

Non-Destructive Prediction of pH in Khormai and Khoni Plums Using a Hyperspectral Imaging System and Machine Learning Methods

M. Latifi-Amoghin¹ , Y. Abbaspour-Gilandeh^{1*} 

1- Department of Biosystems Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

(*- Corresponding Author Email: abbaspour@uma.ac.ir)

<https://doi.org/10.22067/jam.2025.95325.1427>

Introduction

Traditional methods for evaluating fruit quality, such as pH measurement, are often destructive, time-consuming, and costly, leading to product loss and reduced efficiency in the supply chain. The growing need for rapid, accurate, and non-destructive methods makes the use of technologies like Hyperspectral Imaging (HSI) essential. HSI combines two-dimensional imaging with spectroscopy to simultaneously acquire spatial and spectral information from an object. Numerous studies have shown that this method is capable of accurately estimating internal fruit parameters in a non-destructive manner. The objective of this research was to develop a fast and reliable method for the non-destructive estimation of pH in two plum cultivars using HSI and machine learning algorithms such as Partial Least Squares Regression (PLSR) and Artificial Neural Networks (ANN). This study aims to overcome the limitations of conventional methods by leveraging the power of advanced imaging and computational techniques, providing a sustainable and efficient solution for the fruit industry.

Materials and Methods

In this study, 80 samples from each of the Khormaei and Khoni plum cultivars were used, which were purchased from local orchards. The samples were uniform in size, shape, and colour and were free from any physical damage. Hyperspectral images of the samples were acquired using a rotating hyperspectral imaging system in the range of 418 to 1072 nm. The pH of each fruit juice sample was measured using a digital pH meter. In the analysis of spectral data, the initial part of the spectrum was first removed due to high noise, and then the remaining data were processed with preprocessing methods such as a Gaussian filter and Multiplicative Scatter Correction (MSC). To select effective wavelengths (EWs), a hybrid approach using a Decision Tree (DT) and five metaheuristic algorithms was employed, with the Particle Swarm Optimisation (PSO) algorithm showing the best performance. Finally, pH modelling was performed on the selected wavelengths using PLSR and ANN. This comprehensive methodology ensures that the models are trained on high-quality data and are optimised for maximum accuracy.

Results and Discussion

Spectral analysis showed that the reflectance spectra of the Khoni and Khormaei plums had a high degree of variation, which is related to the differences in their chemical composition and structure. Descriptive statistics indicated that the average pH of Khormaei plum (3.909) was higher than that of Khoni plum (3.7375), and the pH range of Khoni plum (3.15 to 4.44) was wider than that of Khormaei plum (3.6 to 4.2). The results showed that modelling with ANN on the wavelengths selected by PSO, especially for Khoni plum, significantly increased prediction accuracy. The best ANN model for Khoni plum achieved an R^2 of 0.9834 and an RPD of 8.01, which indicates the outstanding accuracy of this method. For the Khormaei plum, the best ANN model also reached an R^2 of approximately 0.76 and a Ratio of Performance to Deviation (RPD) of 2.12, showing a considerable improvement over the PLSR model. The superior performance of the ANN models can be attributed to their ability to capture complex, non-linear relationships between spectral data and pH values, which linear models like PLSR may miss.

Conclusion

This research successfully demonstrated that hyperspectral imaging, in combination with machine learning algorithms, particularly ANN and PSO, can be an accurate and reliable method for the non-destructive prediction of pH in different plum cultivars. The hybrid approach used in this study, which combined DT for initial feature selection with PSO for optimal wavelength selection, enabled the models to predict pH values with very high accuracy, especially for the Khoni plum cultivar. This method can be used as an efficient tool in post-harvest quality control processes, helping to reduce waste and improve efficiency in the fruit supply chain. This work

paves the way for the development of smart grading and sorting systems that can quickly and accurately assess fruit quality, benefiting both producers and consumers.

Keywords: Hyperspectral imaging, Metaheuristic algorithms, Particle Swarm Optimisation, pH, Plum

پیش‌بینی غیرمخرب pH آلو خرمایی و خونی با به‌کارگیری سامانه تصویربرداری فراطیفی و روش‌های یادگیری ماشین

میثم لطیفی عموقین^۱، یوسف عباس‌پور گیلانده^{۱*}

تاریخ دریافت: ۱۴۰۴/۰۶/۱۹

تاریخ پذیرش: ۱۴۰۴/۰۸/۰۴

چکیده

این پژوهش با هدف توسعه و ارزیابی یک روش غیرمخرب برای پیش‌بینی pH دو رقم آلو (خرمایی و خونی) با استفاده از سامانه تصویربرداری فراطیفی در حالت بازتابی انجام شد. اثر پیش‌پردازش‌های مختلف طیفی بر عملکرد مدل رگرسیون حداقل مربعات جزئی بر روی طیف کامل داده‌ها بررسی شد. در هر دو رقم، بهترین عملکرد این الگوریتم با فیلتر گوسی در مرحله آزمون با $R^2=0/8203$ ، $RMSE=0/059$ و $RPD=2/44$ برای رقم خرمایی و $R^2=0/8819$ ، $RMSE=0/1411$ و $RPD=3/01$ برای رقم خونی به ثبت رسید. در ادامه به‌عنوان یک نوآوری از ترکیب الگوریتم درخت تصمیم با پنج الگوریتم فراابتکاری برای کاهش ابعاد داده‌ها استفاده شد که میان آن‌ها الگوریتم بهینه‌سازی ازدحام ذرات عملکرد بهتری در انتخاب ویژگی داشت. در نهایت، مدل‌سازی pH بر روی طول‌موج‌های منتخب با استفاده از مدل‌های رگرسیون حداقل مربعات جزئی و شبکه‌ی عصبی مصنوعی صورت پذیرفت. نتایج نشان داد که شبکه‌ی عصبی مصنوعی بر روی طول‌موج‌های منتخب الگوریتم بهینه‌سازی ازدحام ذرات، به‌ویژه برای آلو خونی، منجر به دقت‌های پیش‌بینی به‌مراتب بالاتری شد؛ به‌طوری‌که بهترین عملکرد این الگوریتم برای آلو خونی (با فیلتر گوسی) در مرحله آزمون به $R^2=0/9834$ و $RMSE=0/0512$ و $RPD=8/01$ دست‌یافت، که نشان‌دهنده توانایی قابل‌توجه این روش برای اندازه‌گیری کمی دقیق است. برای آلو خرمایی نیز، بهترین عملکرد این الگوریتم (با داده‌های بدون پیش‌پردازش و فیلتر میانه) در مرحله آزمون به $R^2=0/76$ و $RPD=2/12$ رسید که بهبود قابل‌توجهی نسبت به الگوریتم رگرسیون حداقل مربعات جزئی نشان داد. این نتایج، نشان‌دهنده پتانسیل بالا این روش جهت استقرار در خطوط فرآوری برای کنترل کیفیت میوه‌ها طی عملیات پس از برداشت می‌باشد.

واژه‌های کلیدی: آلو، الگوریتم‌های فراابتکاری، بهینه‌سازی ازدحام ذرات، تصویربرداری فراطیفی، pH

مقدمه

آلو سرشار از ویتامین‌ها و مواد معدنی مفید است به‌طور خاص منبع خوبی از ویتامین‌های A و C و پتاسیم به‌شمار می‌آید (Ayub et al., 2023). یکی از پارامترهای مهم آلو که نقش زیادی در ماندگاری و نگهداری دارد pH یا اسیدیته می‌باشد کاهش اسیدیته می‌تواند علاوه بر تغییر رنگ، روند رشد میکروارگانیسم‌ها را نیز تسریع کند و به همین دلیل کنترل pH در طول نگهداری از اهمیت ویژه‌ای برای حفظ کیفیت و ماندگاری آلو برخوردار است (Majeed & Jawandha, 2016).

pH به‌عنوان تابعی مستقیم از یون‌های هیدروژن آزاد موجود در یک ماده غذایی تعریف می‌شود و معیاری اساسی برای اسیدیته آزاد است. این پارامتر ذاتی، به‌طور قابل‌توجهی بر جنبه‌های مختلف کیفیت میوه، از جمله طعم، بافت، ماندگاری و مناسب بودن برای فرآوری تأثیر می‌گذارد (McGlynn, 2010). این پارامتر ذاتی، یک شاخص کلیدی

آلو (*Prunus domestica* L.)، گیاهی از تیره گل‌سرخ‌ها (*Rosaceae*) و جنس *Prunus* می‌باشد این درختچه یا درخت کوچک بومی مناطق معتدل آسیا و اروپا می‌باشد و به‌طور گسترده در مناطق معتدل جهان کشت می‌شود در تیره گل‌سرخ‌ها انواع مختلف میوه‌های هسته‌دار از جمله زردآلو، هلو، شلیل و آلبالو جای دارند و گونه‌های دیگری مانند *P. salicina* (آلوی ژاپنی) و *P. americana* (آلوی آمریکایی) نیز به‌عنوان آلو شناخته می‌شوند (Trendafilova et al., 2022). آلو یک میوه مهم باغبانی است که به‌صورت تازه مصرف می‌شود و بخش عمده‌ای از تولید آن به‌صورت خشک (آلوچه یا prune) و فرآورده‌های جانبی عرضه می‌گردد. برای نمونه، در ایالات متحده گونه‌های اروپایی آلو عمدتاً برای تولید آلوچه به‌کار می‌روند این میوه از نظر تغذیه‌ای دارای ارزش بالایی است همچنین

(*)- نویسنده مسئول: (Email: abbaspour@uma.ac.ir)

<https://doi.org/10.22067/jam.2025.95325.1427>

۱- گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران

در سال‌های اخیر روش‌های طیفی غیرمخرب مانند تصویربرداری فراطیفی (HSI)، که تصویربرداری دو بعدی سنتی را با طیف‌سنجی ترکیب می‌کند، به‌طور گسترده‌ای برای بررسی ویژگی‌های داخلی میوه‌ها به‌کار گرفته شده‌اند. استفاده از این روش‌ها موجب حفظ یکپارچگی نمونه و جلوگیری از ضایعات محصول می‌شود (Huang, Liu, & Ngadi, 2014).

مطالعات متعددی نشان داده‌اند که HSI قادر به تخمین غیرمخرب پارامترهای داخلی میوه‌ها با دقت بالا است. به‌عنوان نمونه، گل‌محمدی و همکاران (Golmohammadi, Tahmasebi, & Razavi, 2022) با استفاده از HSI بازتابی در محدوده ۴۰۰-۱۱۰۰ نانومتر موفق شدند pH سیب رد دلشیز را طی دوره نگهداری با خطای بسیار کم ($RMSE=0.02$) و ضریب تعیین تقریباً ۰/۹۸ مدل‌سازی کنند همچنین لیو و همکاران (Liu et al., 2021) با به‌کارگیری روش تصویربرداری فراطیفی و الگوریتم‌های یادگیری ماشین ضریب همبستگی $R=0.829$ و $RMSE=0.07$ را برای پیش‌بینی اسیدیته (pH) آلو سبز گزارش کرده‌اند که نشان‌دهنده توانایی بالای این فناوری در تشخیص غیرمخرب اسیدیته است به‌طور مشابه، وانگ و همکاران (Wang et al., 2022) با استفاده از سیستم HSI فلورسانس، pH میوه کیوی را با مدل SVR چندهسته‌ای پیش‌بینی کردند و $R^2=0.851$ را برای مجموعه پیش‌بینی به‌دست آوردند و همچنین به‌طور مختصر می‌توان به تعدادی از پژوهش‌ها که برای تشخیص یا پیش‌بینی pH میوه‌ها با استفاده از تصویربرداری فراطیفی انجام گرفته‌اند اشاره کرد؛ از جمله سیب فوجی (Zhiming, Wenqian, Liping, Yankun, & Xiu, 2014)، انبه (Purwanto et al., 2013)، میوه لیچی (Pu, Liu, Wang, & Sun, 2016)، سیب قرمز و زرد (Hasanzadeh et al., 2022)، گلابی (Zhang et al., 2024)، کیوی (Zhu et al., 2017)، گوجه (Rahman et al., 2017)، لنگور قرمز (Gao & Xie, 2024)، توت‌فرنگی (Weng, Yu, Guo, Tang, & Liang, 2020) که قابلیت اتکای این روش را در تعیین غیرمخرب pH تایید می‌کنند در کارهای دیگر نیز از تصویربرداری فراطیفی به‌منظور پیش‌بینی میزان جامدات محلول^۲ (SSC)، سفتی بافت و رطوبت در انواع میوه‌ها (از جمله بلوبری، سیب، هلو، انبه، ملون و گلابی) استفاده شده است (Wang et al., 2022).

با توجه به تحقیقات انجام‌گرفته در این زمینه، هدف این مطالعه، توسعه یک روش سریع و غیرمخرب برای تخمین pH میوه ارقام مختلف آلو با استفاده از تصویربرداری فراطیفی و به‌کارگیری الگوریتم‌های یادگیری ماشین از جمله رگرسیون حداقل مربعات

بلوغ و توسعه طعم است؛ به این صورت که در طول فرآیند رسیدن، pH میوه به‌طور کلی افزایش و اسیدیته کل کاهش می‌یابد، که نشان‌دهنده‌ی تبدیل اسیدهای آلی به قندها و تغییرات در ویژگی‌های حسی می‌باشد. همچنین، pH به‌طور مستقیم بر ماندگاری، بافت و مناسب بودن برای فرآوری تأثیر می‌گذارد، زیرا کاهش اسیدیته می‌تواند روند رشد میکروارگانیسم‌ها و تغییرات رنگ را تسریع کند؛ بنابراین، نظارت بر pH برای کنترل کیفیت مؤثر در سراسر زنجیره تأمین آلو حیاتی است (Majeed & Jawandha, 2016).

روش‌های سنتی برای ارزیابی پارامترهای کیفیت داخلی میوه مانند pH ذاتاً مخرب هستند. این امر معمولاً شامل نفوذ فیزیکی، له کردن یا آب‌گیری میوه است. ماهیت مخرب به این معنی است که هر میوه آزمایش‌شده غیرقابل فروش می‌شود و منجر به از دست دادن مستقیم محصول و افزایش ضایعات در زنجیره تأمین می‌شود. این روش‌ها همچنین معمولاً زمان‌بر و کار فشرده هستند و اغلب به پرسنل ماهر نیاز دارند. این امر آن‌ها را برای ارزیابی کیفیت سریع و با توان بالا در خطوط بسته‌بندی تجاری یا در تأسیسات فرآوری در مقیاس بزرگ غیرعملی می‌کند (Hasanzadeh et al., 2022). توصیف مداوم روش‌های سنتی به‌عنوان پرهزینه، زمان‌بر و مخرب نشان می‌دهد که این‌ها نه تنها ناراحتی‌های جزئی، بلکه موانع اقتصادی و عملیاتی قابل‌توجهی هستند. از دست دادن محصول به دلیل تخریب، همراه با تجزیه و تحلیل آهسته، مستقیماً بر سودآوری تأثیر می‌گذارد و مانع از مرتب‌سازی و درجه‌بندی کارآمد در زمان واقعی در زنجیره تأمین می‌شود. این امر منجر به استفاده ناکارآمد از منابع و افزایش ضایعات می‌شود. بنابراین، محدودیت‌های ذاتی روش‌های مخرب ارزیابی کیفیت، یک الزام تجاری قوی برای اتخاذ فناوری‌های غیرمخرب ایجاد می‌کند، زیرا آن‌ها پتانسیل کاهش قابل توجه ضایعات، بهبود کارایی عملیاتی و افزایش سودآوری کلی در تولید و توزیع میوه در مقیاس بزرگ را ارائه می‌دهند (Liu et al., 2025).

