

Journal Pre-proofs

Non-destructive detection of peroxidase enzyme activity in different plum varieties using VIS/NIR spectroscopy and machine learning methods

Meysam Latifi-Amoghin, Yousef Abbaspour-Gilandeh

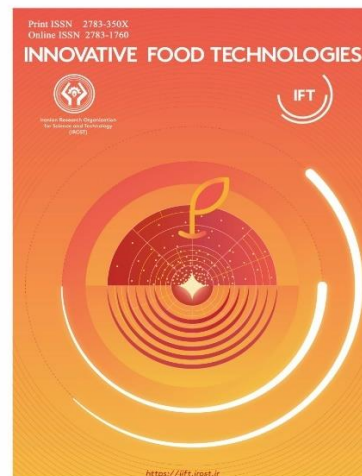
DOI: <https://doi.org/10.22104/ift.2025.7883.2239>

To appear in: Innovative Food Technologies (IFT)

Received Date: 20 September 2025

Revised Date: 28 October 2025

Accepted Date: 9 November 2025



Please cite this article as: Meysam Latifi-Amoghin, Yousef Abbaspour-Gilandeh, Non-destructive detection of peroxidase enzyme activity in different plum varieties using VIS/NIR spectroscopy and machine learning methods , *Innovative Food Technologies* (2025), doi: <https://doi.org/10.22104/ift.2025.7883.2239>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 The Author(s). Published by irost.org.

تشخیص غیرمخرب فعالیت آنزیم پراکسیداز ارقام مختلف آلو با استفاده از طیف‌سنجی VIS/NIR و روش‌های یادگیری ماشین

میثم لطیفی عموقین^۱، یوسف عباس‌پور گیلانده^{۲*}

^۱ دانش‌آموخته‌ی دکتری، گروه مهندسی مکانیک بیوسیستم، دانشکده علوم و فناوری کشاورزی، دانشگاه محقق اردبیلی، اردبیل، ایران

^۲ استاد، گروه مهندسی مکانیک بیوسیستم، دانشکده علوم و فناوری کشاورزی، دانشگاه محقق اردبیلی، اردبیل، ایران (abbaspour@uma.ac.ir)

چکیده

این پژوهش با هدف توسعه و ارزیابی یک روش غیرمخرب برای پیش‌بینی فعالیت آنزیم پراکسیداز (POD) در دو رقم آلو خرمایی و خونی با استفاده از طیف‌سنجی Vis/NIR و روش‌های یادگیری ماشین انجام شد. پس از جمع‌آوری طیف‌های جذبی، داده‌ها با استفاده از الگوریتم‌های رگرسیون حداقل مربعات جزئی (PLSR) و ماشین‌بردار پشتیبان (SVM) در محدوده کامل طیفی و ابعاد کاهش‌یافته مدل‌سازی شدند و عملکرد آن‌ها با معیارهای ضریب تعیین (R^2)، ریشه میانگین مربعات خطا (RMSE) و نسبت عملکرد به انحراف (RPD) مورد ارزیابی قرار گرفت. به‌منظور کاهش ابعاد داده‌ها و افزایش کارایی مدل‌ها، از تلفیق مدل SVM با الگوریتم‌های فراابتکاری برای انتخاب طول موج‌های مؤثر استفاده شد. از میان این الگوریتم‌ها الگوریتم بهینه‌سازی ازدحام ذرات (PSO) به‌عنوان بهترین الگوریتم انتخاب شد. نتایج نشان داد که استراتژی کاهش ابعاد با حذف نویز و اطلاعات اضافی، توانایی مدل‌ها را در پیش‌بینی دقیق بهبود می‌بخشد. یافته‌ها حاکی از آن است که بهترین مدل برای هر رقم متفاوت بود. بعد از کاهش ابعاد و انتخاب طول موج‌های مؤثر برای رقم خرمایی، مدل SVM-R با هسته rbf و پیش‌پردازش فیلتر میانه بهترین عملکرد را ارائه داد و با $RPD=2/687$ در رده مدل‌های عالی قرار گرفت که بر اهمیت مدل‌های غیرخطی برای این رقم تأکید دارد. در مقابل، برای رقم خونی، بهترین نتیجه با مدل خطی PLSR و پیش‌پردازش نرمال‌سازی حاصل شد که با $RPD=2/29$ در رده مدل‌های بسیار خوب جای گرفت. این تفاوت نشان می‌دهد که انتخاب بهینه مدل به ویژگی‌های ذاتی هر محصول بستگی دارد. در نهایت، این پژوهش تأیید می‌کند که طیف‌سنجی Vis/NIR ابزاری قدرتمند برای پایش سریع و غیرمخرب کیفیت پس از برداشت محصولات کشاورزی است.

واژه‌های کلیدی: آلو، آنزیم پراکسیداز، طیف‌سنجی، تشخیص غیرمخرب، یادگیری ماشین.

۱- مقدمه

آلو (سرده *Prunus* از خانواده گل‌سرخ‌سانان) یکی از مهم‌ترین میوه‌های هسته‌دار مناطق معتدل جهان است که گونه‌های اصلی آن شامل آلو اروپایی (*Prunus domestica* L.) و آلو ژاپنی (*Prunus salicina* L.) می‌شود. این میوه‌ها به دلیل ارزش غذایی بالا منبع مهمی از ویتامین‌های گروه B، پتاسیم و منیزیم و نیز ترکیبات آنتی‌اکسیدانی متنوع (فنل‌ها و فلاونوئیدها) شناخته می‌شوند که فواید قلبی-عروقی و گوارشی را فراهم می‌کنند. مصرف آلو به صورت تازه و نیز در محصولات فرآوری شده (مثل مربا، آبمیوه،

کمپوت و براندی) رایج است. تولید جهانی آلو در سال ۲۰۲۲ حدود ۱۲ میلیون تن برآورد شده است. در ایران نیز براساس آمار سال ۲۰۲۲ حدود ۳۳۲ هزار تن آلو تولید شده است [۱]. گوشت آلو حاوی ترکیبات متنوعی از جمله پلی‌فنل‌ها، قندها، اسیدهای آلی و رنگدانه‌هاست. مطالعات شیمیایی نشان داده‌اند اسیدهای کلروژنیک (به‌ویژه ۳-کافئویل کوئینیک اسید) و پروآنتوسیانیدین‌ها از ترکیبات فنولی غالب در آلو هستند [۲]. همچنین قندهای محلول اصلی در آلو شامل گلوکز، فروکتوز، ساکارز و سوربیتول استو اسیدهای مالیک و کینیک عمده اسیدهای آلی گوشت این میوه محسوب می‌شوند [۳]. علاوه بر این، آنزیم‌هایی مانند پراکسیداز (POD) در واکنش‌های اکسیداسیون ترکیبات فنلی نقش دارند و فعالیت این آنزیم یکی از پارامترهای مهم کیفیت پس از برداشت میوه‌ها محسوب می‌شود [۴].

جنبه‌های کیفیت و ایمنی مواد غذایی به یک نگرانی عمده برای تامین‌کنندگان و مصرف‌کنندگان مواد غذایی تبدیل شده است. با نیاز روزافزون به تولید و فرآوری با توان بالا، صنایع غذایی با چالش‌های متعددی مواجه است که مدیریت کیفیت بیشترین تأثیر را بر بازارپسندی دارد. کیفیت غذا به طور کلی با ویژگی‌های ترکیبی و فیزیکی آن تعریف می‌شود اگرچه بازرسی‌های دستی برای ارزیابی‌های کوچک مقرون به صرفه هستند، اما همیشه خطر نادیده گرفتن ویژگی‌های کیفیت و ایمنی حیاتی وجود دارد. این امر نیاز به رویکردهای بلندپایه، سریع و غیرمخرب برای تعیین کمیت هر یک از پارامترهای کیفیت را به طور همزمان ضروری می‌کند تا ایمنی غذا را تضمین کند [۵].

