

# Analysis of Water and Soil Data in Estimating Crop Yield in Pressurized Irrigation Systems Using Data Mining Methods (Case Study: Mehdol Hazelnut-Gilan Province)

Masoumeh Yeganeh<sup>1\*</sup>, Farzaneh Lotfi<sup>2</sup>, Afshin Ashrafzadeh<sup>3</sup>

<sup>1</sup>MSc Graduated, Department of Water Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

<sup>2</sup>Associate Professor, Department of Water Engineering, Faculty of Agriculture and Natural Resources, Imam Khomeini International University, Qazvin, Iran

<sup>3</sup>MSc Graduated, Department of Water Engineering, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

Corresponding Author's E-mail: [masomeh.yeganeh@alumni.um.ac.ir](mailto:masomeh.yeganeh@alumni.um.ac.ir)

Received: 02 Nov. 2023    Accepted: 25 Dec. 2024  
Revised: 24 Dec. 2024    Published: 24 Apr. 2025



Homepage: [ijwer.uoz.ac.ir](http://ijwer.uoz.ac.ir)

**Abstract:** Hazelnuts are one of Iran's agricultural products, which are 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 effectiveness of advanced artificial neural network data mining methods, K-nearest neighbor, CART decision tree and multi linear regression in estimating hazelnut yield in Gilan province using water and soil characteristics. The data of this research was obtained from the field measurement of water and soil information and yield of hazelnut orchards equipped with pressurized 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, sand percentage, water electrical conductivity and water reaction index and irrigation volume were selected as model inputs and model output performance. The results showed that the leading artificial neural network model performs better than the other three models due to higher correlation coefficient statistics (0.98) and higher Nash-Sancliff coefficient (0.96). Also, CART decision tree has correlation coefficient (0.93), Nash Sutcliffe coefficient (0.93) and K-nearest neighbor method has correlation coefficient (0.9), Nash-Sancliff coefficient (0.67), Nash-Sancliff coefficient (0.47), which indicates higher accuracy of CART decision tree method. Therefore, the advanced artificial neural network model can act as a powerful tool in estimating hazelnut performance.

**Keywords:** Multi linear regression, Classification and Regression Tree, Feed Forward backdrop Neural Network,, K-Nearest Neighbor

**Introduction:** Predicting crop yield using data has an important impact on socio-economic issues and political decision-making at the regional scale. Guilan province is considered as the hub of hazelnut production in the country. In the science of statistics, there are various methods for categorizing, recognizing patterns, forecasting and modeling data. To predict the performance of agricultural products with data mining, various methods such as K-nearest neighbor (KNN), Classification and Regression Tree (CART), neural network feedforward back prop (FFB) and multi linear regression (MLR) are leading. Since the prediction of hazelnut yield using data mining methods is very important for the decision makers of the agricultural sector in terms of knowing the positive and negative factors in the yield. This research aims to answer this need by using all these methods for hazelnut orchards equipped with drip irrigation system in Guilan province.

**Methodology:** The information used in this study includes the amount of water consumption and yield of hazelnut and the parameters related to these two indicators in hazelnut orchards equipped with pressurized irrigation systems under the management of operators in the cities of Guilan province, including Roodsar, Amlash, Siyahkol and Roodbar, which in a research The measured quantities included the salinity of the irrigation water, the salinity of the saturated soil extract, the number of times of irrigation, the yield of corn, and the volume of water consumed per tree in each irrigation round. To determine the characteristics of the soil in each garden, the depth of activity The roots (0-30 cm) were sampled from the soil.

At the same time, water samples were also obtained from the gardens of the region and the electrical conductivity of the water samples was measured. The volume of water needed in each round of irrigation for each tree was obtained by measuring the irrigation water. The percentage of sand, silt and clay was also obtained by determining the soil texture by hydrometric method. The amount of yield in one hectare was also obtained after the harvest at the end of the harvest in the year 2021-2022.

**Results and Discussion:** To model the input variables including daily maximum evaporation and transpiration data, electrical conductivity and soil acidity, percentage of clay and percentage of silt and percentage of sand, electrical conductivity and acidity and irrigation volume for each tree in each irrigation round and hazelnut yield as the output of the model. Work was done. The results obtained from each of the models were compared with the measured values using statistical indices of correlation coefficient (R), root mean square (RMSE), Nash Sutcliffe coefficient (NSE) and the best model was selected. The variables of volume of water required by the tree in the irrigation cycle, irrigation water reaction index, soil silt percentage and soil reaction index showed a positive correlation with crop yield. Also Irrigation hours, electrical conductivity of irrigation water, sand percentage, clay percentage, sand percentage, soil electrical conductivity and maximum daily evaporation and transpiration showed a negative correlation with crop yield. The results showed that the leading artificial neural network model performs better than the other three models due to higher correlation coefficient statistics (0.98) and higher Nash-Sancliffé coefficient (0.96). Also, CART decision tree has correlation coefficient (0.93), Nash Sutcliffe coefficient (0.93) and K-nearest neighbor method has correlation coefficient (0.9), Nash-Sancliffé coefficient (0.7) and multiple linear regression method. The variable has correlation coefficient (0.67), Nash-Sancliffé coefficient (0.47), which indicates higher accuracy of CART decision tree method

**Conclusion:** By evaluating the results, it was found that the forward neural network method with the error backpropagation method provides more accurate results than the decision tree, K-nearest neighbor, and multivariate linear regression models. The CART decision tree method was more accurate than the K-nearest neighbor method, which is the reason for the higher efficiency of the decision tree model due to the uncertainty in the phenomena related to water and soil or the approximations of the measured values. The multivariate linear regression model also cannot provide accurate results due to the non-linearity of water and soil characteristics. Therefore, artificial intelligence techniques can be introduced as an efficient tool for developing correct management plans in the field of product performance.

© 2023 University of Zabol, Zabol, Iran.



This is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (<http://creativecommons.org/licenses/by/4.0/>)

# تحلیل داده‌های آب و خاک در برآورد عملکرد محصول در سامانه‌های آبیاری تحت فشار با استفاده از روش‌های داده‌کاوی (مطالعه موردی: محصول فندق – استان گیلان)

معصومه یگانه<sup>۱\*</sup>، فرزانه لطفی<sup>۲</sup>، افشین اشرف زاده<sup>۳</sup>

<sup>۱</sup>فارغ‌التحصیل کارشناسی ارشد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران.  
<sup>۲</sup>دانشیار، گروه مهندسی آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه بین‌المللی قزوین، قزوین، ایران  
<sup>۳</sup>فارغ‌التحصیل کارشناسی ارشد، گروه مهندسی آب، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت، ایران

پست الکترونیکی نگارنده مسئول: [masomeh.yeganeh@alumni.um.ac.ir](mailto:masomeh.yeganeh@alumni.um.ac.ir)

وبگاه نشریه: [ijwer.uoz.ac.ir](http://ijwer.uoz.ac.ir)



تاریخ دریافت: ۱۴۰۲/۰۸/۱۱ تاریخ پذیرش: ۱۴۰۳/۱۰/۰۵

تاریخ ویرایش: ۱۴۰۳/۱۰/۰۴ تاریخ چاپ: ۱۴۰۴/۰۲/۰۴

**چکیده:** روش‌های داده‌کاوی به‌خوبی قادر هستند تا در زمینه مدل‌سازی عملکرد محصول اطلاعات لازم را در اختیار تولیدکنندگان قرار دهند. این تحقیق به بررسی کارایی روش‌های داده‌کاوی شبکه عصبی مصنوعی پیشرو، K-نزدیک‌ترین همسایگی، درخت تصمیم CART و رگرسیون خطی چندمتغیره در برآورد عملکرد محصول فندق در استان گیلان با استفاده از خصوصیات آب و خاک پرداخته است. داده‌های این پژوهش از اندازه‌گیری میدانی اطلاعات آب و خاک و عملکرد محصول فندق 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 شن، هدایت الکتریکی آب و شاخص واکنش آب و حجم آبیاری به‌عنوان ورودی‌های مدل و عملکرد خروجی مدل انتخاب گردید. نتایج نشان داد که مدل شبکه عصبی مصنوعی پیشرو به دلیل آماره‌های ضریب همبستگی بیشتر (۰/۹۸) و ضریب نش-سانکلیف بیشتر (۰/۹۶) عملکرد بهتری نسبت به سه مدل دیگر دارد. همچنین درخت تصمیم CART دارای ضریب همبستگی (۰/۹۳)، ضریب نش-سانکلیف (۰/۹۳) و روش K-نزدیک‌ترین همسایگی دارای ضریب همبستگی (۰/۹۶)، ضریب نش-سانکلیف (۰/۹۳) و روش رگرسیون خطی چندمتغیره دارای ضریب همبستگی (۰/۶۷)، ضریب نش-سانکلیف (۰/۴۷) بوده که حاکی از دقت بالاتر روش درخت تصمیم CART است. بنابراین، مدل شبکه عصبی مصنوعی پیشرو می‌تواند به‌عنوان یک ابزار قدرتمند در تخمین عملکرد فندق عمل نماید.

