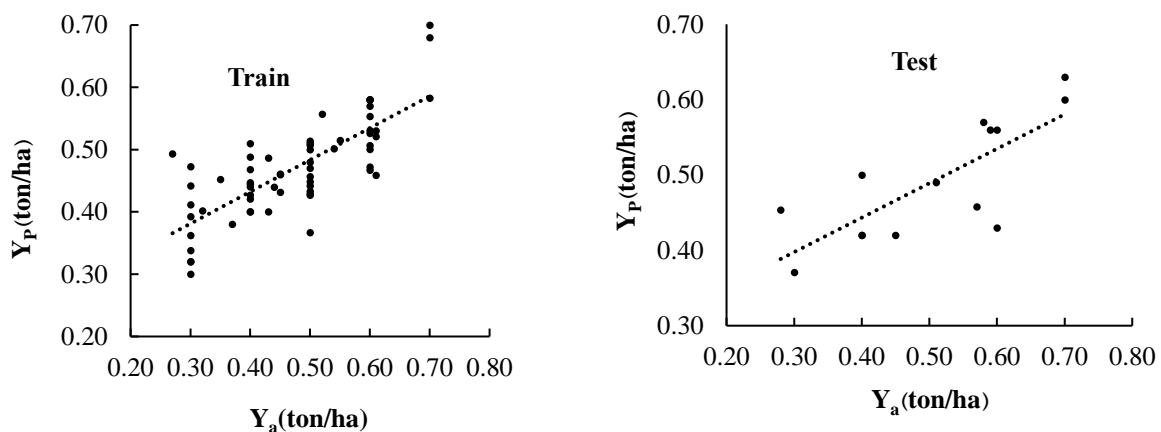
مدل رگرسیونی تخمین عملکرد کیوی در جدول ۳ ارائه شده است، با توجه به مدل مشخص است ویژگی‌های اثرگذار بر عملکرد کیوی شامل هدایت الکتریکی آب آبیاری، درصد رس، درصد سیلت، درصد شن و هدایت الکتریکی خاک می‌باشد. برای داده‌های واسنجی در این حالت مقادیر شاخص‌های آماری شامل R^2 و RMSE به ترتیب برابر ۰/۵۹ و ۰/۰۷۴ می‌باشد، این نتایج با نتایج محققان دیگر نیز همخوانی دارد. بابایی و همکاران (Babaei et al., 2016) با به کارگیری روش رگرسیون مدلی برای بیان رابطه عملکرد دانه گندم دیم و ویژگی‌های خاک ارائه کردند، درصد سیلت مهمترین عامل بر عملکرد گندم دیم بود. بهمنی و همکاران (Bahmani et al., 2020) بیان نمودند از بین ویژگی‌های خاک درصد سیلت خاک از پارامترهای مهم در تخمین عملکرد گل محمدی می‌باشد.

پورمحمدعلی و همکاران (Pour-Mohammadali et al., 2018) برای مدل‌سازی عملکرد پسته از روش MLR استفاده کردند. نتایج آنها نشان که روش MLR تنها تغییرات اندکی از عملکرد را توجیه می‌نماید. گمرکچی (Gomrokchi et al., 2015) در بررسی تخمین عملکرد گندم با مدل شبکه عصبی مصنوعی و مدل رگرسیون خطی چند متغیره ضریب تبیین مدلها به ترتیب برابر ۰/۶۷ و ۰/۵۷ به دست آورد. دهمرده و همکاران (Dehmardeh et al., 2021) در پیش‌بینی عملکرد محصول گندم با روش MLR نشان دادند که این مدل تنها توانسته ۴۰ درصد عملکرد کل را تبیین نماید. در شکل ۲ عملکرد مدل MLR برای داده‌های واسنجی و صحت‌سنجی نشان داده شده است. مقدار عملکرد اندازه‌گیری شده Y_a و مقدار عملکرد محصول پیش‌بینی شده Y_p نامگذاری شده است.