رنگی که روی سطح محصول وجود دارد را می‌توان به فعالیت آنزیم POD و وجود رنگدانه‌های دیگر مانند کلروفیل و کاروتنوئیدها نسبت داد. فعالیت آنزیم POD به شدت به عواملی همچون بلوغ میوه، محتوای فنلی و رقم بستگی دارد. همچنین این آنزیم‌ها بر ویژگی‌های حسی و تغذیه‌ای محصولات تازه و فرآوری شده تأثیر منفی می‌گذارند و در نتیجه پذیرش مصرف‌کنندگان را کاهش می‌دهند [۶]. آنزیم‌های POD متعلق به گروه بزرگ آنزیم‌های اکسیدوردوکتاز^۱ با شماره دسته‌بندی EC 1.11.1.7 هستند که حضور گسترده‌ای در موجودات زنده، به ویژه در گیاهان و میوه‌ها، دارند. این آنزیم‌ها نقشی محوری و چندمنظوره در متابولیسم گیاهی ایفا می‌کنند. نقش بیوشیمیایی اصلی آنزیم‌های پراکسیداز، کاتالیز کردن واکنش اکسیداسیون سوبسترهای مختلف (دهنده‌های الکترون) با استفاده از هیدروژن پراکسید (H_2O_2) به عنوان پذیرنده الکترون است. این واکنش موجب تجزیه H_2O_2 می‌شود که یک ترکیب فعال اکسیژنی مضر برای سلول است. از دیدگاه صنایع غذایی و مدیریت پس از برداشت، تغییرات در فعالیت POD به طور مستقیم با فرایندهای تخریب کیفیت میوه‌هایی مانند آلو مرتبط است. افزایش ناگهانی یا تغییرات در الگوی فعالیت این آنزیم اغلب به عنوان یک شاخص بیوشیمیایی مهم برای ارزیابی وضعیت فیزیولوژیکی، تعیین زمان بهینه برداشت و پیش‌بینی عمر انبارمانی میوه در نظر گرفته می‌شود. همچنین، فعالیت آنزیمی بالا می‌تواند منجر به بروز سریع‌تر و شدیدتر قهوه‌ای شدن آنزیمی شود که عامل اصلی کاهش کیفیت ظاهری و ارزش تجاری میوه‌های تازه است [۷].

کاهش کیفیت محصولات کشاورزی در دوره پس‌از برداشت اغلب به دلیل واکنش‌های بیوشیمیایی مختلف از جمله قهوه‌ای شدن داخلی رخ می‌دهد. یکی از مهم‌ترین عوامل بروز این قهوه‌ای شدن، فعالیت آنزیم POD است که در فرآیندهای اکسیداسیونی نقش کلیدی ایفا می‌کند. پایش دقیق و به‌هنگام این تغییرات آنزیمی برای حفظ کیفیت، افزایش عمر انباری و کاهش ضایعات محصولات تازه اهمیت بسزایی دارد [۵].

روش‌های سنتی اندازه‌گیری فعالیت آنزیمی مبتنی بر آزمون‌های شیمیایی و استخراج عصاره‌های بافتی هستند که علاوه بر زمان‌بر بودن و نیاز به آماده‌سازی نمونه، باعث تخریب محصول می‌شوند و دقت نتایج آن‌ها نیز به عوامل مختلفی وابسته است. به همین دلیل، توسعه فناوری‌های نوین و غیرمخرب برای شناسایی سریع تغییرات ناشی از فعالیت آنزیم‌های مسئول قهوه‌ای‌شدن، به‌ویژه POD، مورد توجه محققان قرار گرفته است. برای ارزیابی غیرمخرب کیفیت میوه‌ها روش‌های تصویربرداری و طیف‌سنجی (خصوصاً در حوزه مرئی و نزدیک‌فروسرخ) به کار می‌روند. طیف‌سنجی در بازه VIS/NIR ابزاری سریع و غیرتهاجمی است که می‌تواند اطلاعاتی از خواص داخلی میوه (محتوای قند، اسید، رنگدانه‌ها و تغییرات فیزیولوژیک) فراهم کند. استفاده از این ابزارها مزایایی چون سرعت بالا، تکرارپذیری مناسب، عدم آسیب به میوه و هزینه‌ی پایین را برای کاربران (کشاورزان و واحدهای فرآوری) فراهم می‌آورد [۸].

تعدادی از پژوهش‌هایی که در رابطه با ارزیابی غیرمخرب فعالیت آنزیم POD و سایر پارامترهای کیفی در میوه‌ها و سبزیجات انجام شده می‌توان به مواردی از جمله تشخیص سریع فعالیت پراکسیداز در برگ‌های گوجه‌فرنگی آلوده به *Botrytis cinerea* [۹]، ارزیابی غیرمخرب فعالیت آنزیم‌های POD و پلی‌فنل اکسیداز در ارقام مختلف فلفل دلمه‌ای [۵]، پیش‌بینی کیفیت میوه آلو ماریان بدون تأثیر وضعیت محیطی توسط طیف‌سنجی مادون قرمز [۱۰]، پیش‌بینی فعالیت POD با استفاده از تصویربرداری فراطیفی مادون قرمز نزدیک در میوه سیب رد دلشیز در طول زمان نگهداری [۴] اشاره کرد. با توجه به این قابلیت‌ها، استفاده از طیف‌سنجی نزدیک‌به‌فروسرخ به منظور شناسایی و پیش‌بینی فعالیت آنزیم POD می‌تواند رویکردی نوین و کارآمد در پایش کیفیت محصولات کشاورزی باشد و به ارتقای مدیریت پس‌از برداشت و کاهش ضایعات کمک شایانی کند. هدف اصلی این پژوهش، توسعه یک روش غیرمخرب با استفاده از طیف‌سنجی فروسرخ نزدیک (NIR) برای تعیین سطح فعالیت آنزیم POD در ارقام مختلف آلو است. در این راستا، از روش‌های یادگیری ماشین برای مدل‌سازی و تحلیل داده‌های طیفی حاصل از طیف‌سنجی NIR استفاده خواهیم کرد.

۲- مواد و روش‌ها

در این پژوهش، دو رقم آلو، شامل خرمایی و خونی، از باغات محلی تهیه شدند. از هر رقم، تعداد ۸۰ نمونه با ویژگی‌های ظاهری یکنواخت (اندازه، شکل و رنگ) و بدون هیچ‌گونه آسیب مکانیکی یا نشانه‌ای از پوسیدگی قارچی برای انجام آزمایش‌ها انتخاب گردیدند (شکل ۱). این انتخاب با هدف اطمینان از همگنی نمونه‌ها و کاهش متغیرهای ناخواسته صورت گرفت.

آلو خرمایی، که با نام قره آلو نیز شناخته می‌شود و به دلیل ویژگی‌های بافتی و طعم مطلوب، به طور گسترده برای تولید خشکبار آلو (صادراتی و داخلی) فرآوری می‌شود و از نظر اقتصادی اهمیت بالایی دارد. در مقابل، آلو خونی، که با نام‌های توسرخ، خاکی، خاکی خونی و آلو قرمز نیز معروف است، دارای میوه‌ای نسبتاً درشت با رنگ قرمز تیره تا بنفش است و طعمی شیرین و بسیار خوش‌خوراک دارد. کیفیت میوه این رقم در مقایسه با بسیاری از ارقام آلو مطلوب‌تر است و به همین دلیل جایگاه مناسبی در بازار میوه تازه ایران دارد و یک رقم پرسود تجاری محسوب می‌شود. انتخاب این دو رقم آلو به دلیل تفاوت‌های فیزیولوژیکی و کیفیتی بارز آن‌ها انجام شد تا امکان بررسی گسترده‌تر و جامع‌تر میزان فعالیت POD میوه آلو فراهم گردد. پیش از انجام اندازه‌گیری‌ها، نمونه‌ها به مدت دو ساعت در دمای ۲۵ درجه سلسیوس (دمای اتاق) نگهداری شدند تا دمای آن‌ها با دمای محیط آزمایشگاه به تعادل برسد. پس از طیف‌سنجی نمونه‌ها با استفاده از دستگاه طیف‌سنج، بلافاصله اندازه‌گیری مخرب میزان فعالیت POD آن‌ها انجام شد.