رشد مداوم تجارت جهانی میوه و انتظارات فزاینده مصرف‌کنندگان برای طعم، تازگی و ایمنی، نیاز به تغییر به سمت تکنیک‌های ارزیابی سریع، دقیق و غیرمخرب را ضروری می‌سازد که بتوانند کیفیت میوه را با دقت در هر مرحله، از کشت در باغ تا توزیع در بازار، مدیریت کنند. فراتر از کارایی صنعتی، تغییر به سمت آزمایش غیرمخرب مستقیماً به خواسته‌های مصرف‌کننده برای کیفیت، تازگی و ایمنی پایدار پاسخ می‌دهد. اگر هر میوه بدون آسیب ارزیابی شود، تضمین می‌کند که تنها محصولات با معیارهای کیفی خاصی را برآورده می‌کنند به بازار می‌رسند و منجر به تجربه مصرف‌کننده سازگارتر و رضایت‌بخش‌تری می‌شوند (Liu et al., 2025).

تصویربرداری فراطیفی

تصاویر فراطیفی نمونه‌ها با استفاده از سامانه تصویربرداری فراطیفی چرخشی ساخت شرکت فناوریان فیزیک ایران تهیه شدند. این سامانه شامل یک موتور دوار افقی، دوربین CCD با لنز مناسب، نرم‌افزار دریافت تصویر و رایانه پردازشگر می‌باشد. محدوده طیفی تحت پوشش سیستم از ۴۱۸ تا ۱۰۷۲ نانومتر با گام طیفی ۱/۴۳ نانومتر بوده و هر تصویر شامل ۴۲۰ باند طیفی است. فاصله دوربین از نمونه‌ها ۱۰۰ سانتی‌متر در نظر گرفته شد تا نمونه به‌طور کامل در میدان دید دوربین قرار گیرد. به‌منظور جلوگیری از ایجاد سایه و بهبود یکنواختی نور، دو منبع نور ۷ واتی پایه استارتی برند افراتاب با زاویه تابش ۴۵ درجه نسبت به سطح نمونه‌ها استفاده شد شکل ۱ تصویر سامانه تصویربرداری فراطیفی و نحوه قرارگیری بخش‌های مختلف سامانه را نشان می‌دهد (Hasanzadeh et al., 2022).

برای اطمینان از یکنواختی داده‌های طیفی و کاهش خطای موقعیت‌دهی، استانداردسازی موقعیت نمونه‌ها با دقت انجام شد. در تمامی آزمایش‌ها، هر نمونه آلو به گونه‌ای روی صفحه نمونه قرار گرفت که محور طولی آن به‌صورت عمودی باشد و بزرگترین سطح جانبی میوه (پهلوی) مستقیماً رو به لنز دوربین قرار گیرد. این روش موقعیت‌دهی ثابت تضمین کرد که بازتاب‌های طیفی از ناحیه استاندارد شده و قابل تکرار میوه (پهلوی) ثبت شوند و تأثیر جهت‌گیری نمونه بر نتایج به حداقل برسد.



شکل ۱- تصویر سامانه‌ی تصویربرداری فراطیفی و بخش‌های مختلف سامانه

Fig. 1. Image of the hyperspectral imaging system and the different parts of the system

مقداری از آب میوه با استفاده از آب‌میوه‌گیر گرفته شد و در یک فالكون ریخته شد ابتدا pH متر توسط مایع با pH مشخص کالیبره شد و با قرار دادن الکترود در آب‌میوه، مقدار pH اندازه‌گیری شد (Khodabakhshian & Baghbani, 2025; Latifi Amoghini et al.)

3- Particle Swarm Optimisation

جزئی^۱ (PLSR) و شبکه عصبی مصنوعی^۲ (ANN) می‌باشد؛ بنابراین، در این پژوهش فرض بر این است که ترکیب الگوریتم بهینه‌سازی ازدحام ذرات^۳ (PSO) با مدل شبکه عصبی مصنوعی می‌تواند دقت پیش‌بینی pH را نسبت به روش‌های کلاسیک (PLSR) بهبود دهد تا مدلی دقیق و قابل اعتماد برای پیش‌بینی pH ارائه دهد که بتواند در مراحل پس‌از برداشت و فرآیندهای کنترل کیفیت میوه مورد استفاده قرار گیرد.

مواد و روش‌ها

انتخاب نمونه‌ها

در این پژوهش، برای دستیابی به درکی عمیق از رفتار pH در آلو، دو گونه محبوب و بومی آلو خرمایی (قره آلو) و آلو خونی (توسرخ یا آلو قرمز) از باغ‌های محلی خریداری شدند. از هر رقم، ۸۰ میوه با دقت انتخاب گردید؛ تنها نمونه‌هایی که از نظر اندازه، شکل و رنگ کاملاً یکنواخت و عاری از هرگونه آسیب فیزیکی یا نشانه‌های پوسیدگی بودند، به مرحله بعد راه یافتند. پیش از آغاز اندازه‌گیری‌های دقیق، تمامی نمونه‌ها به مدت دو ساعت در دمای ۲۵ درجه سلسیوس نگهداری شدند تا دمایشان با محیط آزمایشگاه یکسان شود. پس از اتمام مرحله تصویربرداری فراطیفی (HSI)، بلافاصله وارد فاز اندازه‌گیری مخرب pH شدیم.

آزمون غیرمخرب

آزمون مخرب

pH

به‌منظور اندازه‌گیری pH آب‌میوه، از pH متر دیجیتالی (Metrohm 827, Switzerland) در دمای محیط استفاده شد.

1- Partial Least Squares Regression

2- Artificial Neural Network

(al., 2024b).

تجزیه و تحلیل داده‌های طیفی

فرآیند تحلیل ویژگی‌های طیفی آلو با گردآوری داده‌های طیفی در بازه ۴۱۸ تا ۱۰۷۲ نانومتر آغاز شد که این محدوده، بخش قابل توجهی از ناحیه مرئی (Vis) و مادون قرمز نزدیک (NIR) را دربرمی‌گیرد. به منظور ارتقای کیفیت داده‌ها و کاهش اثرات نویز شدید، بخش ابتدایی طیف (از ۴۱۸ تا حدود ۴۲۰ نانومتر) به دلیل افت محسوس نسبت سیگنال به نویز از مجموعه داده‌ها حذف شد. این گام، برای هر دو رقم آلو به طور یکسان انجام گرفت تا پایایی نتایج آتی تضمین شود. برای اطمینان از شفافیت در روش تحلیل داده‌ها، لازم به ذکر است که مجموعه داده‌های طیفی نهایی مورد استفاده در مدل‌سازی، شامل ۱۶۰ نمونه و ۱۲۶۶ ویژگی (طول موج‌ها) بود. این حجم داده از ترکیب ۸۰ نمونه آلو خرمایی و ۸۰ نمونه آلو خونی به دست آمد؛ بنابراین، ماتریس داده‌های طیفی خام نهایی دارای ابعاد 160×1266 [بود.

در ادامه، داده‌های اصلاح‌شده به نرم‌افزار MATLAB نسخه ۲۰۲۲a منتقل گردیدند تا فرایند مدل‌سازی پیش‌بینی آغاز شود. مدل‌سازی بر مبنای PLSR صورت پذیرفت و حداکثر تعداد ۱۰ مؤلفه در فرایند مدل‌سازی در نظر گرفته شد. انتخاب PLSR با توجه به توانایی بالای آن در مدیریت داده‌های حجیم و استخراج ساختارهای ارتباطی میان متغیرهای مستقل و پاسخ توجیه‌پذیر بود. به منظور ارتقای قابلیت اعتماد مدل‌ها و کاهش تأثیرات ناشی از تقسیم‌بندی تصادفی داده‌ها، رویکردی مبتنی بر تکرارهای چندباره مورد استفاده قرار گرفت. در این رویکرد، فرآیند تقسیم تصادفی داده‌ها به سه مجموعه آموزش (۶۰٪)، اعتبارسنجی (۲۰٪) و آزمون (۲۰٪) در مجموع ۲۰۰ بار به طور مستقل انجام شد. در هر تکرار، مدل PLSR بدون استفاده از هیچ‌گونه پیش‌پردازش بر داده‌های خام توسعه یافت و شاخص‌های ارزیابی شامل ضریب تعیین (R^2)، ریشه میانگین مربعات خطا^۲ (RMSE) و نسبت عملکرد به انحراف^۳ (RPD) در هر سه مجموعه محاسبه گردید. از میان ۲۰۰ تکرار، تقسیم‌بندی‌ای که ترکیبی از بهترین مقادیر این شاخص‌ها و تعداد بهینه مؤلفه‌ها را ارائه می‌داد، به عنوان تقسیم‌بندی مرجع انتخاب و تثبیت شد.

پس از تعیین ساختار بهینه داده‌ها، این تقسیم‌بندی ثابت نگه داشته شد تا تأثیر هر روش پیش‌پردازش به طور مستقل و بدون دخالت عوامل تصادفی مورد بررسی قرار گیرد. در این مرحله، مجموعه‌ای

متنوع از تکنیک‌های پیش‌پردازش شامل نرمال‌سازی^۴، متغیر نرمال استاندارد^۵ (SNV)، تصحیح پراکندگی چنگانه^۶ (MSC)، مشتق اول^۷، مشتق دوم^۸، فیلتر میانگین‌گیری متحرک^۹، فیلتر گوسی^{۱۰}، فیلتر میانه^{۱۱}، صاف‌سازی Savitzky-Golay، Detrend و میانگین مرکزی کردن^{۱۲} بر روی داده‌ها اعمال شد. به‌ازای هر روش، مدل PLSR مجدداً آموزش داده شد و عملکرد آن با دقت ارزیابی گردید تا بهترین راهکار پیش‌پردازش برای ارتقای دقت پیش‌بینی شناسایی شود.

برای تضمین قابلیت بازتولید نتایج، پارامترهای فنی روش‌های نیازمند تنظیم به شرح زیر استفاده شدند: صاف‌سازی ساویتزکی-گولای با مرتبه چندجمله‌ای ۲ و اندازه پنجره ۱۱ به کار رفت. محاسبه مشتق اول و دوم نیز با استفاده از فیلتر ساویتزکی-گولای انجام شد و برای هر دو، مرتبه چندجمله‌ای ۳ و اندازه پنجره ۱۱ در نظر گرفته شد. سایر فیلترهای صاف‌سازی شامل فیلتر میانگین‌گیری متحرک و فیلتر میانه، هر دو با اندازه پنجره ۵ اعمال شدند. همچنین، فیلتر گوسی با هسته‌ای دارای انحراف معیار ۲ اعمال شد.

ارزیابی عملکرد مدل‌های پیش‌بینی

به منظور سنجش دقت و قابلیت اعتماد مدل‌های توسعه‌یافته، مجموعه‌ای از شاخص‌های آماری کلیدی شامل R^2 ، RMSE و RPD در سه مرحله آموزش، اعتبارسنجی و آزمون محاسبه گردید. فرمول محاسبه این مقادیر به ترتیب در روابط (۱)، (۲) و (۳) نشان داده شده است. کلیه تحلیل‌ها، شامل اعمال روش‌های پیش‌پردازش و مدل‌سازی، در محیط نرم‌افزاری MATLAB نسخه ۲۰۲۲a انجام شد که یکی از ابزارهای معتبر و پرکاربرد در حوزه پردازش داده‌های طیفی و مدل‌سازی شیمی‌سنجی به‌شمار می‌رود (Latifi Amoghini, Abbaspour-Gilandeh, Tahmasebi, & Arribas, 2024a). شاخص RPD که به صورت نسبت انحراف معیار مقادیر اندازه‌گیری‌شده مرجع به ریشه میانگین مربعات خطای مجموعه آزمون تعریف می‌شود، معیاری برای سنجش توانایی مدل در پیش‌بینی خارج از داده‌های آموزشی است. هرچه RPD بالاتر باشد، قدرت پیش‌بینی مدل بیشتر است. تفسیر مقادیر RPD بر مبنای طبقه‌بندی ارائه‌شده توسط چانگ و همکاران (Chang, Laird, Mausbach, & Hurburgh, 2001) صورت گرفت؛ به‌طوری‌که RPD کمتر از ۱ نشان‌دهنده مدلی بسیار ضعیف و فاقد کارایی لازم است، مقادیر بین ۱ تا ۱/۴ معرف مدل ضعیف، بازه ۱/۴ تا ۱/۸ بیانگر مدلی قابل قبول برای برآوردهای مقدماتی، محدوده ۱/۸ تا ۲ نمایانگر مدلی خوب با

- 7- 1st Derivative
- 8- 2st Derivative
- 9- Moving Average
- 10- Gaussian Filter
- 11- Median Filter
- 12- Mean Centering

- 1- Coefficient of Determination
- 2- Root Mean Square Error
- 3- Ratio of Performance to Deviation
- 4- Normalise
- 5- Standard Normal Variate
- 6- Multiplicative Scatter Correction

ضعف را شناسایی کرده و از آن‌ها برای بهینه‌سازی استفاده کند (Kashan, 2014). الگوریتم PSO از رفتار اجتماعی گروهی مانند پرندگان یا ماهی‌ها الهام گرفته است. در این روش، هر ذره (فرد) موقعیت خود را بر اساس تجربیات گذشته و اطلاعات به‌دست‌آمده از همسایگانش به‌روز می‌کند این الگوریتم بسیار سریع است و برای مسائل بهینه‌سازی پیوسته مناسب می‌باشد (Alam, Dobbie, Koh, Riddle, & Rehman, 2014). الگوریتم ACO این روش از رفتار مورچه‌ها برای یافتن کوتاه‌ترین مسیر بین غذا و لانه الهام گرفته است و برای حل مسائل گسسته مانند مسئله فروشنده دوره‌گرد استفاده می‌شود (Dorigo, Birattari, & Stutzle, 2006). الگوریتم LA این روش مبتنی بر یادگیری تدریجی و تطبیق با محیط برای انتخاب بهینه است. کاربرد آن بیشتر در مسائل گسسته دیده می‌شود (Narendra & Thathachar, 2012).