جدول ۲. مدل رگرسیون پیش‌بینی عملکرد محصول کیوی

Table 2. Regression model for predicting the yield of kiwi fruit

ویژگی	ET (mm/day)	ECs (ds/m)	pHsoil	Silt (%)	Clay (%)	OC (%)	ECw (ds/m)	Phwater	H (hr)	V (m ³)	ثابت
ضریب رگرسیون	0.177	0.134	0.348	0.328	-0.077	-0.191	-0.08	-0.064	0.144	-0.013	0.16

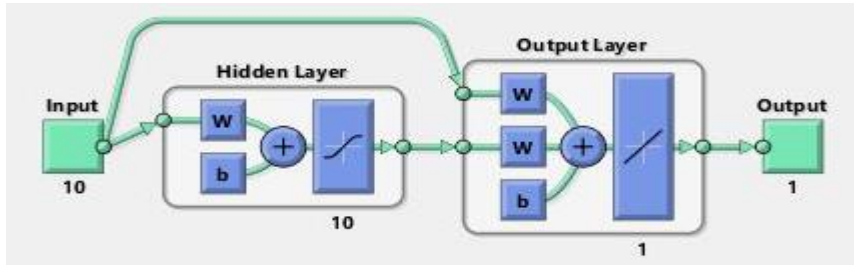


شکل ۲. مقایسه عملکرد مدل MLR برای داده‌های واسنجی و صحت‌سنجی

Fig. 2. Comparison of MLR model performance for train and test data

جهت مدل‌سازی عملکرد محصول کیوی مدل شبکه عصبی پیشرو با پس‌انتشار خطا مورد استفاده قرار گرفت. در جدول ۳ مشخصات لایه‌ها و پارامترهای تابع انتقالی مدل FFB آورده شده است. در شکل ۳ معماری شبکه عصبی پیشرو آورده شده است. در شکل ۴ عملکرد مدل FFB برای داده‌های واسنجی و صحت‌سنجی نشان داده شده است. برای داده‌های واسنجی در این حالت مقادیر شاخص‌های آماری شامل R^2 و RMSE به ترتیب برابر ۰/۹۸ و ۰/۰۱ می‌باشد.

نتایج نشان می‌دهد مدل FFB از دقت بالایی در تخمین عملکرد محصول برخوردار است، این نتایج با نتایج محققان دیگر نیز همخوانی دارد. در تحقیق پورمحمدعلی و همکاران (Pour-Mohammadali et al., 2018) دقت مدل شبکه عصبی برای پیش‌بینی عملکرد پسته، براساس متغیرهای ورودی، به میزان ۹۰٪ بوده است.



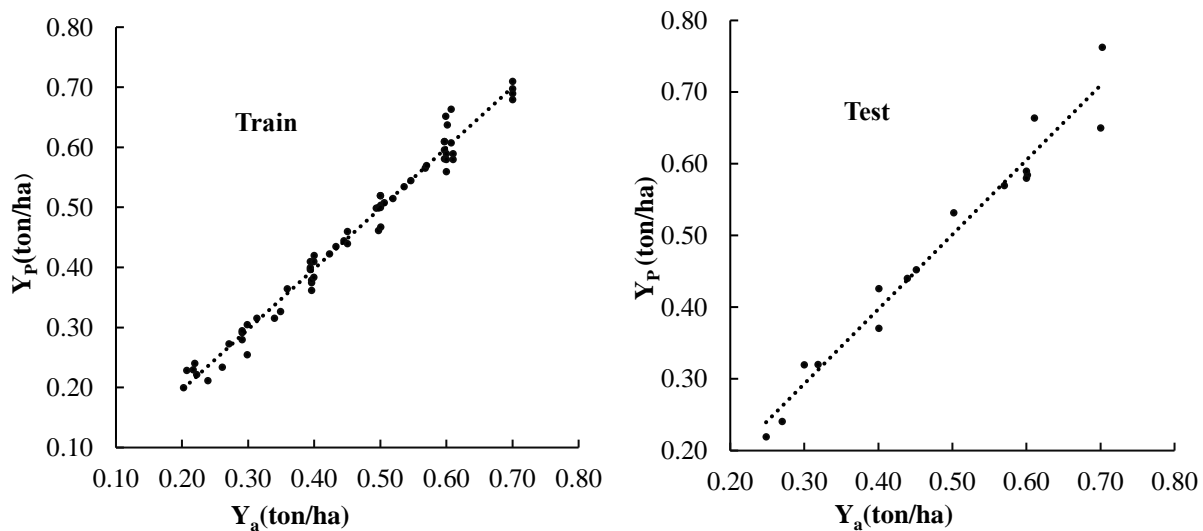
شکل ۳. معماری شبکه عصبی پیشرو با پس‌انتشار خطا

