



Using VIS/NIR Spectroscopy for Non-destructive Detection of Soluble Solids Content in Plum Cultivars

Meysam Latifi-Amoghin¹ | Yousef Abbaspour-Gilandeh²

1. Department of Biosystems Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: m.latifi@uma.ac.ir

2. Corresponding Author, Department of Biosystems Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: abbaspour@uma.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Sep. 3, 2025

Revised: Oct. 24, 2025

Accepted: Nov. 10, 2025

Published online: Winter 2026

Keywords:

*Non-destructive,
Plum,
Soluble Solids Content,
VIS/NIR spectroscopy.*

ABSTRACT

In this study, a non-destructive method for evaluating soluble solids content in plum cultivars was developed using visible/near-infrared (VIS/NIR) spectroscopy. For this purpose, absorption spectra of 80 samples from two plum cultivars (Khormaei and Khoni) in the 350-1100 nm range were collected, and their SSC values were measured using a refractometer as a reference method. To reduce data dimensionality and identify effective wavelengths, a set of five metaheuristic algorithms (including LCA, GA, PSO, ACO, and ICA) was employed in combination with the SVM model, with the main innovation lying in the integration of the League Championship Algorithm (LCA), as the strongest feature selector, with Artificial Neural Network (ANN) regression models. Subsequently, Partial Least Squares Regression (PLSR) and Artificial Neural Network (ANN) models were developed to predict SSC on both full and reduced spectra. To optimize model accuracy, more than 12 pre-processing methods were applied and evaluated. The results indicated that the ANN model based on the selected wavelengths and with Moving Average pre-processing achieved the highest prediction accuracy, particularly for the Khormaei cultivar, with a coefficient of determination of $R^2=0.989$ and a ratio of performance to deviation of $RPD=6.08$, indicating the excellent capability of the model for quantitative prediction. This research demonstrates that the combination of VIS/NIR spectroscopy with novel metaheuristic feature selection methods like LCA provides a new and efficient approach for the non-destructive determination of SSC. By significantly reducing computational requirements, this methodology offers a practical, accurate, and reliable tool for real-time quality control in post-harvest and fruit processing industries.

Cite this article: Latifi-Amoghin, M., Abbaspour-Gilandeh, Y., (2026). Using VIS/NIR Spectroscopy for Non-destructive Detection of Soluble Solids Content in Plum Cultivars., *Iranian Journal of Biosystem Engineering*, 56 (4), 25-46.
<https://doi.org/10.22059/ijbse.2025.401742.665611>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijbse.2025.401742.665611>



EXTENDED ABSTRACT

Objective

The primary objective of this research was to develop and evaluate a rapid, non-destructive method for predicting the soluble solids content (SSC) in two distinct plum cultivars, namely Khormaei and Khoni, using visible and near-infrared (VIS/NIR) spectroscopy. The study aimed to overcome the limitations of traditional destructive methods, such as the refractometer, which are time-consuming and unsuitable for real-time quality control in the post-harvest industry. Furthermore, a key goal was to enhance the predictive performance of models by employing a novel combination of advanced meta-heuristic algorithms and machine learning techniques for optimal wavelength selection and model development.

Research Method

A total of 80 plum samples, consisting of 40 samples from each of the Khormaei and Khoni cultivars, were utilized for this study. The fruit samples were subjected to VIS/NIR spectroscopy in the range of 350 to 1100 nm to capture their absorption spectra. The corresponding SSC values for each sample were simultaneously measured using a digital refractometer as the reference method. To address the challenge of high dimensionality and multicollinearity inherent in spectral data, a comprehensive feature selection strategy was implemented. A suite of five meta-heuristic algorithms, including Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), League Championship Algorithm (LCA), and Imperialist Competitive Algorithm (ICA), was employed in conjunction with a Support Vector Machine (SVM) model. The LCA algorithm was found to be the most effective in identifying a subset of key wavelengths that were most influential in predicting SSC. For model development and comparison, both Partial Least Squares Regression (PLSR) and Artificial Neural Network (ANN) models were employed. To ensure the robustness and generalizability of the models, data splitting was performed 200 times with a ratio of 60% for training, 20% for validation, and 20% for testing. Additionally, the impact of more than 12 different pre-processing techniques, such as Moving Average, Savitzky-Golay filtering, and Standard Normal Variate (SNV), on model accuracy was systematically investigated. The parameters for the ANN model, including the number of hidden layers and neurons, were optimized through a systematic process to prevent overfitting and achieve the best performance.

Findings

The results demonstrated that the non-destructive VIS/NIR method combined with chemometric modeling is highly effective for predicting SSC in both plum cultivars. The performance of the models varied depending on the cultivar, the pre-processing method, and the model type. The ANN model consistently outperformed the PLSR model, especially when applied to the reduced spectral data set. The highest prediction accuracy was achieved by the ANN model using the key wavelengths selected by the LCA algorithm, with the Moving Average pre-processing method. For the Khormaei cultivar, this model yielded a coefficient of determination of $R^2=0.989$ and a ratio of performance to deviation of $RPD=6.08$, indicating excellent predictive capability. Similarly, for the Khoni cultivar, the model's performance was also highly satisfactory. The selection of optimal wavelengths by the meta-heuristic algorithms significantly improved model performance while reducing computational complexity. The LCA-SVM combination proved to be particularly powerful in isolating the most relevant spectral information for SSC prediction.

Conclusion

The study successfully demonstrated that VIS/NIR spectroscopy coupled with advanced meta-heuristic feature selection and ANN modeling is a highly accurate and reliable method for the non-destructive determination of SSC in plum cultivars. The combination of the LCA algorithm for wavelength selection with an ANN model offers a novel and powerful approach for creating robust and efficient predictive models. This methodology not only improves prediction accuracy but also reduces the computational requirements, making it a viable and practical tool for real-time quality control applications in the post-harvest handling and food processing industries. The findings confirm that this non-destructive technique is a promising alternative to traditional destructive methods and can be effectively implemented to ensure product quality and consistency.

Author Contributions

Conceptualization, M.L.-A. and Y.A.-G.; methodology, M.L.-A.; software, M.L.-A.; validation, Y.A.-G.; formal analysis, M.L.-A. and Y.A.-G.; investigation, M.L.-A.; resources, M.L.-A.; data curation, M.L.-A. and Y.A.-G.; writing—original draft preparation, M.L.-A. and Y.A.-G.; writing—review and editing, Y.A.-G.; visualization, M.L.-A.; supervision, Y.A.-G.; project administration, Y.A.-G.; funding acquisition, Y.A.-G. All authors have read and agreed to the published version of the manuscript.

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgements

This research was supported by the University of Mohaghegh Ardabili, which is greatly appreciated.

Ethical considerations

Not applicable.

Conflict of interest

The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

استفاده از طیف‌سنجی VIS/NIR به منظور تشخیص غیرمخرب مواد جامد محلول در ارقام آلو

میثم لطیفی عموقین^۱ | یوسف عباس‌پور گیلانده^۲^۱ گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران. رایانامه: m.latifi@uma.ac.ir^۲ نویسنده مسئول، گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران. رایانامه:abbaspour@uma.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	در این پژوهش، روشی غیرمخرب برای ارزیابی میزان مواد جامد محلول (SSC) در ارقام آلو با استفاده از طیف‌سنجی مرئی/مادون قرمز نزدیک (VIS/NIR) توسعه یافت. بدین منظور، طیف‌های جذبی ۸۰ نمونه از دو رقم آلو خرمایی و خونی در محدوده ۳۵۰ تا ۱۱۰۰ نانومتر جمع‌آوری و مقادیر SSC آن‌ها با استفاده از رفاکتومتر به‌عنوان روش مرجع اندازه‌گیری شد. برای کاهش ابعاد داده‌ها و شناسایی طول‌موج‌های مؤثر، مجموعه‌ای از پنج الگوریتم فراابتکاری (شامل LCA، GA، PSO، ACO و ICA) در ترکیب با مدل ماشین‌بردار پشتیبان (SVM) به کار گرفته شد که نوآوری اصلی در تلفیق الگوریتم قهرمانی لیگ (LCA) به‌عنوان قوی‌ترین انتخاب‌گر ویژگی، با مدل‌های رگرسیونی شبکه عصبی مصنوعی (ANN) نهفته است. سپس، مدل‌های رگرسیون حداقل مربعات جزئی (PLSR) و ANN برای پیش‌بینی SSC بر روی طیف‌های کامل و کاهش‌یافته توسعه یافتند. برای بهینه‌سازی دقت مدل‌ها، بیش از ۱۲ روش پیش‌پردازش بر روی طیف‌ها اعمال و تأثیر هر یک ارزیابی شد. نتایج نشان داد که مدل ANN بر پایه طول‌موج‌های منتخب و با اعمال پیش‌پردازش میانگین متحرک، به‌ویژه برای رقم خرمایی، به بالاترین دقت پیش‌بینی ($R^2 = 0.989$ و $RPD = 6.08$) دست یافت که حاکی از قابلیت عالی مدل برای پیش‌بینی کمی است. این پژوهش ثابت می‌کند که ترکیب طیف‌سنجی VIS/NIR با روش‌های نوین انتخاب ویژگی فراابتکاری مانند LCA، یک رویکرد جدید و کارآمد برای تعیین غیرمخرب SSC ارائه می‌دهد. این متدولوژی با کاهش چشمگیر نیازهای محاسباتی، یک ابزار عملی، دقیق و قابل اطمینان برای کنترل کیفیت در لحظه در صنایع پس از برداشت و فرآوری میوه است.
واژه‌های کلیدی: آلو، طیف‌سنجی VIS/NIR، مواد جامد محلول، پیش‌پردازش، شبکه عصبی مصنوعی.	

استناد: لطیفی عموقین، میثم؛ عباس‌پور گیلانده، یوسف (۱۴۰۴). استفاده از طیف‌سنجی VIS/NIR جهت تشخیص غیرمخرب مواد جامد محلول در ارقام آلو، مجله

<https://doi.org/10.22059/ijbse.2025.401742.665611> مهندسی بیوسیستم ایران/ایران، ۵۶ (۴)، ۴۶-۲۵.

© نویسنده‌گان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2025.401742.665611>

مقدمه

آلو (*Prunus salicina* Lindl.) یکی از میوه‌های مهم از تیره *Prunus* و خانواده گل‌سرخیان (Rosaceae) است که در بازه زمانی متفاوتی با توجه به ارقام زودرس یا دیررس به بلوغ می‌رسد (Weekley et al., 2010). این میوه به دلیل طعم شیرین و غنی بودن از ترکیبات مغذی مانند قندها، اسیدها، اسیدهای آمینه، و ویتامین‌ها جایگاه ویژه‌ای در رژیم غذایی انسان دارد (Martínez-Esplá et al., 2018). انواع مختلف آلو به‌طور گسترده‌ای در سراسر جهان کشت می‌شوند (Butac et al., 2011) و نقش مهمی در تأمین فیبر غذایی، مواد معدنی، آنتی‌اکسیدان‌ها و ترکیبات زیست‌فعال ایفا می‌کنند. مطالعات مختلف نشان داده‌اند که مصرف منظم میوه‌ها، ازجمله آلو، در پیشگیری از بیماری‌های مزمن، کاهش اختلالات دژنراتیو نظیر بیماری‌های قلبی-عروقی و سرطان، و حفظ سلامت عمومی بدن مؤثر است (Rasmussen et al., 2006). این اثرات مثبت عمدتاً به دلیل حضور ترکیبات فیتوشیمیایی مانند فلاونوئیدها، آنتوسیانین‌ها، پلی‌فنول‌ها و اسید آسکوربیک است که با خاصیت آنتی‌اکسیدانی خود به خنثی‌سازی رادیکال‌های آزاد کمک می‌کنند.

کیفیت میوه تحت تأثیر عوامل مختلفی ازجمله رقم، منشأ، مرحله بلوغ، شرایط برداشت و نگهداری، و زمان و دمای ذخیره‌سازی قرار دارد. با افزایش مدت ماندگاری میوه روی درخت، تغییرات فیزیولوژیکی قابل‌توجهی در آن رخ می‌دهد که بر ترکیب شیمیایی آن، به‌ویژه جامدات محلول کل، اسید آسکوربیک، pH و اسیدیته قابل تیتراسیون اثر می‌گذارد (Antoniali et al., 2007). همچنین فرآیند رسیدن در میوه‌ها پس از برداشت نیز ادامه دارد که باعث تغییر در غلظت رنگدانه‌ها و ترکیبات داخلی آن‌ها می‌شود (Deli, 2002). محتوای جامدات محلول^۱ (SSC) یکی از مهم‌ترین شاخص‌های کیفی میوه‌ها محسوب می‌شود که نقش تعیین‌کننده‌ای در طعم، میزان شیرینی، ارزش تغذیه‌ای و پذیرش بازار دارد. SSC عمدتاً شامل قندها، اسیدهای آلی، ویتامین‌ها و سایر ترکیبات محلول است که به‌طور مستقیم با مرحله رسیدگی و کیفیت داخلی محصول مرتبط‌اند. مقدار SSC معمولاً برحسب درصد بریکس بیان می‌شود و به‌طور سنتی با استفاده از رفاکتومتر از طریق عصاره‌گیری تخریبی اندازه‌گیری می‌شود. با توجه به ماهیت زمان‌بر و مخرب این روش‌های مرسوم، توسعه و به‌کارگیری فناوری‌های نوین غیرمخرب به‌منظور پیش‌بینی سریع و دقیق SSC در میوه‌ها اهمیت روزافزونی یافته است (Zhan et al., 2023).