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

Figure 1. Samples used (a) Khouni plum and (b) Khormaei plum

۲-۱-۲- آزمون غیرمخرب

۲-۱-۲- طیف‌سنجی Vis/NIR

اندازه‌گیری‌های غیرمخرب با استفاده از طیف‌سنج Vis/NIR (مرئی/نزدیک مادون قرمز) صورت پذیرفت. این طیف‌سنج، مدل PS-100 از شرکت (Apogee Instruments, INC., Logan, Utah, USA)، مجهز به یک آشکارساز CCD با ۲۰۴۸ پیکسل و قدرت تفکیک ۱ نانومتر بود. به عنوان منبع نور از یک لامپ هالوژن-تنگستن استفاده شد و طیف‌ها در محدوده طول‌موجی ۳۵۰ تا ۱۱۰۰ نانومتر جمع‌آوری شدند. شکل ۲ چگونگی نمونه‌برداری با سامانه طیف‌سنج را نشان می‌دهد. طیف‌سنج از طریق کابل USB به رایانه متصل شد و نرم‌افزار SpectraWiz برای جمع‌آوری و ذخیره‌سازی طیف‌ها به کار رفت. پیش از آغاز طیف‌سنجی نمونه‌ها، جهت کالیبراسیون و نرمال‌سازی داده‌ها، طیف‌های مرجع تیره و سفید ثبت شدند. برای ثبت طیف تیره، منبع نور خاموش گردید. سپس، با روشن کردن منبع نور، از یک دیسک تفلون استاندارد به عنوان مرجع برای ثبت طیف سفید استفاده شد [۱۱]. برای هر نمونه آلو، از چهار نقطه مختلف طیف‌سنجی انجام شد تا میانگین‌گیری و کاهش خطای اندازه‌گیری میسر گردد. داده‌های طیفی حاصل از این چهار نقطه، پس از میانگین‌گیری، به عنوان طیف جذب نهایی هر نمونه مورد استفاده قرار گرفتند. در نهایت، ۸۰ داده طیفی برای هر رقم آلو (در مجموع ۱۶۰ داده طیفی) برای تحلیل‌های بعدی استخراج گردیدند.

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

Figure 2. View of the VIS/NIR spectroscopy system used for data collection

۲-۲- آزمون مخرب

۲-۲-۱- بررسی خواص آنزیمی

برای بررسی خواص آنزیمی، POD در نمونه‌های آلو اندازه‌گیری شد. این بخش شامل دو مرحله اصلی، یعنی تهیه عصاره آنزیمی و اندازه‌گیری POD است.

۲-۲-۱-۱- تهیه عصاره آنزیمی

برای تهیه عصاره آنزیمی، ابتدا ۱۰ گرم از پالپ میوه به دقت درون یک مخلوط‌کن ریخته و کاملاً هموژن شد. سپس، پالپ هموژن‌شده به ۲۰ میلی‌لیتر محلول استخراج آنزیمی اضافه گردید. این محلول استخراج حاوی بافر سدیم فسفات ۰/۴ مولار با pH=۶/۵ تهیه شده از دی‌سدیم هیدروژن فسفات (Na_2HPO_4) و سدیم دی‌هیدروژن فسفات (NaH_2PO_4) از شرکت (Merck, Germany)، چهار درصد (وزنی/حجمی) پلی‌ونیل‌پیرولیدون (PVP) (Merck, Germany) و یک درصد (حجمی/حجمی) ترایتون ایکس-۱۰۰ (Triton X-100) (Merck, Germany) بود. مخلوط حاصل با استفاده از ورتکس (Labtron LS-100, Iran) به طور کامل مخلوط شد تا از توزیع یکنواخت اجزا اطمینان حاصل شود. در ادامه، محلول مخلوط شده در دمای ۴ درجه سلسیوس به مدت ۱۰ دقیقه با سرعت ۴۰۰۰ دور در دقیقه توسط سانتریفیوژ (LISA 2.5L centrifuge AFI, France) سانتریفیوژ گردید. ماده شناور (سوپرناتانت) حاصل از این مرحله، به عنوان عصاره آنزیمی برای بررسی فعالیت آنزیمی مورد استفاده قرار گرفت [۱۲].

۲-۱-۲-۲- فعالیت آنزیم پراکسیداز (POD)

POD با استفاده از روش رنگ‌سنجی و اندازه‌گیری تغییرات جذب در زمان انجام شد. برای این منظور، ۵۰۰ میکرولیتر از عصاره آنزیمی با ۱ میلی‌لیتر بافر سدیم فسفات ۰/۰۵ مولار با pH=۶/۵ مخلوط گردید. برای آغاز واکنش آنزیمی، ۱ میلی‌لیتر از بافر سدیم فسفات ۰/۰۵ مولار (pH=۶/۵) حاوی ۵۰ میکرولیتر هیدروژن پراکسید (H₂O₂) ۱/۵ درصد (حجمی/حجمی) (Merck, Germany) و ۱ میلی‌لیتر پی-فنیل‌دی‌آمین (p-phenylenediamine) یک درصد (وزنی/حجمی) (Merck, Germany) به مخلوط واکنش اضافه شد. یک نمونه شاهد (کنترل) نیز به روش مشابه تهیه گردید، با این تفاوت که به جای عصاره آنزیمی، از آب استفاده شد. میزان جذب در طول موج ۴۸۵ نانومتر به مدت ۱۰ دقیقه و در دمای ۲۵ درجه سلسیوس به صورت سینتیک (تغییرات جذب بر حسب زمان) با استفاده از اسپکتروفتومتر نانودراپ (NanoDrop™ OneC, Thermo Fisher Scientific, USA) اندازه‌گیری شد. فعالیت آنزیمی به صورت تغییرات جذب در دقیقه در هر گرم از نمونه بیان گردید [۱۳ و ۱۴].

۲-۳- تجزیه و تحلیل داده‌ها

۲-۳-۱- پیش‌پردازش داده‌های طیفی

در ابتدا، داده‌های طیفی خام نمونه‌ها که در محدوده ۳۵۰ تا ۱۱۰۰ نانومتر (شامل بخش‌های مرئی و نزدیک مادون قرمز) جمع‌آوری شده بودند، مورد بررسی قرار گرفتند. به منظور افزایش کیفیت و پایداری داده‌ها برای تحلیل‌های بعدی، بخش ابتدایی طیف‌ها، یعنی محدوده ۳۵۰ تا حدود ۵۱۰ نانومتر، حذف شد. این حذف به دلیل وجود نویز قابل توجه ناشی از کاهش نسبت سیگنال به نویز و محدودیت حساسیت آشکارساز در این بازه صورت گرفت. این فرآیند پاک‌سازی داده‌ها به‌طور یکسان برای هر دو رقم آلو (خرمایی و خونی) اعمال شد.

پس از اصلاح اولیه، داده‌های طیفی به نرم‌افزار MATLAB نسخه ۲۰۲۲a منتقل شدند. برای بررسی تأثیر روش‌های مختلف پیش‌پردازش بر عملکرد مدل‌های پیش‌بینی، چندین روش پیش‌پردازش بر روی داده‌های طیفی اعمال گردید. این روش‌ها شامل: نرمال‌سازی^۱، استانداردسازی بردار نرمال^۲ (SNV)، تصحیح پراکندگی چندگانه^۳ (MSC)، فیلتر میانگین‌گیری متحرک^۴، فیلتر گوسی^۵، فیلتر میانه^۶، حذف روند^۷ و میانگین مرکزی کردن^۸ بودند. هدف از به کارگیری این روش‌ها، کاهش اثرات نویز، تصحیح تغییرات ناشی از پراکندگی نور و افزایش وضوح اطلاعات طیفی مرتبط با فعالیت آنزیمی بود.