Fig. 3. The architecture of the neural network Feed Forward Backpropagation algorithm

جدول ۴. مشخصات لایه‌های پیشرو با پس‌انتشار خطا

Table 4. Characteristics of the Feed Forward Backpropagation algorithm

ویژگی مقدار	لایه ورودی	لایه مخفی	لایه خروجی (خطی)	گام زمانی	شیب تابع کارایی	حداقل گرادیان	تعداد تکرار
	10	10	1	12	0.075	0.048	1000



شکل ۴. مقایسه عملکرد مدل FFB برای داده‌های واسنجی و صحت‌سنجی

Fig. 4. Comparison of FFB model performance for train and test data

بهمنی و همکاران (Bahmani et al., 2020) از یک شبکه پس‌انتشار برای پیش‌بینی عملکرد گل محمدی به کمک ویژگی‌های آب و خاک استفاده کردند، مدل شبکه عصبی مصنوعی به دلیل لحاظ کردن روابط غیرخطی بین عملکرد و عوامل تأثیرگذار بر آن، کارایی بهتری نسبت به رگرسیون چند متغیره در تخمین عملکرد داشت.

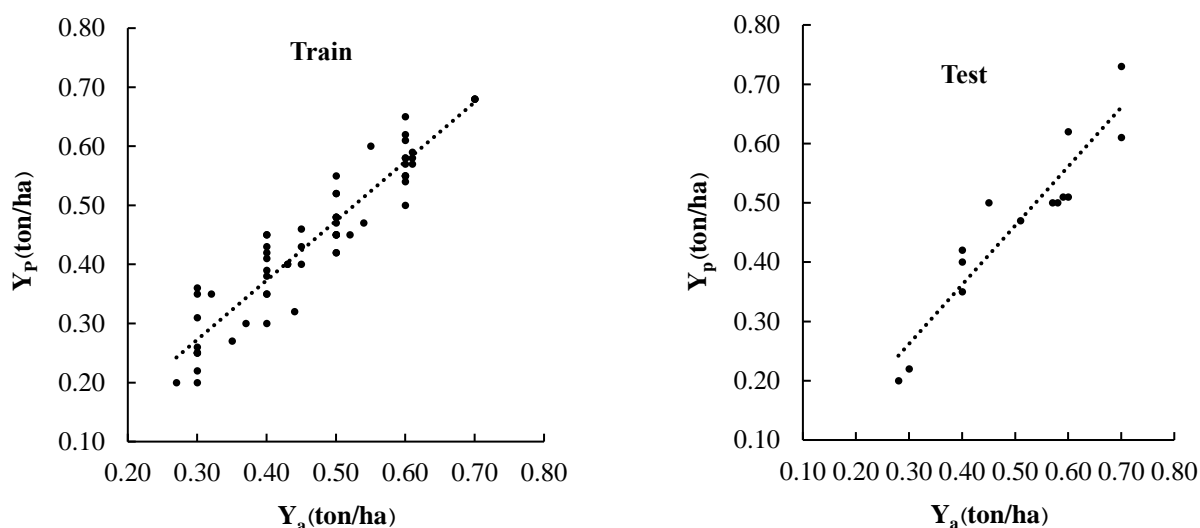
در تحقیق جعفری و همکاران (Jafari et al., 2023) روش پس‌انتشار با ترکیب ورودی‌های بارش و رطوبت نسبی با مقادیر ضریب تبیین ۰/۸۵ و جذر میانگین مربعات خطا داده‌های آزمون به ترتیب برابر ۰/۸۵ و ۰/۰۹ تن در هکتار به عنوان مدل بهینه در تخمین عملکرد گندم دیم شناخته شد. دهمرده و همکاران (Dehmardeh et al., 2021) با مقایسه مدل‌های مورد استفاده در پیش‌بینی اجزاء عملکرد گندم با استفاده از ویژگی‌های خاک نشان داد که بالاترین ضریب تبیین (۰/۷۶) و کمترین جذر میانگین مربعات خطا مربوط به

روش شبکه عصبی بود. در تحقیق گمرگچی و همکاران (Gomrokchi et al., 2022) ضریب تبیین مدل پیش‌بینی عملکرد محصول گندم در مدل شبکه عصبی مصنوعی و مدل رگرسیون خطی چند متغیره به ترتیب برابر ۰/۶۷ و ۰/۵۷ بود.