در فرآیند برداشت، درجه‌بندی و بازاریابی محصولات باغی، ارزیابی دقیق کیفیت داخلی از اهمیت بالایی برخوردار است. بااین‌حال، بسیاری از سامانه‌های درجه‌بندی موجود عمدتاً بر مبنای ویژگی‌های ظاهری مانند رنگ، اندازه و شکل عمل می‌کنند و توانایی ارزیابی ویژگی‌های کیفی درونی را ندارند. به همین دلیل، توسعه فناوری‌های نوین برای ارزیابی غیرمخرب کیفیت میوه‌ها موردتوجه بسیاری از پژوهشگران قرار گرفته است. روش‌های سنتی ارزیابی کیفیت ازجمله آنالیزهای فیزیکوشیمیایی و حسی، معمولاً مخرب، زمان‌بر و محدود به نمونه‌برداری‌های کوچک هستند و امکان کاربرد آن‌ها به‌صورت برخط وجود ندارد. در این راستا، استفاده از فناوری‌های پیشرفته مانند بینایی ماشین، حسگرهای هوشمند، طیف‌سنجی نوری و روش‌های غیرمخرب به‌ویژه در ناحیه مرئی/مادون‌قرمز نزدیک^۲ (VIS/NIR)، به‌عنوان رویکردی نوین و مؤثر برای تعیین کیفیت داخلی میوه‌ها در حال گسترش است (Guo et al., 2022). طیف‌سنجی نزدیک فروسرخ^۴ (NIR) یک تکنیک مناسب برای تجزیه و تحلیل کمی و کیفی در پزشکی، کشاورزی، شیمی و سایر علوم است. این تکنیک، به دلیل سرعت بالا، عدم نیاز به آماده‌سازی نمونه، قابلیت استفاده برخط، دقت مناسب، سازگاری با محیط‌زیست و مقرون‌به‌صرفه بودن، به‌عنوان یکی از ابزارهای امیدوارکننده در ارزیابی غیرمخرب پارامترهای کیفی میوه‌ها ازجمله میزان مواد جامد محلول (SSC) شناخته می‌شود (Jamshidi et al., 2015). استفاده از این فناوری در ارقام مختلف آلو می‌تواند به بهینه‌سازی فرآیندهای برداشت، طبقه‌بندی، و بازاریابی محصول کمک شایانی کند.

در فرآیند برداشت، درجه‌بندی و بازاریابی محصولات باغی، ارزیابی دقیق کیفیت داخلی از اهمیت بالایی برخوردار است. بااین‌حال، بسیاری از سامانه‌های درجه‌بندی موجود عمدتاً بر مبنای ویژگی‌های ظاهری مانند رنگ، اندازه و شکل عمل می‌کنند و توانایی ارزیابی ویژگی‌های کیفی درونی را ندارند. روش‌های سنتی ارزیابی کیفیت شامل آنالیزهای فیزیکوشیمیایی و حسی، مخرب، زمان‌بر و محدود به نمونه‌برداری‌های کوچک هستند و امکان کاربرد آن‌ها به‌صورت برخط در خطوط تولید و درجه‌بندی با سرعت بالا وجود ندارد. این محدودیت‌ها، انگیزه اصلی برای توسعه فناوری‌هایی است که بتوانند ارزیابی کیفی ۱۰۰ درصد محصولات را بدون تخریب انجام دهند. در

1. Soluble Solids Content

2. %Brix

3. Visible/Near-Infrared Spectroscopy

4. Near-Infrared Spectroscopy

این راستا، استفاده از فناوری‌های پیشرفته مانند بینایی ماشین، حسگرهای هوشمند، طیف‌سنجی نوری و روش‌های غیرمخرب به‌ویژه در ناحیه VIS/NIR، به‌عنوان رویکردی نوین و مؤثر برای تعیین کیفیت داخلی میوه‌ها در حال گسترش است (Gao et al., 2022). طیف‌سنجی NIR به دلیل سرعت بالا، عدم نیاز به آماده‌سازی نمونه، سازگاری با محیط‌زیست و مقرون‌به‌صرفه‌بودن، به ابزاری کلیدی برای ارزیابی غیرمخرب پارامترهای کیفی میوه‌ها از جمله میزان SSC تبدیل شده است.

اخیراً، از طیف‌سنجی در VIS/NIR به‌طور گسترده‌ای برای ارزیابی غیرمخرب ویژگی‌های کیفی محصولات باغی بهره گرفته شده است. از جمله این تحقیقات می‌توان به پیش‌بینی SSC در توت‌فرنگی (Agulheiro-Santos et al., 2022)، ارزیابی SSC در میوه‌هایی با ضخامت‌های پوستی مختلف (Pratiwi et al., 2023)، پیش‌بینی کیفیت میوه آلو ماریان بدون تأثیر شرایط محیطی (Posom et al., 2020)، تعیین دوره نگهداری آلوهای برداشت‌شده (Guo et al., 2022)، پیش‌بینی میزان و تنوع آنتوسیانین در عصاره آلو (Johnson et al., 2022)، ارزیابی مقدار SSC و سفتی در سیب (Ma et al., 2021)، بهبود صحت تشخیص SSC گوجه‌فرنگی گیلانی (Zheng et al., 2024)، تشخیص SSC در گلابی (Zhan et al., 2023)، تعیین مقدار SSC در سیب (Xu et al., 2019) اشاره کرد. این تحقیقات نشان می‌دهند که فناوری VIS/NIR می‌تواند به‌عنوان ابزاری سریع، دقیق و غیرمخرب برای تعیین SSC در میوه‌هایی با ساختار و ترکیب متفاوت، از جمله ارقام مختلف آلو، مورد استفاده قرار گیرد. با وجود پیشرفت‌های قابل توجه در استفاده از VIS/NIR برای ارزیابی کیفیت میوه‌ها، اغلب مطالعات بر روش‌های سنتی‌تر انتخاب طول موج مانند (PCA و SPA) یا مدل‌های رگرسیونی عمومی‌تر (PLSR و ANN) متمرکز بوده‌اند. این روش‌ها در حالی که نتایج قابل قبولی ارائه داده‌اند، ممکن است در بهینه‌سازی دقیق‌تر انتخاب ویژگی‌ها و افزایش کارایی مدل‌ها با چالش‌هایی مواجه باشند، به خصوص زمانی که با حجم زیادی از داده‌های طیفی و نویزهای احتمالی سروکار داریم. به عنوان مثال، مطالعات پیشین اگرچه به کاربرد VIS/NIR در میوه‌های مختلف اشاره کرده‌اند، اما به ندرت به بررسی رویکردهای نوآورانه در انتخاب طول موج‌های بهینه و تلفیق آن‌ها با مدل‌های یادگیری ماشینی پیشرفته برای میوه‌های خاصی مانند آلو پرداخته‌اند. شکاف دانشی موجود در این زمینه، لزوم کاوش روش‌های فراابتکاری پیشرفته‌تر را برای شناسایی موثرترین طول موج‌ها که قادر به افزایش دقت و پایداری مدل‌های پیش‌بینی SSC در آلو باشند، برجسته می‌سازد. در این راستا، استفاده از الگوریتم‌های فراابتکاری به عنوان ابزاری قدرتمند برای حل مسائل بهینه‌سازی پیچیده مطرح است. الگوریتم قهرمانی لیگ (LCA) به عنوان یک رویکرد نوین و کارآمد برای انتخاب طول موج‌های کلیدی از داده‌های طیف‌سنجی VIS/NIR آلو به کار گرفته شده است. تلفیق نوآورانه LCA با مدل ماشین‌بردار پشتیبان^۲ (SVM) برای شناسایی زیرمجموعه‌ای از طول موج‌ها، امکان توسعه مدل‌های پیش‌بینی دقیق‌تر و با پیچیدگی محاسباتی کمتر را فراهم می‌آورد که این امر یک گام رو به جلو در ارزیابی غیرمخرب کیفیت آلو در صنایع غذایی و کشاورزی محسوب می‌شود و به طور خاص به چالش‌های موجود در این حوزه پاسخ می‌دهد. با توجه به اهمیت ارزیابی سریع و غیرمخرب کیفیت داخلی میوه آلو، این پژوهش با هدف پیش‌بینی میزان SSC با استفاده از VIS/NIR انجام شد. بدین منظور، انواع روش‌های پیش‌پردازش داده‌های طیفی، مدل‌سازی رگرسیون حداقل مربعات جزئی^۳ (PLSR)، شناسایی طول‌موج‌های مؤثر و توسعه مدل شبکه عصبی مصنوعی^۴ (ANN) به کار گرفته شد و نتایج با روش‌های مرجع مقایسه شده تا کارآمدترین رویکرد ارزیابی SSC معرفی گردد.

مواد و روش‌ها

انتخاب نمونه‌ها

در این پژوهش دو رقم آلو شامل خرمایی و خونی از باغات محلی خریداری شد و از هر رقم، تعداد ۸۰ نمونه با اندازه، شکل و رنگ یکنواخت و بدون هیچ‌گونه نشانه‌ای از آسیب مکانیکی یا پوسیدگی قارچی برای انجام آزمایش‌ها انتخاب گردید (شکل ۱). آلو خرمایی که با نام قره آلو نیز شناخته می‌شود، میوه‌ای سیاه‌رنگ، سفت، شیرین و آبدار است که از نظر ظاهر و طعم شباهت زیادی به آلو شابلون دارد. این رقم به دلیل ویژگی‌های بافتی و طعم مطلوب، در اروپا به‌طور گسترده به‌صورت خشک‌بار مصرف شده و از محبوبیت زیادی برخوردار است. آلو خونی که به نام‌های دیگری مانند توسرخ، آلو خاکی، آلو خاکی خونی و آلو قرمز نیز معروف است، دارای میوه‌ای نسبتاً درشت با رنگ قرمز تیره تا بنفش و طعمی شیرین و بسیار خوش‌خوراک می‌باشد. کیفیت میوه این رقم در مقایسه با بسیاری از ارقام آلو مطلوب‌تر

1. League Championship Algorithm
2. Support Vector Machine
3. Partial Least Squares Regression
4. Artificial Neural Network

بوده و به همین دلیل از جایگاه مناسبی در بازار برخوردار است. میوه آلو خونی به صورت تازه خوری، تهیه خشکبار و سایر فرآورده‌ها مصرف می‌شود. انتخاب این دو رقم به دلیل تفاوت‌های فیزیولوژیکی و کیفیتی آن‌ها انجام شد تا امکان بررسی گسترده‌تر و جامع‌تر مقدار SSC میوه آلو فراهم شود. نمونه‌ها پیش از اندازه‌گیری به مدت دو ساعت در دمای ۲۵ درجه سلسیوس نگهداری شدند تا دمای آن‌ها با دمای محیط آزمایشگاه متعادل شود. پس از تصویربرداری نمونه‌ها با استفاده از دستگاه طیف‌سنج، بلافاصله اندازه‌گیری مخرب میزان SSC آن‌ها انجام گرفت.



شکل ۱: نمونه‌های مورد استفاده (الف) آلو خرمایی و (ب) آلو خونی

آزمون غیرمخرب

طیف‌سنجی Vis/NIR

آزمون‌های طیف‌سنجی Vis/NIR با استفاده از یک طیف‌سنج مدل PS-100 (Apogee Instruments, INC., Logan, Utah, USA) با آشکارساز افزاره با بارجفت‌شده CCD، ۲۰۴۸ پیکسل، با قدرت تفکیک ۱ نانومتر و منبع نور هالوژن-تنگستن در محدوده طول‌موج‌های ۳۵۰-۱۱۰۰ نانومتر صورت پذیرفت. شکل ۲ چگونگی نمونه‌برداری با سامانه طیف‌سنج را نشان می‌دهد. طیف‌سنج از طریق کابل USB به رایانه متصل و طیف‌های حاصل توسط نرم‌افزار Spectra Wiz در آن ذخیره شدند. پیش از طیف‌سنجی، ابتدا طیف‌های تیره و سفید به عنوان مرجع تعریف و ذخیره شدند. به این ترتیب که ابتدا با خاموش کردن منبع نور طیف تیره گرفته، سپس در حالت روشن بودن منبع، از یک دیسک تفلون استاندارد برای دستیابی به طیف مرجع استفاده شد. برای تمامی نمونه‌ها از چهار نقطه مختلف هر نمونه طیف‌سنجی صورت گرفته و داده‌های مورد نظر ثبت شدند سپس طیف جذب هر کدام از نمونه‌ها پس از میانگین‌گیری به دست آمد در نهایت ۸۰ داده طیفی برای هر رقم جهت آنالیز استخراج گردید (Amoghini et al., 2024).

آزمون مخرب

مواد جامد محلول (SSC)

میزان SSC برحسب درصد بریکس توسط دستگاه رفاکتومتر دیجیتالی (PrismaTech BPTR50, Iran) با محدوده اندازه‌گیری ۰ تا ۸۵ درصد بریکس و دقت اندازه‌گیری ± 0.2 درصد بریکس و با ریختن یک قطره آب‌میوه بر روی صفحه شیشه‌ای رفاکتومتر انجام شد (Latifi Amoghini et al., 2024).



شکل ۲: نمایی از سیستم طیف‌سنجی VIS/NIR مورد استفاده برای جمع‌آوری داده‌ها

تجزیه و تحلیل داده‌های طیفی

داده‌های طیفی خام نمونه‌ها در محدوده ۳۵۰ تا ۱۱۰۰ نانومتر جمع‌آوری شدند که این محدوده بخش‌هایی از Vis/NIR را شامل می‌شود. قسمت ابتدایی داده‌های طیفی، یعنی محدوده ۳۵۰ تا حدود ۵۱۰ نانومتر، به دلیل وجود نویز قابل توجه ناشی از کاهش نسبت سیگنال به نویز و محدودیت حساسیت آشکارساز در این بازه، حذف گردید. این فرآیند پاک‌سازی داده‌ها به‌طور مشابه برای هر دو رقم مورد مطالعه اعمال شد تا داده‌های آماده‌سازی‌شده برای تحلیل‌های بعدی، از کیفیت و ثبات بیشتری برخوردار باشند.

پس از اصلاح اولیه، داده‌های طیفی به نرم‌افزار MATLAB نسخه ۲۰۲۲a منتقل شدند و فرآیند مدل‌سازی با استفاده از روش PLSR حداکثر با ۱۰ مؤلفه آغاز گردید. الگوریتم PLSR به دلیل توانایی بالا در استخراج روابط میان داده‌های با ابعاد زیاد و متغیرهای پاسخ، به عنوان رویکرد اصلی مدل‌سازی انتخاب شد.