**کلیدواژه‌ها:** رگرسیون خطی چندمتغیره، درخت تصمیم CART، شبکه عصبی پیشرو، K-نزدیک‌ترین همسایگی

## ۱- مقدمه

محصول فندق یکی از محصولات ارزشمند باغی استان گیلان است که با دارا بودن سطح کشت معادل ۱۶۴۰۰ هکتار و میانگین تولید سالانه ۱۴۰۰۰ تن به عنوان قطب تولید فندق در کشور به شمار می‌رود. مهم‌ترین مسئله برای کشاورزان، کسب درآمد و نیل به حداکثر محصول است. این موضوع، نیازمند درک کامل روابط بین مقدار عملکرد با ویژگی‌های باغ‌ها است. برآورد عملکرد محصول به عنوان یک ویژگی دیریافت از روی ویژگی‌های زودیافت (مانند ویژگی‌های فیزیکی و شیمیایی خاک و آب) از اهمیت فراوانی برخوردار است. 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 چندمتغیره، شبکه‌های عصبی مصنوعی، روش K-نزدیک‌ترین همسایه، درخت تصمیم است.

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، 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 تاثیر هر یک از متغیرهای ورودی و خروجی را برای کاربر مشکل می‌سازد (Bahmani, 2020). در برخی پژوهش‌هایی که با روش‌های رگرسیون چندمتغیره به تخمین عملکرد پرداخته‌اند، نتایج کارایی نه‌چندان قوی این مدل‌ها را نشان می‌دهد. (Tashakkori, 2021).

در پژوهش (Ayoubi et al (2009) که رابطه بین عملکرد دانه جو را با ویژگی‌های خاک از طریق به‌کارگیری رگرسیون چندمتغیره بررسی نمودند، بیان نمودند روابط رگرسیونی می‌تواند ۸۰ درصد از عملکرد دانه را توجیه کند.

یکی دیگر از روش‌ها برای پیش‌بینی تولید محصول، مدل شبکه عصبی مصنوعی<sup>۱</sup> است. 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 غیرخطی است که بدون داشتن اطلاعات کامل از نحوه ارتباط پارامترها و قوانین حاکم بین آن‌ها سامانه‌های مختلف را شبیه‌سازی می‌کند. (Hafezi et al (2021) در پیش‌بینی عملکرد نیشکر مدل شبکه عصبی استفاده نمودند، نتایج حاکی از عملکرد دقیق این مدل‌ها است. (Jafari et al (2023) نشان دادند که مدل شبکه عصبی بر مبنای خصوصیات هواشناسی قادر به پیش‌بینی عملکرد گندم با دقت مناسب است.

(Emamgholizadeh et al (2015) با مطالعه بر عملکرد کنگد در اصفهان بیان نمودند که مدل شبکه عصبی کارایی بیشتری نسبت به رگرسیون چندگانه دارد. طی پژوهشی که (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. همچنین مدل شبکه عصبی پیشنهادی (Emami et al (2021) برای عملکرد محصول شلیل مغان برازش خوبی با داده‌های واقعی داشت. (Ahmadi et al (2023) در شبیه‌سازی عملکرد گندم با استفاده از شاخص‌های خشک‌سالی و مدل شبکه عصبی نشان داد نتایج آن‌ها از کارایی بهتر، ضریب دقت بیشتر و خطای کمتری نسبت به سایر مدل‌ها است. تحقیقات Tashakkori et al (2021) نشان داد که مدل شبکه عصبی بر پایه ویژگی‌های خاک قادر به پیش‌بینی عملکرد گل زعفران با کمترین خطا است. یکی از پرکاربردترین فرآیند آموزشی شبکه عصبی مصنوعی، روش پیشرو با پس‌انتشار خطا<sup>۲</sup> است، در این روش یادگیری الگو در طول آموزش روی می‌دهد که موجب افزایش دقت مدل و کاهش خطا می‌شود. (Moodi et al (2022) در تخمین مقاومت فشاری با

<sup>2</sup> Feed Forward Backprop (FFB)

<sup>1</sup> Artificial neural network (ANN)

آن استنتاج می‌کنند، سپس درخت می‌تواند به صورت مجموعه قوانینی برای پیش‌بینی نتیجه ویژگی‌های معلوم استفاده شود و این مدل به لحاظ سادگی و ایجاد قوانین پیش‌بینی ارجحیت دارد.

در پژوهش Ansari (2023) درخت تصمیم پیشنهادی با پارامترهای ورودی کمتر نسبت به مدل شبکه عصبی مصنوعی توانایی بالاتری در تخمین ویژگی‌های هیدرولیکی دارد. Arumugam (2017) در مطالعه‌ای عملکرد محصول برنج را با استفاده از الگوریتم درخت تصمیم تخمین زد و نشان داد درخت تصمیم دقت بالایی دارد. در پژوهش Sadr and Eslami (2021) درخت تصمیم پیشنهادی توانست روابط پیچیده بین متغیرهای مختلف اقلیمی با عملکرد پسته را بهتر از رگرسیون خطی استخراج کند و مهم‌ترین متغیرهای موثر را با دقت قابل قبولی مشخص کند.

Monjezi (2020) نشان می‌دهد که روش درخت تصمیم قادر است با استفاده از پارامترهای موثر، مقاومت نفوذ خاک مزارع را با دقت بالایی پیش‌بینی کند. در علوم کشاورزی با توجه به وجود روابط بین غیرخطی و پیچیده پارامترها، استفاده از مدل‌های شبکه عصبی مصنوعی و درخت تصمیم بسیار سودمند است. Mesri et al (2020) با توجه به اهمیت موضوع تخمین عملکرد محصول، این تحقیق باهدف بررسی عملکرد روش‌های داده‌کاوی، K- نزدیک‌ترین همسایگی شبکه عصبی مصنوعی پیشرو، درخت تصمیم و مدل رگرسیونی چندمتغیره برای محصول فندق در باغات مجهز به سیستم آبیاری قطره‌ای در استان گیلان انجام می‌شود.

## ۲- مواد و روش‌ها

### ۲-۱- منطقه مطالعاتی و داده‌های مورد استفاده

**شکل ۱** موقعیت جغرافیایی مناطق مورد مطالعه را نشان می‌دهد. در این پژوهش باغات فندق در شهرستان‌های رودسر، املش، سیاهکل و رودبار (مناطق اصلی تولیدکننده فندق کشور) انتخاب گردید. اطلاعات مورد استفاده شامل مقادیر حجم آب مصرفی و عملکرد فندق و کمیت‌های

شبکه عصبی پیشرو با پس‌انتشار خطا دریافتند شبکه پس‌انتشار خطا باعث کاهش خطا نسبت به مدل‌های تحلیلی 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.

یکی دیگر روش‌های متداول برای پیش‌بینی تولید محصول روش K- نزدیک‌ترین همسایه<sup>۱</sup> است (Zheng, 2017). این روش مزایای دارد از جمله: سادگی، کارایی، شهودی و دارای عملکرد طبقه‌بندی رقابتی، دقت مناسب. در پیش‌بینی عملکرد گندم توسط Ansari (2023) با روش‌های فراکاوشی 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. (Salajegheh et al (2010) با استفاده از مدل-K نزدیک‌ترین همسایه مقادیر دبی را در رودخانه کرج را پیش‌بینی کردند. نتایج نشان داد که این عملکرد مناسبی در تخمین دبی دارد.

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. 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.

Modaresi et al (2022) با مقایسه شبیه‌سازی جریان ماهانه ورودی با خصوصیات دما و بارش در سد کرخه نشان دادند که K- نزدیک‌ترین همسایه نسبت به مدل شبکه عصبی دقت پایین‌تری دارد اما در تعیین میانگین سالانه بارش توسط 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 داده‌کاوی درخت تصمیم<sup>۲</sup> است. درختان تصمیم مجموعه‌ای از داده‌ها را می‌گیرند و یک درخت تصمیم را از

<sup>۱</sup> K-Nearest Neighbors (KNN)

<sup>۲</sup> Decision Tree

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 سیگموئید 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).