در این پژوهش جهت مدل‌سازی عملکرد محصول کیوی مدل KNN استفاده گردید. این روش پیش‌بینی با توجه به تعداد تاخیرها و همچنین شعاع همسایگی انجام می‌شود.

البته انتخاب منطقی تعداد همسایگی نیز از پارامترهای مهم در پیش‌بینی، توسط این روش می‌باشد. نکته قابل توجه، آموزش سریع این شبکه و کاهش زمان آزمون و خطا برای انتخاب پارامتر بهینه آن است. در این راستا با فرض بهترین تعداد همسایگی برابر ۵ توسط تابع فاصله‌سنجی اقلیدسی به عنوان ترکیب کارآمد مشخص گردید.

در شکل ۵ عملکرد مدل KNN برای داده‌های واسنجی و صحت‌سنجی به ازای بهینه‌ترین ساختار مدل نشان داده شده است. برای داده‌های واسنجی در این حالت مقادیر شاخص‌های آماری شامل R^2 و RMSE به ترتیب برابر ۰/۸۷ و ۰/۰۵ می‌باشد.



شکل ۵. مقایسه عملکرد مدل KNN برای داده‌های واسنجی و صحت‌سنجی
 Fig. 5. Comparison of KNN model performance for train and test data

نتایج این تحقیق با نتایج محققان دیگر همخوانی دارد. نتایج تحقیق انصاری (Ansari (2023) نشان داد روش KNN نسبت به روشهای ماشین بردار پشتیبان و برنامه‌ریزی ژنتیک بهترین دقت را دارد و معیارهای ارزیابی ضریب همبستگی و جذر میانگین مربعات خطا به ترتیب از ۰/۸۴ تا ۰/۹۲، ۰/۲۱ تا ۰/۲۴ تن در هکتار متغیر بود. (Hafezi et al., 2021) به منظور پیش‌بینی عملکرد نیشکر از پارامترهای مؤثری نظیر مشخصات گیاه، هدایت الکتریکی آب و خاک، حداکثر درجه حرارت و میانگین ساعات آفتابی در طول دوره رشد محصول و در مقیاس زمانی هفت ساله استفاده گردید و نشان داد عملکرد مدل KNN، ۹۵/۵٪ بوده است.

در این پژوهش جهت مدل‌سازی عملکرد محصول کیوی از مدل برنامه‌ریزی ژنتیک (GP) استفاده گردید. در تحقیق حاضر سعی شده است که با استفاده از پارامترهای ذکر شده فرمولی جهت شبیه‌سازی عملکرد محصول کیوی ارائه گردد که بیشترین انطباق را با داده‌های مشاهداتی داشته باشد. جدول ۵ تنظیمات اولیه مقادیر پارامترهای مورد استفاده در مدل GP را نشان می‌دهد.

یکی از قابلیت‌های مهم روش برنامه‌ریزی ژنتیک، ارائه رابطه ریاضی برای متغیرهای مورد بررسی می‌باشد. در مدل GP انتخاب جمعیت‌های اولیه که همان الگوهای ورودی می‌باشد از اهمیت بالایی برخوردار است. رابطه ۷، سازگارترین رابطه به‌دست‌آمده از مدل GP را نشان می‌دهد. آنالیز حساسیت پارامترهای مورد استفاده در مدل GP در جدول ۶ آورده شده است.

جدول ۵. مقادیر پارامترهای مورد استفاده در مدل GP
Table 5. The values of the parameters used in the GP model