تکرار تصادفی تقسیم‌بندی داده‌ها

برای کاهش تأثیر تغییرات ناشی از تقسیم‌بندی تصادفی نمونه‌ها و افزایش اعتبار نتایج، فرآیند تقسیم داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمون به‌طور مستقل ۲۰۰ بار تکرار شد. در هر تکرار، داده‌های خام بدون پیش‌پردازش به‌صورت تصادفی در نسبت‌های ۶۰ درصد آموزش، ۲۰ درصد اعتبارسنجی و ۲۰ درصد آزمون تفکیک شدند. پس از انجام تکرارها، بهترین تقسیم‌بندی بر اساس معیارهای آماری (شامل ضریب تعیین (R^2) ، ریشه میانگین مربعات خطا $(RMSE)$ و نسبت عملکرد به انحراف $(RPD)^3$) در مراحل آموزش، اعتبارسنجی و آزمون) و بهترین تعداد مؤلفه انتخاب گردید. این تقسیم‌بندی ابتدا بر روی محدوده کامل طیفی و مدل‌سازی با روش PLSR و سپس بر روی داده‌های کاهش داده‌شده و مدل‌سازی با روش‌های PLSR و ANN انجام گرفت.

اعمال پیش‌پردازش‌ها و مدل‌سازی

به‌منظور بررسی تأثیر پیش‌پردازش‌های مختلف بر عملکرد مدل، تمامی روش‌های پیش‌پردازش صرفاً بر روی همان تقسیم‌بندی منتخب از داده‌ها اعمال شد. به‌عبارت‌دیگر، پس از انتخاب بهترین تقسیم‌بندی تصادفی داده‌های خام، این ترتیب ثابت نگه داشته شد تا اثر هر پیش‌پردازش مستقل از تغییرات ناشی از تقسیم‌بندی باشد و مقایسه دقیق‌تری فراهم گردد. این پیش‌پردازش‌ها شامل نرمال‌سازی^۴، استانداردسازی بردار نرمال^۵ (SNV)، تصحیح پراکندگی چندگانه^۶ (MSC)، مشتق اول^۷، مشتق دوم^۸، فیلتر میانگین‌گیری متحرک^۹،

1. Coefficient of Determination
2. Root Mean Square Error
3. Ratio of Performance to Deviation
4. Normalization
5. Standard Normal Variate
6. Multiplicative Scatter Correction
7. First Derivative
8. Second Derivative
9. Moving Average

فیلتر گوسی^۱، فیلتر میانه^۲، صاف‌سازی ساویتزکی-گولای^۳، حذف روند^۴ و میانگین مرکزی کردن^۵ بودند. برای هر روش، مدل PLSR مجدداً آموزش داده شد و عملکرد آن مورد ارزیابی قرار گرفت (Hasanzadeh et al., 2022).

ارزیابی عملکرد مدل‌ها

به منظور ارزیابی و مقایسه دقت مدل‌ها، شاخص‌های آماری شامل R^2 ، RMSE و RPD در مراحل آموزش، اعتبارسنجی و آزمون (روابط ۱، ۲ و ۳) محاسبه گردیدند (Razavi et al., 2025).

$$R^2 = 1 - \frac{\sum_{i=1}^N (p_i - d_i)^2}{\sum_{i=1}^N (d_i - \bar{d})^2} \quad \text{(رابطه ۱)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (p_i - d_i)^2}{N}} \quad \text{(رابطه ۲)}$$

$$RPD = \frac{SD}{RMSE} \quad \text{(رابطه ۳)}$$

در این معادلات، p_i مقدار SSC پیش‌بینی شده برای مؤلفه- i ام و d_i مقدار SSC واقعی (مرجع) اندازه‌گیری شده برای مؤلفه- i ام را نشان می‌دهد. N تعداد کل نمونه‌های استفاده شده در مجموعه داده (مجموعه کالیبراسیون یا پیش‌بینی) می‌باشد. علاوه بر این، $\bar{d}$ میانگین مقادیر واقعی SSC نمونه‌ها، و SD انحراف استاندارد داده‌های پیش‌بینی شده SSC در مجموعه داده مربوطه است.

در $RPD < 1$ ، مدل بسیار ضعیف است و توصیه نمی‌شود. $1 < RPD < 1.4$ یک مدل ضعیف است، $1.4 < RPD < 1.8$ مدل مناسبی است و می‌تواند برای ارزیابی و پیش‌بینی استفاده شود. $1.8 < RPD < 2$ مدل خوبی است و برای پیش‌بینی‌های کمی امکان‌پذیر است. $RPD > 2$ مدل بسیار خوب و $RPD < 2.5$ یک مدل عالی است (Chang et al., 2001).

انتخاب طول موج‌های مؤثر و مدل‌سازی مجدد بر روی داده‌های کاهش‌یافته

به منظور کاهش ابعاد داده‌های طیفی و استخراج طول موج‌های مؤثر در پیش‌بینی، از تلفیق الگوریتم SVM با مجموعه‌ای از الگوریتم‌های فراابتکاری شامل الگوریتم قهرمانی لیگ^۶ (LCA)، الگوریتم ژنتیک^۷ (GA)، الگوریتم بهینه‌سازی ازدحام ذرات^۸ (PSO)، الگوریتم کلونی مورچگان^۹ (ACO) و الگوریتم رقابت استعماری^{۱۰} (ICA) استفاده شد (مشخصات پارامترهای هر الگوریتم در جدول ۱ ارائه شده است) (Masoudi-Sobhanzadeh et al., 2019). پارامترهای هر یک از الگوریتم‌های فراابتکاری (ACO, PSO, GA, LCA, ICA) که در جدول ۱ ارائه شده‌اند، پس از انجام آزمایش‌های اولیه و با الهام از مطالعات مشابه در این حوزه تنظیم گردیدند تا از همگرایی بهینه و کارایی مناسب الگوریتم‌ها در جستجوی فضای ویژگی‌ها اطمینان حاصل شود. این تنظیمات با هدف دستیابی به بهترین زیرمجموعه از طول موج‌ها برای پیش‌بینی SSC انجام گرفت. در این فرآیند، هر الگوریتم به طور مستقل مجموعه‌ای از طول موج‌های بهینه را شناسایی نمود و عملکرد خروجی‌های حاصل با توجه به R^2 و RMSE ارزیابی گردید تا بهترین الگوریتم با بالاترین دقت پیش‌بینی و کمترین میزان خطا انتخاب شود. ترکیب الگوریتم‌های فراابتکاری با SVM به عنوان یک رویکرد دو مرحله‌ای عمل کرد: ابتدا، الگوریتم‌ها به عنوان یک بهینه‌ساز فراابتکاری برای جستجو در فضای وسیع طول موج‌ها به منظور یافتن زیرمجموعه‌ای بهینه از ویژگی‌ها عمل نمود. تابع هدف LCA و بقیه الگوریتم‌ها در این فرآیند، میزان خطای پیش‌بینی مدل SVM بودند. به این ترتیب، این الگوریتم‌ها طول موج‌هایی را انتخاب می‌کردند که منجر به بهبود عملکرد SVM در طبقه‌بندی یا رگرسیون داده‌های طیفی می‌شد. این تلفیق، مزایای جستجوی کارآمد و سراسری این الگوریتم‌ها را با توانایی بالای SVM در مدل‌سازی روابط پیچیده ترکیب می‌کند و به نتایج جدیدی از جمله کاهش ابعاد داده‌ها بدون دست دادن اطلاعات حیاتی و افزایش قابل توجه دقت پیش‌بینی SSC منجر شد. این رویکرد، در مقایسه با روش‌های انتخاب طول موج تک‌الگوریتمی، به دلیل قابلیت جستجوی هوشمندانه و هدفمند این الگوریتم‌ها در فضای ویژگی‌ها و استفاده از SVM به عنوان ارزیابی‌کننده

1. Gaussian Filter
2. Median Filter
3. Savitzky-Golay
4. Detrend
5. Mean Centering
6. League Championship Algorithm
7. Genetic Algorithm
8. Particle Swarm Optimization
9. Ant Colony Optimization
10. Imperialist Competitive Algorithm

کیفیت انتخاب، نتایج بهینه‌تری را در شناسایی الگوهای مرتبط با SSC ارائه داد. این بهبود در شناسایی دقیق‌تر طول موج‌های مؤثر، به نوبه خود، منجر به ساخت مدل‌های پیش‌بینی قوی‌تر و پایدارتر برای ارزیابی میزان مواد جامد محلول در ارقام آلو گردید. پس از تعیین طول‌موج‌های مؤثر و کاهش حجم داده‌ها، داده‌های کاهش‌یافته با استفاده از روش PLSR و ANN مدل‌سازی گردید. از شبکه عصبی پرسپترون چندلایه بر پایه الگوریتم پس انتشار شامل یک لایه ورودی با تعداد نورون برابر با تعداد طول‌موج‌های مؤثر و لایه خروجی با یک نورون (مقدار SSC) و از سه لایه پنهان به ترتیب با ۱۰، ۱۵، ۲۰ نورون استفاده شد. جهت تقسیم‌بندی داده‌ها از نسبت ۶۰ درصد - ۲۰ درصد - ۲۰ درصد برای آموزش، اعتبارسنجی و آزمایش شبکه استفاده شد همچنین در این پژوهش از الگوریتم لونبرگ-مارکوارت برای به هنگام‌سازی وزن‌های ANN استفاده شد که به علت آموزش سریع شبکه و به حداقل رساندن سطح خطا از الگوریتم‌های پرکاربرد در این زمینه است (Latifi-Amoghini et al., 2025).

انتخاب تعداد نورون‌ها در هر یک از لایه‌های پنهان (۱۰، ۱۵، ۲۰) و همچنین پارامترهای الگوریتم لونبرگ-مارکوارت، از طریق یک فرآیند بهینه‌سازی سیستماتیک (مانند جستجوی شبکه‌ای) با هدف دستیابی به بهترین عملکرد مدل و جلوگیری از بیش‌برازش صورت گرفت. این فرآیند شامل ارزیابی عملکرد مدل ANN با پیکربندی‌های مختلف و انتخاب بهینه‌ترین ساختار بر اساس معیارهای R^2 و RMSE بر روی مجموعه اعتبارسنجی بود. مشابه مرحله پیشین، جهت افزایش اعتبار نتایج، تقسیم‌بندی داده‌ها به صورت تصادفی ۲۰۰ بار تکرار شد که این انتخاب تصادفی تنها بر روی داده‌های خام کاهش‌یافته انجام گرفت. پس از انتخاب بهترین تقسیم‌بندی بر اساس شاخص‌های آماری، تمامی پیش‌پردازش‌های موردنظر بر روی همان ترتیب ثابت داده‌ها اعمال شد تا تأثیر واقعی هر روش پیش‌پردازش بر عملکرد مدل بدون دخالت تغییرات ناشی از تقسیم‌بندی تصادفی ارزیابی شود.

جدول ۱: مشخصات الگوریتم‌های فراابتکاری مورد استفاده برای انتخاب طول‌موج‌های مؤثر

ردیف	الگوریتم	پارامترها
۱	LCA	تعداد تیم‌ها: ۸، تعداد فصل‌ها: ۱۰، زمان مسابقه: ۲
۲	GA	اندازه جمعیت: ۱۲۰، تعداد نسل‌ها: ۱۰، نرخ ترکیب (Crossover): ۳۰ درصد، نرخ جهش (Mutation): ۳۰ درصد
۳	PSO	تعداد ذرات: ۴۰، تعداد تکرارها (Episodes): ۱۰، تعداد بهترین‌های محلی (P Bests): ۸
۴	ACO	تعداد مورچه‌ها: ۴۰، تعداد تکرارها: ۱۰، نرخ تبخیر: ۴
۵	ICA	تعداد کشورها: ۴۰، تعداد استعمارگران: ۱۱

نتایج و بحث

تحلیل داده‌های طیفی در ناحیه مرئی-مادون قرمز نزدیک (۵۱۰-۱۱۰۰ نانومتر)

شکل ۳، نمودار پراکندگی طیف‌های جذب نمونه‌های دو رقم آلو خرمایی (الف) و خونی (ب) را در محدوده طول‌موج ۵۱۰ تا ۱۱۰۰ نانومتر نمایش می‌دهد. همان‌طور که در هر دو نمودار مشاهده می‌شود، الگوهای طیفی با تنوع قابل توجهی همراه هستند که این امر نشان‌دهنده تفاوت‌های ترکیبی و ساختاری میان نمونه‌ها در مراحل مختلف رسیدگی و همچنین اختلافات جزئی در رنگ یا بافت هر میوه است.

در هر دو رقم، طیف‌ها در ناحیه مرئی (تقریباً ۵۱۰ تا ۷۰۰ نانومتر) دارای میزان جذب نسبتاً بالایی هستند. این بخش از طیف عمدتاً تحت تأثیر حضور رنگ‌دانه‌های مختلف در پوست و گوشت آلو قرار دارد. به‌ویژه، پس از طول‌موج حدود ۶۷۰ نانومتر، کاهش مشخصی در میزان جذب مشاهده می‌شود که معرف پایان ناحیه جذب کلروفیل (در صورت وجود) و ورود به ناحیه مادون قرمز نزدیک است.

در محدوده مادون قرمز نزدیک (حدود ۷۰۰ تا ۱۱۰۰ نانومتر)، الگوهای طیفی مشخص‌تری ظاهر می‌شوند. یک پیک جذب شاخص در حدود ۹۷۰ نانومتر در هر دو رقم به‌وضوح قابل شناسایی است که به‌طور معمول به دومین هارمونیک ارتعاشات پیوندهای O-H مولکول‌های آب نسبت داده می‌شود و بیانگر محتوای آب موجود در نمونه‌هاست (Rajkumar et al., 2022). علاوه بر این، تغییراتی در بازتاب یا جذب در حدود ۱۰۳۰ نانومتر مشاهده می‌شود که می‌تواند با ارتعاشات پیوندهای C-H در قندها و سایر ترکیبات آلی مرتبط باشد (Currà et al., 2024). این ویژگی‌های طیفی، قابلیت طیف‌سنجی NIR را در شناسایی و اندازه‌گیری پارامترهای شیمیایی داخلی میوه‌ها از جمله محتوای قند و رطوبت به‌خوبی برجسته می‌کند.