در این فرآیند، هر الگوریتم فراابتکاری به‌طور مستقل مجموعه‌ای از طول‌موج‌های بهینه را شناسایی کرد. عملکرد خروجی‌های حاصل از هر الگوریتم با توجه به ضریب تعیین R^2 و RMSE ارزیابی شد. هدف نهایی این مرحله، انتخاب الگوریتمی بود که بالاترین دقت پیش‌بینی و کمترین میزان خطا را ارائه دهد. پس از تعیین طول‌موج‌های مؤثر و در نتیجه کاهش حجم داده‌ها، داده‌های حاصله با استفاده از دو روش قدرتمند PLSR و ANN مدل‌سازی مجدد شدند. برای بخش ANN، از یک شبکه عصبی پرسپترون چندلایه^۸ مبتنی بر الگوریتم پس‌انتشار استفاده شد. این شبکه شامل یک لایه ورودی بود که تعداد نورون‌های آن برابر با تعداد طول‌موج‌های مؤثر انتخاب‌شده بود و یک لایه خروجی با تنها یک نورون که مقدار pH را پیش‌بینی می‌کرد. به‌منظور بررسی تأثیر ساختار شبکه بر عملکرد مدل، معماری‌های مختلفی طراحی و ارزیابی شدند. این معماری‌ها شامل یک، دو و سه لایه پنهان با تعداد نورون‌های متفاوت بودند (به‌عنوان مثال: [۵]، [۱۰]، [۱۵]، [۲۰] برای یک لایه؛ [۵ ۵]، [۵ ۱۰]، [۵ ۱۵]، [۵ ۲۰]، [۱۰ ۱۰]، [۱۵ ۱۵]، [۲۰ ۲۰] برای دو لایه؛ و [۲۰ ۱۵ ۱۰] برای سه لایه). جهت تقسیم‌بندی داده‌ها برای آموزش، اعتبارسنجی و آزمون شبکه، به‌ترتیب از نسبت‌های ۶۰٪، ۲۰٪ و ۲۰٪ استفاده شد. پارامترهای آموزشی ثابت برای تضمین باز تولیدپذیری نتایج به شرح زیر تنظیم شدند: از الگوریتم لونبرگ-مارکوارت^۹ به دلیل سرعت بالای همگرایی و کارایی در مسائل رگرسیون استفاده شد (Latifi Amoghini et al., 2024b). برای لایه‌های پنهان، از تابع فعال‌سازی تانژانت سیگموئید و برای لایه خروجی، از تابع خطی استفاده گردید. به‌منظور کمینه‌سازی

توانایی پیش‌بینی کمی مناسب، مقادیر بین ۲ تا ۲/۵ بیانگر مدلی بسیار خوب و مقادیر بالاتر از ۲/۵ معرف مدلی دقیق و کاملاً قابل اعتماد در کاربردهای پیش‌بینی کمی است.

$$R^2 = 1 - \frac{\sum_{i=1}^N (p_i - d_i)^2}{\sum_{i=1}^N (d_i - \bar{d})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (p_i - d_i)^2}{N}} \quad (2)$$

$$RPD = \frac{SD}{RMSE} \quad (3)$$

در این روابط SD انحراف استاندارد داده‌های پیش‌بینی‌شده و RMSE ریشه میانگین مربعات خطا، p_i و d_i به ترتیب مقادیر واقعی و پیش‌بینی‌شده مؤلفه‌های نام هستند، $\bar{d}$ میانگین مقادیر واقعی و N تعداد مؤلفه‌ها است.

انتخاب طول‌موج‌های مؤثر و مدل‌سازی مجدد بر روی داده‌های کاهش‌یافته

برای کاهش ابعاد داده‌های طیفی و شناسایی طول‌موج‌های مؤثر که بیشترین نقش را در پیش‌بینی دقیق دارند، رویکردی ترکیبی به کار گرفته شد. این رویکرد شامل تلفیق روش درخت تصمیم^۱ (DT) با مجموعه‌ای از الگوریتم‌های فراابتکاری^۲ بود. این الگوریتم‌ها که به دلیل توانایی‌شان در جستجوی بهینه در فضاها پیچیده شناخته شده‌اند، شامل الگوریتم مسابقه رقابتی جهانی^۳ (WCC)، الگوریتم قهرمانی لیگ^۴ (LCA)، الگوریتم بهینه‌سازی ازدحام ذرات^۵ (PSO)، الگوریتم کلونی مورچه^۶ (ACO) و الگوریتم اتوماتای یادگیری^۷ (LA) بودند (Masoudi-Sobhanzadeh, Motieghader, & Masoudi, 2019). مشخصات دقیق پارامترهای هر یک از این الگوریتم‌ها در جدول ۱ ارائه شده است.

الگوریتم WCC بر اساس شبیه‌سازی رقابت‌های جهانی طراحی شده است. در این الگوریتم، ایده رقابت بین تیم‌ها یا کشورها برای بهینه‌سازی یک هدف خاص مورد استفاده قرار می‌گیرد. این روش بیشتر برای مسائل پیچیده و چندهدفه مناسب است و تلاش می‌کند تا از استراتژی‌های متنوع رقبا بهره‌برداری کند (Masoudi-Sobhanzadeh & Motieghader, 2016). الگوریتم LCA از رقابت بین تیم‌های ورزشی در یک لیگ الهام گرفته شده است. تیم‌ها در طول مسابقات بهبود عملکرد خود را به کمک تحلیل نتایج و تنظیم استراتژی دنبال می‌کنند. این الگوریتم تلاش می‌کند تا نقاط قوت و

6- Ant Colony Optimisation
7- Learning Automata
8- Multilayer Perceptron
9- Levenberg-Marquardt

1- Decision Tree
2- Metaheuristic Algorithms
3- World Competitive Contest
4- League Championship Algorithm
5- Particle Swarm Optimisation

تقسیم‌بندی تصادفی داده‌ها ۲۰۰ بار تکرار شد. این انتخاب تصادفی تنها بر روی داده‌های خام کاهش‌یافته انجام گرفت. پس از انتخاب بهترین تقسیم‌بندی بر اساس شاخص‌های آماری، تمامی پیش‌پردازش‌های مورد نظر بر روی همان ترتیب ثابت داده‌ها اعمال شدند. این رویکرد تضمین کرد که تأثیر واقعی هر روش پیش‌پردازش بر عملکرد مدل، بدون دخالت تغییرات ناشی از تقسیم‌بندی تصادفی، به دقت ارزیابی شود.

خطای مدل، تابع میانگین مربعات خطا (MSE) به کار رفت. برای جلوگیری مؤثر از بیش‌برازش، از روش توقف زودهنگام^۱ استفاده شد که در آن حداکثر تکرار^۲ برابر با ۳۰۰ و معیار عدم بهبود عملکرد در مجموعه اعتبارسنجی برابر با ۱۰ تنظیم گردید. همچنین، مجموعه آموزشی اصلی (۶۰٪ کل داده‌ها) به طور داخلی در مرحله آموزش به نسبت ۸۰٪ آموزش، ۱۰٪ اعتبارسنجی و ۱۰٪ تست تقسیم شد تا فرآیند توقف زودهنگام به طور مؤثر انجام گیرد. مشابه مراحل قبلی، برای افزایش اعتبار و پایداری نتایج، فرآیند

جدول ۱- مشخصات الگوریتم‌های فراابتکاری مورد استفاده برای انتخاب طول موج‌های مؤثر

Table 1- Parameters of the metaheuristic algorithms used for selecting effective wavelengths

ردیف No	الگوریتم Algorithm	پارامترها Parameters
1	WCC	تعداد تیم‌ها: ۶، تعداد گروه‌ها: ۳، تعداد فصل‌ها: ۱۰، زمان مسابقه: ۲ Teams: 6, Groups: 3, Seasons: 10, Match Time: 2
2	LCA	تعداد تیم‌ها: ۸، تعداد فصل‌ها: ۱۰، زمان مسابقه: ۲ Teams: 8, Seasons: 10, Match Time: 2
3	PSO	تعداد ذرات: ۴۰، تعداد تکرارها (Episodes): ۱۰، تعداد بهترین‌های محلی (P Bests): ۸ Particles: 40, Episodes: 10, P Bests: 8
4	ACO	تعداد مورچه‌ها: ۴۰، تعداد تکرارها: ۱۰، نرخ تبخیر: ۴ Ants: 40, Episodes: 10, Evaporation Rate: 4
5	LA	تعداد نسل‌ها: ۱۰، تعداد تلاش‌ها: ۴۰ Generations: 10, Tries: 40

(نزدیک به نور قرمز) دیده می‌شود که ممکن است به حضور کلروفیل (حتی مقادیر اندک آن در مراحل پایانی رسیدگی) یا سایر رنگدانه‌ها مرتبط باشد. در بخش مادون قرمز نزدیک (بالا تر از ۷۵۰ نانومتر)، شدت بازتاب به طور کلی کاهش می‌یابد. باین حال، تفاوت‌های محسوسی نیز وجود دارد؛ به عنوان مثال، در آلو خونی (ب)، به نظر می‌رسد شدت بازتاب کلی در مقایسه با آلو خرمایی (الف) بالاتر است، که می‌تواند نشان‌دهنده تفاوت در سطح و ساختار پوست یا محتوای آب باشد. این اختلافات طیفی زیربنای شناسایی و کمی‌سازی خواص فیزیکوشیمیایی میوه‌ها با استفاده از روش‌های فراطیفی هستند.

جدول ۲، آماره‌های توصیفی مقدار pH را برای نمونه‌های دو رقم آلو مورد مطالعه، خرمایی و خونی، ارائه می‌دهد. مشاهده می‌شود که هر دو رقم، طیف وسیعی از مقادیر pH را در بر می‌گیرند: برای آلو خرمایی، pH از ۳/۶ تا ۴/۲ متغیر است، در حالی که برای آلو خونی این دامنه بین ۳/۱۵ تا ۴/۴۴ قرار دارد. این دامنه وسیع مقادیر pH در هر دو رقم، یک مزیت کلیدی برای توسعه مدل‌های رگرسیونی به شمار می‌رود، چرا که به مدل‌ها اجازه می‌دهد تا ارتباط پیچیده بین الگوهای طیف بازتابی و سطوح متنوع pH را به طور جامع‌تری بیاموزند و در نتیجه، از قوت پیش‌بینی بالاتری برخوردار باشند. با بررسی دقیق‌تر،

نتایج و بحث

تحلیل داده‌های طیفی

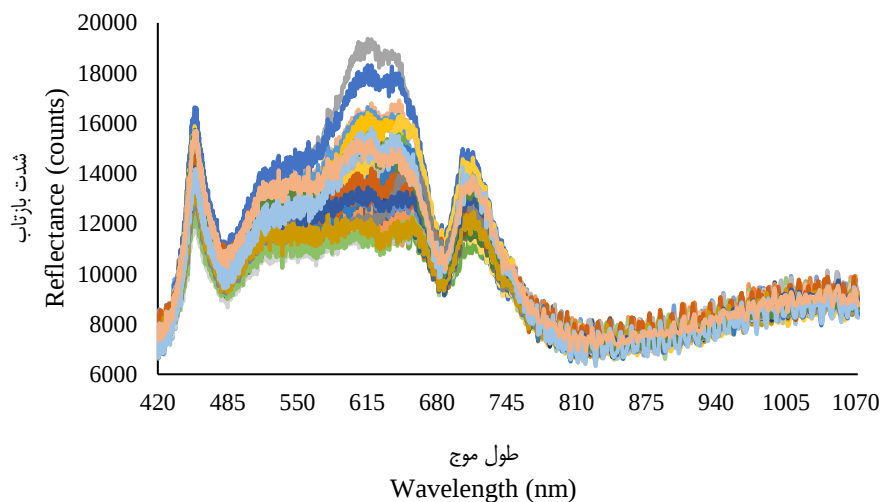
در این پژوهش، تحلیل جامع داده‌های طیف بازتابی و مقادیر pH نمونه‌های آلو خرمایی و آلو خونی، اطلاعات ارزشمندی را در خصوص ویژگی‌های شیمیایی و فیزیکی این دو رقم و پتانسیل تصویربرداری فراطیفی برای ارزیابی غیرمخرب آن‌ها فراهم آورد.

شکل ۲، طیف‌های بازتابی (Reflectance Spectra) نمونه‌های دو رقم آلو شامل آلو خرمایی (الف) و آلو خونی (ب) را در گستره طول موج ۴۲۰ تا ۱۰۷۲ نانومتر به تصویر می‌کشد. همان‌طور که از پراکندگی وسیع طیف‌ها در هر دو نمودار (الف و ب) مشهود است، الگوهای طیفی تنوع قابل توجهی را نشان می‌دهند. این تنوع در طیف‌های بازتابی عمدتاً به تفاوت‌های ترکیبات شیمیایی و ساختاری درون هر نمونه آلو، از جمله مراحل مختلف رسیدگی، و همچنین به اختلافات ظریف در رنگدانه و بافت هر میوه نسبت داده می‌شود. با دقت بیشتر به نمودارها، می‌توان الگوهای کلی مشترکی را مشاهده کرد: در هر دو رقم، یک پیک بازتابی در محدوده ۶۱۵ نانومتر (نزدیک به نور سبز) و یک ناحیه جذب قوی‌تر در محدوده ۶۹۰-۶۸۰ نانومتر

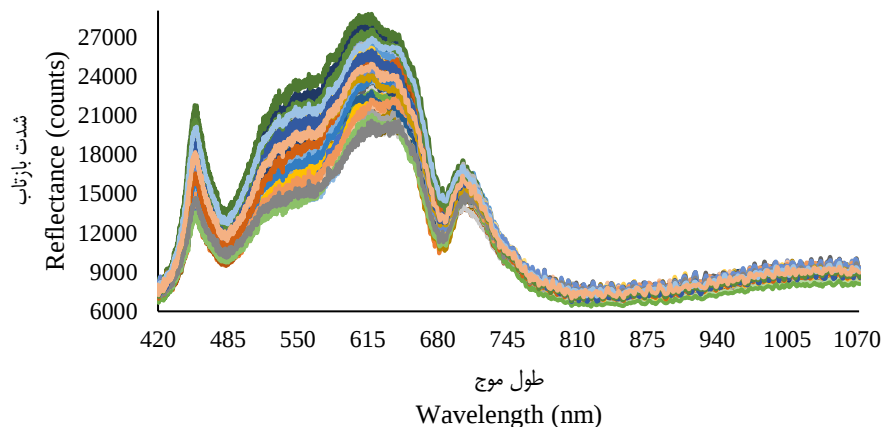
ویژگی‌های طیفی (مثلاً در نواحی جذب رنگدانه‌ها یا جذب آب) بین دو رقم آلو، تا حدی می‌تولندناشی از همین اختلافات در میانگین و پراکندگی pH باشد. اسیدیته میوه (که با pH نشان داده می‌شود) بر ساختار سلولی، فعالیت آنزیمی و در نتیجه بر ویژگی‌های نوری و بازتابی میوه تأثیر می‌گذارد. این همبستگی، اهمیت حیاتی بهره‌گیری از کل بازه طیفی مرئی-مادون قرمز نزدیک را برای ساخت مدل‌های پیش‌بینی pH تأیید می‌کند. استفاده از طیف وسیع داده‌های طیفی، به مدل این قابلیت را می‌دهد که ویژگی‌های شیمیایی و فیزیکی مرتبط با pH را، با وجود تنوع بیولوژیکی، به‌طور مؤثرتری استخراج کرده و در نهایت به ساخت مدل‌های پیش‌بینی قابل اعتماد و عمومی‌تر برای هر دو رقم آلو منجر شود.