1 . Normalize
2 . Standard Normal Variate
3 . Multiplicative Scatter Correction
4 . Moving Average
5 . Gaussian Filter
6 . Median Filter
7 . Detrend
8 . Mean Centering

۲-۳-۲- مدل سازی داده های طیفی با استفاده از رگرسیون حداقل مربعات جزئی^۱ (PLSR) و روش ماشین بردار پشتیبان^۲ (SVM)

پس از مرحله پیش پردازش، داده های طیفی برای مدل سازی پیش بینی فعالیت POD به نرم افزار MATLAB نسخه ۲۰۲۲a منتقل شدند. در این مطالعه، دو الگوریتم قدرتمند شیمی سنجی شامل PLSR و SVM برای ساخت مدل ها مورد استفاده قرار گرفتند. الگوریتم PLSR به دلیل توانایی برجسته اش در مدیریت داده های با ابعاد بالا و وجود همبستگی داخلی بین متغیرها (که در داده های طیفی رایج است) انتخاب شد. این الگوریتم قادر است روابط پنهان بین طیف های چندمتغیره و متغیر پاسخ (میزان فعالیت POD) را استخراج کند. در این پژوهش، مدل PLSR با حداکثر ۱۰ مؤلفه پنهان مورد بررسی قرار گرفت.

در کنار PLSR، مدل SVM نیز به عنوان یک الگوریتم یادگیری ماشین قدرتمند، برای مدل سازی به کار رفت. در مدل SVM، تأثیر سه نوع کرنل (تابع هسته) مختلف شامل کرنل خطی^۳، کرنل چندجمله ای^۴ و کرنل پایه شعاعی^۵ (RBF) بر عملکرد مدل مورد ارزیابی و مقایسه قرار گرفت. انتخاب نوع کرنل مناسب برای SVM نقش کلیدی در توانایی مدل برای شناسایی الگوهای غیرخطی در داده ها دارد.

برای افزایش اعتبار نتایج و کاهش تأثیر نوسانات ناشی از تقسیم بندی تصادفی داده ها، فرآیند تقسیم داده ها به مجموعه های آموزش^۶، اعتبارسنجی^۷ و آزمون^۸ به طور مستقل ۲۰۰ بار تکرار شد. در هر تکرار، داده های طیفی خام (قبل از اعمال هرگونه پیش پردازش) به صورت تصادفی با نسبت های ۶۰٪ برای آموزش، ۲۰٪ برای اعتبارسنجی و ۲۰٪ برای آزمون تفکیک شدند. هدف از تقسیم به مجموعه اعتبارسنجی، بهینه سازی پارامترهای مدل (مانند تعداد مؤلفه های PLSR یا پارامترهای کرنل در SVM) و جلوگیری از بیش برآزش^۹ بود، در حالی که مجموعه آزمون برای ارزیابی نهایی عملکرد مدل بر روی داده های جدید و ندیده شده استفاده شد.

پس از انجام ۲۰۰ تکرار، بهترین تقسیم بندی داده ها بر اساس معیارهای آماری عملکرد مدل ها (شامل ضریب تعیین^{۱۰} (R^2)، ریشه میانگین مربعات خطا^{۱۱} (RMSE) و نسبت عملکرد به انحراف^{۱۲} (RPD) در هر سه مرحله آموزش، اعتبارسنجی و آزمون) انتخاب گردید. علاوه بر این، بهترین تعداد مؤلفه های پنهان برای PLSR نیز در این مرحله تعیین شد. به منظور اطمینان از مقایسه دقیق و مستقل اثر هر روش پیش پردازش، تمامی روش های پیش پردازش (که در بخش قبل توضیح داده شد) صرفاً بر روی همان تقسیم بندی منتخب و ثابت شده از داده ها اعمال شدند. این رویکرد تضمین می کند که هرگونه بهبود یا کاهش در عملکرد مدل ها صرفاً ناشی از تأثیر پیش پردازش بوده و به دلیل تغییرات ناشی از تقسیم بندی تصادفی نمونه ها نمی باشد. تمامی تحلیل ها، اعمال پیش پردازش ها و فرآیندهای مدل سازی در محیط MATLAB نسخه ۲۰۲۲a انجام شد که به عنوان یکی از نرم افزارهای معتبر و پرکاربرد در پردازش داده های طیفی و مدل سازی شیمی سنجی شناخته می شود [۱۵].

1 . Partial Least Squares Regression

2 . Support Vector Machine

3 . Linear

4 . Polynomial

5 . Radial Basis Function

6 . Training

7 . Validation

8 . Test

9 . Overfitting

10 . Coefficient of Determination

11 . Root Mean Square Error

12 . Ratio of Performance to Deviation

معیارهای ارزیابی عملکرد مدل عملکرد مدل‌های پیش‌بینی توسعه‌یافته با استفاده از معیارهای آماری استاندارد شامل R^2 ، RMSE و RPD ارزیابی شد. R^2 از رابطه (۱) محاسبه می‌شود، نشان‌دهنده میزان تبیین تغییرات متغیر پاسخ توسط مدل است؛ مقادیر نزدیک به ۱ نشان‌دهنده برازش بهتر مدل هستند. RMSE میزان دقت پیش‌بینی مدل را بر حسب واحد متغیر پاسخ بیان می‌کند (رابطه ۲)؛ مقادیر کمتر RMSE به معنای خطای کمتر و دقت بالاتر مدل است. در نهایت، RPD، که از رابطه (۳) محاسبه می‌شود توانایی پیش‌بینی کمی مدل را نشان می‌دهد. در این روابط، d_i در این معادلات، p_i مقدار POD پیش‌بینی شده برای مؤلفه- i ام و d_i مقدار POD 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 POD نمونه‌ها و SD انحراف استاندارد داده‌های پیش‌بینی شده POD در مجموعه داده مربوطه است. RPD همچنین با استفاده از سیستم طبقه‌بندی توصیه‌شده توسط Chang و همکاران [۱۶] تفسیر شد: مدلهایی با RPD کمتر از ۱ بسیار ضعیف تلقی شده و توصیه نمی‌شوند. RPD بین ۱ تا ۱/۴ نشان‌دهنده یک مدل ضعیف است، در حالی که مقادیر بین ۱/۴ تا ۱/۸ برای ارزیابی و پیش‌بینی مناسب هستند. مدلهایی با RPD بین ۱/۸ تا ۲ خوب ارزیابی شده و برای پیش‌بینی‌های کمی قابل قبول می‌باشند. RPD بین ۲ تا ۲/۵ نشان‌دهنده یک مدل بسیار خوب، و RPD بالای ۲/۵ نشانگر عملکرد عالی مدل است.

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

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

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

۲-۳-۳- انتخاب طول موج مؤثر و مدل‌سازی مجدد

به‌منظور کاهش ابعاد داده‌های طیفی و شناسایی طول‌موج‌های مؤثر که بیشترین ارتباط را با فعالیت POD دارند، از تلفیق SVM با شش الگوریتم بهینه‌سازی فراابتکاری استفاده شد. این رویکرد با هدف حذف نویز و اطلاعات اضافی، افزایش کارایی محاسباتی، و در نهایت بهبود دقت مدل‌های پیش‌بینی به کار گرفته شد. الگوریتم‌های بهینه‌سازی مورد استفاده شامل مسابقه رقابتی جهانی (WCC)، بهینه‌سازی ازدحام ذرات (PSO)، الگوریتم بهینه‌سازی انتقال حرارت (HTS)، الگوریتم بهینه‌سازی جنگل (FOA)، جستجوی ارگانیسم‌های همزیست گسسته (DSOS) و بهینه‌سازی فاخته (CUK) بودند [۱۷]. پارامترهای تنظیمی خاص برای هر یک از این الگوریتم‌ها که نقش کلیدی در عملکرد آن‌ها داشتند، در جدول ۱ ارائه شده است.