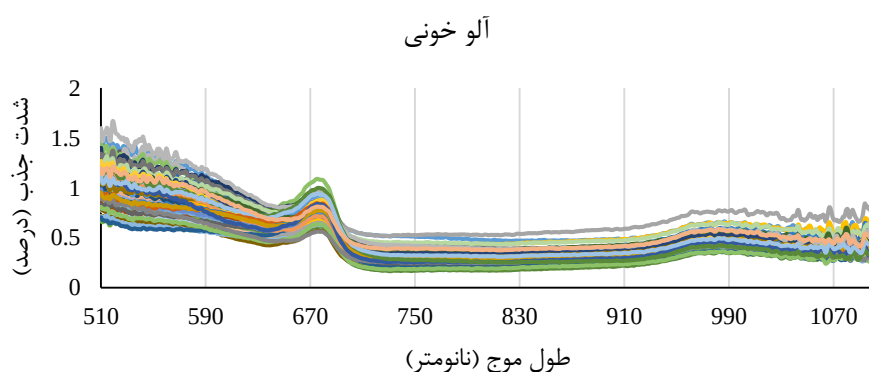
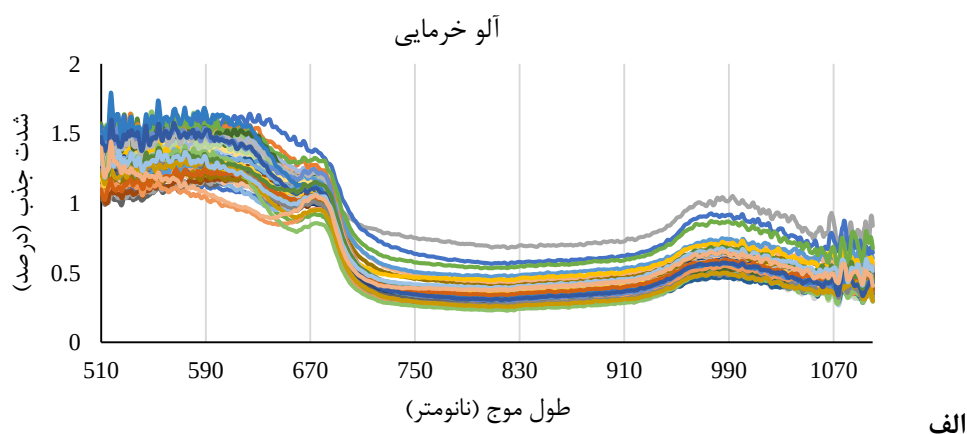
اگرچه شکل کلی طیف‌ها در دو رقم مشابه است، تفاوت‌هایی در شدت جذب و شیب منحنی‌ها در بخش‌هایی از طیف قابل مشاهده است که احتمالاً ناشی از اختلافات فیزیولوژیکی و شیمیایی بین آلوهای خرمایی و خونی است. این تفاوت‌ها می‌توانند در نهایت بر کیفیت و ویژگی‌های حسی میوه‌ها تأثیر بگذارند.

تحلیل میزان مواد جامد محلول (SSC) در ارقام آلو

جدول ۲ آمار توصیفی میزان SSC را برحسب درصد بریکس برای دو رقم آلو خرمایی و خونی نشان می‌دهد. این مقادیر با استفاده از روش مرجع رفاکتومتری اندازه‌گیری شده‌اند و به‌عنوان مبنای اعتبارسنجی مدل‌های پیش‌بینی طیف‌سنجی استفاده شده‌اند. با توجه به این جدول آلو خرمایی دارای میانگین SSC برابر با ۱۶/۵۸ درصد بریکس است. دامنه تغییرات این رقم از حداقل ۱۰/۶ درصد بریکس تا حداکثر ۱۸/۷ درصد بریکس متغیر بوده است. انحراف معیار ۲/۸۳ نشان‌دهنده پراکندگی نسبتاً بالا و ناهمگنی در میزان SSC نمونه‌های این رقم می‌باشد.

آلو خونی دارای میانگین SSC معادل ۱۵/۵۹ درصد بریکس بوده که اندکی کمتر از رقم خرمایی است. دامنه تغییرات SSC برای این رقم از ۱۱/۵ تا ۱۹/۶ درصد بریکس مشاهده شد. انحراف معیار ۲/۲۶ بیانگر تنوع کمتر مقادیر SSC در مقایسه با رقم خرمایی است. مقایسه این دو رقم نشان می‌دهد که اگرچه میانگین SSC در آلو خرمایی کمی بالاتر است، اما بیشینه مقدار SSC در آلو خونی (۱۹/۶ درصد بریکس) بالاتر از حداکثر SSC رقم خرمایی بوده است. این امر حاکی از آن است که برخی نمونه‌های آلو خونی دارای محتوای قند بسیار بالاتری نسبت به سایرین بوده‌اند. از طرفی، تنوع بیشتر SSC در آلو خرمایی می‌تواند ناشی از تفاوت‌های رسیدگی یا شرایط برداشت باشد.

وجود دامنه وسیع مقادیر SSC در هر دو رقم، فرصت مناسبی برای توسعه مدل‌های رگرسیونی دقیق فراهم می‌کند؛ چراکه مدل‌ها قادر خواهند بود ارتباط طیف‌های بازتابی را با سطوح متنوع SSC به‌خوبی یاد بگیرند. به‌علاوه، تفاوت‌های مشاهده‌شده در الگوهای طیفی (شکل ۳) میان دو رقم تا حدی می‌تواند به اختلاف میانگین و پراکندگی SSC (جدول ۲) مربوط باشد. این موضوع اهمیت بهره‌گیری از کل بازه طیفی VIS/NIR را برای ساخت مدل‌های پیش‌بینی قابل اعتماد و عمومی‌تر نشان می‌دهد.



شکل ۳: طیف‌های جذبی نمونه‌های آلو خرمایی (الف) و آلو خونی (ب) اندازه‌گیری شده با طیف‌سنج VIS/NIR

جدول ۲: آمار توصیفی میزان SSC برحسب درصد بریکس در دو رقم آلو

رقم آلو	میانگین	انحراف معیار	حداقل	حداکثر
خرمایی	۱۶/۵۸	۲/۸۳	۱۰/۶	۱۸/۷
خونی	۱۵/۵۹	۲/۲۶	۱۱/۵	۱۹/۶

مقادیر SSC بر اساس اندازه‌گیری رفرکتومتری به‌دست‌آمده‌اند. تعداد نمونه‌ها برای هر رقم: ۸۰ عدد.

مدل PLSR و تأثیر پیش‌پردازش‌ها و تعداد مؤلفه‌ها روی دقت مدل

در این مطالعه، مدل‌سازی PLSR برای پیش‌بینی ویژگی‌های دو رقم آلو با استفاده از داده‌های طیفی انجام شد. نتایج حاصل از ارزیابی مدل در سه مجموعه آموزش، اعتبارسنجی و آزمون با استفاده از شاخص‌های معیار R^2 ، RMSE و RPD بررسی گردید. برای اطمینان از اعتبار و قابلیت تعمیم مدل‌های توسعه‌یافته، علاوه بر معیارهای آماری (R^2 ، RMSE و RPD)، توجه ویژه‌ای به کنترل بیش‌برازش شد. تکرار ۲۰۰ باره تقسیم‌بندی تصادفی داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمون، روشی مؤثر برای ارزیابی پایداری و تعمیم‌پذیری مدل در مواجهه با داده‌های جدید بود. همچنین، استفاده از مجموعه اعتبارسنجی برای تنظیم هایپرپارامترها و تعیین تعداد مؤلفه‌های بهینه در PLSR و ساختار ANN، به جلوگیری از بیش‌برازش کمک شایانی نمود. این رویکرد تضمین کرد که مدل‌ها نه تنها بر روی داده‌های آموزش عملکرد خوبی دارند، بلکه قادر به پیش‌بینی دقیق بر روی داده‌های دیده نشده نیز هستند.

انتخاب تعداد مؤلفه‌ها و تأثیر آن

یکی از مراحل کلیدی در این روش، تعیین تعداد مؤلفه‌های بهینه است که می‌تواند تأثیر بسزایی بر دقت و توانایی پیش‌بینی داده‌های جدید مدل داشته باشد. برای این منظور، ابتدا تعداد مؤلفه‌ها تا سقف ۱۰ افزایش داده شد تا با بررسی معیارهای مختلف، بهترین تعداد مؤلفه تعیین شود. معیارهای اصلی انتخاب مؤلفه‌ها شامل R^2 ، RMSE و RPD بودند. R^2 بالاتر نشان‌دهنده تناسب بهتر مدل با داده‌ها است، RMSE کمتر بیانگر دقت پیش‌بینی بالاتر است، و RPD به عنوان نسبت انحراف معیار داده‌ها به RMSE، معیاری برای کیفیت مدل به شمار می‌رود. معمولاً مقادیر RPD بالاتر از ۲ نشان‌دهنده مدلی با دقت خوب تا عالی هستند.

مطابق نتایج گزارش‌شده در جدول ۳ که به رقم خرمایی اختصاص دارد، تعداد مؤلفه‌های بهینه برابر با ۵ انتخاب شد. این تعداد مؤلفه توانست تعادل مناسبی بین پیچیدگی مدل و دقت پیش‌بینی برقرار کند. به‌طوری‌که بدون پیش‌پردازش مدل روی داده آموزش R^2 برابر با ۰/۸۸۷ و RPD برابر با ۳/۰۰ داشت که نشان‌دهنده کیفیت خوب مدل است. همچنین روی مجموعه اعتبارسنجی نیز R^2 به ۰/۹۰ و RPD به ۳/۴۱ رسید که تأییدکننده توانایی تعمیم‌پذیری مدل است. اما در مجموعه آزمون عملکرد مدل با افت محسوس تا R^2 برابر ۰/۶۵ و RPD حدود ۰/۸۰ کاهش یافت.

برای رقم آلو خونی مطابق جدول ۴، بهترین تعداد مؤلفه ۸ بود که موجب افزایش قابل توجه دقت مدل شد. مدل بدون پیش‌پردازش روی مجموعه آموزش $R^2=0/98$ و $RPD=7/31$ داشت که بیانگر دقت بسیار بالا و مدل‌سازی موفق است. همچنین در مجموعه اعتبارسنجی نیز $R^2=0/97$ و $RPD=6/25$ نشان‌دهنده عملکرد عالی مدل است. البته در مجموعه آزمون با افتی در R^2 تا ۰/۷۴ مواجه شدیم که همچنان از نظر RPD (۲/۰۱) مدل قابل قبولی است.

این نتایج نشان می‌دهد که انتخاب تعداد مؤلفه‌های مناسب، نقش کلیدی در جلوگیری از بیش‌برازش و بهبود تعمیم‌پذیری مدل دارد. افزایش بیش‌ازحد تعداد مؤلفه‌ها ممکن است منجر به یادگیری نویز و کاهش دقت روی داده‌های جدید شود، درحالی‌که تعداد کمتر ممکن است باعث از دست رفتن اطلاعات مهم گردد. این نتایج با مطالعات لطیفی عمیقین و همکاران و همچنین مطالعات تقی‌نژاد و همکاران که از روش PLSR جهت مدل‌سازی‌های خود استفاده کرده‌اند مطابقت دارد (Taghinezhad et al., 2025; Latifi-Amoghini et al., 2025).

اثر پیش‌پردازش‌ها روی عملکرد مدل

بررسی روش‌های پیش‌پردازش نشان می‌دهد که در هر دو رقم، مدل‌های بدون پیش‌پردازش و برخی پیش‌پردازش‌های ساده مانند مشتق اول و دوم، مرکزکردن میانگین، و نرمال‌سازی نتایج بهتری ارائه دادند. برای رقم خرمایی، روش‌هایی مانند SNV و MSC باعث کاهش

قابل توجه R^2 و RPD شدند، به خصوص در مجموعه‌های اعتبارسنجی و آزمون که نشان‌دهنده تأثیر منفی این پیش‌پردازش‌ها بر کیفیت مدل بود. این می‌تواند به علت تغییرات غیرمطلوب در سیگنال طیفی یا حذف بخش‌های مهم اطلاعات باشد. روش‌های فیلترینگ (فیلتر میانه، فیلتر گوسی و میانگین متحرک) عملکرد نسبتاً خوبی داشتند، اما به‌طور کلی نتوانستند از مدل بدون پیش‌پردازش پیشی بگیرند. برای رقم خونی نیز همین روند قابل مشاهده است؛ SNV و MSC باعث افت قابل توجه عملکرد مدل شدند. بهترین نتایج مربوط به مدل بدون پیش‌پردازش، مشتق اول و دوم، نرمال‌سازی و مرکز کردن میانگین بودند که در آن‌ها R^2 و RPD در مجموعه‌های آموزش و اعتبارسنجی بسیار بالا باقی ماند. این نشان می‌دهد که در این داده‌ها، پیش‌پردازش شدید یا تغییر شکل طیف‌ها باعث کاهش دقت پیش‌بینی شده است.