تفاوت‌های مشخصی در آماره‌های pH بین دو رقم مشاهده می‌شود. آلو خرمایی دارای میانگین pH بالاتری (۳/۹۰۹) نسبت به آلو خونی (۳/۷۳۷۵) است. این اختلاف در میانگین pH می‌تواند به تفاوت‌های فیزیولوژیکی و ترکیبات اسیدی موجود در این دو رقم مرتبط باشد. علاوه بر این، انحراف معیار برای آلو خونی (۰/۴۱۰۲) به مراتب بیشتر از آلو خرمایی (۰/۱۷۶۸) است. این نشان‌دهنده پراکنش (تنوع) بیشتر در مقادیر pH بین نمونه‌های آلو خونی است، در حالی که آلو خرمایی همگنی بیشتری در pH خود دارد. این پراکندگی بالاتر در آلو خونی ممکن است به تفاوت‌های بیشتر در مرحله رسیدگی یا عوامل محیطی مختلف در زمان برداشت این رقم بازگردد.

ارتباط میان این تفاوت‌های pH در جدول ۲ و الگوهای طیفی مشاهده‌شده در شکل ۲، قابل تأمل است. تفاوت‌های فاحش در



(a) الف



(b) ب

شکل ۲- نمودار طیف‌های بازتابی نمونه‌های (الف) آلو خرمایی و (ب) خونی اندازه‌گیری شده با سامانه تصویربرداری فراطیفی

Fig. 2. Reflectance spectra of (a) Khormaei and (b) Khoni plum samples measured by hyperspectral imaging system

جدول ۲- آمار توصیفی مقدار pH در دو رقم آلو

Table 2- Descriptive statistics of pH values in two varieties of plums

رقم Variety	مینیم Minimum	ماکزیم Maximum	انحراف معیار Standard deviation	میانگین Mean
خرمایی Khormaei	3.6	4.2	0.1768	3.909
خونی Khoni	3.15	4.44	0.4102	3.7375

گردیده است، که RPD را به ۲/۲۴ افزایش می‌دهد. این مقدار RPD مدل را در دسته بسیار خوب ($2 < RPD < 2.5$) قرار می‌دهد. همچنین، فیلتر گوسی با $R^2 = 0.8203$ و $RMSE = 0.0590$ و $RPD = 2.44$ بهترین عملکرد را در مرحله آزمون برای رقم خرمایی به نمایش گذاشته است و آن را در آستانه یک مدل بسیار خوب قرار می‌دهد. این نشان می‌دهد که هموارسازی (smoothing) طیف‌ها می‌تولند به طور مؤثری نویز را کاهش داده و اطلاعات مفید مربوط به pH را آشکار سازد. عملکرد بهترین مدل PLSR در مرحله آزمون رقم آلو خرمایی مربوط به فیلتر گوسی به عنوان یک معیار پایه برای مقایسه با مدل‌های غیرخطی، در شکل ۳-الف به صورت بصری نشان داده شده است.

برخی پیش‌پردازش‌ها مانند SNV، MSC، فیلتر میانه، Savitzky-Golay و Detrend نیز عملکرد مدل را در مرحله آزمون بهبود بخشیده‌اند (با RPD بالای ۲). با این حال، پیش‌پردازش‌های مشتق اول و مشتق دوم و مرکز کردن میانگین (Mean Centering) (ردیف‌های ۴، ۵ و ۱۲) عملاً تغییری در نتایج ایجاد نکرده‌اند که احتمالاً به دلیل عدم نیاز به تغییرات شدید در دامنه طیفی یا حذف روند در این مجموعه داده خاص بوده است. همچنین، نرمال‌سازی در مقایسه با سایر روش‌ها عملکرد کمی ضعیف‌تر ($RPD = 1.76$) را در مرحله آزمون نشان داده است.

ارزیابی عملکرد مدل رگرسیون حداقل مربعات جزئی و تأثیر پیش‌پردازش‌های طیفی

پس از تحلیل کیفی طیف‌ها و بررسی توزیع pH، گام بعدی ارزیابی مدل‌های رگرسیونی برای پیش‌بینی کمی این ویژگی می‌باشد به منظور بهینه‌سازی دقت مدل‌سازی pH در آلو، تأثیر انواع روش‌های پیش‌پردازش طیفی بر عملکرد مدل PLSR مورد بررسی قرار گرفت. نتایج حاصل برای رقم خرمایی در جدول ۳ و برای رقم خونی در جدول ۴ خلاصه شده‌اند.

جدول ۳ عملکرد مدل PLSR را برای رقم آلو خرمایی با ۹ مؤلفه بهینه، تحت تأثیر پیش‌پردازش‌های مختلف نشان می‌دهد. مشاهده می‌شود که حتی بدون پیش‌پردازش (ردیف ۱)، مدل PLSR توانایی قابل‌قبولی در پیش‌بینی pH رقم خرمایی از خود نشان داده است، به طوری که در مرحله آزمون به $R^2 = 0.7045$ و $RMSE = 0.0756$ و $RPD = 1.9$ دست‌یافته است. بر اساس سیستم طبقه‌بندی RPD، این مدل در دسته مناسب تا خوب ($1.8 < RPD < 2$) قرار می‌گیرد که نشان‌دهنده قابلیت استفاده آن برای ارزیابی و پیش‌بینی‌های کمی است.

با اعمال پیش‌پردازش‌های مختلف، بهبودهایی در عملکرد مدل مشاهده می‌شود. به عنوان مثال، پیش‌پردازش فیلتر میانگین متحرک منجر به بهبود R^2 آزمون به ۰/۷۸۸۳ و کاهش $RMSE$ به ۰/۰۶۴۰

جدول ۳- نتایج مدل‌سازی رگرسیون حداقل مربعات جزئی (PLSR) رقم خرمایی با انواع پیش‌پردازش‌های طیفی (۹ مؤلفه)

Table 3- Results of the Partial Least Squares Regression (PLSR) model for the Khormaei plum variety with various spectral preprocessing methods (9 Components)

ردیف No	پیش‌پردازش Preprocessing	آموزش Training			اعتبارسنجی Validation			آزمون Test		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
1	بدون پیش‌پردازش No Preprocessing	0.9769	0.0267	6.65	0.9525	0.0424	4.74	0.7045	0.0756	1.9
2	SNV	0.988	0.0193	9.22	0.8999	0.0616	3.26	0.7591	0.0683	2.1
3	MSC	0.9834	0.0226	7.84	0.8632	0.072	2.79	0.7052	0.0756	1.9
4	مشتق اول First Derivative	0.9769	0.0267	6.65	0.9525	0.0424	4.74	0.7045	0.0756	1.9
5	مشتق دوم Second Derivative	0.9769	0.0267	6.65	0.9525	0.0424	4.74	0.7045	0.0756	1.9

6	نرمال سازی Normalisation	0.9826	0.0232	7.67	0.9432	0.0464	4.33	0.6561	0.0816	1.76
7	میانگین متحرک Moving Average	0.9742	0.0282	6.29	0.9084	0.0589	3.41	0.7883	0.064	2.24
8	فیلتر گوسی Gaussian Filter	0.9655	0.0326	5.44	0.8886	0.065	3.09	0.8203	0.059	2.44
9	فیلتر میانه Median Filter	0.9889	0.0185	9.6	0.9152	0.0567	3.55	0.7176	0.074	1.94
10	ساویتزکی-گولای Savitzky-Golay	0.9792	0.0253	7.01	0.9008	0.0613	3.28	0.7716	0.0665	2.16
11	حذف روند Detrending	0.985	0.0215	8.26	0.9419	0.0469	4.28	0.7401	0.0709	2.03
12	مرکز کردن میانگین Mean Centering	0.9769	0.0267	6.65	0.9525	0.0424	4.74	0.7045	0.0756	1.9

برای مقایسه با مدل‌های غیرخطی، در شکل ۳-ب به صورت بصری نشان داده شده است.

به طور کلی، مدل سازی pH برای آلو خونی (با بهترین RPD حدود ۲/۴۴) دقت بالاتری را نشان داده است. این تفاوت در دقت می‌تواند به دلیل تفاوت در ویژگی‌های طیفی ذاتی و دامنه pH وسیع‌تر آلو خونی باشد که فرصت بهتری برای مدل جهت یادگیری و تمایز فراهم می‌کند، یا حتی به دلیل تفاوت در پیچیدگی ماتریس شیمیایی هر رقم باشد. توانایی مدل PLSR در دستیابی به RPDهای بالای ۲ در هر دو رقم، بر پتانسیل بالای تکنیک تصویربرداری فراطیفی در ارزیابی غیرمخرب کیفیت داخلی میوه‌های آلو، به ویژه pH، تأکید می‌کند. در راستای تایید و گسترش یافته‌های این پژوهش، مقایسه نتایج حاصل با پژوهش‌های مشابه در زمینه ارزیابی غیرمخرب میوه‌ها حائز اهمیت است. به عنوان مثال، در مطالعه‌ای دیگر، پتانسیل تصویربرداری فراطیفی در محدوده ۱۰۰۰ تا ۱۵۵۰ نانومتر برای ارزیابی غیرمخرب مؤلفه‌های شیمیایی نظیر رطوبت، pH و مواد جامد محلول (SSC) در گوجه‌فرنگی‌های سالم مورد بررسی قرار گرفت. در آن تحقیق، مدل PLSR مبتنی بر پیش‌پردازش مشتق اول ساویتزکی-گولای بهترین عملکرد را برای pH نشان داد که با ضریب تعیین ۰/۶۹ و ریشه میانگین مربعات خطای پیش‌بینی ۰/۰۶ به دست آمد (Rahman et al., 2017). علاوه بر این، در مطالعه‌ای توسط رضوی و همکاران (Razavi, Sharabiani, Tahmasebi, Grassi, & Szymanek, 2025)، پتانسیل طیف‌سنجی Vis-NIR (در محدوده ۵۴۰ تا ۹۶۰ نانومتر) برای ارزیابی پارامترهای کیفی سیب رقم رد دلشیز^۱، از جمله pH، TA، اسید اسکوربیک، سفتی، محتوای مواد جامد محلول (SSC) و آنتوسیانین‌ها، با استفاده از روش‌های یادگیری ماشین مورد بررسی قرار گرفت. در آن تحقیق، عملکرد قوی برای تمامی رگرسیون‌های

جدول ۴، نتایج مدل‌سازی PLSR را برای رقم آلو خونی با ۱۰ مؤلفه بهینه ارائه می‌دهد. مشاهده می‌شود که رقم خونی به طور کلی دارای دقت پیش‌بینی بالاتری نسبت به رقم خرمایی است. حتی بدون پیش‌پردازش (ردیف ۱)، مدل PLSR برای آلو خونی به $R^2=0.8276$ و $RMSE=0.1705$ و $RPD=2.49$ در مرحله آزمون دست یافته است. این مقدار RPD مدل را در دسته بسیار خوب قرار می‌دهد.

در میان پیش‌پردازش‌ها، نرمال‌سازی بهترین عملکرد را برای آلو خونی به ارمغان آورده است، با $R^2=0.8612$ و $RMSE=0.1530$ و $RPD=2.77$ در مرحله آزمون. این RPD نشان‌دهنده یک مدل عالی ($RPD > 2/5$) است که توانایی بسیار بالایی در پیش‌بینی pH رقم خونی دارد. این نتیجه حاکی از آن است که حذف تأثیرات دامنه یا مقیاس از سیگنال طیفی برای این رقم بسیار مؤثر بوده است.

سایر پیش‌پردازش‌ها مانند میانگین متحرک ($RPD=2.61$)، فیلتر گوسی ($RPD=3.01$)، فیلتر میانه ($RPD=2.64$)، Savitzky-Golay ($RPD=2.96$) و Detrend ($RPD=2.66$) نیز عملکرد بسیار خوبی را در مرحله آزمون از خود نشان داده‌اند و مدل‌ها را در دسته‌های بسیار خوب تا عالی قرار داده‌اند. در این رقم نیز، پیش‌پردازش‌های مشتق اول، مشتق دوم و مرکز کردن میانگین تأثیر چشمگیری بر نتایج آزمون نداشته‌اند.

نتایج حاصل از جداول ۳ و ۴ به وضوح نشان می‌دهند که پیش‌پردازش داده‌های طیفی نقش حیاتی در بهبود دقت مدل‌های پیش‌بینی pH با استفاده از PLSR دارد. انتخاب نوع پیش‌پردازش بهینه، بسته به ویژگی‌های خاص طیف و رقم مورد مطالعه، متفاوت است؛ در حالی که میانگین متحرک و فیلتر گوسی برای آلو خرمایی بهترین نتایج را داشت، نرمال‌سازی و فیلتر گوسی بهترین عملکرد را برای آلو خونی نشان دادند. عملکرد بهترین مدل PLSR در مرحله آزمون رقم آلو خونی مربوط به فیلتر گوسی به عنوان یک معیار پایه

آزمایشی (ANN و PLSR, PCR, MLR, SVM-R) در فازهای اعتبارسنجی متقابل، اعتبارسنجی و آزمون به‌دست آمد. به‌طور خاص، مدل SVM-R در پیش‌بینی pH سیب به $R^2=0.935$ و $R^2=0.936$ به دست یافت.

این مقایسه‌ها نشان می‌دهد که هر دو تحقیق مورد بررسی، مشابه پژوهش حاضر، بر کارایی بالای طیف‌سنجی در ارزیابی غیرمخرب میوه‌ها تأکید دارند، هرچند در گونه میوه، محدوده طیفی و پارامترهای هدف متفاوت بوده‌اند. RPD حاصل برای pH در مطالعه رضوی بر روی سیب (۳/۶۳۶) و ضریب تعیین ۰/۶۹ برای pH گوجه‌فرنگی در مطالعه دیگر، نشان‌دهنده مدل‌های قوی و قابل قبول هستند. در پژوهش حاضر نیز، مدل PLSR برای آلو خونی با بهترین RPD حدود ۳/۰۱ و برای آلو خرمایی با RPD حدود ۲/۴۴، عملکرد بسیار خوبی در پیش‌بینی pH از خود نشان داد. این نتایج در راستای یک روند تحقیقاتی موفق قرار دارد که بر اهمیت انتخاب طول موج و استفاده از روش‌های هوشمند برای بهینه‌سازی مدل‌ها و کاربردهای صنعتی تأکید دارد.