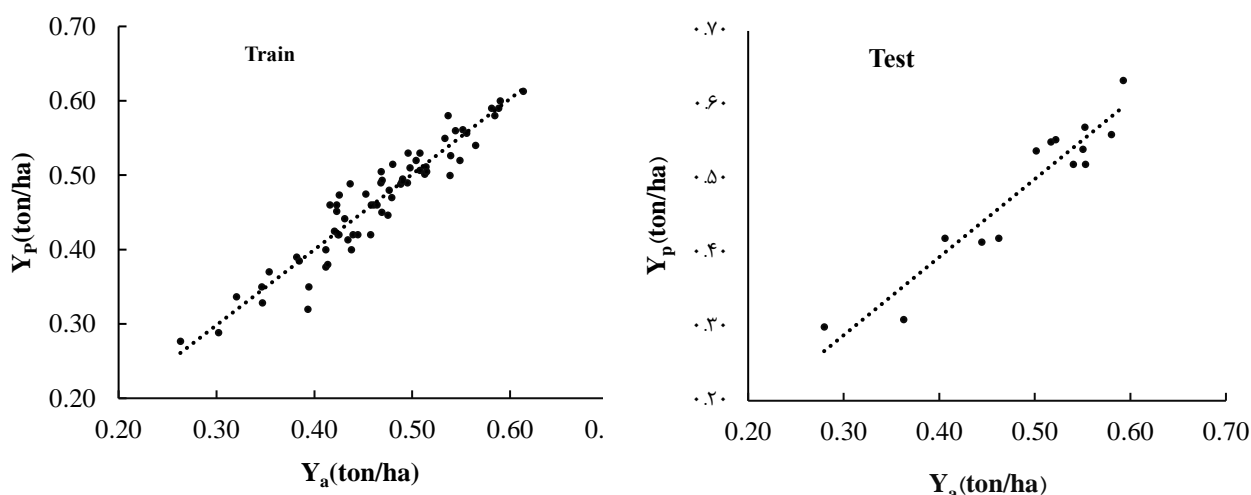
ویژگی	زمان (دقیقه)	تولید نسل ^۱	نرخ تولید نسل در ثانیه	مقدار ارزیابی ^۲	پایداری ^۳ (درصد)	همگرایی (درصد)
مقدار	22	213747	168.3	3.7e9	55.5	100

$$Y = 0.207 + V \times EC_w \times Silt + 0.567 \times ET^2 + 1.435 \times H \times Silt \times PH_s - H \times Clay \times ET \quad (10)$$

جدول ۶. آنالیز حساسیت پارامترهای مورد استفاده در مدل GP
Table 6. sensitivity analysis of the parameters used in the GP model

ویژگی	H(hr)	ECW (ds/m)	V(m3)	Clay (%)	Silt (%)	pH soil	ET (mm/day)
مقدار حساسیت	0.18	0.31	0.52	0.53	0.54	0.59	0.7

مشاهده می‌گردد مهم‌ترین ویژگی اثرگذار بر عملکرد کیوی، تبخیر و تعرق روزانه می‌باشد. همچنین ویژگی‌های شاخص واکنش خاک، درصد سیلت، درصد رس، حجم آب آبیاری، هدایت الکتریکی آب و زمان آبیاری در مرتبه‌های بعدی قرار دارند. در شکل ۶ عملکرد مدل GP برای داده‌های واسنجی و صحت‌سنجی نشان داده شده است.



شکل ۶. مقایسه عملکرد مدل GP برای داده‌های واسنجی و صحت‌سنجی
Fig. 6. Comparison of GP model performance for train and test data

برای داده‌های واسنجی در این حالت مقادیر شاخص‌های آماری شامل R2 و RMSE به ترتیب برابر ۰/۹ و ۰/۰۲ می‌باشد. نتایج این تحقیق با نتایج محققان دیگر همخوانی دارد. در تحقیق شریفی و همکاران (Sharafi et al., 2021) در پیش‌بینی بارش ماهانه، مدل GP با میانگین خطای مطلق ۰/۸ و ضریب همبستگی ۰/۹۷ عملکرد خوبی داشت. در تحقیق احمدی و همکاران (Ahmadi et al., 2021) در پیش‌بینی جریان روزانه رودخانه مدل GP پیشنهادی با ضریب تبیین ۰/۹۷ دارای بیشترین دقت در مرحله صحت‌سنجی بود. همچنین نتایج تحقیق قربانی و همکاران (Ghorbani et al., 2010) در پیش‌بینی نوسانات سطح آب نشان می‌دهد که در اکثر حالتها نتایج حاصل از برنامه‌ریزی ژنتیک از دقت بیشتری نسبت به شبکه‌های عصبی مصنوعی برخوردار است. در جدول ۷ نتایج ارزیابی دقت مدل‌های FFB، GP، KNN، MLR برای داده‌های واسنجی و صحت‌سنجی آورده شده است. برای داده‌های صحت‌سنجی مقادیر شاخص RMSE به ترتیب ۰/۰۱۹، ۰/۰۳۳، ۰/۰۵۹، ۰/۰۹۳، و مقادیر شاخص R2 به ترتیب ۰/۹۶، ۰/۸۹، ۰/۸۸ و ۰/۵۸ می‌باشد. نتایج نشان می‌دهد که دقت مدل شبکه عصبی پیشرو (FFB) نسبت به سایر مدلها بسیار بالاتر بوده است.