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

در رابطه (۱)،  $X_s$ : مقدار استاندارد شده،  $X_i$ : مقدار واقعی،  $X_{\text{mean}}$ : متوسط مقادیر،  $X_{\text{max}}$ : مقدار حداکثر،  $X_{\text{min}}$ : مقدار حداقل است. از ۴۰ سری داده در آب و خاک در دسترس، ۷۵ درصد داده‌ها به‌منظور واسنجی مدل و ۲۵ درصد داده‌ها برای صحت‌سنجی استفاده شد. برای انجام مدل‌سازی یک متغیر وابسته یعنی عملکرد محصول و متغیرهای مستقل شامل درصد شن، درصد سیلت، درصد رس، شاخص واکنش خاک، هدایت الکتریکی خاک، شاخص واکنش آب، هدایت الکتریکی آب، حجم آب مصرفی در هر درخت در دور آبیاری، زمان آبیاری و حداکثر تبخیر و تعرق مورد استفاده قرار گرفت.

### ۲-۳- مدل‌سازی درخت تصمیم

یکی از روش‌های دسته‌بندی، درخت تصمیم است که مدل خود را بر اساس یک درخت پیاده‌سازی می‌کند. در این روش با توجه به مجموعه آموزش یک درخت بر اساس ویژگی‌های مخالف آن درست می‌شود که با استفاده از این درخت باید بتوان یک عضو جدید را در دسته خاصی طبقه‌بندی کرد. با جستجو در برگه‌ای درخت می‌توان به یک برگ رسید که آن برگ مشخص‌کننده نوع عضو جدید است. ریشه درخت محل آغازین پیمایش درخت است که به بررسی اولین متغیر می‌پردازد. گره‌ها نیز مثل ریشه به

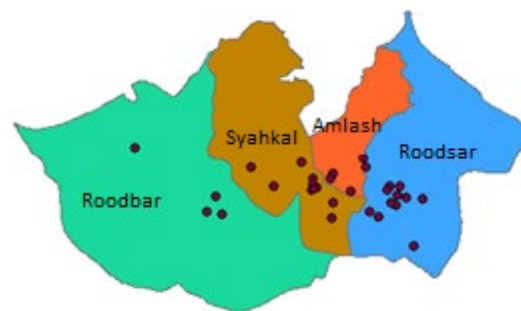


Fig. 1 Location of the study area

شکل ۱ موقعیت مناطق مورد مطالعه

مربوط به این دو شاخص در باغات تحت مدیریت زارعین (تعداد ۴۰ باغ) به‌صورت میدانی برداشت شده است. بدین منظور از عمق ۳۰-۰ سانتی‌متری نمونه خاک برداشت شد. خصوصیات فیزیکی و شیمیایی خاک شامل درصد شن (Sand)، درصد سیلت (Silt)، درصد رس (Clay)، شاخص واکنش خاک ( $PH_s$ )، هدایت الکتریکی خاک ( $EC_s$ ) 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. در نمونه‌های خاک مقادیر  $PH_s$  و  $EC_s$  در عصاره گل اشباع، درصد ذرات معدنی خاک (رس، شن و سیلت) به روش هیدرومتری با قرائت چهارزمانه اندازه‌گیری شدند.

با نمونه‌برداری آب، خصوصیات آب شامل شاخص واکنش آب ( $PH_w$ )، هدایت الکتریکی ( $EC_w$ ) اندازه‌گیری شد. حجم آب مصرفی برای درخت در دور آبیاری ( $V$ )، زمان آبیاری ( $Hr$ ) نیز اندازه‌گیری شد. حداکثر تبخیر و تعرق روزانه ( $ET$ ) از داده‌های هواشناسی هر شهرستان تعیین شد. مقدار عملکرد محصول فندق در هکتار برحسب کیلوگرم در پایان فصل برداشت سال ۱۴۰۰-۱۴۰۱ تعیین شد. این اطلاعات با هماهنگی سازمان جهاد کشاورزی استان گیلان، برای باغات فندق مجهز به سیستم آبیاری تحت فشار جمع‌آوری گردید.

### ۲-۲- نرمال‌سازی داده‌ها

برای جلوگیری از تاثیر نامطلوب مقیاس متفاوت متغیرها بر

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

که در آن  $X$  نماینده داده‌های تعلیم با پارامترهای مشخص ( $X_1$  تا  $X_n$ ) و  $y$  نماینده داده‌های آموزش با همان تعداد پارامترهای مشخص ( $y_1$  تا  $y_n$ ) 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 به شمار می‌آید (Xindung and Kumar., 2009). برای مدل‌سازی عملکرد محصول فندق با روش  $K$ - نزدیک‌ترین همسایه از نرم‌افزار Statistica 13.5.0.17 استفاده گردید.

## ۲-۵- مدل شبکه عصبی پیشرو

شبکه عصبی مصنوعی دارای لایه‌های ورودی، پنهان (میانی) و خروجی است. لایه‌های شبکه عصبی از واحدهای مستقلی به نام گره ساخته شده است. گره‌ها داده‌های ورودی خود را دریافت کرده و با اعمال محاسبات بر روی آن‌ها، مقداری را در خروجی برمی‌گرداند (Brown et al., 2008). در این پژوهش از شبکه عصبی پیشرو جهت مدل‌سازی استفاده گردید. پیشرو بودن به این معناست که نورون‌های عصبی در لایه‌های متوالی قرار گرفته است 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 به دست آوردن

بررسی متغیرها می‌پردازد. نتیجه بررسی آن است که با شاخه‌های آن گره به گره موردنظر دیگری برود و یا به برگ برسد. برگ‌های نهایی پایان درخت هستند. نقاطی که وضعیت داده را مشخص می‌کند (Breiman et al., 1984). در این تحقیق از درخت تصمیم<sup>1</sup> CART استفاده شد. درخت CART متغیرهای ورودی را برای یافتن بهترین نتیجه می‌آزماید تا شاخص ناخالصی حاصل از تجزیه کمترین مقدار باشد. در تجزیه دو زیرگروه تعیین می‌شود و هرکدام در مرحله بعد به دو زیرگروه دیگر تقسیم خواهند شد و این روند ادامه می‌یابد تا یکی از معیارهای توقف برآورده شود. درخت CART بازگشتی دوتایی است که گره‌های والدین را دقیقاً به دو گره منشعب می‌کند و به‌طور بازگشتی منشعب کردن را تا زمانی که انشعاب دیگری نتواند ساخته شود ادامه می‌دهد (Chattamvelli., 2011). در این تحقیق برای مدل‌سازی عملکرد محصول فندق با درخت CART از نرم‌افزار Statistica 13.5.0.17 استفاده گردید.

## ۲-۴- روش K- نزدیک‌ترین همسایه

برخلاف توابع انتقالی کلاسیک، روش نزدیک‌ترین همسایگی (KNN) از هیچ تابع ریاضی از پیش تعریف‌شده‌ای برای تخمین متغیرهای مختلف استفاده نمی‌کند. این روش یک طبقه‌بندی یادگیری با ناظر است 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 روش  $k$  تعداد نمونه مشابه را پیدا می‌کند و بر اساس آن‌ها ارزش نمونه مطالعه‌شده را پیش‌بینی می‌کند (Fang, 2009). 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:

<sup>1</sup> Classification and Regression Tree

$$R = \left[ \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sum_{i=1}^n \sqrt{(O_i - \bar{O})^2} \sum_{i=1}^n \sqrt{(P_i - \bar{P})^2}} \right]^2 \quad (7)$$

در روابط فوق n تعداد داده‌ها،  $O_i$  و  $P_i$ ، به ترتیب مقادیر اندازه‌گیری شده و پیش‌بینی شده عملکرد هستند. هر چه مقدار شاخص NSE به ۱ نزدیک‌تر باشد، خطای مدل کمتر است. هر چه شاخص RMSE به ۰ نزدیک‌تر باشد، نشان‌دهنده خطای کمتر و کارایی بهتر مدل است. مقدار شاخص ضریب همبستگی در دامنه ۱- تا ۱ تغییر می‌کند. مقدار ۱ و ۱- به ترتیب نشان‌دهنده همبستگی کامل و معکوس مقادیر مشاهداتی و شبیه‌سازی شده است. هر چه مقدار این شاخص به ۱ نزدیک‌تر باشد، کارایی مدل بهتر است. مدلی که دارای R و NSE بیشتر، RMSE کمتر باشد دقیق‌تر است (جدول ۱).