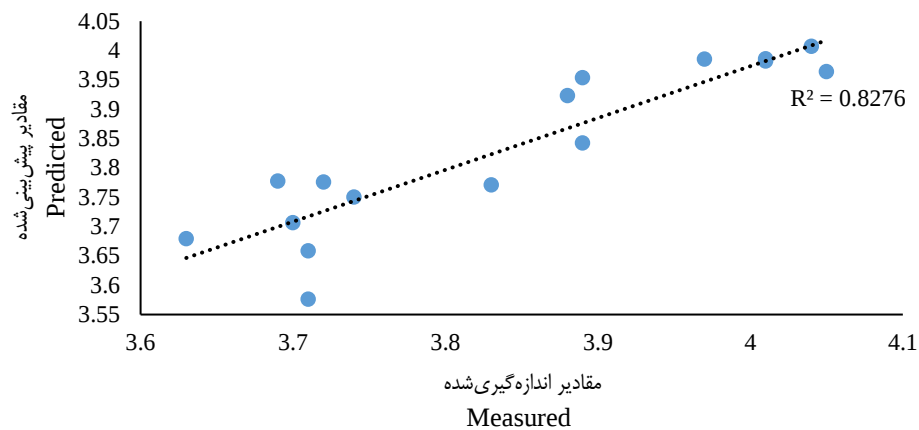
کاهش ابعاد داده‌های طیفی و انتخاب طول‌موج‌های مؤثر

تحلیل داده‌های طیفی، که از بازتاب نمونه‌های آلو خرمایی و خونی در محدوده ۴۲۰ تا ۱۰۷۰ نانومتر (مطابق شکل ۲) به‌دست آمده‌اند، حجم وسیعی از اطلاعات را در قالب تعداد زیادی متغیر (طول موج) شامل می‌شود. این حجم بالای داده‌ها، چالش‌های مهمی نظیر پیچیدگی محاسباتی و پدیده هم‌خطی شدید را در پی دارد که می‌تواند کارایی و دقت مدل‌های پیش‌بینی را کاهش دهد. از این رو، کاهش ابعاد داده‌ها و انتخاب طول‌موج‌های مؤثر، نه تنها برای ساده‌سازی مدل و بهبود دقت پیش‌بینی ضروری است، بلکه با افزایش قابلیت تعمیم‌پذیری و کاهش زمان محاسبات، به پایداری و تفسیرپذیری بهتر مدل نیز کمک شایانی می‌کند. این فرآیند حیاتی، با تمرکز بر استخراج اطلاعات کلیدی از طول‌موج‌هایی که بیشترین ارتباط را با ویژگی‌های هدف (مانند pH) دارند، مسیر را برای مدل‌سازی‌های کارآمد هموار می‌سازد.

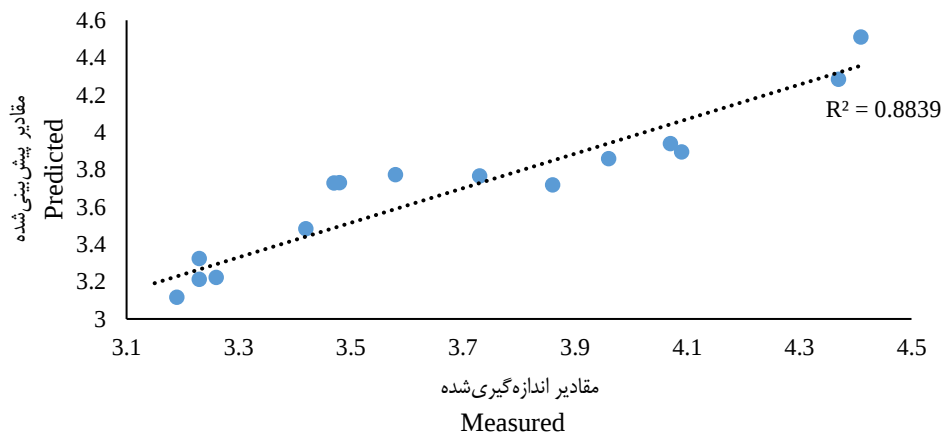
جدول ۴- نتایج مدل‌سازی رگرسیون حداقل مربعات جزئی (PLSR) رقم خونی با انواع پیش‌پردازش‌های طیفی (۱۰ مؤلفه)

Table 4- Results of the Partial Least Squares Regression (PLSR) model for the Khoni plum variety with various spectral preprocessing methods (10 components)

ردیف No	پیش‌پردازش Preprocessing	آموزش Training			اعتبارسنجی Validation			آزمون Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	بدون پیش‌پردازش No Preprocessing	0.9927	0.0342	11.85	0.9787	0.0624	7.07	0.8276	0.1705	2.49
2	SNV	0.9877	0.0445	9.1	0.9108	0.1277	3.46	0.834	0.1673	2.54
3	MSC	0.9878	0.0443	9.13	0.9113	0.1273	3.47	0.8363	0.1661	2.55
4	مشتق اول First Derivative	0.9927	0.0342	11.85	0.9787	0.0624	7.07	0.8276	0.1705	2.49
5	مشتق دوم Second Derivative	0.9927	0.0342	11.85	0.9787	0.0624	7.07	0.8276	0.1705	2.49
6	نرمال‌سازی Normalisation	0.9953	0.0274	14.75	0.9719	0.0717	6.16	0.8612	0.153	2.77
7	میانگین متحرک Moving Average	0.9848	0.0493	8.21	0.9593	0.0862	5.12	0.8435	0.1624	2.61
8	فیلتر گوسی Gaussian Filter	0.9809	0.0554	7.31	0.9238	0.118	3.74	0.8819	0.1411	3.01
9	فیلتر میانه Median Filter	0.9941	0.0307	13.18	0.9648	0.0802	5.5	0.8465	0.1609	2.64
10	ساویتزکی-گولای Savitzky-Golay	0.9864	0.0468	8.65	0.9576	0.088	5.02	0.878	0.1434	2.96
11	حذف روند Detrending	0.9941	0.0308	13.15	0.9727	0.0706	6.25	0.8495	0.1593	2.66
12	مرکز کردن میانگین Mean Centering	0.9927	0.0342	11.85	0.9787	0.0624	7.07	0.8276	0.1705	2.49



(a) الف



(b) ب

شکل ۳- نمودار پراکنش مقادیر pH اندازه‌گیری شده در برابر مقادیر پیش‌بینی شده برای بهترین مدل PLSR در محدوده طیف کامل با

پیش‌پردازش فیلتر گوسی بر روی مجموعه آزمون ارقام: الف) خرمایی، ب) خونی

Fig. 3. Scatter plot of the measured versus predicted pH values for the best PLSR model on the full spectrum range with Gaussian filter preprocessing on the test set for varieties: (a) Khormaei, (b) Khoni

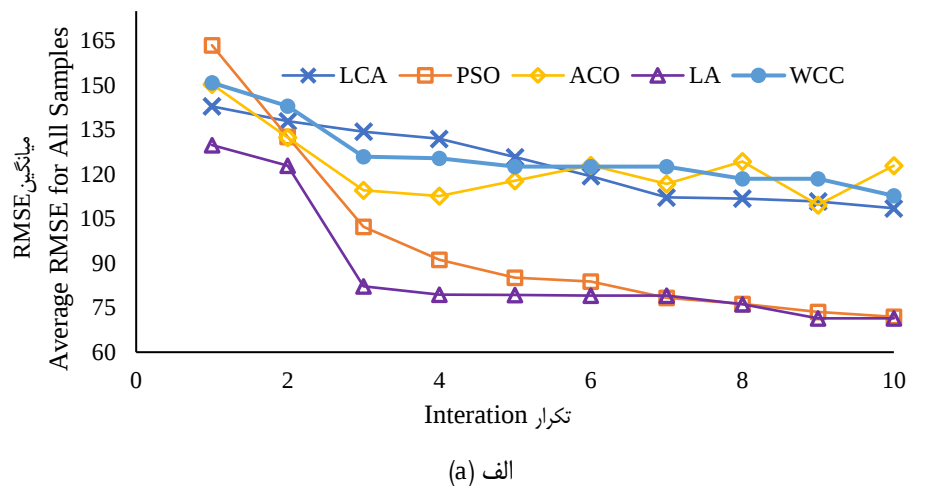
(۰/۹۸۱) و به دنبال آن LCA با عملکردی بسیار نزدیک، دقت بالاتری را به نمایش گذاشتند. با این حال، PSO مجدداً با زمان اجرای ۵/۶۴ ثانیه، سریع‌ترین روش باقی ماند که مزیت آن را در سرعت محاسبات تایید می‌کند طول موج‌های منتخب حاصل از الگوریتم PSO، که بر اساس درجه اهمیت مرتب شده‌اند، در جدول ۵ ارائه شده و در مراحل بعدی مدل‌سازی و تحلیل مورد استفاده قرار گرفته‌اند. در راستای رویکردهای مشابه در انتخاب طول موج‌های مؤثر، پژوهش لطیفی عموقین و همکاران (Latifi Amoghini et al., 2024b) بر روی سه رقم فلفل دلمه‌ای (در محدوده ۱۱۵۰-۳۵۰ نانومتر) نیز از ترکیب ماشین‌های بردار پشتیبان (SVM) با الگوریتم‌های ژنتیک (GA)، بهینه‌سازی ازدحام ذرات (PSO)، کلونی مورچه (ACO) و الگوریتم رقابت امپریالیستی (ICA) برای بهینه‌سازی

به منظور شناسایی بهینه‌ترین طول موج‌ها، پنج الگوریتم فراابتکاری شامل WCC، LCA، PSO، ACO و LA مورد ارزیابی قرار گرفتند. عملکرد این الگوریتم‌ها بر اساس معیارهای میانگین RMSE، میانگین همبستگی و زمان اجرا، برای هر دو نوع آلو تحلیل شد. برای آلو خرمایی، نتایج (مندرج در شکل ۴) نشان دادند که الگوریتم‌های PSO و LA در کاهش خطا و افزایش همبستگی بهترین عملکرد را داشتند؛ PSO با کمترین RMSE نهایی (۷۱/۹۱) و همبستگی بالا (۰/۹۱۵)، و LA با همبستگی نهایی برتر (۰/۹۲۳) و RMSE نزدیک (۷۱/۴۰) برجسته ظاهر شدند. در بعد کارایی محاسباتی، PSO با زمان اجرای تنها ۵/۴۶ ثانیه، برتری قابل توجهی نسبت به سایر الگوریتم‌ها نشان داد. این الگو در مورد آلو خونی نیز تکرار شد (شکل ۵)، هرچند الگوریتم ACO با کمترین RMSE (۴۳/۳۲) و بالاترین همبستگی

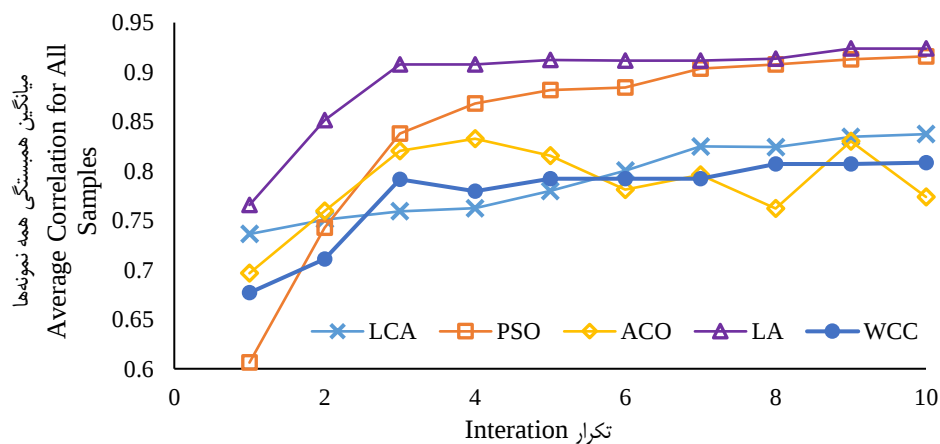
برتر برای انتخاب طول موج‌ها توصیه می‌شود. اگرچه برای آلو خونی، ACO و LCA از نظر دقت کمی بهتر از PSO عمل کردند، اما سرعت بالای PSO آن را به گزینه‌ای بسیار جذاب تبدیل می‌کند. با توجه به این که مدل‌های ANN خود نیازمند منابع محاسباتی هستند، انتخاب الگوریتمی مانند PSO که علاوه بر دقت مناسب، کارایی محاسباتی بالایی در مرحله انتخاب ویژگی دارد، به بهینه‌سازی زمان کلی مدل‌سازی کمک شایانی خواهد کرد؛ بنابراین، با در نظر گرفتن هر دو فاکتور دقت و کارایی محاسباتی، الگوریتم PSO به عنوان روشی بهینه برای استخراج زیرمجموعه‌ای از طول موج‌های کلیدی پیشنهاد می‌گردد. این انتخاب تضمین می‌کند که داده‌ها با حداقل ابعاد ممکن، حاوی اطلاعات کافی برای توسعه مدل‌های ANN دقیق و پایدار باشند و درعین حال، فرآیند کلی مدل‌سازی به لحاظ زمانی نیز بهینه گردد.

انتخاب طول موج استفاده کرده‌اند. در آن مطالعه، SVM-PSO به عنوان روش بهینه انتخاب طول موج بر اساس RPD، RMSE و مقادیر همبستگی شناخته شد و به طور متوسط ۱۵ طول موج مؤثر شناسایی گردید. این همخوانی در نتایج، یعنی مؤثر بودن PSO در انتخاب طول موج‌های کلیدی، حتی با وجود تفاوت در مدل پایه ترکیبی (در پژوهش حاضر DT-PSO و در مطالعه لطیفی SVM-PSO)، بر توانایی بالای این الگوریتم در کاهش ابعاد داده‌های طیفی و استخراج اطلاعات حیاتی تأکید می‌کند.

با در نظر گرفتن این نتایج جامع، انتخاب الگوریتم برتر برای مدل‌سازی‌های بعدی با شبکه‌های عصبی مصنوعی نیازمند توازن دقیق بین دقت پیش‌بینی و کارایی محاسباتی است. برای آلو خرمایی، الگوریتم PSO به دلیل تعادل عالی میان دقت (RMSE پایین و همبستگی بالا) و به‌ویژه زمان اجرای فوق‌العاده پایین، به عنوان گزینه



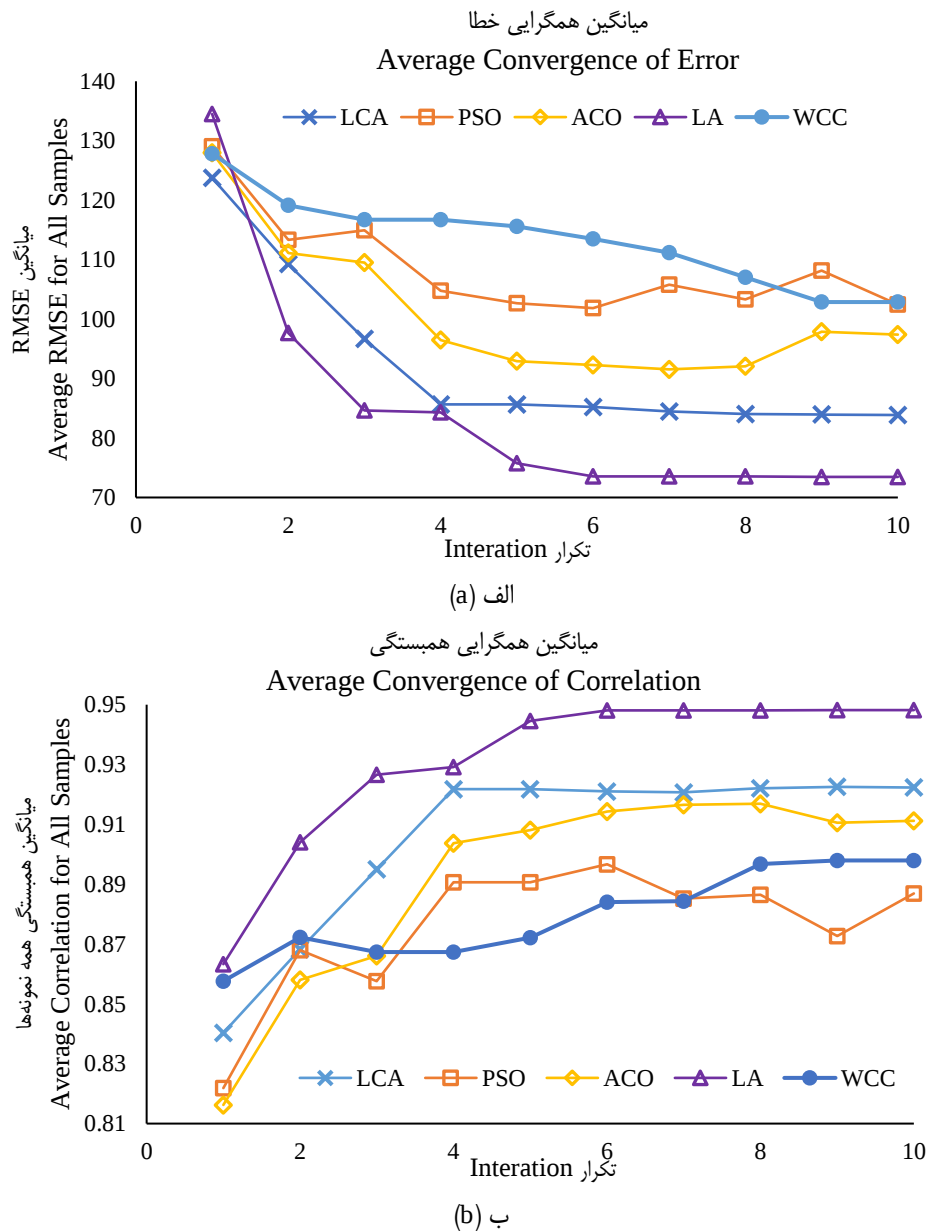
(a) الف



(b) ب

شکل ۴- عملکرد مدل DT همراه با الگوریتم‌های مختلف برای نمونه‌های رقم خرمایی: (الف) میانگین RMSE برای همه نمونه‌ها، (ب) میانگین همبستگی همه نمونه‌ها