جدول ۷. نتایج ارزیابی مدل‌ها
Table 7. Results of model's evaluation

صحت‌سنجی		واسنجی		روش
R ²	RMSE	R ²	RMSE	
0.96	0.019	0.99	0.01	FFB
0.89	0.033	0.90	0.024	GP
0.88	0.059	0.87	0.05	KNN
0.58	0.093	0.59	0.074	MLR

با توجه به پراکنش داده‌ها و آمارهای ارزیابی دقت مدل برنامه‌ریزی ژنتیک (GP) نسبت به مدل روش K-نزدیکترین همسایه (KNN) بالاتر بوده است. همچنین مدل MLR کمترین دقت را در مدلسازی داشت. بطور کلی نتایج کارایی مدلها در ارزیابی تخمین عملکرد محصول گیوی با نتایج محققان دیگر همخوانی دارد.

نتیجه‌گیری

در این پژوهش با استفاده از روشهای داده‌کاوی برنامه‌ریزی ژنتیک، K-نزدیکترین همسایه، شبکه عصبی پیشرو با پس‌انتشار خطا و رگرسیون خطی چند متغیره عملکرد محصول گیوی برای باغات گیوی مجهز به سیستم آبیاری تحت فشار در استان گیلان برآورد گردید. ورودی مدلسازی شامل داده‌های آب و خاک (حداکثر تبخیر و تعرق، ساعات آبیاری، حجم آب آبیاری، شاخص واکنش آب، هدایت الکتریکی آب آبیاری، شاخص واکنش خاک، هدایت الکتریکی خاک، درصد سیلت و درصد رس و درصد ماده آلی خاک بود. مشخص شد که روش شبکه عصبی پیشرو با روش پس‌انتشار خطا نتایج دقیق‌تری نسبت به مدل‌های داده‌کاوی برنامه‌ریزی ژنتیک، K-نزدیکترین همسایه و رگرسیون خطی چند متغیره ارائه می‌دهد. روش برنامه‌ریزی ژنتیک نسبت به روش K-نزدیکترین همسایه دارای دقت بالاتری بود که با توجه به عدم قطعیت در پدیده‌های مرتبط با آب و خاک یا تقریبی بودن مقادیر اندازه‌گیری شده، کارایی بالاتر مدل داده‌کاوی برنامه‌ریزی ژنتیک به همین علت است. مدل رگرسیون خطی چند متغیره نیز به واسطه غیر خطی بودن ویژگی‌های آب و خاک نمی‌تواند نتایج دقیقی ارائه دهد. بنابراین روشهای داده‌کاوی هوشمند می‌تواند به عنوان ابزاری کارآمد برای تدوین برنامه‌های صحیح مدیریتی در زمینه عملکرد محصول معرفی شود.

مراجع

- Ahmadi, F., Aysham, S., Khalili, K., & Javad bemanesh., A. (2015). Application of ANFIS and GP Models to Estimate Monthly Reference Crop Evapotranspiration in Northwest of Iran. *Journal of Water Research in Agriculture*, 29(2), 235-247.
- Ahmadi, M., Kaviani, A., Ramezani Etedali, H., & Tavakoli, A. (2023). Simulation of Rained Wheat Yield Using Drought Indices by Employing Artificial Neural Network, Random Forest and Support Vector Regression Case Study, Saqqez City. *Iran-Water Resources*, 194, 95-109.
- Anaraki, V., Mousavi, M., Mousavi, F., Farzin, S., & karami, h. (2020). Introducing a Nonlinear Model Based on Hybrid Machine Learning for Modeling and Prediction of Precipitation and Comparison with SDSM Method Cases Studies, Shahrekord, Barez, and Yasuj. *Iranian Journal of Soil and Water Research*, 5(12), 325-339.
- Ansari Ghojghar, M. (2023). Development of strategic wheat crop prediction toolkit using machine learning algorithms to reduce food security risks case study, Alborz province. *Iranian Journal of Soil and Water Research*, 53(10), 2277-2294.
- Arumugam, A. (2017). A predictive modeling approach for improving paddy crop productivity using data mining techniques. *Turkish. Journal of Electrical Engineering and Computer Sciences*, 25(6), 4777-4787.
- Ayoubi, SH.A., & Jalalian, A. (2010). *Land Evaluation Agricultural and Natural Resources Second Edition*, Isfahan University of Technology Publication Center, Isfahan, Iran, 385 (In Persian).
- Baba ali, H., Ramak, Z., & Sepahv, R. (2017). The Effectiveness of Genetic Planning Model in rainfall-runoff Simulation process. *Journal of Water Science Engineering*, 7(18), 83-94.
- Babaei, F., Vaezi, A., Teheri, M., Zarrinabadi, E., & Eslami, F. (2016). Development a regression relationship between rainfed wheat yield and soil properties in a semiarid region, Zanjan Province. *Iranian Journal of Soil and Water Research*, 46(4), 715-725.