در این فرآیند، هر یک از الگوریتم‌های بهینه‌سازی به‌طور مستقل مسئول شناسایی مجموعه‌ای از طول‌موج‌های بهینه بودند. عملکرد نتایج حاصل از هر الگوریتم با ارزیابی میانگین R^2 و RMSE سنجیده شد تا الگوریتمی انتخاب شود که بالاترین دقت پیش‌بینی (بیشترین R^2) و کمترین خطای ممکن (کمترین RMSE) را ارائه دهد. پس از تعیین طول‌موج‌های مؤثر و کاهش موفقیت‌آمیز حجم داده‌ها، این داده‌های طیفی کاهش‌یافته مجدداً با استفاده از هر دو روش PLSR و SVM مدل‌سازی شدند. هدف از این مدل‌سازی مجدد، ارزیابی بهبود عملکرد مدل‌ها پس از انتخاب طول‌موج‌های کلیدی و تأیید اهمیت کاهش ابعاد داده‌ها در ساخت مدل‌های کارآمدتر بود. تمامی مراحل این تحلیل، شامل اعمال الگوریتم‌های بهینه‌سازی و مدل‌سازی مجدد، در محیط نرم‌افزاری MATLAB نسخه ۲۰۲۲a انجام پذیرفت.

جدول ۱. مشخصات پارامترهای تنظیمی هر یک از الگوریتم‌های انتخاب طول موج مؤثر

Table 1. Specifications of the tuning parameters for each effective wavelength selection algorithms

۳- نتایج و بحث

۳-۱- ویژگی‌های طیفی و تحلیل داده‌های پراکسیداز

همان‌طور که در شکل ۳ (نمودار پراکندگی طیف‌های جذب نمونه‌های دو رقم آلو خرمایی (الف) و خونی (ب) در محدوده طول موج ۵۱۰ تا ۱۱۰۰ نانومتر) مشاهده می‌شود، الگوهای طیفی وسیعی از تنوع را نشان می‌دهند. این تنوع حاصل از تفاوت‌های بیوشیمیایی و ساختاری در نمونه‌ها، از جمله مراحل مختلف رسیدگی و اختلافات جزئی در رنگ یا بافت هر میوه است. این نوسانات طیفی، مدل‌سازی مستقیم فعالیت POD را به یک چالش تبدیل می‌کند، اما همزمان، وجود اطلاعات حیاتی در طیف را برای تحلیل‌های شیمی‌سنجی تأیید می‌کند.

در هر دو رقم، ناحیه مرئی طیف (تقریباً ۵۱۰ تا ۷۰۰ نانومتر) دارای میزان جذب بالایی است. این جذب عمدتاً تحت تأثیر رنگ‌دانه‌های مختلف در پوست و گوشت آلو قرار دارد. پس از طول موج حدود ۶۷۰ نانومتر، کاهش مشخصی در میزان جذب مشاهده می‌شود که نشان‌دهنده پایان ناحیه جذب کلروفیل 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 در مولکول‌های آب نسبت داده می‌شود [18] و با محتوای آب نمونه‌ها مرتبط است. تغییراتی در جذب در حدود ۱۰۳۰ نانومتر نیز مشاهده می‌شود که می‌تواند با ارتعاشات پیوندهای C-H در قندها و سایر ترکیبات آلی مرتبط باشد [19]. این ویژگی‌های طیفی، قابلیت طیف‌سنجی NIR را در شناسایی و اندازه‌گیری پارامترهای بیوشیمیایی داخلی میوه‌ها، از جمله محتوای قند و رطوبت، به‌خوبی برجسته می‌کند. این ویژگی‌ها به صورت مستقیم فعالیت POD را منعکس نمی‌کنند، اما تغییرات آن‌ها به طور غیرمستقیم با فرایندهای بیوشیمیایی و متابولیسم میوه از جمله فعالیت آنزیمی مرتبط هستند. بنابراین، با استفاده از الگوریتم‌های یادگیری ماشین مانند PLSR و SVM، می‌توان همبستگی‌های پیچیده بین این ویژگی‌های طیفی و فعالیت POD را مدل‌سازی و با دقت بالایی پیش‌بینی کرد.

جدول ۲، آمار توصیفی مربوط به فعالیت POD در ارقام خونی و خرمایی آلو را ارائه می‌دهد. این آمار شامل میانگین، انحراف معیار، مقادیر حداقل و حداکثر است. تحلیل این داده‌ها به درک اولیه از دامنه و توزیع فعالیت آنزیمی در نمونه‌ها کمک می‌کند. برای رقم خونی، فعالیت آنزیم POD در دامنه‌ای از ۰/۰۸۰۸ تا ۰/۷۲۴۴ همچنین برای رقم خرمایی نیز این دامنه از ۰/۰۷۸۷ تا ۰/۷۳۱۴ متغیر است. این مقادیر حداکثر و حداقل نشان می‌دهد که نمونه‌های مورد مطالعه در هر دو رقم، طیف وسیعی از فعالیت آنزیمی از مقادیر بسیار پایین تا مقادیر بسیار بالا را در بر می‌گیرند. این تنوع گسترده در داده‌ها برای یک مدل‌سازی موفق بسیار حیاتی است، زیرا به مدل امکان می‌دهد تا ارتباط بین طیف‌ها و مقادیر مختلف فعالیت آنزیمی را به خوبی یاد بگیرد. اگر دامنه داده‌ها بسیار محدود بود، مدل قادر به پیش‌بینی دقیق مقادیر خارج از آن دامنه نبود.

میانگین فعالیت آنزیم POD برای هر دو رقم بسیار نزدیک به هم است (۰/۴۲۳۶ برای خونی و ۰/۴۳۳۴ برای خرمایی). این نشان می‌دهد که به طور متوسط، تفاوت معناداری در فعالیت آنزیمی بین دو رقم وجود ندارد. انحراف معیار نیز برای هر دو رقم نزدیک

به هم و نسبتاً بالا است (۰/۱۸۶۱ برای خونی و ۰/۲۰۲۹ برای خرمایی). این انحراف معیار بالا، گویای پراکندگی قابل توجه داده‌ها حول میانگین است که مجدداً نشان‌دهنده تنوع بالای نمونه‌ها در هر رقم است.

شکل ۳. طیف‌های جذبی نمونه‌های آلو خرمایی (الف) و آلو خونی (ب)

Figure 3. Absorption spectra of Khormaei plum (a) and Khouni plum (b) samples

جدول ۲. آمار توصیفی میزان فعالیت آنزیم پراکسیداز (absorbance/min.g) ارقام مختلف آلو

Table 2. Descriptive statistics of peroxidase enzyme activity (absorbance/min.g) for different plum varieties

۲-۳- نتایج مدل‌سازی داده‌های طیفی بدون کاهش ابعاد و در محدوده کامل

این بخش از پژوهش بر ارزیابی عملکرد مدل‌های پیش‌بینی با استفاده از الگوریتم‌های رگرسیونی PLSR و SVM و با در نظر گرفتن انواع روش‌های پیش‌پردازش طیفی تمرکز دارد. نتایج به‌دست‌آمده، اطلاعات ارزشمندی درباره‌ی تأثیر نوع مدل، روش پیش‌پردازش و تفاوت بین ارقام مختلف آلو ارائه می‌دهد.

۱-۲-۳- عملکرد مدل PLSR

نتایج جدول ۳ نشان می‌دهد که مدل PLSR برای رقم خرمایی، با وجود R^2 بالا در بخش آموزش (R^2 تا ۰/۹۵۰۶)، عملکرد نسبتاً ضعیفی در بخش آزمون دارد. بهترین عملکرد با پیش‌پردازش Detrend و SNV به دست آمده که به ترتیب RPD برابر با ۱/۶۴ و ۱/۶ را کسب کرده‌اند. این مقادیر، مدل را در طبقه‌بندی مناسب (RPD بین ۱/۴ تا ۱/۸) قرار می‌دهد. این یعنی مدل قادر به انجام پیش‌بینی‌های تقریبی برای ارزیابی کیفی و نه کمی دقیق است.