### ۳- نتایج و بحث

#### ۳-۱- توصیف آماری داده‌ها

خلاصه آماری ویژگی‌های آب و خاک و عملکرد محصول فندق در جدول ۲ آورده شده است. مشاهده می‌گردد متغیرهای حجم آب موردنیاز درخت در دور آبیاری، شاخص واکنش آب آبیاری، درصد سیلت خاک و شاخص واکنش خاک 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. درصد سیلت بیشترین ضریب همبستگی مثبت را دارد. به عبارت دیگر با افزایش درصد سیلت در خاک عملکرد محصول افزایش می‌یابد.

ساعت آبیاری، هدایت الکتریکی آب آبیاری، درصد ماسه، درصد رس، درصد شن و هدایت الکتریکی خاک و حداکثر تبخیر و تعرق روزانه همبستگی منفی با عملکرد محصول نشان دادند؛ این بدان مفهوم است که با کاهش مقدار هریک

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 سپس الگوی فعال‌سازی رسیده به لایه خروجی با الگوی صحیح خروجی مقایسه می‌شود تا سیگنال خطا محاسبه شود. سیگنال خطا برای هر الگوی خروجی هدف از لایه خروجی به نورون‌های ورودی به‌منظور تنظیم وزن لایه‌ها در شبکه به عقب انتشار می‌یابد. پس از مرحله آموزش، شبکه عصبی از الگوی ایجادشده برای طبقه‌بندی الگوهای جدید استفاده می‌شود. جهت مدل‌سازی با شبکه عصبی پیشرو با پس‌انتشار خطا (FFB) از نرم‌افزار Mathwork Matlab2023 استفاده گردید.

#### ۲-۶- مدل‌سازی رگرسیونی

هدف کلی مدل‌سازی رگرسیون خطی چندمتغیره (MLR<sup>۱</sup>)، پیدا کردن رابطه بین چند متغیر مستقل و یک متغیر وابسته است. به‌منظور انجام (MLR) به روش گام‌به‌گام، نرم‌افزار IBM SPSS Statistica 26 مورد استفاده قرار گرفت.

#### ۲-۷- ارزیابی مدل‌ها

برای ارزیابی و مقایسه کارایی مدل‌های بکار رفته در این تحقیق از سه شاخص نش-ساتکلیف<sup>۲</sup> (NSE)، جذر میانگین مربعات<sup>۳</sup> (RMSE) و ضریب همبستگی<sup>۴</sup> (R) استفاده شد که به ترتیب در روابط (۵)، (۶) و (۷) ارائه شده‌اند:

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (5)$$

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

<sup>3</sup> Root Mean Square Error

<sup>4</sup> Correlation Coefficient

<sup>1</sup> Multiple Linear Regression

<sup>2</sup> Nash Sutcliffe Coefficient

## جدول ۱ طبقه‌بندی کارایی مدل‌ها بر اساس شاخص‌های ارزیابی

Table 1 Efficiency classification of models based on evaluation indicators (Diaz-Ramirez et al., 2011)

Efficiency	R	RMSE	NSE
Very well <sup>a</sup>	0.93 < R < 1	0 ≤ RMSE ≤ 0.5SD*	0.75 (NSE < 1)
Well <sup>b</sup>	0.88 < R < 0.92	0.5SD < RMSE ≤ 0.6SD	0.65 < NSE < 0.75
Suitable <sup>c</sup>	0.81 < R < 0.87	0.6SD < RMSE ≤ 0.7SD	0.5 < NSE < 0.65
Unsuitable <sup>d</sup>	R < 0.8	RMSE > 0.7SD	NSE < 0.65

\*SD: Standard deviation, a: Very well, b: Good, c: Appropriate, d: Inappropriate

## جدول ۲ خلاصه‌ای از وضعیت آماری ویژگی‌های مورد بررسی در منطقه مطالعاتی

Table 2 Summary of descriptive statistics of investigated features in the study area

Variable	ET (mm/day)	V (m <sup>3</sup> )	H (hr)	PH <sub>w</sub>	EC <sub>w</sub> (dsm <sup>-1</sup> )	Clay (%)	Silt (%)	Sand (%)	PH <sub>s</sub>	EC <sub>s</sub> (dsm <sup>-1</sup> )	Y (Ton/ha)
Mean	3.46	1.95	116.3	7.47	448.8	24.75	31.92	43.11	7.22	0.61	1.29
Variance	0.47	0.20	4699.6	0.05	9223	94.19	40.48	184.74	0.02	0.02	0.01
Skewness	-0.18	-0.24	3.71	0.27	0.41	-0.60	-0.61	1.27	0.79	-0.05	-0.52
Kurtosis	1.50	2.38	18.48	-0.94	-0.75	-0.74	0.50	1.00	1.33	0.38	0.35
Minimum	2.00	1.00	33.00	7.06	300.0	4.00	17.10	24.60	6.91	0.31	1.00
Maximum	5.40	3.00	471.00	7.86	650.0	38.00	42.50	77.40	7.70	0.95	1.50
Coefficient Variation	19.8	23.08	58.93	3.05	21.40	39.22	19.93	31.53	2.13	20.82	8.03
Pearson's coefficients	-0.033	-0.02	0.170	0.278	-.44**	-0.070	0.296	-0.106	0.139	-0.13	

a: Mean, b: Standard deviation, c: Variance, d: Skewness, e: Erratum, \*: 1% significance level, \*\*: 5% significance level

برخوردار می‌باشند. نمونه‌های خاک برداشت شده دارای شاخص واکنش خاک متوسط ۷/۲۲ می‌باشند. هدایت الکتریکی خاک‌ها به‌طور متوسط در حدود ۰/۶۱ دسی زیمنس بر متر است که حاکی از پائین بودن مقادیر شوری آن‌ها است. میانگین درصد رس، درصد سیلت و درصد شن به ترتیب برابر ۳۱/۹۲ درصد، ۲۴/۷۵ درصد و ۴۳/۱۱ درصد است. همچنین برای نمونه‌های آب برداشت شده شاخص واکنش آب متوسط ۷/۴۷ می‌باشند.

هدایت الکتریکی آب به‌طور متوسط در حدود ۴۴۸/۸۸ دسی زیمنس بر متر است. میانگین حجم آب مورد نیاز برای هر درخت در دور آبیاری، ساعت آبیاری و حداکثر تبخیر و تعرق روزانه در باغات مورد بررسی به ترتیب برابر ۱۱۶/۳۳ مترمکعب، ۱/۹۵ ساعت و ۳/۴۶ میلی‌متر بر روز است. کمترین ضریب تغییرات مربوط به ویژگی شاخص واکنش خاک برابر ۲/۱۳ درصد است و بیشترین ضریب تغییرات مربوط به ویژگی حجم آب مورد نیاز برای هر درخت در دور آبیاری برابر با ۵۸/۹۳ است

در پژوهش Gomrokchi et al (2022) حجم آب مصرفی

از این متغیرها عملکرد محصول کاهش می‌یابد. هرچه ضریب همبستگی به عدد ۱- نزدیک‌تر باشد همبستگی قوی‌تر است؛ بنابراین هدایت الکتریکی آب آبیاری بیشترین ضریب همبستگی منفی را دارد. به عبارت دیگر با افزایش مقدار هدایت الکتریکی آب آبیاری عملکرد محصول افزایش می‌یابد. آب آبیاری عملکرد محصول افزایش می‌یابد.

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 آبیاری بیشترین ضریب همبستگی منفی را دارد. به عبارت دیگر با افزایش مقدار هدایت الکتریکی آب آبیاری عملکرد محصول افزایش می‌یابد. آب آبیاری عملکرد محصول افزایش می‌یابد.

Mohammadi (2018) بیان نمود تمامی ویژگی‌هایی که دارای ضریب چولگی بین ۱- تا ۱ باشند، از توزیع نرمال تبعیت می‌کنند. بر اساس ضریب چولگی به دست آمده، تمام ویژگی‌های محصول فندق به جز زمان آبیاری از توزیع نرمال

عصبی مصنوعی و مدل رگرسیون خطی چندمتغیره به ترتیب برابر  $0/67$  و  $0/57$  بود. (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. مدل‌سازی عملکرد پسته با استفاده از روش MLR نیز مبین این است که روش MLR تنها تغییرات اندکی از عملکرد را توجیه می‌نماید (Pour-Mohammadali et al, 2018). همچنین، در پیش‌بینی عملکرد محصول گندم با روش MLR (Dahmardeh et al, 2021) نشان داد که این مدل تنها توانسته ۴۰ درصد عملکرد کل را تبیین نماید.