جدول ۳: نتایج مدل‌سازی PLSR رقم خرمایی با انواع پیش‌پردازش‌های طیفی (۵ مؤلفه)

ردیف	پیش‌پردازش	آموزش			اعتبارسنجی			آزمون		
		RPD	RMSE	R^2	RPD	RMSE	R^2	RPD	RMSE	R^2
۱	بدون پیش‌پردازش	۰/۷۹۶	۰/۸۸۶	۳/۰۰۴	۰/۹۰۸	۰/۶۵۹	۳/۴۱۴	۰/۶۵۱	۳/۳۷۸	۰/۸۰۴
۲	SNV	۰/۹۵۶	۰/۸۳۶	۲/۵۰۲	۰/۲۱۰	۱/۹۳۶	۱/۱۶۲	۰/۱۶۸	۲/۳۹۶	۱/۱۳۳
۳	MSC	۰/۹۶۲	۰/۸۳۴	۲/۴۸۶	۰/۱۹۲	۱/۹۵۹	۱/۱۴۸	۰/۲۳۲	۲/۳۰۳	۱/۱۷۹
۴	مشتق اول	۰/۷۹۶	۰/۸۸۶	۳/۰۰۴	۰/۹۰۸	۰/۶۵۹	۳/۴۱۴	۰/۶۵۱	۳/۳۷۸	۰/۸۰۴
۵	مشتق دوم	۰/۷۹۶	۰/۸۸۶	۳/۰۰۴	۰/۹۰۸	۰/۶۵۹	۳/۴۱۴	۰/۶۵۱	۳/۳۷۸	۰/۸۰۴
۶	نرمال‌سازی	۰/۷۹۹	۰/۸۸۶	۲/۹۹۵	۰/۹۰۷	۰/۶۶۳	۳/۳۹۴	۰/۵۲۱	۳/۲۴۲	۰/۸۳۷
۷	میانگین متحرک	۰/۸۴۳	۰/۸۷۳	۲/۸۳۸	۰/۹۰۷	۰/۶۶۳	۳/۳۹۶	۰/۵۷۸	۳/۳۰۳	۰/۸۲۲
۸	فیلتر گوسی	۰/۸۷۲	۰/۸۶۴	۲/۷۴۴	۰/۹۰۶	۰/۶۶۶	۳/۳۷۹	۰/۵۳۰	۳/۲۵۲	۰/۸۳۵
۹	فیلتر میانه	۰/۸۱۸	۰/۸۸۰	۲/۹۲۴	۰/۹۰۸	۰/۶۶۰	۳/۴۰۷	۰/۶۲۱	۳/۳۴۸	۰/۸۱۰
۱۰	ساویتزکی-گولای	۰/۸۲۵	۰/۸۷۸	۲/۸۹۹	۰/۹۰۷	۰/۶۶۳	۳/۳۹۵	۰/۶۱۸	۳/۳۴۴	۰/۸۱۲
۱۱	حذف روند	۰/۸۷۱	۰/۸۶۴	۲/۷۴۸	۰/۱۲۵	۲/۰۳۸	۱/۱۰۴	۰/۰۸۹	۲/۵۰۸	۱/۰۸۲
۱۲	مرکز کردن میانگین	۰/۷۹۶	۰/۸۸۶	۳/۰۰۴	۰/۹۰۸	۰/۶۵۹	۳/۴۱۴	۰/۶۵۱	۳/۳۷۰	۰/۸۰۴

جدول ۴: نتایج مدل‌سازی PLSR رقم خونی با انواع پیش‌پردازش‌های طیفی (۸ مؤلفه)

ردیف	پیش‌پردازش	آموزش			اعتبارسنجی			آزمون		
		RPD	RMSE	R^2	RPD	RMSE	R^2	RPD	RMSE	R^2
۱	بدون پیش‌پردازش	۰/۳۰۱	۰/۹۸۰	۷/۳۰۷	۰/۹۷۲	۰/۳۹۷	۶/۲۴۶	۰/۷۳۸	۱/۲۰۰	۲/۰۱۸
۲	SNV	۰/۵۳۹	۰/۹۳۸	۴/۰۸۱	۰/۳۵۸	۱/۹۲۴	۱/۲۹۰	۰/۰۵۰	۲/۲۸۷	۱/۰۶۰
۳	MSC	۰/۵۵۳	۰/۹۳۵	۳/۹۷۸	۰/۳۵۰	۱/۹۳۷	۱/۲۸۲	۰/۰۷۶	۲/۲۵۵	۱/۰۷۴
۴	مشتق اول	۰/۳۰۱	۰/۹۸۰	۷/۳۰۷	۰/۹۷۲	۰/۳۹۷	۶/۲۴۶	۰/۷۳۸	۱/۲۰۰	۲/۰۱۸
۵	مشتق دوم	۰/۳۰۱	۰/۹۸۰	۷/۳۰۷	۰/۹۷۲	۰/۳۹۷	۶/۲۴۶	۰/۷۳۸	۱/۲۰۰	۲/۰۱۸
۶	نرمال‌سازی	۰/۳۳۵	۰/۹۷۶	۶/۵۶۶	۰/۹۵۷	۰/۴۹۴	۵/۰۲۵	۰/۸۳۰	۰/۹۶۶	۲/۵۰۸
۷	میانگین متحرک	۰/۳۲۲	۰/۹۷۸	۶/۸۳۳	۰/۹۷۵	۰/۳۷۳	۶/۶۴۰	۰/۷۲۵	۱/۲۲۹	۱/۹۷۱
۸	فیلتر گوسی	۰/۳۳۴	۰/۹۷۶	۶/۵۸۱	۰/۹۷۶	۰/۳۶۹	۶/۷۱۷	۰/۷۲۰	۱/۲۴۰	۱/۹۵۴
۹	فیلتر میانه	۰/۳۱۱	۰/۹۷۹	۷/۰۶۶	۰/۹۷۵	۰/۳۷۸	۶/۵۵۲	۰/۷۳۰	۱/۲۱۸	۱/۹۸۹
۱۰	ساویتزکی-گولای	۰/۳۱۵	۰/۹۷۹	۶/۹۷۳	۰/۹۷۴	۰/۳۸۵	۶/۴۴۴	۰/۷۲۱	۱/۲۳۸	۱/۹۵۷
۱۱	حذف روند	۰/۳۱۴	۰/۹۷۹	۷/۰۰۷	۰/۸۳۴	۰/۹۷۸	۲/۵۳۸	۰/۴۵۹	۱/۷۲۵	۱/۴۰۵
۱۲	مرکز کردن میانگین	۰/۳۰۱	۰/۹۸۰	۷/۳۰۷	۰/۹۷۲	۰/۳۹۷	۶/۲۴۶	۰/۷۳۸	۱/۲۰۰	۲/۰۱۸

انتخاب طول‌موج‌های مؤثر

یکی از چالش‌های اساسی در تحلیل داده‌های طیفی، حجم بسیار بالای متغیرها (طول‌موج‌ها) است که علاوه بر افزایش پیچیدگی محاسباتی، می‌تواند منجر به بروز هم‌خطی شدید میان متغیرها و کاهش کارایی مدل‌های پیش‌بینی شود. به‌منظور غلبه بر این مشکل، شناسایی و

انتخاب طول موج‌های مؤثر که بیشترین اطلاعات مرتبط با ویژگی هدف را دربردارند، یک گام ضروری به شمار می‌آید. این فرآیند علاوه بر ساده‌سازی مدل، به بهبود دقت پیش‌بینی و افزایش قابلیت تعمیم‌پذیری آن کمک می‌کند.

پس از انجام مدل‌سازی اولیه با کل بازه طیفی، به منظور بهینه‌سازی عملکرد مدل‌ها و شناسایی طول موج‌های کلیدی، از الگوریتم‌های انتخاب ویژگی بهره گرفته شد. این الگوریتم‌ها قادرند طول موج‌هایی را که بیشترین همبستگی با متغیر وابسته داشته یا بیشترین نقش را در تبیین واریانس داده‌ها ایفا می‌کنند، استخراج کنند. طول موج‌های انتخاب‌شده سپس به عنوان ورودی مدل‌های بعدی مورد استفاده قرار گرفتند.

به منظور شناسایی طول موج‌های کلیدی مؤثر بر کیفیت پیش‌بینی داده‌های طیفی دو رقم آلو (خرمایی و خونی)، از ترکیب SVM با پنج الگوریتم فراابتکاری استفاده شد. نتایج اجرای ده تکرار هر الگوریتم برای هر رقم آلو در شکل‌های ۴ و ۵ به صورت نمودارهای روند تغییر میانگین خطای RMSE و میانگین همبستگی نشان داده شده‌اند.

بر اساس نمودارهای مربوط به رقم خرمایی (شکل ۴)، الگوریتم LCA روندی کاملاً نزولی در کاهش میانگین RMSE طی تکرارها نشان داده و توانسته است مقدار خطای پیش‌بینی را به طور پیوسته کاهش دهد. این کاهش تدریجی بیانگر دقت بالا و توان همگرایی این الگوریتم در یافتن ترکیب بهینه طول موج‌هاست. به طور مشابه، الگوریتم PSO نیز عملکرد نسبتاً مطلوبی داشته و اگرچه خطای اولیه بیشتری نسبت به LCA ثبت کرده، اما در انتهای تکرارها به مقدار قابل قبولی از RMSE رسیده است. در مقابل، الگوریتم GA کمترین کارایی را از نظر کاهش خطا داشته و در اغلب تکرارها نوسان و افت‌وخیز بیشتری در مقادیر RMSE مشاهده شده است. این روند نشان می‌دهد که در این مطالعه نسبت به سایر الگوریتم‌ها توان کمتری در همگرایی بهینه داشته است.

در نمودارهای میانگین همبستگی رقم خرمایی نیز همان الگوی مشابه مشاهده می‌شود، به طوری که الگوریتم LCA با روند افزایشی مستمر به بالاترین مقدار همبستگی در تکرارهای پایانی دست‌یافته و عملکرد دقیق‌تری نسبت به سایر روش‌ها داشته است. الگوریتم PSO در جایگاه دوم قرار گرفته و الگوریتم‌های ACO و ICA عملکرد متوسط و نزدیک به هم نشان داده‌اند. الگوریتم GA بار دیگر کمترین سطح همبستگی را ثبت کرده و از منظر پایداری در تکرارها نیز ضعف بیشتری نشان داده است.

در خصوص رقم خونی (شکل ۵)، مجدداً الگوریتم LCA با اختلاف قابل توجهی نسبت به سایر الگوریتم‌ها بهترین روند کاهش خطا را طی تکرارها ارائه کرده است. نمودار RMSE این الگوریتم کاهش تدریجی و یکنواختی را نشان می‌دهد که بیانگر قدرت بالای آن در گزینش طول موج‌های مؤثر بر کیفیت مدل SVM است. روند افزایش همبستگی الگوریتم LCA نیز کاملاً واضح و بدون نوسان شدید بوده و در پایان تکرارها به بالاترین میزان همبستگی رسیده است. سایر الگوریتم‌ها نظیر PSO، ACO و ICA در هر دو نمودار عملکرد میانه و نسبتاً نزدیک به یکدیگر داشته‌اند، در حالی که GA مجدداً ضعیف‌ترین نتایج را هم از نظر کاهش RMSE و هم افزایش همبستگی به نمایش گذاشته است.

به طور کلی، مقایسه روند نمودارهای RMSE و همبستگی در هر دو رقم آلو نشان می‌دهد الگوریتم LCA به دلیل همگرایی پیوسته و پایداری بیشتر در طی تکرارها، بهترین روش انتخاب طول موج محسوب می‌شود. این ویژگی به افزایش دقت مدل SVM و کاهش خطا در فرآیند پیش‌بینی کمک شایانی کرده است. همچنین مشاهده شد که الگوریتم‌های PSO و ACO از نظر توانایی بهبود پارامترهای ارزیابی عملکرد، جایگاه نسبتاً قابل قبولی داشتند و الگوریتم GA در این مسئله خاص کمترین کارایی را نشان داده است. طول موج‌های منتخب حاصل از الگوریتم LCA در جدول ۵ ارائه شده‌اند این طول موج‌ها بر اساس درجه اهمیت مرتب شده و در مراحل بعدی مدل‌سازی و تحلیل مورد استفاده قرار گرفته‌اند.

مدل‌سازی طول موج‌های مؤثر انتخاب‌شده به روش LCA

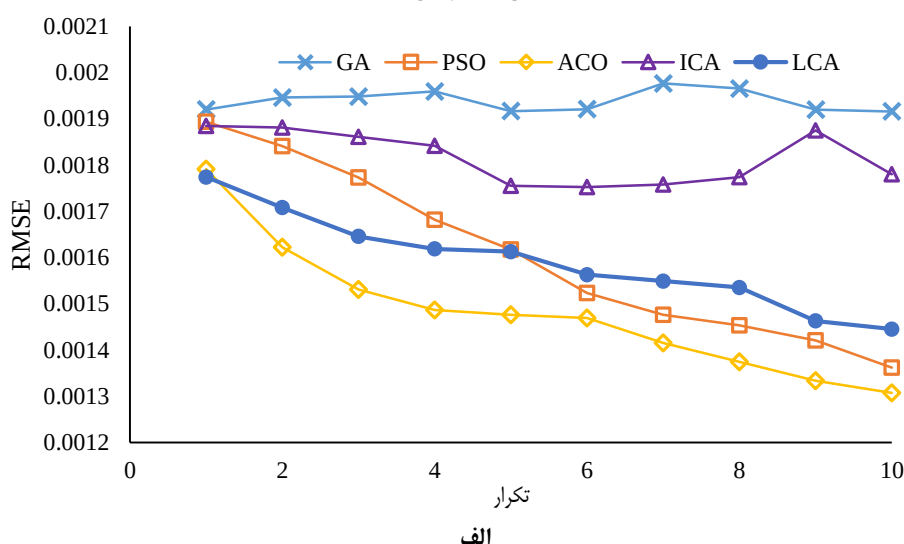
نتایج با R^2 ، RMSE و RPD در سه بخش آموزش، اعتبارسنجی و آزمون ارزیابی گردید. به طور کلی، مقدار بالاتر R^2 و مقدار کمتر RMSE نشان‌دهنده دقت بهتر مدل بوده و RPD بالاتر از ۲ به عنوان مدل خوب تا عالی در نظر گرفته می‌شود؛ همچنین RPD بین ۱/۵ تا ۲ برای اهداف غربالگری مناسب است.