در مقابل، همان‌طور که در جدول ۴ مشاهده می‌شود، مدل PLSR برای رقم خونی به طور قابل توجهی بهتر عمل کرده است. پیش‌پردازش نرمال‌سازی با کسب $RPD=2/16$ و $R^2=0/7722$ در مرحله آزمون، بهترین نتیجه را رقم زده است. این مقدار RPD نشان‌دهنده یک مدل بسیار خوب (RPD بین ۲ تا ۲/۵) است که برای پیش‌بینی‌های کمی و دقیق قابل استفاده است. نتایج سایر پیش‌پردازش‌ها مانند بدون پیش‌پردازش و فیلتر گوسی نیز به RPD بالای ۲ دست یافته‌اند که نشان‌دهنده ثبات و قدرت بالای مدل‌سازی PLSR برای این رقم است. این تفاوت عملکرد بین دو رقم احتمالاً به دلیل ویژگی‌های فیزیکی و شیمیایی متفاوت آن‌هاست که باعث شده اطلاعات طیفی رقم خونی برای مدل PLSR قابل‌تشخیص‌تر باشد.

۲-۲-۳- عملکرد مدل SVM

بررسی نتایج مدل‌سازی SVM در جداول ۵ و ۶، نشان‌دهنده برتری قابل توجه این مدل نسبت به PLSR است. جدول ۵ نتایج مربوط به مدل‌سازی به روش SVM برای رقم خرمایی را نشان می‌دهد بهترین عملکرد مدل SVM برای رقم خرمایی با هسته rbf و پیش‌پردازش فیلتر میانه به دست آمده است. این مدل با کسب $R^2=0/849$ ، $RMSE=0/072$ و $RPD=2/656$ در مرحله آزمون، یک مدل عالی ($RPD > 2/5$) محسوب می‌شود. این نتیجه نشان می‌دهد که مدل SVM با توانایی خود در شناسایی روابط غیرخطی، توانسته است عملکرد ضعیف‌تر مدل خطی PLSR را جبران کرده و به دقت بسیار بالایی دست یابد.

جدول ۶ مقایسه عملکرد مدل SVM با هسته‌ها و پیش‌پردازش‌های مختلف برای پیش‌بینی فعالیت POD در رقم خونی آلو را نشان می‌دهد. نتایج این جدول به وضوح نشان می‌دهد که مدل SVM در ترکیب با پیش‌پردازش‌های مناسب، عملکرد بسیار خوبی در

پیش‌بینی دقیق این پارامتر دارد. در این جدول، مقادیر RPD در مجموعه داده آزمون، به عنوان معیار اصلی ارزیابی، در بسیاری از مدل‌ها بالای ۲ است که طبقه‌بندی مدل‌های بسیار خوب تا عالی را شامل می‌شود. این نشان‌دهنده توانایی بالای طیف‌سنجی VIS/NIR برای اندازه‌گیری این ویژگی در رقم خونی است.

بالاترین دقت پیش‌بینی، با RPD برابر با ۳/۶۴۲، با استفاده از هسته چندجمله‌ای و پیش‌پردازش مرکز‌کردن میانگین به دست آمده است. این نتیجه، نشان‌دهنده یک مدل عالی برای پیش‌بینی کمی دقیق است. این مدل، در مقایسه با تمام مدل‌های PLSR و SVM که تا کنون بررسی کرده‌ایم، بهترین عملکرد را داشته است. همان‌طور که انتظار می‌رفت، هسته‌های غیرخطی مانند rbf و چندجمله‌ای به طور کلی عملکرد بهتری نسبت به هسته خطی از خود نشان داده‌اند. این امر تأکید می‌کند که رابطه بین طیف‌های جذبی و فعالیت POD در رقم خونی، یک رابطه پیچیده و غیرخطی است که توسط الگوریتم‌های خطی به خوبی مدل‌سازی نمی‌شود. همانند نتایج PLSR، انتخاب روش پیش‌پردازش در مدل SVM نیز حیاتی بوده است. در حالی که پیش‌پردازش‌هایی مانند SNV و MSC عملکرد مدل را کاهش داده‌اند، روش‌هایی مانند میانگین متحرک، مرکز‌کردن میانگین، و نرمال‌سازی به طور قابل توجهی به بهبود دقت مدل کمک کرده‌اند. این موضوع نشان می‌دهد که حذف نویزهای تصادفی و اثرات پراکندگی، شرط لازم برای دستیابی به یک مدل پیش‌بینی قوی است.

نتایج به وضوح نشان می‌دهد که روش‌های پیش‌پردازش مانند فیلتر میانه و مرکز‌کردن میانگین، با حذف نویز و اثرات پراکندگی، نقش حیاتی در بهبود عملکرد مدل‌های پیش‌بینی ایفا می‌کنند. مقایسه کلی نتایج نشان می‌دهد که الگوریتم SVM به دلیل توانایی خود در مدل‌سازی روابط غیرخطی، عملکردی به مراتب بهتر از PLSR ارائه می‌دهد. این موضوع اهمیت استفاده از الگوریتم‌های غیرخطی را در تحلیل داده‌های طیفی پیچیده تأیید می‌کند. در هر دو نوع مدل‌سازی، مدل‌های توسعه‌یافته برای رقم خونی به طور کلی دقت بالاتری را نسبت به رقم خرمایی نشان دادند. این تفاوت می‌تواند ناشی از تفاوت در ترکیب شیمیایی یا ساختار فیزیکی دو رقم باشد که بر روی جذب و پراکندگی نور تأثیر می‌گذارد.

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

Table 3. PLSR modeling results for the Khormaei plum variety with various spectral preprocessing methods (5 Component)

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

Table 4. PLSR modeling results for the Khouni plum variety with various spectral preprocessing methods (8 Component)

جدول ۵. مقایسه عملکرد مدل SVM-R رقم خرمایی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 5. Performance comparison of the SVM-R model for the Khormaei plum variety with different kernels and preprocessing methods

جدول ۶. مقایسه عملکرد مدل SVM-R رقم خونی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 6. Performance comparison of the SVM-R model for the Khouni plum variety with different kernels and preprocessing methods

۳-۳- انتخاب طول موج موثر

این بخش از پژوهش بر کاهش ابعاد داده‌های طیفی با استفاده از الگوریتم‌های فراابتکاری تمرکز دارد. هدف از این کار، شناسایی حداقل تعداد طول موج‌های کلیدی با بیشترین اطلاعات برای پیش‌بینی فعالیت POD و در عین حال، کاهش زمان محاسبات و