Fig. 4. Performance of the DT model with various algorithms for Khormaei plum samples: (a) Average RMSE for all samples, (b) Average correlation for all samples



شکل ۵- عملکرد مدل DT همراه با الگوریتم‌های مختلف برای نمونه‌های رقم خونی الف) میانگین RMSE برای همه نمونه‌ها، ب) میانگین همبستگی همه نمونه‌ها

Fig. 5. Performance of the DT model with various algorithms for Khoni plum samples: (a) Average RMSE for all samples, (b) Average correlation for all samples

جدول ۵- طول موج‌های منتخب الگوریتم DT-PSO

Table 5- Selected wavelengths by DT-PSO algorithm

رقم Variety	طول موج‌های موثر انتخاب شده (نانومتر) Selected EWs (nm)
خرمایی Khormaei	519.967, 567.227, 438.989, 471.704, 861.47, 688.055, 571.977, 866.435, 442.203, 746.972, 722.265, 462.389, 507.02, 428.224, 1021.683
خونی Khoni	988.625, 819.5, 432.232, 636.415, 435.922, 630.429, 559.92, 741.437, 527.776, 460.428, 962.516, 950.06, 512.792, 968.704, 794.096

مدل‌سازی طول‌موج‌های منتخب DT-PSO با استفاده از

روش‌های PLSR و ANN

برای شناسایی بهینه‌ترین طول‌موج‌ها و ارزیابی جامع عملکرد مدل‌ها، پنج الگوریتم فراابتکاری به کار گرفته شدند. پس از انتخاب طول‌موج‌های مؤثر توسط PSO (جدول ۵)، نتایج مدل‌سازی با استفاده از روش PLSR و ANN بر روی این طول‌موج‌های منتخب مورد تحلیل قرار گرفت. در بررسی رقم آلو خرمایی، نتایج مدل‌سازی PLSR با طول‌موج‌های منتخب PSO (جدول ۶) حاکی از آن بود که رویکردهایی نظیر بدون پیش‌پردازش، مشتق اول و دوم، و مرکز کردن میانگین با مقادیر تقریباً یکسان در محله آزمون $R^2=0/6404$ و $RPD=1/72$ ، عملکرد متوسطی را در پیش‌بینی برای داده‌های جدید از خود نشان دادند. این RPD که کمتر از ۲ است، اگرچه برای تخمین اولیه مفید است، اما دقت لازم برای کاربردهای حساس‌تر نظیر غربالگری دقیق یا کنترل کیفیت را ندارد و نشان‌دهنده لزوم بهبود بیشتر مدل است. در مقابل، مدل ANN که قابلیت بالاتری در کشف روابط غیرخطی دارد، برای آلو خرمایی (جدول ۸) با پیش‌پردازش‌های بدون پیش‌پردازش و فیلتر میانه به $R^2=0/76$ حدود $RPD=2/12$ دست یافت. این بهبود در RPD، قابلیت ANN را برای غربالگری قابل قبول نشان می‌دهد و توانایی آن در استخراج اطلاعات از طول‌موج‌های منتخب را برجسته می‌کند. نمودار پراکنش مقادیر اندازه‌گیری‌شده در برابر مقادیر پیش‌بینی‌شده برای این مدل در شکل ۶-الف ارائه شده است. در مورد رقم آلو خونی، مدل‌های PLSR با طول‌موج‌های منتخب PSO (جدول ۷) با پیش‌پردازش‌های ساویتزکی-گولای و فیلتر میانه به ترتیب به $RPD=2/85$ و $2/83$ در فاز آزمون دست یافتند. این مقادیر RPD که نزدیک به ۳ هستند، دلالت بر عملکرد قوی مدل برای کاربردهای کنترل کیفیت در سطح قابل اعتماد دارند. باین حال، اوج عملکرد در مدل‌سازی ANN برای آلو خونی (جدول ۹) حاصل شد؛ جایی که پیش‌پردازش فیلتر گوسی با R^2 آزمون $0/9834$ و $RPD=8/01$ ، و بدون پیش‌پردازش با RPD برابر با $7/45$ ، نتایج قابل توجه را به نمایش گذاشتند که به معنی قابلیت این مدل‌ها برای اندازه‌گیری کمی بسیار دقیق و حتی کاربردهای تحقیقاتی پیشرفته است. این نتایج خارق‌العاده به صورت بصری در شکل ۶-ب (با تأکید بر مدل فیلتر گوسی) قابل مشاهده هستند.

معماری و نقش شبکه عصبی مصنوعی در مدل‌سازی

معماری شبکه عصبی مصنوعی، به‌ویژه تعداد لایه‌های پنهان و تعداد نوروها در هر لایه، نقش تعیین‌کننده‌ای در قابلیت مدل برای یادگیری و استخراج الگوهای پیچیده از داده‌ها دارد. همان‌طور که در

ستون بهترین ساختار شبکه جداول ۸ و ۹ مشاهده می‌شود، ساختارهای مختلفی برای هر سناریو بهینه شده‌اند؛ از شبکه‌های ساده تک‌لایه‌ای با تعداد کمی نورو (مانند ۵ یا ۱۰) تا ساختارهای پیچیده‌تر با چندین لایه پنهان و نوروهای بیشتر (مانند [۵ ۵] یا [۱۰ ۱۵]). انتخاب بهینه این ساختار، نه تنها بر توانایی مدل در برازش دقیق داده‌های آموزشی تأثیر می‌گذارد، بلکه بر قابلیت تعمیم‌پذیری آن به داده‌های جدید نیز اثرگذار است. یک معماری نامناسب (بسیار ساده یا بسیار پیچیده) می‌تواند به ترتیب منجر به کم‌برازش^۱ یا بیش‌برازش^۲ شود. کم‌برازش به معنای عدم توانایی مدل در یادگیری الگوهای موجود در داده است، در حالی که بیش‌برازش به معنای یادگیری بیش از حد نویز و جزئیات غیرمفید از داده‌های آموزشی و در نتیجه عملکرد ضعیف بر روی داده‌های آزمون است. نتایج به‌دست‌آمده با ANN در این مطالعه، به‌خصوص برای آلو خونی، به‌خوبی نشان می‌دهد که با وجود کاهش ابعاد داده‌های ورودی، شبکه‌های عصبی با معماری بهینه قادرند روابط غیرخطی پیچیده موجود در طول‌موج‌های منتخب را با دقت فوق‌العاده‌ای استخراج کنند. این توانایی ANN در مدل‌سازی روابط غیرخطی، یک مزیت کلیدی نسبت به روش‌های خطی مانند PLSR است و به همین دلیل در بسیاری از موارد به عملکرد بهتری در فاز آزمون منجر شده است.

در پژوهشی مرتبط با ارزیابی غیرمخرب سیب، قابلیت استفاده از تصویربرداری فراطیفی VIS-NIR با قیمت پایین (در محدوده ۳۸۶-۱۰۲۸ نانومتر) برای پیش‌بینی رطوبت (MC) و pH سیب Pink Lady در سه مرحله رسیدگی مختلف مورد بررسی قرار گرفت. در آن مطالعه، از پنج الگوریتم یادگیری ماشین شامل PLSR، MLR، kNN، DT و ANN برای تحلیل داده‌ها استفاده شد. ویژگی‌های بهینه با استفاده از رویکرد انتخاب ویژگی Bootstrap Random Forest شناسایی شدند. در میان الگوریتم‌های آزمایشی، ANN بهترین عملکرد را برای رطوبت ($R=0/756$ ، $RMSE=0/868$) ارائه داد، اما برای پیش‌بینی pH عملکرد ضعیفی داشت ($R=0/383$ ، $RMSE=0/044$) (Kavuncuoğlu, Çetin, Yildirim, Nadimi, & Paliwal, 2023). این نتیجه در تضاد با یافته‌های برجسته پژوهش حاضر برای pH آلو است؛ جایی که ANN بر روی طول‌موج‌های منتخب توانست pH آلو خونی را با $R^2=0/9834$ و $RPD=8/01$ پیش‌بینی کند. این تفاوت فاحش در عملکرد pH نشان می‌دهد که در حالی که فناوری VIS-NIR HSI پتانسیل زیادی دارد، کارایی آن در پیش‌بینی pH ممکن است به شدت به گونه میوه، دامنه pH مورد بررسی، و به‌ویژه به روش‌های بهینه‌سازی مدل و انتخاب ویژگی بستگی داشته باشد. موفقیت قابل توجه پژوهش حاضر در پیش‌بینی

یافت که به مراتب از بهترین عملکرد PLSR با طیف کامل فراتر است. این امر، بر تولنایی ANN در استخراج روابط پیچیده و غیرخطی از داده‌های کاهش یافته طیفی تأکید می‌کند.

این مقایسه‌های دقیق به وضوح نشان می‌دهند که الگوریتم PSO یک استراتژی بسیار کارآمد برای کاهش ابعاد داده‌های طیفی است. این رویکرد نه تنها منجر به کاهش چشمگیر پیچیدگی محاسباتی می‌شود (که در این تحقیق دلیل اصلی عدم استفاده از ANN روی طیف کامل بود)، بلکه امکان بهره‌برداری از مدل‌های قدرتمندتر و زمان‌بر نظیر شبکه‌های عصبی مصنوعی را نیز فراهم می‌آورد. همان‌طور که مشاهده شد، ANN با طول موج‌های منتخب، به ویژه برای رقم آلو خونی، به دقت‌های پیش‌بینی بسیار بالایی دست یافت که دستیابی به آن‌ها با طیف کامل داده‌ها، بدون کاهش ابعاد، به دلیل نیاز به منابع محاسباتی فراوان و زمان‌بر بودن آموزش، عملاً غیرممکن بود. علاوه بر این، انتخاب صحیح روش پیش‌پردازش طیفی نقش حیاتی در بهینه‌سازی عملکرد مدل‌ها ایفا کرد؛ روش‌های هموارسازی مانند Savitzky-Golay و فیلتر گوسی (که به ترتیب به کاهش نویز و حذف نویزهای فرکانس بالا کمک می‌کنند) و همچنین روش‌های مبتنی بر مشتق (که به حذف خط پایه و افزایش وضوح پیک‌ها منجر می‌شوند) به خصوص در مدل‌سازی آلو خونی نتایج درخشانی داشته‌اند. در مجموع، تلفیق کاهش ابعاد هوشمندانه با PSO و مدل‌سازی پیشرفته با ANN، یک راهکار بهینه و مؤثر برای پیش‌بینی دقیق و کارآمد خواص فیزیکوشیمیایی در نمونه‌های آلو را ارائه می‌دهد که هم از نظر دقت و هم از نظر کارایی محاسباتی برتری چشمگیری دارد.

pH آلو با روش ANN، بر ظرفیت بالای این رویکرد در مدیریت پیچیدگی‌های بیولوژیکی و طیفی و ارائه دقت بیش از انتظار تأکید می‌کند، حتی در شرایطی که مطالعات دیگر برای پارامتر مشابه در میوه‌های دیگر با چالش مواجه بوده‌اند.

برای ارزیابی تأثیر واقعی کاهش ابعاد، نتایج مدل‌سازی با طول موج‌های منتخب PSO با عملکرد مدل‌سازی PLSR بر روی طیف کامل داده‌ها (جدول ۳ و ۴) مقایسه شد. این مقایسه نشان داد که برای رقم آلو خرمایی، بهترین RPD آزمون در مدل PLSR با طیف کامل (جدول ۳) حدود ۲/۴۴ (با پیش‌پردازش میانگین متحرک) بود که اندکی از بهترین RPD در PLSR با طول موج‌های منتخب (۱/۷۲) بالاتر است. این تفاوت جزئی، می‌تواند ناشی از دست دادن برخی اطلاعات ریز در فرآیند کاهش ابعاد برای مدل خطی PLSR باشد. با این حال، همان‌طور که قبلاً اشاره شد، مدل ANN با طول موج‌های منتخب (جدول ۸، RPD ۲/۱۲) توانست این شکاف را تا حد زیادی جبران کرده و عملکردی رقابتی ارائه دهد. در مورد آلو خونی، نتایج حتی چشم‌گیرتر بود؛ بهترین RPD آزمون در PLSR با طیف کامل (جدول ۴) حدود ۳/۰۱ (با فیلتر گوسی) بود که بسیار نزدیک به بهترین RPD در PLSR با طول موج‌های منتخب (۲/۸۵) است. این تطابق نزدیک، نشان‌دهنده قابلیت بالای الگوریتم PSO در انتخاب طول موج‌هایی است که با وجود کاهش چشمگیر ابعاد داده‌ها، حاوی اطلاعات کافی برای حفظ دقت مدل‌های خطی هستند. از همه مهم‌تر، مدل ANN با طول موج‌های منتخب (جدول ۹) برای آلو خونی به RPD قابل توجه ۸/۰۱ دست

جدول ۶- نتایج مدل‌سازی PLSR طول موج‌های منتخب PSO در رقم خرمایی با انواع پیش‌پردازش‌های طیفی (۷ مؤلفه)

Table 6- Results of the PLSR model on DT-PSO selected wavelengths for the Khormaei plum variety with various spectral preprocessing methods (7 components)