مدل‌سازی با استفاده از روش PLSR

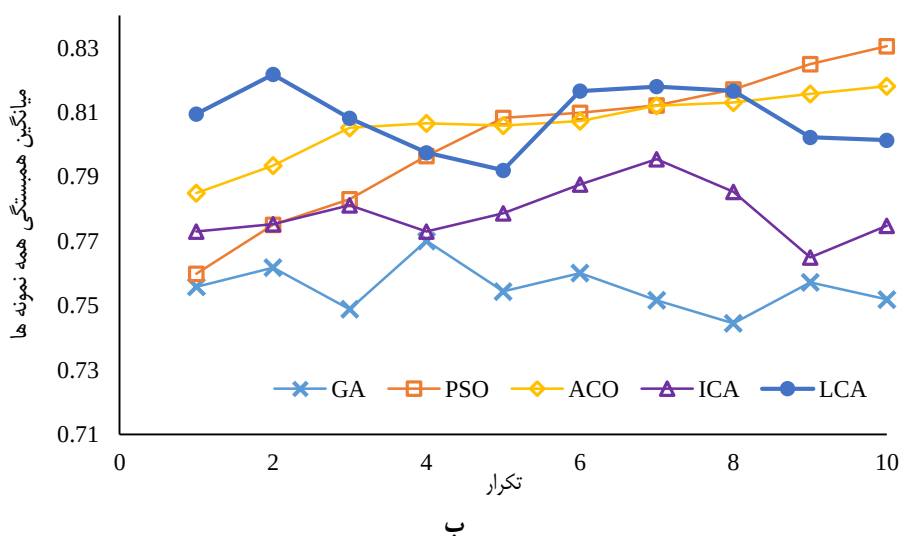
مطابق جدول ۶ که نتایج مربوط به پیش‌بینی مقدار SSC رقم خرمایی با استفاده از روش PLSR و طول موج‌های مؤثر را نشان می‌دهد تعداد ۱۰ مؤلفه به عنوان تعداد بهینه برای مدل انتخاب شد. این مدل در بخش آموزش (با R^2 تا ۰/۸۴ و RPD تا ۲/۵۵) و اعتبارسنجی

(با R^2 تا ۰/۸۷ و RPD تا ۲/۹۶) عملکرد مناسبی از خود نشان داد که بیانگر توانایی مدل در یادگیری و تعمیم روی داده‌های مشابه بود. با این حال، در بخش آزمون عملکرد مدل به طور قابل توجهی کاهش یافت. بهترین نتیجه در بخش آزمون با پیش‌پردازش ساویتزکی-گولای به دست آمد ($R^2=0/856$ و $RPD=2/76$) که نشان‌دهنده کیفیت خوب مدل در پیش‌بینی مقادیر جدید است. این در حالی است که بسیاری از پیش‌پردازش‌های دیگر (مانند بدون پیش‌پردازش، مشتق اول و دوم، نرمال‌سازی، میانگین متحرک، فیلتر گوسی و مرکز کردن میانگین) RPD حدود ۰/۹۸ یا کمتر داشتند که نشان‌دهنده ضعف قابل توجه مدل در پیش‌بینی دقیق داده‌های کاملاً جدید است. همچنین با توجه به جدول ۷ که نتایج مربوط به پیش‌بینی مقدار SSC رقم خونی با استفاده از روش PLSR و طول‌موج‌های مؤثر را نشان می‌دهد بهترین تعداد مؤلفه ۵ تعیین شد. در این حالت، مدل بدون پیش‌پردازش در بخش آموزش ($R^2=0/718$ و $RPD=1/90$) و اعتبارسنجی ($R^2=0/86$ و $RPD=2/76$) عملکرد مطلوبی داشت. در بخش آزمون، با وجود کاهش نسبی دقت، بهترین نتیجه با پیش‌پردازش میانگین متحرک حاصل شد ($R^2=0/583$ و $RPD=1/60$). این مقدار RPD در محدوده ۱/۵ تا ۲ قرار دارد و نشان‌دهنده یک مدل قابل قبول برای اهداف غربالگری است.

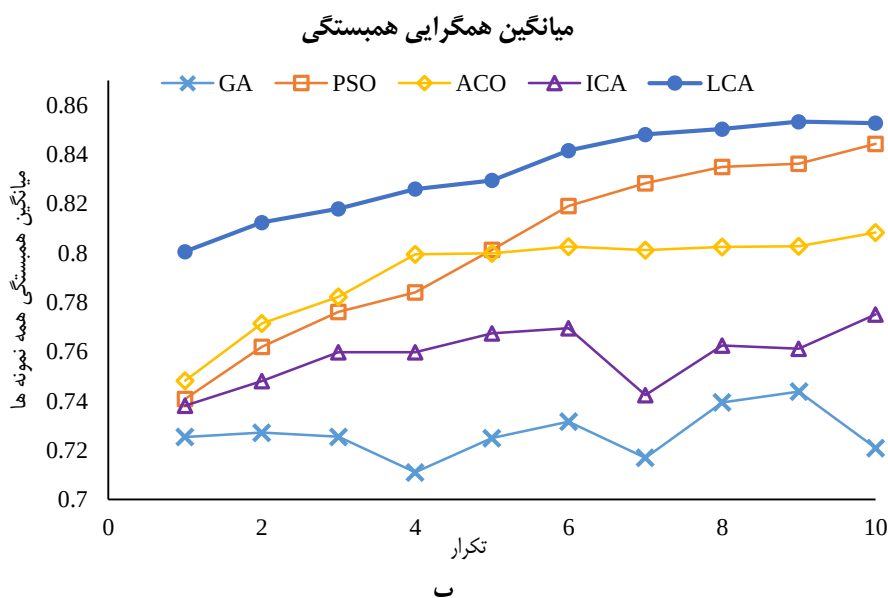
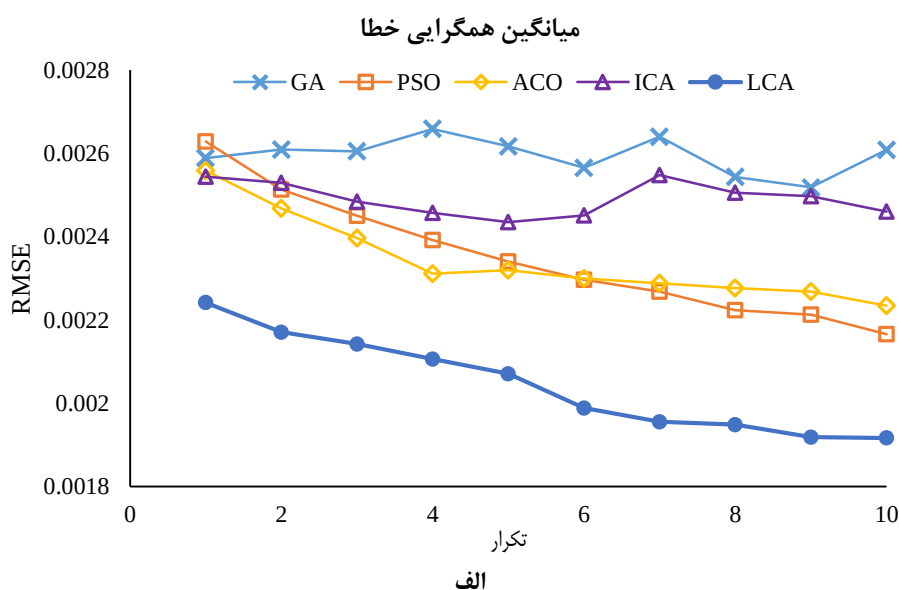
میانگین همگرایی خطا



میانگین همگرایی همبستگی



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



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

جدول ۵: طول موج‌های منتخب الگوریتم SVM-LCA

رقم	طول موج (نانومتر)
خرمایی	۶۸۹، ۹۶۹/۵، ۸۱۳، ۱۱۰۰، ۱۰۹۸/۵، ۸۴۲/۵، ۱۰۸۲/۵، ۱۰۹۰، ۹۴۹/۵، ۹۶۷/۵، ۵۷۰/۵، ۹۵۲/۵، ۸۲۷، ۸۳۷/۵، ۱۰۲۲
خونی	۶۰۶، ۱۰۸۰، ۸۹۰، ۱۰۵۱، ۹۳۶/۵، ۱۰۹۹، ۱۰۷۴/۵، ۷۵۱/۵، ۱۰۸۱، ۱۰۹۴، ۹۰۰، ۸۷۱، ۸۱۲، ۹۵۳/۵، ۱۰۷۱/۵

بررسی اثر پیش‌پردازش‌های مختلف بر عملکرد مدل‌های PLSR در هر دو رقم آلو نشان داد که انتخاب روش مناسب پیش‌پردازش تأثیر قابل توجهی بر دقت پیش‌بینی و تعمیم‌پذیری مدل دارد. در رقم خرمایی، ساولتزکی-گولای بهترین عملکرد را در آزمون با RPD برابر ۲/۷۶ نشان داد، درحالی‌که روش‌های قوی‌تر مثل SNV و MSC موجب کاهش شدید دقت شدند (RPD کمتر از ۱). در رقم خونی، میانگین متحرک بالاترین RPD آزمون (۱/۶۰) را ارائه کرد و سایر روش‌های ساده‌تر نیز نتایج قابل قبولی داشتند، اما باز هم SNV و MSC عملکرد ضعیف‌تری داشتند. این یافته‌ها نشان می‌دهد که انتخاب پیش‌پردازش باید متناسب با رقم و ویژگی‌های طیفی باشد و نمی‌توان یک روش ثابت را برای همه شرایط توصیه کرد.

مدل سازی با استفاده از روش ANN

نتایج ارزیابی مدل های ANN برای دو رقم آلو نشان داد که این مدل ها در استفاده از طول موج های مؤثر توانایی بالایی در پیش بینی دارند. برای رقم خرمایی مطابق جدول ۸، مدل های ANN در داده های آموزش و اعتبارسنجی عملکرد بسیار بالایی داشتند (مثلاً $R^2=0/989$ و $RPD=7/96$ برای میانگین متحرک)، اما در بخش آزمون هم نتایج فوق العاده ای به دست آمد؛ بهترین عملکرد مربوط به میانگین متحرک ($R^2=0/989$ و $RPD=6/08$)، فیلتر گوسی ($RPD=5/47$)، فیلتر میانه ($RPD=5/57$) و ساویتزکی-گولای ($RPD=5/09$) بود که همگی نشان دهنده مدل های عالی برای پیش بینی کمی محسوب می شوند. حتی بدون پیش پردازش، ANN دقت خوبی ($R^2=0/898$ و $RPD=2/31$) ارائه داد. در رقم خونی مطابق جدول ۹، اگرچه عملکرد مدل ها در آموزش و اعتبارسنجی بالا بود، اما در مرحله آزمون مقادیر متغیری را نشان دادند که بهترین نتیجه با SNV به دست آمد ($R^2=0/869$ و $RPD=2/04$) که نشان دهنده یک مدل مناسب پیش بینی کمی است. مشتق اول هم عملکرد قابل قبولی داشت ($RPD=1/56$) و برای غربالگری مناسب بود، اما سایر روش ها RPD کمتر از ۱/۵ داشتند. این نتایج نشان می دهد انتخاب پیش پردازش مناسب در کنار طول موج های منتخب می تواند قدرت پیش بینی مدل های ANN را به ویژه برای رقم خرمایی به طور قابل توجهی افزایش دهد.

مقایسه با نتایج حاصل از محدوده ی طیف کامل

برای مقایسه، نتایج مدل سازی با طول موج های منتخب (جدول های ۶ و ۷) با نتایج مدل سازی با محدوده ی طیف کامل (جدول های ۳ و ۴) مقایسه گردید:

نتایج نشان داد که در مدل های PLSR، استفاده از طیف کامل عملکرد متفاوتی بسته به رقم داشت. برای رقم خرمایی، بهترین نتایج آزمون با طیف کامل برابر با $R^2=0/65$ و $RPD=0/8$ بود که نشان دهنده ضعف مدل در پیش بینی دقیق داده های جدید است. این در حالی بود که با طول موج های منتخب، بهترین RPD آزمون به $2/76$ (یک مدل خوب) رسید. برای رقم خونی نیز، RPD طیف کامل به $2/01$ (یک مدل قابل قبول) و RPD طول موج های منتخب به $1/60$ (قابل قبول برای اهداف غربالگری) به دست آمد. این نشان می دهد که برای رقم خرمایی، انتخاب طول موج ها برای PLSR کارآمدتر بوده، در حالی که برای رقم خونی، هر دو رویکرد نتایج نسبتاً نزدیکی داشتند.

در مقابل، مدل ANN که با طول موج های منتخب کار می کردند، عملکرد به مراتب قوی تری از خود نشان دادند. به ویژه برای رقم خرمایی، RPD آزمون به $6/08$ رسید که نشان دهنده مدلی عالی در پیش بینی کمی است. لازم به ذکر است که مدل سازی ANN با طیف کامل به دلیل پیچیدگی محاسباتی و زمان بر بودن زیاد انجام نشد. با این حال، برتری چشمگیر ANN با طول موج های منتخب، پتانسیل بالای این رویکرد را آشکار می سازد. استفاده از طول موج های مؤثر مزایای عملی فراوانی دارد: ابعاد داده ها را کاهش می دهد، پردازش را تسهیل می کند و راه را برای توسعه ابزارهای طیف سنجی سریع و قابل حمل هموار می سازد. همچنین، این رویکرد به شناسایی طول موج های کلیدی مرتبط با مقدار SSC در آلو کمک می کند.