همچنین جهت مدل‌سازی عملکرد محصول فندق مدل شبکه عصبی پیشرو با پس‌انتشار خطا مورد استفاده قرار گرفت. در جدول ۴ 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. روشن است که مدل شبکه عصبی مصنوعی از توانایی بیشتری نسبت به مدل رگرسیون خطی چندمتغیره در پیش‌بینی عملکرد محصول فندق برخوردار است. پژوهش Gomrokchi et al, (2015) این امر را در پیش‌بینی عملکرد محصول گندم تایید می‌کند. همچنین، تحقیق Pour-Mohammadali et al (2018) دقت مدل شبکه عصبی را برای پیش‌بینی عملکرد پسته، به میزان ۹۰ درصد بیان می‌کند. علاوه بر این، (Bahmani et al (2020) روشن ساخت که مدل شبکه عصبی مصنوعی پس‌انتشار به دلیل لحاظ کردن روابط غیرخطی بین عملکرد و عوامل تاثیرگذار بر آن، کارایی بهتری نسبت به رگرسیون چندمتغیره در تخمین عملکرد گل محمدی دارد. در تحقیق Jafari et al (2023) روش پس‌انتشار با مقادیر ضریب تبیین و جذر میانگین مربعات خطای برای مرحله آزمون به ترتیب

بیشترین همبستگی با عملکرد گندم را داشت. نتایج تحقیق (Dehmardeh et al (2021) نشان داد که رس، ماده آلی و سیلت خاک دارای بیشترین و واکنش خاک کمترین تاثیرگذاری بر اجزاء عملکرد گندم را دارند.

(Cox et al (2006) نیز در مورد شاخص واکنش خاک، کمترین ضریب تغییرات را در بین ویژگی‌های خاک به دست آورده است؛ اما (Bahmani et al (2022) بیشترین ضریب تغییرات را برای هدایت الکتریکی خاک به دست آوردند. مقدار ضریب تغییرات برای ویژگی‌های حداکثر تبخیر و تعرق روزانه، زمان آبیاری، شاخص واکنش آب، هدایت الکتریکی آب، درصد سیلت، درصد شن، شاخص واکنش خاک و هدایت الکتریکی خاک کمتر از ۳۵ درصد و دارای تغییرپذیری کم می‌باشند.

مقدار ضریب تغییرات برای درصد رس بیشتر از ۳۵ درصد و دارای تغییرپذیری زیاد می‌باشند. میانگین عملکرد محصول فندق برابر  $1/29$  تن در هکتار و دارای ضریب تغییرات  $8/03$  درصد است که این مقدار اندکی متاثر از تغییرپذیری زیاد خصوصیات فیزیکی و شیمیایی آب و خاک است.

### ۳-۲- نتایج مدل‌سازی

مدل رگرسیونی تخمین عملکرد فندق، جدول ۳، نشان می‌دهد که، پارامترهای هدایت الکتریکی آب آبیاری، درصد رس، درصد سیلت، درصد شن و هدایت الکتریکی خاک بر عملکرد فندق اثرگذار هستند.

پژوهش‌های پیشین نیز نشان می‌دهد که درصد سیلت خاک، مهم‌ترین عامل موثر بر عملکرد گندم دیم (Babaei et al, 2016) و بر عملکرد گل محمدی (Bahmani et al, 2020) است.

(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.

در تحقیق Gomrokchi et al (2015) نتایج ضریب تبیین مدل‌ها در پیش‌بینی عملکرد محصول گندم در مدل شبکه

جدول ۳ مدل رگرسیون پیش‌بینی عملکرد محصول فندق

Table 3 Regression model of Hazelnut yield prediction

Variable	Constant	ET (mm/day)	V (m <sup>3</sup> )	Hr (hr)	EC <sub>w</sub> (dsm <sup>-1</sup> )	Clay (%)	Silt (%)	Sand (%)	EC <sub>s</sub> (dsm <sup>-1</sup> )	PH <sub>w</sub>	PH <sub>s</sub>
R.C.	4.02	0.10	-0.18	-0.08	-0.24	-2.25	-1.23	-3.23	-0.25	0.07	0.04
Sig.	0.14	0.58	0.69	0.67	0.07	0.15	0.32	0.19	0.15	0.64	0.85

R.C: Regression coefficients, Sig: Values at 5% significance level 5%

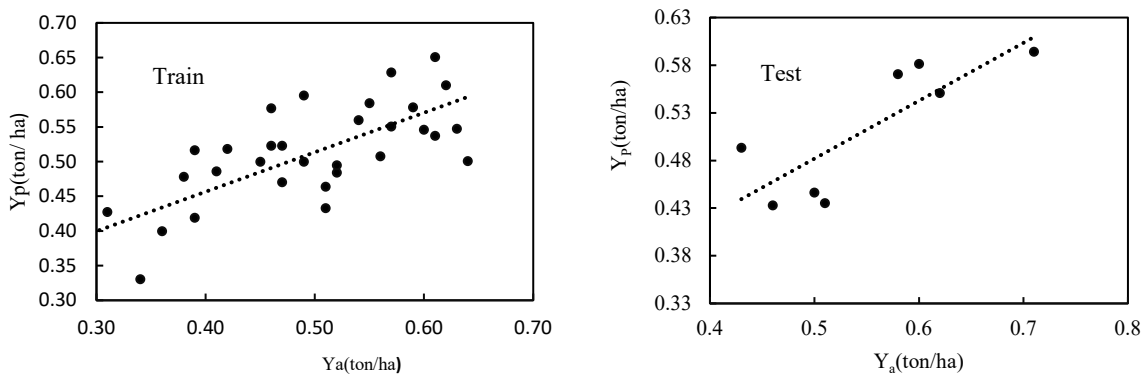


Fig. 2 Performance comparison MLR model for training and test data

شکل ۲ مقایسه عملکرد مدل MLR برای داده‌های واسنجی و صحت‌سنجی

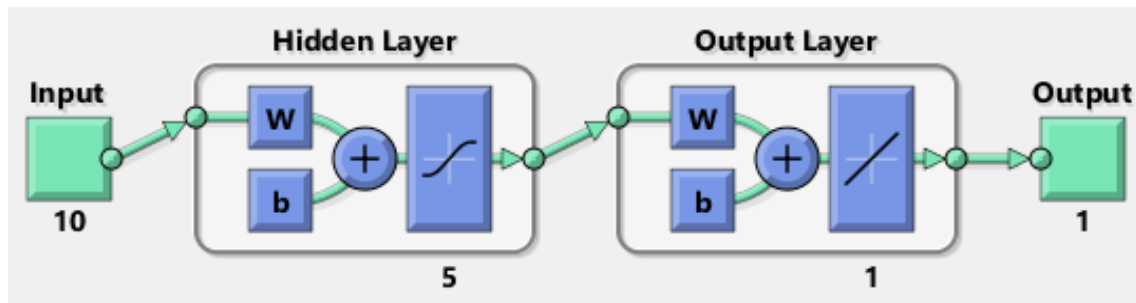


Fig. 3 FFB model architectur

شکل ۳ معماری مدل FFB

جدول ۴ مشخصات مدل FFB

Table 4 FFB model properties

Layers Properties						
Parameter	Efficiency Function	Learning Algorithm	Training Algorithm	layer Output (Linear)	Layer Hidden	layer Input (Sigmund)
Value	Mean Squared Error	Gradient Descent Momentum(GDM)		1	5	2
Function Parameters properties						
Min Gradient	Search	Parameter	Iterations	Slope Efficiency	Min Gradient	Search
Charambos	Charambos	Value	266	0.0029	Charambos	Charambos

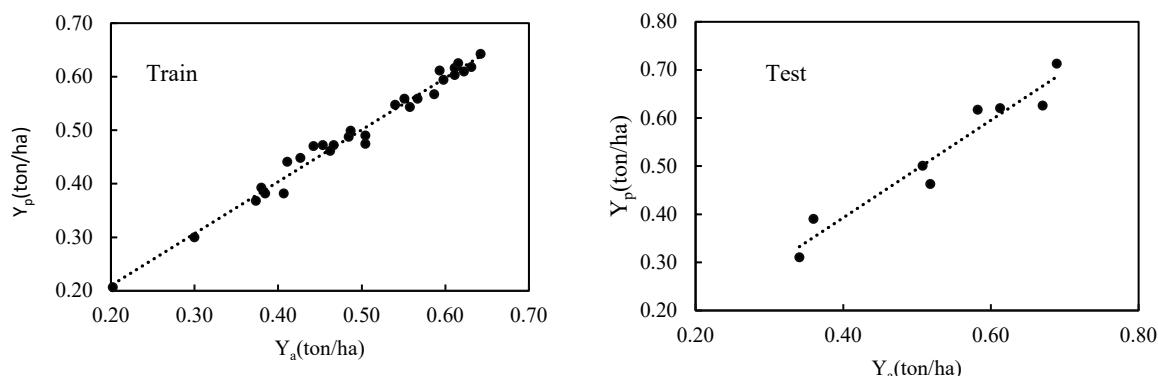


Fig. 4 Performance comparison FFB model for train and test data

شکل ۴ مقایسه عملکرد مدل FFB برای داده‌های واسنجی و صحت‌سنجی

پیش‌بینی عملکرد نیشکر از پارامترهای موثری نظیر مشخصات گیاه، هدایت الکتریکی آب و خاک، حداکثر درجه حرارت و میانگین ساعات آفتابی در طول دوره رشد محصول و در مقیاس زمانی هفت‌ساله استفاده گردید و نشان داد عملکرد مدل KNN، ۹۵/۵ درصد بوده است.