پیچیدگی مدل می‌باشد. شکل‌های ۴ و ۵ به صورت نموداری، عملکرد مدل SVM را پس از انتخاب طول موج‌های موثر توسط الگوریتم‌های مختلف، به ترتیب برای رقم خرمای و خونی نشان می‌دهند. یکی از معیارهای مهم در ارزیابی الگوریتم‌های بهینه‌سازی، سرعت آن‌ها در یافتن پاسخ بهینه است. در این پژوهش، زمان اجرای هر الگوریتم برای یک بار اجرا به شرح زیر است: برای رقم خرمایی، الگوریتم WCC به ۴۳ ثانیه، PSO به ۲/۸۸ ثانیه، HTS به ۲۵ ثانیه، FOA به ۳۲ ثانیه، DSOS به ۴۶ ثانیه و CUK به ۶۱ ثانیه زمان نیاز داشته است. برای رقم خونی، این مقادیر به ترتیب ۴۳، ۲/۸۵، ۲۴، ۳۲، ۴۲ و ۶۱ ثانیه بوده‌اند. بر اساس این داده‌ها، الگوریتم PSO با زمان اجرای حدود ۲/۸۸ ثانیه برای رقم خرمایی و ۲/۸۵ ثانیه برای رقم خونی، به عنوان سریع‌ترین الگوریتم در میان تمام روش‌های مورد بررسی شناسایی شد. این سرعت بالا، یک مزیت بسیار مهم در کاربردهای عملی است که نیاز به تحلیل سریع و بلادرنگ نمونه‌ها دارند. این یافته نشان می‌دهد که PSO به طور مؤثری می‌تواند در زمان بسیار کوتاهی، طول موج‌های بهینه را برای کاهش ابعاد داده‌ها پیدا کند. با توجه به شکل ۴ که میانگین همبستگی و RMSE براساس مدل SVM و الگوریتم‌های فراابتکاری را برای رقم خرمایی نشان می‌دهد از نظر دقت پیش‌بینی، PSO موفق به دستیابی به کمترین میانگین RMSE (برابر با ۰/۰۰۲۰۲) و بالاترین همبستگی (برابر با ۰/۹۲۱۵) شده است. این مقادیر، نشان‌دهنده توانایی بالای الگوریتم در شناسایی طول موج‌هایی است که قوی‌ترین همبستگی را با فعالیت آنزیمی دارند. با کاهش داده‌ها از ۱۱۹۵ طول موج به تنها ۱۵ طول موج، PSO توانسته است نویز و اطلاعات غیرضروری را حذف کرده و دقت مدل را به حداکثر برساند.

برای رقم خونی نیز، الگوریتم PSO با انتخاب ۱۵ طول موج، عملکردی قابل قبول و رقابتی را به نمایش می‌گذارد. اگرچه الگوریتم FOA از نظر تعداد طول موج‌های انتخابی (تنها ۶ طول موج) بهینه بوده، اما با در نظر گرفتن سرعت اجرا و دقت نهایی، انتخاب PSO به عنوان یک رویکرد یکپارچه برای هر دو رقم منطقی است. با توجه به شکل ۵ که عملکرد مدل SVM همراه با الگوریتم‌های مختلف برای نمونه‌های رقم خونی را نشان می‌دهد الگوریتم PSO توانسته با میانگین همگرایی RMSE برابر با ۰/۰۰۵۷۶۴ و همبستگی ۰/۶۱۱۸ به نتایج بسیار خوبی دست یابد. این نتایج نشان‌دهنده کارایی الگوریتم در مدل‌سازی دقیق ویژگی‌های طیفی پیچیده این رقم و ارتباط آن‌ها با فعالیت آنزیمی است.

در مجموع، انتخاب الگوریتم PSO برای هر دو رقم خرمایی و خونی، یک رویکرد هوشمندانه است. این الگوریتم نه تنها به سرعت بالا و دقت بسیار خوب در پیش‌بینی دست می‌یابد، بلکه این عملکرد را با انتخاب تعداد معقولی از طول موج‌ها که پردازش مدل را ساده‌تر می‌کند، حفظ می‌نماید. این نتیجه‌گیری تأکید می‌کند که استفاده از الگوریتم‌های بهینه‌سازی فراابتکاری مانند PSO، یک راهکار مؤثر برای کاهش ابعاد داده‌های طیفی و توسعه مدل‌های پیش‌بینی قوی برای پایش غیرمخرب کیفیت میوه‌ها است.

در نهایت، طول موج‌های نهایی انتخاب شده توسط الگوریتم PSO، که شامل ۱۵ طول موج کلیدی برای هر دو رقم است در جدول ۷ به ترتیب اهمیت تأثیر آن‌ها ارائه شده‌اند. این طول موج‌ها به عنوان مهم‌ترین اطلاعات طیفی برای پیش‌بینی فعالیت POD شناخته شده و می‌توانند مبنای ساخت یک دستگاه اختصاصی و ساده‌تر برای پایش غیرمخرب کیفیت آلو قرار گیرند. این یافته به ما امکان می‌دهد که به جای استفاده از کل طیف، تنها بر روی بخش‌های اطلاعاتی و حیاتی آن تمرکز کنیم.

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

Figure 4. Performance of the SVM model with different algorithms for Khormaei plum variety samples: (a) Average RMSE for all samples, (b) Average correlation for all samples.

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

Figure 5. Performance of the SVM model with different algorithms for Khouni plum variety samples: (a) Average RMSE for all samples, (b) Average correlation for all samples

جدول ۷. طول موج‌های منتخب الگوریتم SVM-PSO

Table 7. Selected wavelengths of the SVM-PSO algorithm

۳-۴- نتایج مدل‌سازی پس از انتخاب طول موج

پس از کاهش ابعاد داده‌ها با استفاده از الگوریتم PSO، مدل‌های PLSR و SVM-R با طول موج‌های منتخب، مجدداً مورد ارزیابی قرار گرفتند. هدف از این مرحله، شناسایی بهترین ترکیب از مدل، پیش‌پردازش و هسته برای هر رقم بود. معیار اصلی ارزیابی، مقادیر R^2 ، RMSE و RPD در مجموعه داده آزمون هستند، که به ترتیب، نشان‌دهنده توانایی مدل در توضیح واریانس، میزان خطای پیش‌بینی و دقت کلی مدل هستند.

۳-۴-۱- تحلیل نتایج مدل‌سازی با الگوریتم PLSR

مدل PLSR برای ارزیابی داده‌های طیفی در هر دو حالت طیف کامل و طول موج‌های منتخب مورد استفاده قرار گرفت. در رقم خرمایی در حالت طیف کامل (جدول ۳)، بهترین عملکرد PLSR با پیش‌پردازش Detrend و SNV به ترتیب RPDهای ۱/۶ و ۱/۶ را به دست آورد که نشان‌دهنده یک مدل قابل قبول اما ضعیف است. پس از انتخاب طول موج (جدول ۸)، بهترین نتیجه با پیش‌پردازش فیلتر میانه با RPD برابر با ۱/۵ به دست آمد که اندکی افت عملکرد را نشان می‌دهد. این امر تأکید می‌کند که مدل PLSR برای این رقم، حتی با طیف کامل نیز عملکرد قوی ندارد و روابط پیچیده‌تر هستند.

در رقم خونی، PLSR عملکرد بهتری داشت. در حالت طیف کامل (جدول ۴)، بهترین نتیجه با پیش‌پردازش نرمال‌سازی با RPD ۲/۱۶ به دست آمد که نشان‌دهنده یک مدل بسیار خوب است. پس از انتخاب طول موج (جدول ۹)، مجدداً پیش‌پردازش نرمال‌سازی با RPD=۲/۲۹ بهترین عملکرد را داشت. این بهبود جزئی در دقت، نشان‌دهنده کارایی فرایند انتخاب طول موج در حذف نویز و بهبود روابط خطی مدل است.

مدل PLSR برای رقم خونی کارایی بسیار بالاتری نسبت به رقم خرمایی از خود نشان داد و ثابت کرد که برای برخی از داده‌ها، مدل‌های خطی با پیش‌پردازش مناسب، می‌توانند نتایج دقیقی ارائه دهند. کاهش ابعاد برای رقم خونی، دقت مدل خطی را حفظ و حتی بهبود بخشید، در حالی که برای رقم خرمایی، تأثیر چندانی بر عملکرد نسبتاً ضعیف آن نداشت.

۳-۴-۲- تحلیل نتایج مدل‌سازی با الگوریتم SVM

مدل SVM، به دلیل توانایی خود در مدل‌سازی روابط غیرخطی، به عنوان جایگزینی برای PLSR مورد بررسی قرار گرفت. در رقم خرمایی همان‌طور که پیش‌بینی می‌شد SVM عملکرد بسیار بهتری نسبت به PLSR از خود نشان داد. در حالت طیف کامل (جدول ۵)، بهترین نتیجه با هسته rbf و پیش‌پردازش فیلتر میانه با RPD=۲/۶۵۶ به دست آمد که یک مدل عالی محسوب

می‌شود. پس از کاهش ابعاد (جدول ۱۰)، این مدل با همان ترکیب، عملکردی تقریباً یکسان با $RPD = 2/687$ از خود نشان داد. این نتیجه نشان‌دهنده موفقیت کامل فرایند انتخاب طول موج برای این رقم است.