جدول ۶: نتایج مدل سازی PLSR طول موج های منتخب LCA در رقم خرمایی با انواع پیش پردازش های طیفی (۱۰ مؤلفه)

ردیف	پیش پردازش	آموزش			اعتبارسنجی			آزمون		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
۱	بدون پیش پردازش	۰/۸۲۳	۰/۹۹۶	۲/۴۰۲	۰/۸۷۸	۰/۷۶۰	۲/۹۶۰	۰/۱۱۹	۲/۷۸۱	۰/۹۷۶
۲	SNV	۰/۳۵۱	۱/۹۰۸	۱/۲۵۵	۰/۶۰۱	۲/۷۵۸	۰/۸۱۶	۰/۴۰۱	۴/۰۷۴	۰/۶۶۶
۳	MSC	۰/۳۴۸	۱/۹۱۲	۱/۲۵۲	۰/۵۲۳	۲/۶۹۰	۰/۸۳۷	۰/۴۸۷	۳/۲۰۶	۰/۸۴۷
۴	مشتق اول	۰/۸۲۳	۰/۹۹۶	۲/۴۰۲	۰/۸۷۸	۰/۷۶۰	۲/۹۶۰	۰/۱۱۹	۲/۷۸۱	۰/۹۷۶
۵	مشتق دوم	۰/۸۲۳	۰/۹۹۶	۲/۴۰۲	۰/۸۷۸	۰/۷۶۰	۲/۹۶۰	۰/۱۱۹	۲/۷۸۱	۰/۹۷۶
۶	نرمال سازی	۰/۸۲۳	۰/۹۹۵	۲/۴۰۵	۰/۸۷۴	۰/۷۷۲	۲/۹۱۴	۰/۰۹۱	۲/۷۴۷	۰/۹۸۸
۷	میانگین متحرک	۰/۸۱۶	۱/۰۱۴	۲/۳۵۹	۰/۸۷۲	۰/۷۷۹	۲/۸۸۷	۰/۱۹۰	۲/۸۶۹	۰/۹۴۶
۸	فیلتر گوسی	۰/۸۱۵	۱/۰۱۶	۲/۳۵۵	۰/۸۴۳	۰/۸۶۱	۲/۶۱۳	۰/۰۱۷	۲/۶۵۲	۱/۰۲۴
۹	فیلتر میانه	۰/۸۴۲	۰/۹۳۹	۲/۵۴۹	۰/۷۳۶	۱/۱۱۸	۲/۰۱۲	۰/۳۰۱	۲/۹۹۹	۰/۹۰۵
۱۰	ساویتزکی-گولای	۰/۸۰۴	۱/۰۴۸	۲/۲۸۲	۰/۶۵۴	۱/۲۸۱	۱/۷۵۷	۰/۸۵۶	۰/۹۸۲	۲/۷۵۸
۱۱	حذف روند	۰/۳۴۵	۱/۹۱۶	۱/۲۴۹	۰/۳۷۷	۲/۵۵۸	۰/۸۷۹	۰/۲۹۶	۲/۲۰۴	۱/۲۳۲
۱۲	مرکز کردن میانگین	۰/۸۲۳	۰/۹۹۶	۲/۴۰۲	۰/۸۷۸	۰/۷۶۰	۲/۹۶۰	۰/۱۱۹	۲/۷۸۱	۰/۹۷۶

جدول ۷: نتایج مدل سازی PLSR طول موج های منتخب LCA در رقم خونی با انواع پیش پردازش های طیفی (۵ مؤلفه)

ردیف	پیش پردازش	آموزش			اعتبار سنجی			آزمون		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
۱	بدون پیش پردازش	۰/۷۱۸	۱/۲۲۱	۱/۹۰۴	۰/۸۵۹	۰/۶۷۶	۲/۷۵۹	۰/۵۵۳	۱/۵۲۹	۱/۴۴۹
۲	SNV	۰/۵۴۴	۱/۵۵۳	۱/۴۹۷	۰/۲۵۳	۱/۵۶۱	۱/۱۹۵	۰/۳۱۲	۱/۸۹۷	۱/۲۴۵
۳	MSC	۰/۵۰۶	۱/۶۱۶	۱/۴۳۸	۰/۱۹۳	۱/۶۲۲	۱/۱۵۰	۰/۳۹۰	۱/۷۸۶	۱/۳۳۳
۴	مشتق اول	۰/۷۱۸	۱/۲۲۱	۱/۹۰۴	۰/۸۵۹	۰/۶۷۶	۲/۷۵۹	۰/۵۵۳	۱/۵۲۹	۱/۴۴۹
۵	مشتق دوم	۰/۷۱۸	۱/۲۲۱	۱/۹۰۴	۰/۸۵۹	۰/۶۷۶	۲/۷۵۹	۰/۵۵۳	۱/۵۲۹	۱/۴۴۹
۶	نرمال سازی	۰/۷۱۹	۱/۲۱۷	۱/۹۰۹	۰/۸۴۴	۰/۷۱۲	۲/۶۱۸	۰/۵۳۴	۱/۵۶۰	۱/۵۱۴
۷	میانگین متحرک	۰/۶۵۱	۱/۳۵۷	۱/۷۱۳	۰/۸۳۹	۰/۷۲۴	۲/۵۷۷	۰/۵۸۳	۱/۴۷۷	۱/۵۹۴
۸	فیلتر گوسی	۰/۶۱۸	۱/۴۲۰	۱/۶۳۷	۰/۸۲۸	۰/۷۴۸	۲/۴۹۳	۰/۵۸۱	۱/۴۷۹	۱/۵۹۷
۹	فیلتر میانه	۰/۳۷۹	۱/۸۱۳	۱/۲۸۳	۰/۳۵۱	۲/۲۱۰	۰/۸۸۸	۰/۲۸۷	۱/۹۳۱	۱/۲۲۳
۱۰	ساویترکی-گولای	۰/۵۹۰	۱/۴۷۳	۱/۵۷۸	۰/۷۶۷	۰/۸۷۰	۲/۱۴۴	۰/۵۵۲	۱/۵۳۱	۱/۵۴۳
۱۱	حذف روند	۰/۷۱۲	۱/۲۳۴	۱/۸۸۴	۰/۶۹۴	۰/۹۹۸	۱/۸۶۷	۰/۴۹۳	۱/۶۲۸	۱/۴۵۱
۱۲	مرکز کردن میانگین	۰/۷۱۸	۱/۲۲۱	۱/۹۰۴	۰/۸۵۹	۰/۶۷۶	۲/۷۵۹	۰/۵۵۳	۱/۵۲۹	۱/۴۴۹

جدول ۸: نتایج مدل سازی ANN طول موج های منتخب LCA در رقم خرمایی با انواع پیش پردازش های طیفی

ردیف	پیش پردازش	آموزش			اعتبار سنجی			آزمون		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
۱	بدون پیش پردازش	۰/۸۹۸	۱/۴۹۱	۱/۵۵۹	۰/۸۳۲	۱/۳۹۸	۱/۶۵۸	۰/۸۹۸	۱/۲۳۶	۲/۳۱۵
۲	SNV	۰/۹۷۷	۰/۸۵۷	۲/۷۱۳	۰/۹۱۴	۱/۹۳۷	۱/۲۱۰	۰/۹۷۷	۱/۰۲۷	۲/۷۸۵
۳	MSC	۰/۳۸۹	۶/۹۳۶	۰/۳۳۵	۰/۰۳۰	۷/۳۳۶	۰/۳۱۹	۰/۳۸۹	۳/۷۸۸	۰/۷۵۵
۴	مشتق اول	۰/۴۷۰	۳/۲۴۱	۰/۷۱۷	۰/۴۷۱	۲/۶۱۶	۰/۸۹۶	۰/۴۷۰	۲/۸۴۲	۱/۰۰۷
۵	مشتق دوم	۰/۰۸۳	۴/۱۶۶	۰/۵۵۸	۰/۰۹۲	۳/۳۵۵	۰/۶۹۹	۰/۰۸۳	۳/۵۹۵	۰/۷۹۶
۶	نرمال سازی	۰/۱۸۵	۶/۲۰۰	۰/۳۷۵	۰/۰۵۲	۳/۷۱۳	۰/۶۳۱	۰/۱۸۵	۳/۱۱۹	۰/۹۱۷
۷	میانگین متحرک	۰/۹۸۹	۰/۴۱۴	۵/۶۱۱	۰/۹۹۱	۰/۲۹۴	۷/۹۶۱	۰/۹۸۹	۰/۴۷۰	۶/۰۸۲
۸	فیلتر گوسی	۰/۹۸۴	۱/۱۷۴	۱/۹۷۹	۰/۹۷۴	۰/۵۲۲	۴/۴۸۵	۰/۹۸۴	۰/۵۲۳	۵/۴۶۶
۹	فیلتر میانه	۰/۹۸۳	۰/۵۴۵	۴/۲۶۰	۰/۹۹۱	۰/۳۳۳	۷/۲۵۲	۰/۹۸۳	۰/۵۱۳	۵/۵۷۱
۱۰	ساویترکی-گولای	۰/۹۸۱	۰/۷۳۵	۳/۱۶۲	۰/۹۶۰	۰/۶۴۶	۳/۶۲۶	۰/۹۸۱	۰/۵۶۱	۵/۰۹۵
۱۱	حذف روند	۰/۸۴۷	۱/۶۱۶	۱/۴۳۹	۰/۸۸۹	۲/۵۳۷	۰/۹۲۷	۰/۸۴۷	۱/۹۱۴	۱/۴۹۴
۱۲	مرکز کردن میانگین	۰/۹۸۶	۱/۶۹۰	۱/۳۷۶	۰/۹۱۹	۲/۳۹۶	۰/۹۹۸	۰/۹۸۶	۱/۱۸۱	۲/۴۲۶

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

ردیف	پیش پردازش	آموزش			اعتبار سنجی			آزمون		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
۱	بدون پیش پردازش	۰/۸۲۲	۱/۸۸۸	۱/۷۲۲	۰/۷۳۸	۲/۸۷۹	۰/۸۷۳	۰/۵۹۲	۲/۳۰۶	۱/۱۰۹
۲	SNV	۰/۸۲۰	۱/۷۳۳	۱/۸۳۱	۰/۷۶۵	۱/۷۹۸	۱/۳۹۷	۰/۸۶۹	۱/۲۵۵	۲/۰۳۷
۳	MSC	۰/۱۱۹	۳/۰۲۰	۰/۷۶۴	۰/۱۰۲	۴/۱۵۵	۰/۶۰۵	۰/۵۴۰	۵/۵۰۰	۰/۴۶۵
۴	مشتق اول	۰/۴۷۰	۳/۰۱۰	۱/۲۶۶	۰/۸۶۴	۱/۵۸۶	۱/۵۸۴	۰/۷۹۹	۱/۶۳۴	۱/۵۶۴
۵	مشتق دوم	۰/۲۴۴	۳/۵۶۱	۱/۱۴۸	۰/۵۷۰	۲/۲۰۸	۱/۱۳۸	۰/۴۴۰	۲/۷۵۴	۰/۹۲۸
۶	نرمال سازی	۰/۵۹۴	۲/۹۰۰	۱/۲۹۶	۰/۵۷۰	۲/۶۹۸	۰/۹۳۱	۰/۷۰۰	۱/۹۶۵	۱/۳۰۱
۷	میانگین متحرک	۰/۴۳۲	۲/۷۰۴	۱/۳۵۳	۰/۶۷۱	۲/۱۴۸	۱/۶۹۸	۰/۲۵۵	۳/۰۰۰	۰/۸۵۲
۸	فیلتر گوسی	۰/۷۹۸	۱/۶۹۸	۱/۸۵۹	۰/۴۷۲	۲/۸۳۲	۰/۸۹۰	۰/۵۲۶	۲/۸۴۷	۰/۸۹۸
۹	فیلتر میانه	۰/۴۱۴	۳/۲۶۶	۱/۲۰۶	۰/۶۰۵	۳/۷۱۳	۰/۶۷۶	۰/۳۹۵	۴/۱۵۱	۰/۶۱۶
۱۰	ساویترکی-گولای	۰/۷۵۲	۲/۱۰۱	۱/۵۹۸	۰/۴۶۲	۲/۵۰۵	۱/۰۰۳	۰/۶۷۰	۱/۹۹۸	۱/۲۸۰
۱۱	حذف روند	۰/۸۰۵	۲/۵۳۵	۱/۴۱۰	۰/۵۱۱	۲/۴۸۳	۱/۰۱۲	۰/۶۷۳	۲/۱۷۱	۱/۱۷۸
۱۲	مرکز کردن میانگین	۰/۸۰۶	۲/۲۰۴	۱/۶۳۷	۰/۷۴۳	۲/۲۰۸	۱/۲۰۵	۰/۷۰۱	۲/۲۰۴	۱/۲۵۲

تحلیل مقایسه‌ای نتایج

مقایسه با مدل‌های خطی و میوه‌های با نتایج ضعیف‌تر

عملکرد عالی مدل بهینه ما (ANN با انتخاب ویژگی LCA) در پیش‌بینی SSC با $R^2 = 0.989$ و $RPD = 6.08$ در مقایسه با بسیاری از مطالعات انجام‌شده بر روی میوه‌های مختلف، قابل توجه است. به عنوان مثال، در پژوهشی که توسط Pratiwi و همکاران (۲۰۲۳) برای تعیین SSC میوه‌هایی مانند موز و ساپودیل با استفاده از مدل PLSR انجام شد، حداکثر R^2 در پیش‌بینی به ترتیب ۰/۸۸ و ۰/۹۰ گزارش گردید. همچنین، Agulheiro-Santos و همکاران (۲۰۲۲) در ارزیابی SSC توت فرنگی با طیف‌سنجی VIS/NIR، نتایج مدل PLSR خود را با R^2 اعتبارسنجی متقاطع ۰/۵۲ گزارش کردند (Agulheiro-Santos et al., 2022) که نشان‌دهنده پیچیدگی بالای مدل‌سازی در برخی میوه‌ها با ساختار نوری چالش‌برانگیز است. دستاورد ما نشان می‌دهد که در محیط طیفی آلو، اتکاء به مدل‌های خطی یا طیف کامل کارایی لازم را ندارد و استفاده از یک رویکرد غیرخطی و قوی برای فیلتر کردن سیگنال‌های اضافی، امری ضروری است.