Poorsalehi et al (2019) بیان نمودند مدل درخت تصمیم نسبت به مدل KNN توانایی بیشتری در پیش‌بینی عملکرد دارد. ساختار درخت CART ایجاد شده جهت مدل‌سازی عملکرد محصول فندق در شکل ۶ نمایش داده شده است.

این درخت از ۱۳ گره تشکیل شده است. متغیرهای هدایت الکتریکی آب، حجم آب مصرفی برای هر درخت، تبخیر و تعرق روزانه و درصد شن در ساختار درخت ظاهر شده است. اولین متغیر انتخاب شده برای ایجاد شاخه در درخت هدایت الکتریکی آب است و مهم‌ترین پارامتر اثرگذار شناخته می‌شود.

در شاخه دوم از ویژگی‌های حجم آب مصرفی برای هر درخت، تبخیر و تعرق روزانه استفاده شده است. در شکل ۷ عملکرد مدل CART برای داده‌های واسنجی و صحت‌سنجی نشان داده شده است. نتایج این تحقیق با تحقیقات مشابه همخوانی دارد. در تحقیق Zakidizaji et al (2019) بر روی پیش‌بینی عملکرد نیشکر در استان خوزستان دقت مدل درخت CART برای داده‌های آزمایش ۸۱ درصد بود.

برابر ۰/۸۵ و ۰/۰۹ تن در هکتار به عنوان مدل بهینه در تخمین عملکرد گندم دیم شناخته شد.

Dahmardeh et al (2021) نتایج مقایسه مدل‌های مورداستفاده در پیش‌بینی اجزاء عملکرد گندم با استفاده از ویژگی‌های خاک نشان داد که بالاترین ضریب تبیین و کمترین ریشه میانگین مربع‌های خطا در تخمین عملکرد گندم مربوط به روش شبکه عصبی بود (ضریب تبیین برابر ۰/۷۶).

در تحقیق Gomrokchi et al (2022) نشان داد ضریب تبیین مدل پیش‌بینی عملکرد محصول گندم در مدل شبکه عصبی مصنوعی و مدل رگرسیون خطی چندمتغیره به ترتیب برابر ۰/۶۷ و ۰/۵۷ بود.

در این پژوهش جهت مدل‌سازی عملکرد محصول فندق مدل KNN با فرض بهترین تعداد همسایگی برابر ۳ توسط تابع فاصله‌سنجی اقلیدسی ترکیب کارآمد مشخص گردید.

در شکل ۵ عملکرد مدل KNN برای داده‌های واسنجی و صحت‌سنجی نشان داده شده است. ارزیابی نتایج این تحقیق با نتایج محققان دیگر همخوانی دارد. نتایج تحقیق Ansari (2023) نشان داد روش KNN نسبت به روش‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن بهترین دقت را داشت 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) به‌منظور