- Badde, D. S., Gupta, A. K., & Patki, V. K. (2013). Cascade and feed forward back propagation artificial neural network models for prediction of compressive strength of ready mix concrete. *IOSR Journal of Mechanical and Civil Engineering*, 3(1), 1-6.
- Bahmani, M., Mohammadi, J., Esfandiarpour Boroujeni, I., & Mottaghian, H. (2020). Assessment of relationships between rose yield and soil and topography properties using multivariate regression, artificial neural network and structure equation modeling. *Applied Soil Research*, 8(2), 142-157.
- Breiman, L., Friedman, J., Olshen, R.A., & Stone, C.J. (1984). *Classification and Regression Trees* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>.
- Brown, M. E., Lary, D. J., Vrieling, A., Stathakis, D., & Mussa, H. (2008). Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. *International Journal of Remote Sensing*, 29(24), 7141-7158.
- Cox, M. S., Gerard, P. D., & Abshire, M. J. (2006). Selected Soil Properties' variability and Their Relationships with Yield In Three Mississippi Fields. *Soil science*, 171(7), 541-551.
- Dahmardeh, A. R., Shahriari, A., Pahlavan Rad, M., Shabani, A., & Ghorbani, M. (2021). Modeling wheat yield using some soil properties at the field scale (Case study: Sistan dam research farm, university of Zabol). *Agricultural Engineering (Scientific Journal of Agriculture)*, 44, 1. (In Persian).
- Diaz-Ramirez, J. N., McAnally, W. H., & Martin, J. L. (2011). Analysis of hydrological processes applying the HSPF model in selected watersheds in Alabama, Mississippi, and Puerto Rico. *Applied Engineering in Agriculture*, 27(6), 937-954.
- Emamgholizadeh, S., Parsacian, M., & Baradaran, M. (2015). Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*, 68, 89-96.
- Emami, S., Dehghanisani, H., & Rezaie, M. (2021). Application of artificial neural network models in estimating nectarine crop yield under two-sided furrow irrigation. *Water and Soil Management and Modelling*, 1(2), 47-59.
- Eslami, R., Azarnoosh, M., Kialashki, A., & Kazemnejad, F. (2021). Assessing the Probability of Forest Fire Occurring using Dong Model, Artificial Neural Network and K Nearest Neighbors in Babolrood Basin, Mazandaran. *Ecology of Iranian Forest*, 9(17), 185-195 (In Persian).
- Ghorbani, M. A., Kisi, O., & Aalinezhad, M. (2010). A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods. *Applied Mathematical Modelling*, 34(12), 4050-4057.
- Gomrokchi, A. U., Baghani, J., & Abbasi, F. (2021). Evaluating the Capability of Data Mining Models in Predicting Irrigated Wheat Yield in Iran. *Water and Soil*, 35(2), 189-202.
- Hafezi, N., Bahrami, H., Sheikh Davoodi, M. J., & Alavi, S. E. (2020). Hybrid artificial neural network with meta-heuristic algorithms for predicting sugarcane yield. *Iranian Journal of Biosystems Engineering*, 51(3), 515-526.
- Jafari, M. M., Liaghat, A., & Mohtashami, S. (2023). Comparison and Evaluation of LM, BR and GD Algorithms of Artificial Neural Network in Estimating Rainfed Wheat Yield Based on Meteorological Parameters (Case Study: Kermanshah Province). *Iranian Journal of Irrigation & Drainage*, 17(3), 529-544.
- Jagtap, S. S., Lall, U., Jones, J. W., Gijsman, A. J., & Ritchie, J. T. (2004). Dynamic nearest-neighbor method for estimating soil water parameters. *Transactions of the ASAE*, 47(5), 1437-1444.
- Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of irrigation and drainage engineering*, 128(4), 224-233.
- Monjezi, N. (2020). Diagnose and Prioritizing of Effective Managerial and Executive Factors on Water Productivity in Sugarcane Production and Providing Practical Solutions to Increase It. *Iranian Journal of Biosystems Engineering*, 51(1), 99-112.
- Moodi, Y., Eslami, E., Mousavi, S. R., Dizangian, B., & Mirshekar, H. (2022). Applying neural networks for estimating the compressive strength of confined circular concrete columns with FRP sheets. *Journal of Structural and Construction Engineering*, 9(5), 58-77.
- Modaresi, F., Ebrahimi, K., & Araghinejad, S. (2022). Ranking Evaluation of Data-driven and Conceptual Modelling of Rainfall-Runoff Process in Monthly Time Scale. *Irrigation and Water Engineering*, 12(4), 258-273.
- Parviz, L., Ahadi, R., & Seyyed Milani, F. (2022). Genetic Algorithm Application for Soil pH Spatial Distribution Estimation with Geostatistics and Gene Expression Programming Combination. *Applied Soil Research*, 10(3), 117-129.
- Pourmohamadali, B., Salehi, M. H., Hosseinfard, S. J., Shirani, H., & Esfandiarpour Boroujeni, I. (2018). Pistachio yield prediction using multiple linear regression and artificial neural network (a case study: rafsanjan and anar regions, kerman province). *Agricultural Engineering*, 40(2), 71-87.
- Salajegheh, A., Fathabadi, A., & Gholami, H. (2010). Predict river discharge using the nearest neighbor. In *5th national conference on science and management engineering Iran*. Gorgan University of Agricultural Sciences and Natural Resources. (In Persian).