ردیف No	پیش‌پردازش Preprocessing	آموزش Training			اعتبارسنجی Validation			آزمون Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	بدون پیش‌پردازش No Preprocessing	0.8652	0.0585	2.75	0.9264	0.0534	3.81	0.6404	0.118	1.72
2	SNV	0.7719	0.0761	2.12	0.7365	0.1011	2.01	0.6802	0.1113	1.83
3	MSC	0.7675	0.0768	2.1	0.7088	0.1062	1.91	0.6447	0.1173	1.73
4	مشتق اول First Derivative	0.8652	0.0585	2.75	0.9264	0.0534	3.81	0.6404	0.118	1.72
5	مشتق دوم Second Derivative	0.8652	0.0585	2.75	0.9264	0.0534	3.81	0.6404	0.118	1.72
6	نرمال‌سازی Normalisation	0.8623	0.0591	2.72	0.9294	0.0523	3.89	0.6379	0.1184	1.72
7	میانگین متحرک Moving Average	0.8574	0.0602	2.68	0.9135	0.0579	3.51	0.6208	0.1211	1.68
8	فیلتر گوسی Gaussian Filter	0.7634	0.0775	2.08	0.8338	0.0802	2.53	0.5922	0.1256	1.62
9	فیلتر میانه Median Filter	0.8333	0.065	2.48	0.8613	0.0733	2.77	0.693	0.109	1.86

10	ساویتزکی-گولای Savitzky-Golay	0.7947	0.0722	2.23	0.8869	0.0662	3.07	0.6985	0.108	1.88
11	حذف روند Detrending	0.8616	0.0593	2.72	0.9009	0.062	3.28	0.6251	0.1205	1.69
12	مرکز کردن میانگین Mean Centering	0.8652	0.0585	2.75	0.9264	0.0534	3.81	0.6404	0.118	1.72

جدول ۷- نتایج مدل‌سازی PLSR طول‌موج‌های منتخب PSO در رقم خونی با انواع پیش‌پردازش‌های طیفی (۸ مؤلفه)

Table 7- Results of the PLSR model on DT-PSO selected wavelengths for the Khoni plum variety with various spectral preprocessing methods (8 components)

ردیف No	پیش‌پردازش Preprocessing	آموزش Training			اعتبارسنجی Validation			آزمون Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	بدون پیش‌پردازش No Preprocessing	0.9711	0.0681	5.95	0.9771	0.0647	6.83	0.8519	0.158	2.68
2	SNV	0.8776	0.1402	2.89	0.9006	0.1348	3.28	0.6474	0.2438	1.74
3	MSC	0.8764	0.1409	2.87	0.8921	0.1405	3.14	0.6734	0.2347	1.81
4	مشتق اول First Derivative	0.9711	0.0681	5.95	0.9771	0.0647	6.83	0.8519	0.158	2.68
5	مشتق دوم Second Derivative	0.9711	0.0681	5.95	0.9771	0.0647	6.83	0.8519	0.158	2.68
6	نرمال‌سازی Normalisation	0.9688	0.0707	5.72	0.9673	0.0773	5.71	0.8041	0.1817	2.33
7	میانگین متحرک Moving Average	0.9663	0.0735	5.51	0.9781	0.0633	6.97	0.8558	0.1559	2.72
8	فیلتر گوسی Gaussian Filter	0.9499	0.0897	4.51	0.9752	0.0673	6.56	0.8592	0.154	2.75
9	فیلتر میانه Median Filter	0.6685	0.2307	1.76	0.7864	0.1976	2.23	0.8665	0.15	2.83
10	ساویتزکی-گولای Savitzky-Golay	0.9529	0.0869	4.66	0.9669	0.0778	5.68	0.8684	0.1489	2.85
11	حذف روند Detrending	0.953	0.0869	4.66	0.9563	0.0894	4.94	0.7655	0.1989	2.13
12	مرکز کردن میانگین Mean Centering	0.9711	0.0681	5.95	0.9771	0.0647	6.83	0.8519	0.158	2.68

جدول ۸- نتایج مدل‌سازی ANN طول‌موج‌های منتخب PSO در رقم خرمایی با انواع پیش‌پردازش‌های طیفی

Table 8- Results of the ANN model on DT-PSO selected wavelengths for the Khormaei plum variety with various spectral preprocessing methods

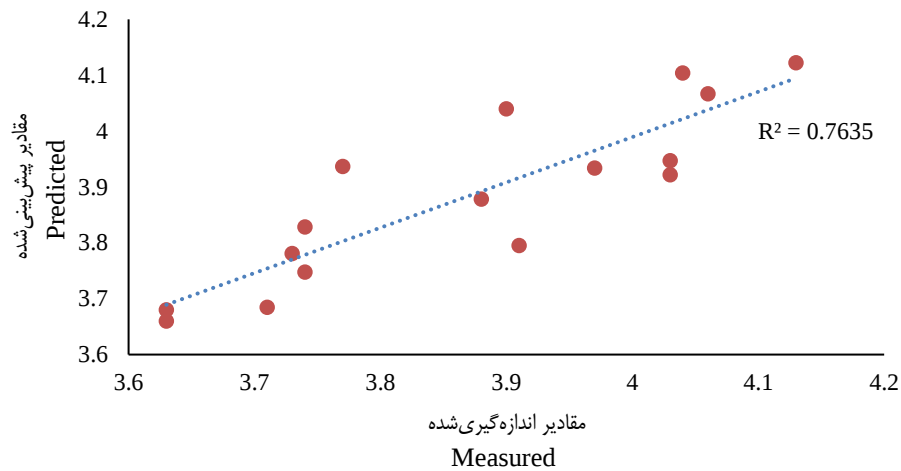
ردیف No	پیش‌پردازش Preprocessing	بهترین ساختار شبکه Best network topology	آموزش Training			اعتبارسنجی Validation			آزمون Test		
			R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	بدون پیش‌پردازش No Preprocessing	10	0.9061	0.0556	3.3	0.77	0.0837	2.15	0.7636	0.0769	2.12
2	SNV	[5 5]	0.6931	0.1005	1.82	0.5664	0.115	1.57	0.7168	0.0842	1.94
3	MSC	[10 15 20]	0.7089	0.0979	1.87	0.6819	0.0985	1.83	0.3934	0.1232	1.33
4	مشتق اول First Derivative	[10 5]	0.8761	0.0639	2.87	0.2986	0.1462	1.23	0.4863	0.1134	1.44

5	مشتق دوم Second Derivative	5	0.8513	0.07	2.62	0.6064	0.1095	1.65	0.5551	0.1055	1.55
6	نرمال سازی Normalisation	5	0.5963	0.1153	1.59	0.6315	0.106	1.7	0.548	0.1064	1.54
7	میانگین متحرک Moving Average	[5 5]	0.8585	0.0683	2.69	0.7666	0.0843	2.14	0.698	0.087	1.88
8	فیلتر گوسی Gaussian Filter	[10 10]	0.8972	0.0582	3.15	0.7017	0.0953	1.89	0.5976	0.1004	1.63
9	فیلتر میانه Median Filter	5	0.8412	0.0723	2.54	0.6377	0.1051	1.72	0.7624	0.0771	2.12
10	ساویتزکی-گولای Savitzky-Golay	[15 15]	0.8885	0.0606	3.03	0.56	0.1158	1.56	0.6078	0.0991	1.65
11	حذف روند Detrending	[15 15]	0.9167	0.0524	3.5	0.6407	0.1046	1.72	0.6433	0.0945	1.73
12	مرکز کردن میانگین Mean Centering	[5 5]	0.9089	0.0548	3.35	0.7673	0.0842	2.14	0.7291	0.0823	1.98

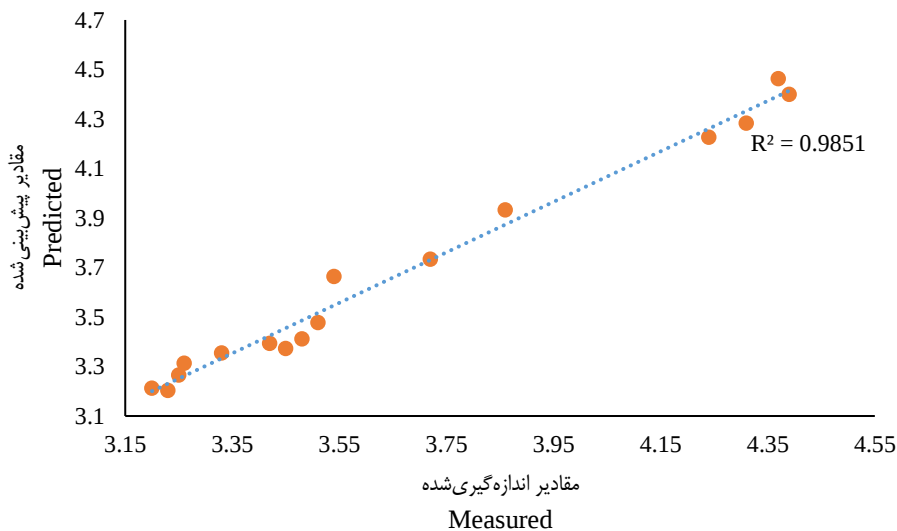
جدول ۹- نتایج مدل سازی ANN طول موج های منتخب PSO در رقم خونی با انواع پیش پردازش های طیفی

Table 9- Results of the ANN model on DT-PSO selected wavelengths for the Khoni plum variety with various spectral preprocessing methods

ردیف No	پیش پردازش Preprocessing	بهترین ساختار شبکه Best network topology	آموزش Training			اعتبار سنجی Validation			آزمون Test		
			R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	بدون پیش پردازش No Preprocessing	[20 5]	0.9889	0.0444	9.6	0.9674	0.0632	5.72	0.9808	0.0585	7.45
2	SNV	[5 5]	0.9944	0.0316	13.49	0.8877	0.1175	3.08	0.9621	0.082	5.31
3	MSC	[10 10]	0.7944	0.1912	2.23	0.2903	0.2952	1.23	0.7587	0.2071	2.1
4	مشتق اول First Derivative	[15 5]	0.9842	0.053	8.04	0.5508	0.2348	1.54	0.9332	0.109	4
5	مشتق دوم Second Derivative	[20 5]	0.9765	0.0647	6.59	0.8437	0.1385	2.61	0.9157	0.1224	3.56
6	نرمال سازی Normalisation	[10 10]	0.8451	0.166	2.57	0.6301	0.2131	1.7	0.8371	0.1701	2.56
7	میانگین متحرک Moving Average	5	0.9762	0.0651	6.55	0.946	0.0814	4.45	0.9804	0.0591	7.37
8	فیلتر گوسی Gaussian Filter	[20 5]	0.9722	0.0702	6.07	0.9012	0.1101	3.29	0.9834	0.0543	8.01
9	فیلتر میانه Median Filter	[10 5]	0.8573	0.1593	2.67	0.1093	0.3307	1.09	0.8395	0.1689	2.58
10	ساویتزکی-گولای Savitzky-Golay	[15 5]	0.9854	0.051	8.36	0.9629	0.0675	5.36	0.9826	0.0555	7.84
11	حذف روند Detrending	[15 5]	0.9861	0.0497	8.58	0.8305	0.1443	2.51	0.9536	0.0908	4.8
12	مرکز کردن میانگین Mean Centering	[10 15 20]	0.9555	0.0889	4.79	0.8349	0.1424	2.54	0.9731	0.0692	6.29



(a) الف



(b) ب

شکل ۶- نمودار پراکنش مقادیر pH اندازه‌گیری شده در برابر مقادیر پیش‌بینی شده برای بهترین مدل ANN بر روی مجموعه آزمون الف) ساختار معماری شبکه 10 بدون پیش‌پردازش رقم خرمایی ب) ساختار معماری شبکه [20 5] رقم خونی

Fig. 6. Scatter plot of the measured versus predicted pH values for the best ANN model on the test set for: (a) Khormaei variety (network architecture 10, no preprocessing), (b) Khoni variety (network architecture [20 5])

مقایسه با پژوهش‌های مشابه

به‌منظور تحلیل عمق دستاوردهای مدل‌های توسعه‌یافته در این پژوهش، به‌ویژه عملکرد برجسته مدل ANN بر روی رقم آلو خونی نتایج با تحقیقات پیشین و جدید در حوزه ارزیابی غیرمخرب pH در میوه‌ها مقایسه شده است. این مقایسه برتری‌های روش ترکیبی انتخاب ویژگی (DT-PSO) و قدرت مدل غیرخطی ANN را در

مقایسه با روش‌های سنتی‌تر (مانند PLSR) آشکار می‌سازد.

مقایسه با مطالعاتی که از همان پارامتر کیفی (pH) و تکنیک مشابه (HSI) استفاده کرده‌اند، از اهمیت بالایی برخوردار است بنابراین به بررسی نتایج یکسری از پژوهش اخیر در این زمینه پرداخته شد.

پژوهشی توسط گل‌محمدی و همکاران (Golmohammadi et al., 2022) برای تخمین pH سیب رقم Red Delicious در محدوده ۴۰۰-۱۱۰۰ نانومتر انجام شد. این محققان پس از تعیین ۹ طول موج

نتایج ضعیف‌تری شامل ضریب همبستگی $R^2=0/55$ کسب کردند. عملکرد ضعیف‌تر در پیش‌بینی pH لیچی که حتی به حد قابل قبول برای ارزیابی روتین هم نمی‌رسد چالش ذاتی پیش‌بینی pH در میوه‌های گرمسیری یا هسته‌دار (مانند آلو و لیچی) را برجسته می‌سازد. علی‌رغم استفاده از محدوده طیفی طولانی‌تر که اطلاعات غنی‌تری در مورد پیوندهای شیمیایی را در اختیار می‌گذارد، استفاده از مدل PLSR خطی موجب شده است که مدل نتواند واریانس‌های pH را به‌طور مؤثر پوشش دهد. این موضوع قاطعانه تأکید می‌کند که در مواردی که هدف دستیابی به دقت بالا است، لازم است از مدل‌های غیرخطی پیشرفته‌ای مانند ANN در کنار تکنیک‌های دقیق انتخاب ویژگی، نظیر رویکرد DT-PSO در پژوهش حاضر، استفاده شود.

مقایسه با مجموعه‌ای از مطالعات کلیدی در زمینه ارزیابی pH میوه‌ها به‌وضوح نشان داد که دستیابی به دقت بالا در پیش‌بینی pH یک چالش بزرگ در میوه‌شناسی غیرمخرب باقی مانده است. عملکرد مدل ANN در پژوهش حاضر جایگاه متمایزی را برای این تحقیق فراهم می‌آورد. عملکرد مدل ANN، نه تنها مدل‌های خطی را پشت سر می‌گذارد، بلکه برتری چشمگیری نسبت به اکثر نتایج گزارش شده در ادبیات موضوع برای پارامتر pH در میوه‌های با خواص اسیدی مشابه دارد. این نتیجه، اعتبار روش ترکیبی DT-PSO را در بهینه‌سازی مدل‌های یادگیری ماشین برای اندازه‌گیری کیفی میوه‌ها تأیید می‌کند و نشان می‌دهد که تلفیق مناسب الگوریتم‌های فراابتکاری با قابلیت‌های غیرخطی ANN، راهکار مؤثری برای حل چالش‌های طیف‌سنجی در میوه‌های دارای pH متغیر است.