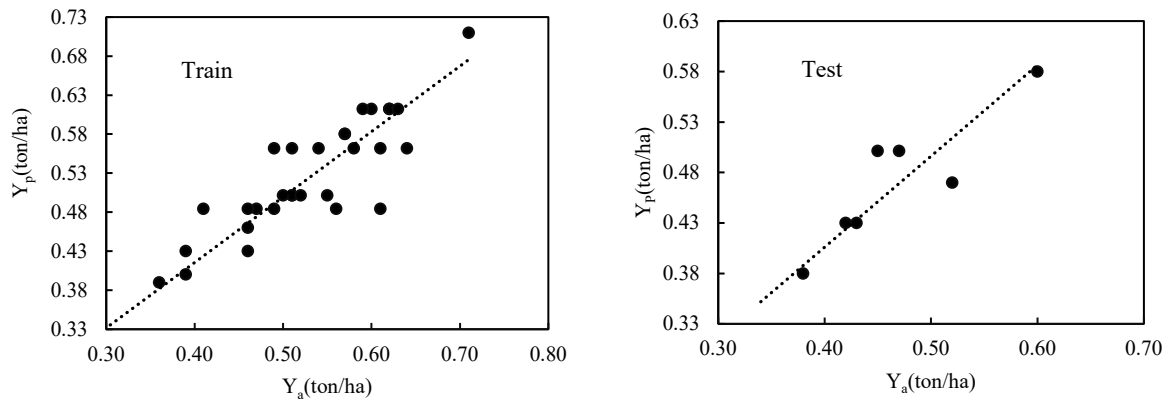


Fig. 5 Performance comparison KNN model for training and test data

شکل ۵ مقایسه عملکرد مدل KNN برای داده‌های واسنجی و صحت‌سنجی

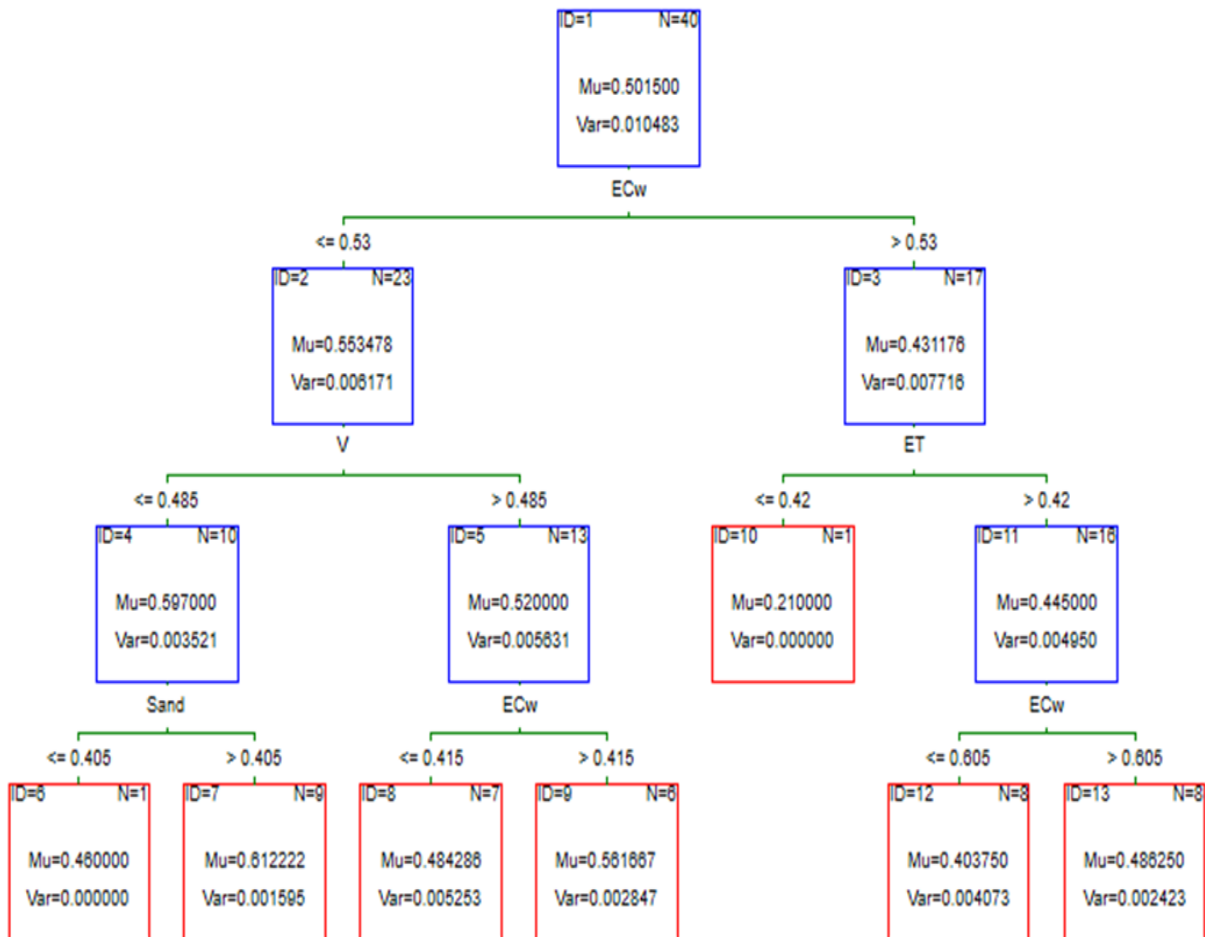


Fig. 6 Tree structure of the model CART

شکل ۶ ساختار مدل CART

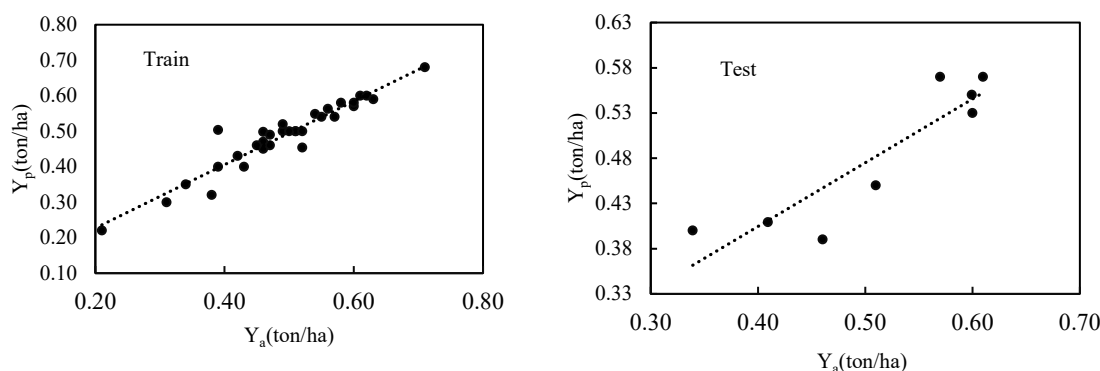


Fig. 7 Performance comparison CART model for training and test data

شکل ۷ مقایسه عملکرد مدل CART برای داده‌های واسنجی و صحت‌سنجی

#### جدول ۵ ارزیابی مدل‌ها

Table 5 Evaluation of models

Parameter	Train			Test		
	NSE	RMSE	R	NSE	RMSE	R
KNN	0.9	0.031	0.95	0.7	0.018	0.9
CART	0.86	0.041	0.92	0.85	0.029	0.93
FFB	0.98	0.014	0.99	0.96	0.023	0.98
MLR	0.72	0.71	0.57	0.47	0.63	0.67

به‌خوبی بر هم منطبق شده‌اند. مدل MLR کمترین دقت را در مدل‌سازی عملکرد محصول فندق داشت. به‌طور کلی نتایج کارایی مدل‌ها در ارزیابی تخمین عملکرد محصول فندق با نتایج محققان دیگر همخوانی دارد.

#### ۴- نتیجه‌گیری

در این پژوهش با استفاده از روش‌های داده‌کاوی درخت تصمیم CART، K- نزدیک‌ترین همسایه، شبکه عصبی پیشرو با پسانتشار خطا و رگرسیون خطی چندمتغیره عملکرد محصول فندق برای باغات فندق مجهز به سیستم آبیاری تحت فشار در استان گیلان برآورد گردید. ورودی روش‌های داده‌کاوی شامل داده‌های آب و خاک (حداکثر تبخیر و تعرق، ساعات آبیاری، حجم آب آبیاری در هر دور آبیاری درخت، شاخص واکنش خاک و هدایت الکتریکی آب آبیاری، شاخص واکنش و هدایت الکتریکی خاک، درصد سیلت و درصد رس و درصد ماسه بود. با ارزیابی نتایج مشخص شد که روش شبکه عصبی پیشرو با روش پسانتشار خطا، نتایج دقیق‌تری نسبت به مدل‌های درخت تصمیم، K- نزدیک‌ترین همسایه و رگرسیون خطی

نتایج حاصل از پژوهش Sadr and Eslami (۲۰۲۱) نشان داد که درخت تصمیم می‌تواند روابط پیچیده بین متغیرهای مختلف اقلیمی با عملکرد محصول پسته را بهتر از رگرسیون خطی استخراج کند و مهم‌ترین متغیرهای موثر را با دقت قابل قبولی مشخص کند

همچنین در تحقیق Monjezi (۲۰۲۱) مدل درختی CART توانست با ضریب همبستگی ۰/۹۵ عملکرد بهتری در پیش‌بینی مقاومت نفوذ خاک مزارع نیشکر داشته است. در [جدول ۵](#)، نتایج ارزیابی دقت مدل‌ها KNN، FFB، CART، MLR برای داده‌های واسنجی و صحت‌سنجی آورده شده است.

برای داده‌های صحت‌سنجی مقادیر شاخص NSE به ترتیب ۰/۹۶، ۰/۸۵، ۰/۷، ۰/۴۷ و مقادیر شاخص RMSE به ترتیب ۰/۰۱۸، ۰/۰۲۳، ۰/۰۲۹، ۰/۰۶۳، ۰/۰۶۷ و مقادیر شاخص R به ترتیب ۰/۹۸، ۰/۹۳، ۰/۹۰ و ۰/۶۷ است. نتایج نشان می‌دهد که عملکرد مدل‌های FFB، KNN، CART خوب بوده اما مدل FFB در مدل‌سازی بسیار خوب بوده به‌طوری‌که داده‌های پیش‌بینی شده و اندازه‌گیری شده

## ۵- فهرست نشانه‌ها

$X_S$	مقدار استاندارد شده
$X_i$	مقدار واقعی
$Y_a$	عملکرد اندازه‌گیری شده (Ton/ha)
$Y_p$	عملکرد محصول پیش‌بینی شده (Ton/ha)
$PH_s$	شاخص واکنش خاک
$PH_w$	شاخص واکنش آب
$EC_s$	هدایت الکتریکی خاک ( $dsm^{-1}$ )
$EC_w$	هدایت الکتریکی آب ( $dsm^{-1}$ )
Hr	ساعات آبیاری (hr)
V	حجم آب مصرفی ( $m^3$ )
ET	حداکثر تبخیر و تعرق روزانه (mm/day)

چندمتغیره ارائه می‌کند. روش درخت تصمیم CART 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 درخت تصمیم به همین علت است. مدل رگرسیون خطی چندمتغیره نیز به واسطه غیرخطی بودن ویژگی‌های آب و خاک نمی‌تواند نتایج دقیقی ارائه کند؛ بنابراین روش‌های داده‌کاوی به‌خصوص روش‌های هوش مصنوعی مانند شبکه عصبی پیشرو می‌تواند به عنوان ابزاری قوی در زمینه عملکرد محصول معرفی شود.

## Acknowledgments

We would like to express our deepest gratitude and appreciation for the cooperation of the Agricultural Jihad Organization of Guilan Province in providing the information required for this research.

## Credit Authorship Contribution Statement

Conceptualization, M.Y., F.L., and A.A.; methodology, M. Y., F.L., and A.A.; data analysis, M.Y., F.L., and A.A.; writing—original draft preparation, M.Y. and A.A.; writing—review and editing, M. Y., F.L., and A.A.; supervision, F.L.. All authors have read and agreed to the published version of the manuscript.

## Funding

No financial support was received from any organization for this research.

## Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors upon request.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

Ahmadi, M. Kaviani, A. Ramezani Etedali, H. and Tavakoli, A. (2023). Simulation of Rainfed Wheat Yield Using Drought Indices by

Employing Artificial Neural Network, Random Forest and Support Vector Regression(Case Study: Saqez City). Iran-Water Resources Research.19(4),95-109, [DOI: 10.22034/jwrr.2023.177606](https://doi.org/10.22034/jwrr.2023.177606).

Arumugam, A. (2017). A predictive modeling approach for improving paddy crop productivity using data mining techniques. Turkish. Journal of Electrical Engineering & Computer Sciences.25(6), 4777-4787, [DOI: 10.3906/elk-1612-361](https://doi.org/10.3906/elk-1612-361).

Ayoubi, SH.A. and Jalalian, A. (2010). Land Evaluation (Agricultural and Natural Resources Second Edition), Isfahan University of Technology Publication Center, Isfahan, Iran, 385p. (In Persian)

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), [Doi: 10.22059/ijswr.2022.342638.669260](https://doi.org/10.22059/ijswr.2022.342638.669260).