- Sadr, S., & Eslami, M. (2021). Determination of effective weather variables on pistachio yield using C&R decision tree algorithm. *Journal of Agricultural Meteorology*, 9(1), 53-62.
- Sattari, M. T., Joudi, A. R., & Kusiak, A. (2016). Estimation of water quality parameters with data-driven model. *Journal-American Water Works Association*, 108(4), E232-E239.
- Sharafi, M., Samadianfard, S., & Hashemi, S. (2021). Monthly rainfall Forecasting using genetic programming and support vector machine. *Journal of Rainwater Catchment Systems*, 8(4), 63-71.
- Sivaneasan, B., Yu, C. Y., & Goh, K. P. (2017). Solar forecasting using ANN with fuzzy logic pre-processing. *Energy procedia*, 143, 727-732.
- Soltani, A., Gorbani, M. A., Fakheri Fard, A., Darbandi, S., & Farsadizadeh, D. (2011). Genetic programming and its application in rainfall-runoff modeling. *Water and Soil Science*, 20(4), 62-71.
- Tashakkori, F., Mohammadi Torkashvand, A., Ahmadi, A., & Esfandiari, M. (2021). Prediction of Saffron Yield based on Soil properties Using Regression and Artificial Neural Networks Models in the Vamenan Region of Golestan Province. *Saffron Agronomy and Technology*, 9(2), 159-175.
- Xindong, W., & Kumar, V. (2009). The Top Ten Algorithms in Data Mining. *Chaman and Hall/CRC, Data Mining and Knowledge Discovery Series, Taylor & Francis Group, LLC*.
- Yue, J., Feng, H., Yang, G., & Li, Z. (2018). A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sensing*, 10(1), 66.
- Zahiri, A., & Azamathulla, H. M. (2014). Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. *Neural Computing and Applications*, 24, 413-420. (In Persian).
- Zheng, K., Si, G., Diao, L., Zhou, Z., Chen, J., & Yue, W. (2017, May). Applications of support vector machine and improved k-Nearest Neighbor algorithm in fault diagnosis and fault degree evaluation of gas insulated switchgear. In *2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE)* (pp. 364-368). IEEE.