نتیجه‌گیری

این پژوهش با هدف توسعه و ارزیابی یک روش غیرمخرب و دقیق برای پیش‌بینی pH در دو رقم مهم آلو (خرمایی و خونی) با استفاده از فناوری تصویربرداری فراطیفی در محدوده مرئی-مادون قرمز نزدیک (Vis-NIR) به انجام رسید. دستاوردهای این مطالعه به‌وضوح نشان‌دهنده پتانسیل بالای این تکنیک، در کنار کاربرد مؤثر روش‌های پیش‌پردازش طیفی و الگوریتم‌های هوش مصنوعی، برای ارزیابی سریع و کارآمد خواص کیفی میوه‌هاست.

نتایج حاصل از بررسی تأثیر روش‌های مختلف پیش‌پردازش طیفی بر عملکرد مدل‌های رگرسیون حداقل مربعات جزئی (PLSR) به‌وضوح نشان داد که انتخاب بهینه پیش‌پردازش، نقشی حیاتی در بهبود دقت پیش‌بینی ایفا می‌کند. یکی از چالش‌های اصلی در تحلیل داده‌های طیفی، پدیده هم‌خطی و حجم بالای داده‌هاست. این پژوهش با موفقیت از الگوریتم بهینه‌سازی ازدحام ذرات (PSO) در ترکیب با

مؤثر، از مدل ANN استفاده کردند و بهترین نتایج در مرحله اعتبارسنجی شامل $R^2=0/98$ و $RMSE=0/018$ به‌دست آمد. در حالی که R^2 این پژوهش بسیار نزدیک به R^2 آزمون مدل ANN آلو خونی است، اما معیاری که برتری مدل حاضر را به‌وضوح نشان می‌دهد، شاخص RPD است. با دستیابی به $RPD=8/01$ ، مدل ما به یک سطح دقت و قدرت پیش‌بینی رسیده که به مراتب بالاتر از سطح RPD مورد نیاز برای کنترل کیفیت دقیق صنعتی است. این برتری قابل‌ملاحظه، احتمالاً ناشی از کارایی بالاتر روش ترکیبی DT-PSO در انتخاب بهینه‌ترین زیرمجموعه از ویژگی‌های طیفی در مقایسه با روش‌های صرفاً مبتنی بر ضریب رگرسیون است.

ژیمینگ و همکاران (Zhiming et al., 2014) با استفاده از تصویربرداری فراطیفی فروسرخ موج کوتاه، در محدوده ۱۰۰۰-۲۵۰۰ نانومتر برای پیش‌بینی pH سیب فوجی، به بهترین نتیجه با مدل رگرسیون حداقل مربعات جزئی بازه‌های هم‌افزایی^۱ (siPLS) شامل $R^2=0/718$ و $RMSEP=0/0398$ دست یافتند. عملکرد ضعیف‌تر مدل siPLS در مقایسه با مدل ANN حاضر، بر اثربخشی بالاتر مدل‌های غیرخطی برای پیش‌بینی پارامترهای شیمیایی پیچیده میوه‌ها تأکید می‌کند. اگرچه محدوده طیفی^۲ SWIR (موج بلند) حاوی اطلاعات مهمی درباره پیوندهای O-H و C-H است، اما مدل‌های خطی مانند siPLS و PLSR قادر به درک روابط غیرخطی میان این اطلاعات طیفی و pH میوه با دقت لازم نیستند؛ موضوعی که مدل ANN در پژوهش حاضر با موفقیت آن را مدیریت کرده است.

پوروانتو و همکاران (Purwanto et al., 2013) در پژوهش خود، با استفاده از طیف‌سنجی NIR در حالت عبور در بازه ۱۰۰۰-۲۴۰۰ نانومتر و مدل PLS را برای پیش‌بینی pH انبه رقم Gedong Gincu توسعه دادند و به ضریب همبستگی $R^2=0/817$ دست یافتند. اگرچه این نتیجه برای پیش‌بینی pH انبه، در محدوده مناسب برای غربالگری و سورتینگ مقدماتی قرار می‌گیرد، اما به‌مراتب پایین‌تر از دقت مدل ANN حاضر در مرحله آزمون است. این تفاوت در دقت، نه تنها به ماهیت غیرخطی روابط طیفی pH اشاره دارد، بلکه بر اهمیت حیاتی کاهش دقیق ابعاد تأکید می‌کند. مدل PLS، به‌عنوان یک مدل خطی، نمی‌تواند پیچیدگی سیگنال‌های مربوط به اسیدهای آلی و ساختار قند-اسید (که مستقیماً بر pH تأثیر می‌گذارند) را در محدوده NIR به‌طور کامل درک کند. در مقابل، مدل ما با استفاده از DT-PSO، تنها اطلاعات غیرهم‌خط و پر قدرت را به ANN وارد کرده است، که ANN را قادر می‌سازد تا این روابط پیچیده را با دقت فراتر از مدل‌های خطی، مدل‌سازی کند.

پو و همکاران (Pu et al., 2016) در ارزیابی pH میوه لیچی با استفاده از HSI در محدوده ۱۰۰۰-۲۵۰۰ نانومتر و مدل PLSR،

با تلفیق نوآورانه تصویربرداری فراطیفی، کاهش ابعاد هوشمندانه با DT-PSO و مدل‌سازی پیشرفته غیرخطی با ANN، یک راهکار دقیق و کارآمد برای ارزیابی غیرمخرب pH ارقام مختلف آلو ارائه می‌دهد.

مشارکت نویسندگان

میثم لطیفی عموقین: مفهوم‌سازی، روش‌شناسی، جمع‌آوری داده، پردازش داده، تصویرسازی نتایج، استخراج و تهیه متن اولیه
یوسف عباس‌پور گیلانده: نظارت و مدیریت، مفهوم‌سازی، مشاوره فنی، تحلیل آماری، اعتبارسنجی، ویرایش متن

درخت تصمیم (DT-PSO) به‌عنوان یک روش بسیار کارآمد برای کاهش ابعاد داده‌ها و انتخاب طول‌موج‌های مؤثر بهره برد. الگوریتم PSO با وجود کاهش چشمگیر تعداد متغیرها، توانست طول‌موج‌هایی را انتخاب کند که اطلاعات کلیدی را در خود حفظ کرده و به دقت پیش‌بینی بالا در مدل‌های بعدی منجر شدند. در نهایت، مقایسه عملکرد PLSR و شبکه‌های عصبی مصنوعی (ANN) بر روی طول‌موج‌های منتخب، برتری چشمگیر ANN را در مدل‌سازی روابط پیچیده و غیرخطی آشکار ساخت. در حالی که مدل‌های PLSR با طول‌موج‌های منتخب عملکرد قابل‌قبولی ارائه دادند، ANN توانست به دقت‌های بسیار بالاتری دست یابد. همچنین، معماری بهینه شبکه عصبی، شامل تعداد لایه‌های پنهان و نورون‌ها، نقشی کلیدی در به حداکثر رساندن دقت این مدل‌ها ایفا کرد. به‌طور خلاصه، این مطالعه

References

1. Alam, S., Dobbie, G., Koh, Y. S., Riddle, P., & Rehman, S. U. (2014). Research on particle swarm optimization based clustering: a systematic review of literature and techniques. *Swarm and Evolutionary Computation*, 17, 1-13. <https://doi.org/10.1016/j.swevo.2014.02.001>
2. Ayub, H., Nadeem, M., Mohsin, M., Ambreen, S., Khan, F. A., Oranab, S., Rahim, M. A., Zubair khalid, M., Zongo, E., & Zarlash, M. (2023). A comprehensive review on the availability of bioactive compounds, phytochemicals, and antioxidant potential of plum (*Prunus Domestica*). *International Journal of Food Properties*, 26(1), 2388-2406. <https://doi.org/10.1080/10942912.2023.2249254>
3. Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2), 480-490. <https://doi.org/10.2136/sssaj2001.652480x>
4. Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE computational intelligence magazine*, 1(4), 28-39.
5. Kashan, A. H. (2014). League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships. *Applied Soft Computing*, 16, 171-200. <https://doi.org/10.1016/j.asoc.2013.12.005>
6. Gao, S., & Xie, W. (2024). SSC and pH prediction and maturity classification of grapes based on hyperspectral imaging. *Smart Agricultural Technology*, 8, 100457. <https://doi.org/10.1016/j.atech.2024.100457>
7. Golmohammadi, A., Tahmasebi, M., & Razavi, M. S. (2022). Near infrared hyperspectral imaging for non-destructive determination of pH value in red delicious apple fruit during shelf life. *Innovative Food Technologies*, 9(2), 99-111. <https://doi.org/10.22104/jift.2021.5124.2059>
8. Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Hernández-Hernández, M., Gallardo-Bernal, I., & Hernández-Hernández, J. L. (2022). Non-destructive detection of fruit quality parameters using hyperspectral imaging, multiple regression analysis and artificial intelligence. *Horticulturae*, 8(7), 598. <https://doi.org/10.3390/horticulturae8070598>
9. Huang, H., Liu, L., & Ngadi, M. O. (2014). Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4), 7248-7276. <https://doi.org/10.3390/s140407248>
10. Kashan, A. H. (2014). League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships. *Applied Soft Computing*, 16, 171-200. <https://doi.org/10.1016/j.asoc.2013.12.005>
11. Kavuncuoğlu, E., Çetin, N., Yildirim, B., Nadimi, M., & Paliwal, J. (2023). Exploration of machine learning algorithms for pH and moisture estimation in apples using VIS-NIR imaging. *Applied Sciences*, 13(14), 8391. <https://doi.org/10.3390/app13148391>
12. Khodabakhshian, R., & Baghbani, R. (2025). Non-destructive Internal Quality Evaluation of Apple Fruit Using X-ray CT. *Journal of Agricultural Machinery*. (in Press). <https://doi.org/10.22067/jam.2025.90983.1317>
13. Latifi Amoghin, M., Abbaspour-Gilandeh, Y., Tahmasebi, M., & Arribas, J. I. (2024a). Automatic non-destructive estimation of polyphenol oxidase and peroxidase enzyme activity levels in three bell pepper varieties by Vis/NIR spectroscopy imaging data based on machine learning methods. *Chemometrics and Intelligent Laboratory Systems*, 105137. <https://doi.org/10.1016/j.chemolab.2024.105137>
14. Latifi Amoghin, M., Abbaspour-Gilandeh, Y., Tahmasebi, M., Kaveh, M., El-Mesery, H. S., Szymanek, M., & Sprawka, M. (2024b). VIS/NIR Spectroscopy as a Non-Destructive Method for Evaluation of Quality Parameters of Three Bell Pepper Varieties Based on Soft Computing Methods. *Applied Sciences*, 14(23), 10855. <https://doi.org/10.3390/app142310855>

14. Liu, J., Sun, J., Wang, Y., Liu, X., Zhang, Y., & Fu, H. (2025). Non-Destructive Detection of Fruit Quality: Technologies, Applications and Prospects. *Foods*, 14(12), 2137. <https://doi.org/10.3390/foods14122137>
15. Liu, Y., Wang, H., Fei, Y., Liu, Y., Shen, L., Zhuang, Z., & Zhang, X. (2021). Research on the prediction of green plum acidity based on improved XGBoost. *Sensors*, 21(3), 930. <https://doi.org/10.3390/s21030930>
16. Majeed, R., & Jawandha, S. (2016). Enzymatic changes in plum (*Prunus salicina* Lindl.) subjected to some chemical treatments and cold storage. *Journal of Food Science and Technology*, 53, 2372-2379. <https://doi.org/10.1007/s13197-016-2209-9>
17. Masoudi-Sobhanzadeh, Y., & Motieghader, H. (2016). World Competitive Contests (WCC) algorithm: A novel intelligent optimization algorithm for biological and non-biological problems. *Informatics in Medicine Unlocked*, 3, 15-28. <https://doi.org/10.1016/j.imu.2016.06.002>
18. Masoudi-Sobhanzadeh, Y., Motieghader, H., & Masoudi-Nejad, A. (2019). FeatureSelect: a software for feature selection based on machine learning approaches. *BMC Bioinformatics*, 20, 1-17. <https://doi.org/10.1186/s12859-019-2754-0>
19. McGlynn, W. (2010). Importance of food pH in commercial canning operations [superseded].
20. Narendra, K. S., & Thathachar, M. A. (2012). Learning automata: an introduction. Courier corporation.
21. Pu, H., Liu, D., Wang, L., & Sun, D.-W. (2016). Soluble solids content and pH prediction and maturity discrimination of lychee fruits using visible and near infrared hyperspectral imaging. *Food Analytical Methods*, 9, 235-244. <https://doi.org/10.1007/s12161-015-0186-7>
22. Purwanto, Y. A., Zainal, P. W., Ahmad, U., Sutrisno, M., Makino, Y., Oshita, S., Kawagoe, Y., & Kuroki, S. (2013). Non destructive prediction of pH in mango fruit cv. Gedong Gincu using NIR spectroscopy. *International Journal of Engineering and Technology*, 13, 70-73.
23. Rahman, A., Kandpal, L. M., Lohumi, S., Kim, M. S., Lee, H., Mo, C., & Cho, B.-K. (2017). Nondestructive estimation of moisture content, pH and soluble solid contents in intact tomatoes using hyperspectral imaging. *Applied Sciences*, 7(1), 109. <https://doi.org/10.3390/app7010109>
24. Razavi, M. S., Sharabiani, V. R., Tahmasebi, M., Grassi, S., & Szymanek, M. (2025). Chemometric and meta-heuristic algorithms to find optimal wavelengths and predict 'Red Delicious' apples traits using Vis-NIR. *Applied Food Research*, 5(1), 100853. <https://doi.org/10.1016/j.afres.2025.100853>
25. Trendafilova, A., Ivanova, V., Trusheva, B., Kamenova-Nacheva, M., Tabakov, S., & Simova, S. (2022). Chemical composition and antioxidant capacity of the fruits of European plum cultivar "Čačanska Lepotica" influenced by different rootstocks. *Foods*, 11(18), 2844. <https://doi.org/10.3390/foods11182844>
26. Wang, X., Xu, L., Chen, H., Zou, Z., Huang, P., & Xin, B. (2022). Non-destructive detection of pH value of kiwifruit based on hyperspectral fluorescence imaging technology. *Agriculture*, 12(2), 208. <https://doi.org/10.3390/agriculture12020208>
27. Weng, S., Yu, S., Guo, B., Tang, P., & Liang, D. (2020). Non-destructive detection of strawberry quality using multi-features of hyperspectral imaging and multivariate methods. *Sensors*, 20(11), 3074. <https://doi.org/10.3390/s20113074>
28. Zhang, Z., Cheng, H., Chen, M., Zhang, L., Cheng, Y., Geng, W., & Guan, J. (2024). Detection of Pear Quality Using Hyperspectral Imaging Technology and Machine Learning Analysis. *Foods*, 13(23), 3956. <https://doi.org/10.3390/foods13233956>
29. Zhiming, G., Wenqian, H., Liping, C., Yankun, P., & Xiu, W. (2014). Shortwave infrared hyperspectral imaging for detection of pH value in Fuji apple. *International Journal of Agricultural and Biological Engineering*, 7(2), 130-137. <https://doi.org/10.3965/j.ijabe.20140702.016>
30. Zhu, H., Chu, B., Fan, Y., Tao, X., Yin, W., & He, Y. (2017). Hyperspectral imaging for predicting the internal quality of kiwifruits based on variable selection algorithms and chemometric models. *Scientific Reports*, 7(1), 7845. <https://doi.org/10.1038/s41598-017-08509-6>