Anaraki, V. Mousavi, M. Mousavi, F. Farzin, S. and 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.51(2),325-339. [DOI: 20.1001.1.2008479.1399.51.2.4.5](https://doi.org/20.1001.1.2008479.1399.51.2.4.5).

Badde, D.S. Gupta, A.k. and 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 (IOSR-JMCE)*, 1-6.
- Bahmani M, Mohammadi J, Esfandiarpour Borouleni I and Mottaghian H.R. (2020). Assessment of relationship between Rose yield and soil and topography properties by using multivariate regression, artificial neural network and structure equation modeling. *Applied Soil Research*, 8(2), 142-157.
- Breiman, L. Friedman, J. Olshen, R. and Stone, Ch. (1984). *Classification and regression trees*. Boca Raton: Chapman & Hall/CRC.
- Babaei, F. Vaezi A. Teheri, M. Zarrinabadi, E. and 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, DOI: [10.22059/ijswr.2015.56795](https://doi.org/10.22059/ijswr.2015.56795).
- Brown, ME. Lary, DJ. Vrieling, A. Stathakis, D. and Mussan, H. (2008). Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. *The International Journal of Remote Sensing*, 29,7141-7158, DOI: [10.1080/01431160802238435](https://doi.org/10.1080/01431160802238435).
- Chattamvelli, R. (2011). *Data Mining Algorithms*, First Edition. Alpha Science International. 8: 274-290.
- Dahmardeh, A.R. Shahriari, A. Pahlavan-Rad, M. Shabani, A. and 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), DOI: [10.22055/AGEN.2021.35343](https://doi.org/10.22055/AGEN.2021.35343).
- Diaz-Ramirez, J. N., W. H. McAnally, J. L. Martin. (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. November 2011 DOI: [10.13031/2013.40627](https://doi.org/10.13031/2013.40627)
- Emamgholizadeh, S. Parsaeian, M. and Baradaran M. (2015). Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*. 68, 89-96, DOI: [10.1016/j.eja.2015.04.010](https://doi.org/10.1016/j.eja.2015.04.010).
- Eslami, R. Azarnoosh, M. Kialashki, A. and 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, DOI: [10.52547/ifej.9.17.185](https://doi.org/10.52547/ifej.9.17.185)
- SID. <https://sid.ir/paper/385504/en>.
- Emami, S. Dehghanisani, H. and Rezaie, M. (2021). Application of artificial neural network models in estimating nectarine crop yield under two-sided furrow irrigation. *Water and Soil Management and Modeling*.1(2), 47-59, DOI: [10.22098/mmws.2021.9090.1019](https://doi.org/10.22098/mmws.2021.9090.1019).
- Fang, X. (2009). Are you becoming a diabetic a data mining approach. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. Tianjin, China: IEEE, 18-22.
- Hafezi, N. Bahrami, H. Seikh Davoodi, M. Alavi, E. (2021). Hybrid Artificial Neural Network with Meta-heuristic Algorithms for Predicting Sugarcane Yield, *Iranian Journal of Biosystems Engineering*.51(3), 515-526, DOI: [10.22059/ijbse.2020.290905.665234](https://doi.org/10.22059/ijbse.2020.290905.665234).
- Gomrokchi, A.U. Baghani, J. Abbasi, F. (2022). Evaluating the Capability of Data Mining Models in Predicting Irrigated Wheat Yield in Iran. *Journal of Water and Soil*. (35) 2, 189-202, DOI: [20.1001.1.20084757.1400.35.2.4.1](https://doi.org/20.1001.1.20084757.1400.35.2.4.1).
- Jafari, M.M. Liaghat, A. Mohtasham. (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 and Drainage*. 3(17), 529-544, DOI: [20.1001.1.20087942.1402.17.3.10.8](https://doi.org/20.1001.1.20087942.1402.17.3.10.8).
- Kumar, M. Raghuvanshi, N.S. Wallender, W.W. and Pruitt, W.O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*. 128(4), 224-233, DOI: [10.1061/\(ASCE\)0733-9437\(2007\)133:2\(83\)](https://doi.org/10.1061/(ASCE)0733-9437(2007)133:2(83)).
- Jagtap SS. Lall U. Jones, JW. Gijsman AJ. Ritchie JT. (2004). Dynamic nearest-neighbor method for estimating soil water parameters. *Trans. ASAE*. 47,1437-1444. DOI: [10.13031/2013.17623](https://doi.org/10.13031/2013.17623)
- Lukas, F. Zuzana, S. Maria, D. Beata, H. and Tatian, P. (2015). Application of Neural Network Models in Modelling Economic Time Series with Non-Constant Volatility, *Business Economics and Management*. Conference BEM2015, *Procedia Economics and Finance*. 34, 600 – 607, DOI: [10.1016/S2212-5671\(15\)01674-3](https://doi.org/10.1016/S2212-5671(15)01674-3).
- 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, DOI: [10.22059/IJBSE.2019.289835.665228.CV](https://doi.org/10.22059/IJBSE.2019.289835.665228.CV).
- Moodi, Y. Eslami, E. Mousavi, R. Dizangian, B. and 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), 58-7710,

- [DOI:22065/JSCE.2021.178649.1825.0](https://doi.org/10.22065/JSCE.2021.178649.1825.0)  
Mesri, S. Ghorbani Dashtaki, Sh. Shirani, H. Kamkar Rohani, A and Motaghian, H.R. (2020). Separating Effective Soil Properties on Moisture Characteristic Curve Using Decision tree. *J. of Soil Management and Sustainab.*110(3), [DOI: 10.22069/ejsms.2021.17673.1928](https://doi.org/10.22069/ejsms.2021.17673.1928).
- Mohammadi J. (2018). *Pedomining, Volume 9, Structure Equation Modeling*. Pelk Publishing, 396 pages. (In Persian).
- Pourmohamadali, B. Salehi, M.H. Hosseinifard, S.J Shirani, H. Esfandiarpour Borujeni, I. (2018). Pistachio yield prediction using multiple linear regression and artificial neural network (A Case Study: Rafsanjan and Anar regions, Kerman Province). *journal of Agricultural Engineering*.40(2), 71-78, [DOI:10.22055/agen.2018.17814.1274](https://doi.org/10.22055/agen.2018.17814.1274).
- Sivaneasan, B, Yu, C.Y, Goh, K.P, (2017). Solar Forecasting Using ANN with Fuzzy Logic Pre-Processing, *Energy Procedia*. 143, 727-732. [DOI:10.1016/j.egypro.2017.12.753](https://doi.org/10.1016/j.egypro.2017.12.753)
- Salajegheh, A. Fathabadi, A. and Gholami, H. (2010). Predict river discharge using the nearest neighbor. 5th national conference on science and management engineering Iran. Gorgan University of Agricultural Sciences and Natural Resources. (In Persian).
- Sattari, MT. Rezazadeh Joudi, A. and Kusiak, A. (2015). Estimation of water quality parameters with data-driven models. *American Water Works Association*.108(4), 232-239 [DOI:10.5942/jawwa.2016.108.0012](https://doi.org/10.5942/jawwa.2016.108.0012)
- Zheng, K. G. Q. L. J. Z. J. X. W. M. (2017). Applications of support vector machine and improved k-Nearest Neighbor algorithm in fault diagnosis and fault degree evaluation of gas insulated switchgear. 1st International Conference on Electrical Materials and Power Equipment, Xian, Pepler R China.
- Sadr, S. and Eslami, M. Determination of effective weather variables on pistachio yield using C and R decision tree algorithm, *Journal of Agricultural Meteorology*9(1):53-2, [DOI: 10.22125/agmj.2021.248249.1102](https://doi.org/10.22125/agmj.2021.248249.1102).
- Tashakkori, F. Mohammadi, Torkashvand, A. Ahmadi, A. and Esfandiari, M. (2021). Prediction of Saffron Yield based on Soil Properties Using Regression and Artificial Neural Network Models in the Vamenan Region of the Golestan Province. *Saffron Agronomy and Technology*. 9(2),159-175, [DOI:10.22048/jsat.2020.240519.1404](https://doi.org/10.22048/jsat.2020.240519.1404).
- Xindung, W. and Kumar, V. (2009). *Top Ten Algorithm in Data Mining*, First Edition, Taylor & Francis Group, USA.
- Yue, J. Feng, H. Yang, G. and Li, Z. (2018). A comparison of regression techniques for estimation of aboveground winter wheat biomass using near-surface spectroscopy. *Remote Sensing*. 10(1),66, [DOI:10.3390/rs10010066](https://doi.org/10.3390/rs10010066).