مقایسه با روش‌های پیشرفته و اختصاصی

تازگی علمی این تحقیق نه تنها در دقت بالا، بلکه در روش‌شناسی به کار رفته برای انتخاب ویژگی نهفته است. در سال‌های اخیر، تحقیقات متعددی تلاش کرده‌اند دقت مدل‌سازی SSC را با استفاده از تکنیک‌های پیچیده‌تر بهبود بخشند؛ نظیر استفاده از طیف‌سنجی با تفکیک مکانی برای سبب (Ma et al., 2021) یا ترکیب روش‌های مختلف انتخاب ویژگی مانند عملگر انقباض و انتخاب مطلق حداقل (LASSO) و فیلتر گوسی برای گلابی (Zhan et al., 2023). با این حال، در پژوهش ما، الگوریتم LCA که یک تکنیک فراابتکاری است، توانست با تکیه بر داده‌های طیف‌سنجی ساده و غیرمخرب (نیاز به تجهیزات کم‌تر تخصصی)، به دقتی معادل یا بالاتر دست یابد. این امر اثبات می‌کند که نوآوری در حوزه کمومتریکس می‌تواند حتی از لحاظ اقتصادی به صرفه‌تر و مؤثرتر از سرمایه‌گذاری صرف بر روی تجهیزات پیشرفته یا سنسورهای پیچیده باشد.

جمع‌بندی برتری

دستیابی به معیار $RPD = 6.08$ برای رقم خرمایی، که به مراتب بالاتر از آستانه $RPD > 3$ (مناسب برای کنترل کیفیت روزمره) و حتی فراتر از آستانه $RPD > 5$ (مناسب برای تحقیقات و مدل‌های دقیق) است، نشان‌دهنده قابلیت عالی و اعتمادپذیری خوب مدل پیشنهادی LCA-ANN در پیش‌بینی کمی SSC است. مقایسه این نتیجه با پژوهش‌های دیگری که بر میوه‌های کوچک‌تر مانند گوجه گیلاسی (Zheng et al., 2024) متمرکز بوده‌اند، نشان می‌دهد که رویکرد ترکیبی ما به یک استراتژی جهانی برای بهبود دقت مدل‌سازی در میوه‌های مختلف نزدیک است. در نهایت، این نتایج، توجیه‌گر قوی‌تری برای بکارگیری مدل‌های غیرخطی پیشرفته و روش‌های بهینه‌سازی فراابتکاری به جای مدل‌های خطی سنتی در محیط‌های صنعتی پس از برداشت هستند، زیرا دقت و اطمینان لازم برای درجه‌بندی خودکار را فراهم می‌آورند.

نتیجه‌گیری

در پژوهش حاضر به بررسی پتانسیل طیف‌سنجی مرئی / مادون‌قرمز نزدیک (VIS/NIR) برای تشخیص غیرمخرب میزان جامد محلول (SSC) در دو رقم آلو (خرمایی و خونی) پرداخته شد. نتایج نشان دادند که طیف‌سنجی VIS/NIR قابلیت بالایی در ارتباط با محتوای SSC ارقام آلو دارد. در این میان، استفاده از الگوریتم قهرمانی لیگ (LCA) برای انتخاب طول‌موج‌های مؤثر، راهکاری کارآمد برای کاهش ابعاد داده‌ها و بهبود عملکرد مدل‌ها بود. طول‌موج‌های منتخب LCA، اطلاعات کلیدی مرتبط با SSC را به خوبی حفظ کردند و پتانسیل پیاده‌سازی این فناوری را در ابزارهای طیف‌سنجی کوچک و قابل حمل افزایش دادند. نقش پیش‌پردازش داده‌های طیفی نیز به عنوان یک گام حیاتی در بهبود دقت مدل‌سازی تأیید شد. بررسی‌ها نشان داد که بهترین روش پیش‌پردازش به نوع مدل و رقم آلو وابسته است. درحالی‌که مدل PLSR با استفاده از طیف کامل اغلب بهترین عملکرد را بدون هیچ‌گونه پیش‌پردازشی نشان داد، مدل ANN با طول‌موج‌های منتخب به شدت به پیش‌پردازش‌های خاصی پاسخ مثبت داد. به‌طور خاص، روش‌های فیلترینگ مانند میانگین متحرک و سواپتزکی-گولای و همچنین SNV توانستند دقت پیش‌بینی را به شکل چشمگیری افزایش دهند. مدل‌سازی با استفاده از ANN روی

طول موج‌های منتخب، به ویژه برای رقم خرمایی، دقت پیش‌بینی قابل توجهی را به نمایش گذاشت. این امر با دستیابی به بالاترین RPD آزمون برابر با ۶/۰۸ به مشهود بود. این در حالی است که مدل‌های PLSR، اگرچه با طیف کامل برای رقم خونی عملکرد قابل قبولی (RPD تا ۲/۰۱) و برای رقم خرمایی عملکرد ضعیفی (RPD تا ۰/۸۰) داشتند، در حالت کاهش یافته با طول موج‌های منتخب نیز دقت کمتری نسبت به ANN ارائه دادند (برای مثال بهترین RPD آزمون PLSR با طول موج‌های منتخب برای خرمایی ۲/۷۶ و برای خونی ۱/۶۰ بود). این برتری ANN نشان‌دهنده توانایی بالای این مدل در استخراج روابط غیرخطی پیچیده از داده‌های طیفی کاهش یافته است. همچنین، مشخص شد که انتخاب مناسب پیش‌پردازش (مانند SNV و میانگین متحرک برای ANN، و ساویتزکی-گولای و میانگین متحرک برای PLSR) نقش مهمی در بهینه‌سازی عملکرد هر دو نوع مدل دارد.

بر اساس این یافته‌ها، ما قویاً اجرای این متدولوژی LCA-ANN را به عنوان یک ابزار عملی کنترل کیفیت در لحظه در خطوط درجه‌بندی تجاری پس از برداشت توصیه می‌کنیم. کاهش چشمگیر بار محاسباتی به دلیل انتخاب طول موج‌ها، این سیستم را برای ادغام در ماشین‌آلات سورتینگ خودکار با سرعت بالا مناسب می‌سازد. محدودیت‌های اصلی این مطالعه شامل تمرکز آن بر تنها دو رقم آلو (خرمایی و خونی) و محدود بودن جمع‌آوری داده‌ها به یک فصل برداشت واحد و شرایط نگهداری خاص است. علاوه بر این، دقت مدل ممکن است تحت تأثیر تغییرات اندازه میوه و رنگ پوست در طول سال‌های مختلف قرار گیرد. برای گسترش کاربرد این تحقیق، مطالعات آتی باید بر مواردی همچون اعتبارسنجی مدل LCA-ANN در دامنه گسترده‌تری از ارقام آلو و محیط‌های رشد/فصول مختلف، بررسی پیش‌بینی همزمان چندین شاخص کیفی (مانند سفتی، اسیدیته و SSC) با استفاده از یک طیف VIS/NIR واحد و توسعه یک دستگاه قابل حمل و کم هزینه بر اساس طول موج‌های کلیدی شناسایی شده برای ارزیابی کیفیت در سطح مزرعه یا خرده‌فروشی تمرکز داشته باشد. در مجموع، این مطالعه تأیید می‌کند که تلفیق طیف‌سنجی VIS/NIR با الگوریتم‌های هوشمند انتخاب ویژگی و مدل‌سازی (به ویژه ANN)، رویکردی بسیار کارآمد و غیرمخرب برای ارزیابی دقیق SSC در میوه آلو فراهم می‌کند. این دستاورد می‌تواند پیامدهای مهمی برای سیستم‌های کنترل کیفیت در صنعت فرآوری و عرضه میوه داشته باشد، به بهبود مدیریت پس از برداشت کمک کرده و ارزش تجاری محصولات را افزایش دهد.

REFERENCES

- Agulheiro-Santos, A. C., Ricardo-Rodrigues, S., Laranjo, M., Melgão, C., & Velázquez, R. (2022). Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy. *Journal of the Science of Food and Agriculture*, 102(11), 4866-4872.
- Amoghin, M. L., Abbaspour-Gilandeh, Y., Tahmasebi, M., & Arribas, J. I. (2024). 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.
- Antoniali, S., Leal, P. A. M., Magalhães, A. M. d., Fuziki, R. T., & Sanches, J. (2007). Physico-chemical characterization of 'Zarco HS' yellow bell pepper for different ripeness stages. *Scientia Agricola*, 64, 19-22.
- Butac, M., Bozhkova, V., Zhivondov, A., Milosevic, N., Bellini, E., Nencetti, V., Blazek, J., Balsemin, E., Lafarque, B., & Kaufmane, E. (2011). Overview of plum breeding in Europe. II Balkan Symposium on Fruit Growing 981.
- 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.
- Currà, A., Gasbarrone, R., Gattabria, D., Bonifazi, G., Serranti, S., Greco, D., Missori, P., Fattapposta, F., Picciano, A., & Maffucci, A. (2024). In Vivo Insights: Near-Infrared Photon Sampling of Reflectance Spectra from Cranial and Extracranial Sites in Healthy Individuals and Patients with Essential Tremor. *Photonics*.
- Deli, J., & Molnar, P. (2002). Paprika carotenoids: analysis, isolation, structure elucidation. *Current Organic Chemistry*, 6, 1197e1219.
- Gao, Q., Wang, P., Niu, T., He, D., Wang, M., Yang, H., & Zhao, X. (2022). Soluble solid content and firmness index assessment and maturity discrimination of *Malus micromalus* Makino based on near-infrared hyperspectral imaging. *Food Chemistry*, 370, 131013.

- Guo, H., Yan, F., Li, P., & Li, M. (2022). Determination of storage period of harvested plums by near-infrared spectroscopy and quality attributes. *Journal of Food Processing and Preservation*, 46(9), e16504 .
- Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Cruz-Gámez, E. D. L., Hernández-Hernández, J. L., & Martínez-Arroyo, M. (2022). Non-destructive measurement of quality parameters of apple fruit by using visible/near-infrared spectroscopy and multivariate regression analysis. *Sustainability*, 14(22), 14918 .
- Jamshidi, B., Mohajerani, E., Jamshidi, J., Minaei, S., & Sharifi, A. (2015). Non-destructive detection of pesticide residues in cucumber using visible/near-infrared spectroscopy. *Food Additives & Contaminants: Part A*, 32(6), 857-863 .
- Johnson, J. B., El Orche, A., & Naiker, M. (2022). Prediction of anthocyanin content and variety in plum extracts using ATR-FTIR spectroscopy and chemometrics. *Vibrational Spectroscopy*, 121, 103406 .
- Latifi-Amoghin, M., Abbaspour-Gilandeh, Y., Tahmasebi, M., Kisalaei, A., Hernández-Hernández, J. L., Hernández-Hernández, M., & Cruz-Gámez, E. D. L. (2025). Analyzing the Nitrate Content in Various Bell Pepper Varieties Through Non-Destructive Methods Using Vis/NIR Spectroscopy Enhanced by Metaheuristic Algorithms. *Processes*, 13(6), 1731 .
- Latifi Amoghin, M., Abbaspour-Gilandeh, Y., Tahmasebi, M., Kaveh, M., El-Mesery, H. S., Szymanek, M., & Sprawka, M. (2024). 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 .
- Ma, T., Xia, Y., Inagaki, T., & Tsuchikawa, S. (2021). Rapid and nondestructive evaluation of soluble solids content (SSC) and firmness in apple using Vis–NIR spatially resolved spectroscopy. *Postharvest Biology and Technology*, 173, 111417 .
- Martínez-Esplá, A., Zapata, P. J., Valero, D., Martínez-Romero, D., Díaz-Mula, H. M., & Serrano, M. (2018). Preharvest treatments with salicylates enhance nutrient and antioxidant compounds in plum at harvest and after storage. *Journal of the Science of Food and Agriculture*, 98(7), 2742-2750 .
- 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 .
- Posom, J., Klaprachan, J., Rattanasopa, K., Sirisomboon, P., Saengprachatanarug, K., & Wongpichet, S. (2020). Predicting marian plum fruit quality without environmental condition impact by handheld visible–near-infrared spectroscopy. *ACS omega*, 5(43), 279.27921-09
- Pratiwi, E. Z. D., Pahlawan, M. F., Rahmi, D. N., Amanah, H. Z., & Masithoh, R. E. (2023). Non-destructive evaluation of soluble solid content in fruits with various skin thicknesses using visible–shortwave near-infrared spectroscopy. *Open Agriculture*, 8(1), 20220183 .
- Rajkumar, D., Künemeyer, R., Kaur, H., Longdell, J., & McGlone, A. (2022). Interactions of linearly polarized and unpolarized light on kiwifruit using aquaphotomics. *Molecules*, 27(2), 494 .
- Rasmussen, M., Krølner, R., Klepp, K.-I., Lytle, L., Brug, J., Bere, E., & Due, P. (2006). Determinants of fruit and vegetable consumption among children and adolescents: a review of the literature. Part I: quantitative studies. *International journal of behavioral nutrition and physical activity*, 3, 1-19 .
- 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.
- Taghinezhad, E., Szumny, A., Figiel, A., Amoghin, M. L., Mirzazadeh, A., Blasco, J., Mazurek, S., & Castillo-Gironés, S. (2025). The Potential Application of HSI and VIS/NIR Spectroscopy for Non-Invasive Detection of Starch Gelatinization and Head Rice Yield during Parboiling and Drying Process. *Journal of Food Composition and Analysis*, 107443 .
- Weekley, C. W., Zaya, D. N., Menges, E. S., & Faivre, A. E. (2010). Multiple causes of seedling rarity in scrub plum, *Prunus geniculata* (Rosaceae), an endangered shrub of the Florida scrub. *American journal of botany*, 97(1), 144-155 .
- Xu, X., Mo, J., Xie, L., & Ying, Y. (2019). Influences of detection position and double detection regions on determining soluble solids content (SSC) for apples using on-line visible/near-infrared (Vis/NIR)



spectroscopy. *Food Analytical Methods*, 12, 2078-2085 .

- Zhan, B., Li, P., Li, M., Luo, W., & Zhang, H. (2023). Detection of Soluble Solids Content (SSC) in Pears Using Near-Infrared Spectroscopy Combined with LASSO–GWF–PLS Model. *Agriculture*, 13(8), 1491 .
- Zheng, Y., Liu, P., Zheng, Y., & Xie, L. (2024). Improving SSC detection accuracy of cherry tomatoes by feature synergy and complementary spectral bands combination. *Postharvest Biology and Technology*, 213, 112922 .