در رقم خونی نیز عملکرد بسیار قوی‌تری نسبت به PLSR داشت. در حالت طیف کامل (جدول ۶)، بهترین نتیجه با هسته چندجمله‌ای و پیش‌پردازش مرکزکردن میانگین با RPD عالی $3/642$ به دست آمد. با این حال، پس از انتخاب طول موج (جدول ۱۱)، بهترین مدل SVM (هسته خطی با $RPD=2/049$) عملکردی ضعیف‌تر از مدل طیف کامل داشت. با این حال، این نتیجه همچنان در رده مدل‌های بسیار خوب قرار می‌گیرد.

مدل SVM به دلیل توانایی در مدل‌سازی روابط غیرخطی، به طور کلی دقت بالاتری نسبت به PLSR داشت. برای رقم خرمایی، SVM با طول موج‌های منتخب به همان دقت طیف کامل دست یافت که نشان‌دهنده کارایی بالای کاهش ابعاد است. برای رقم خونی، اگرچه دقت مدل طیف کامل به مراتب بالاتر بود، اما مدل با طول موج‌های منتخب نیز نتایج بسیار خوبی ارائه کرد که ارزش استفاده از این رویکرد را در کاربردهای عملی تأیید می‌کند.

تحلیل جامع جداول ۸ تا ۱۱ نشان می‌دهد که انتخاب بهترین مدل و پیش‌پردازش به ویژگی‌های ذاتی هر رقم آلو بستگی دارد. انتخاب طول موج‌های مؤثر توسط الگوریتم PSO، یک استراتژی موفق برای کاهش ابعاد داده‌ها بود که تأثیر متفاوتی بر روی عملکرد مدل‌ها برای هر دو رقم داشت. برای رقم خرمایی، این کاهش ابعاد نه تنها به ساده‌سازی و تسریع فرایند مدل‌سازی کمک کرد، بلکه دقت مدل SVM-R را نیز در مقایسه با مدل طیف کامل، حفظ و حتی اندکی بهبود بخشید (RPD از $2/686$ به $2/687$ افزایش یافت). این امر نشان می‌دهد که برای این رقم، ۱۵ طول موج منتخب توانستند تمام اطلاعات حیاتی طیف را به خوبی نمایندگی کنند و نویزهای غیرضروری را حذف نمایند. در مقابل، برای رقم خونی، مدل‌سازی با طیف کامل همچنان عملکرد برتری نسبت به مدل با طول موج‌های منتخب داشت (RPD از $3/642$ به $2/29$). با این حال، با وجود این تفاوت، عملکرد مدل با طول موج‌های منتخب با RPD برابر با $2/29$ همچنان در رده مدل‌های بسیار خوب قرار می‌گیرد و نشان می‌دهد که این فرایند کاهش ابعاد، بخش قابل توجهی از اطلاعات اصلی را حفظ کرده و قادر به ارائه یک پیش‌بینی قوی و قابل اتکا است. این مقایسه تأیید می‌کند که استراتژی کاهش ابعاد داده برای هر نوع محصول متفاوت عمل می‌کند، اما حتی در مواردی که به حداکثر دقت نمی‌رسد، همچنان می‌تواند یک مدل کارآمد و بسیار خوب برای کاربردهای عملی فراهم کند.

جدول ۸. نتایج مدل‌سازی PLSR رقم خرمایی با انواع پیش‌پردازش‌های طیفی (۶ مؤلفه)

Table 8. PLSR modeling results for the Khormaei plum variety with various spectral preprocessing methods (6 components)

جدول ۹. نتایج مدل‌سازی PLSR رقم خونی با انواع پیش‌پردازش‌های طیفی (۱۰ مؤلفه)

Table 9. PLSR modeling results for the Khouni plum variety with various spectral preprocessing methods (10 components)

جدول ۱۰. مقایسه عملکرد مدل SVM-R رقم خرمایی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 10. Performance comparison of the SVM-R model for the Khormaei plum variety with different kernels and preprocessing methods

جدول ۱۱. مقایسه عملکرد مدل SVM-R رقم خونی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 11. Performance comparison of the SVM-R model for the Khouni plum variety with different kernels and preprocessing methods

این تحقیق با هدف توسعه یک پروتکل غیرمخرب بر پایه طیفسنجی Vis/NIR و روش‌های یادگیری ماشین برای پایش سریع و دقیق فعالیت POD در ارقام آلو خرمایی و خونی انجام شد. فعالیت این آنزیم یک شاخص بیوشیمیایی کلیدی برای ارزیابی پیری و کاهش کیفیت پس از برداشت در محصولات کشاورزی است. نتایج مدل‌سازی به‌دست‌آمده، کارایی بالای این رویکرد را در ارزیابی سریع و غیرمخرب کیفیت آلو تأیید می‌کند. نتایج پژوهش حاضر به وضوح نشان داد که عملکرد بهینه مدل پیش‌بینی، وابسته به رقم است. برای رقم خرمایی، بهترین نتیجه با مدل غیرخطی SVM در ترکیب با الگوریتم بهینه‌سازی PSO حاصل شد که با نسبت عملکرد به انحراف (RPD) معادل ۲/۶۸۷، به طبقه‌بندی عالی دست یافت. این برتری مدل غیرخطی SVM بر رگرسیون خطی PLSR، نشان می‌دهد که رابطه بین فعالیت آنزیم POD و الگوهای طیف جذبی در این رقم، پیچیده و ماهیت غیرخطی دارد. طیف‌های Vis/NIR نه تنها تغییرات آنزیم، بلکه ترکیبات پیچیده دیگری مانند ترکیبات فنولی (سوبسترای POD) و رنگدانه‌ها را نیز منعکس می‌کنند که تعامل آن‌ها منجر به ایجاد یک رابطه غیرخطی می‌شود. در مقابل، برای رقم خونی، مدل PLSR با پیش‌پردازش نرمال‌سازی عملکردی بسیار خوب ($RPD = 2/29$) ارائه داد. این امر حاکی از آن است که تغییرات آنزیمی در رقم خونی ممکن است به‌صورت خطی‌تری با سیگنال‌های طیفی اولیه مرتبط باشد، یا اینکه پیش‌پردازش اعمال شده توانسته است اثرات غیرخطی را تعدیل نماید. با این حال، استفاده از مدل‌های خطی با وجود سادگی، در مقایسه با مدل‌های غیرخطی در رقم خرمایی (که پیچیدگی بیوشیمیایی بیشتری نشان می‌دهد)، دقت کمتری داشت. یافته‌های این تحقیق در مورد ضرورت استفاده از مدل‌های غیرخطی برای ویژگی‌های بیوشیمیایی پیچیده، با مطالعات مشابه همخوانی دارد. لطیفی عموقین و همکاران در پژوهشی بر روی فلفل دلمه‌ای، عملکرد ضعیف‌تر مدل PLSR را در پیش‌بینی فعالیت آنزیم POD گزارش کرده و برتری مدل‌های غیرخطی SVM و استفاده از الگوریتم PSO را برای انتخاب طول موج‌های مؤثر تأیید کردند. این همسویی، اهمیت الگوریتم PSO را به عنوان یک ابزار قدرتمند برای انتخاب ویژگی در داده‌های طیفسنجی بیوشیمیایی، برجسته می‌سازد [۵]. همچنین، گل محمدی و همکاران در پیش‌بینی فعالیت POD در سیب رد دلشیز با استفاده از تصویربرداری فراطیفی، پس از انتخاب طول موج با الگوریتم SPA، بهترین نتیجه را با مدل غیرخطی شبکه عصبی مصنوعی (ANN) کسب کردند ($R^2_v = 0/94$) [۴]. این مقایسه‌ها، جایگاه پژوهش حاضر را که از مدل غیرخطی SVM (با RPD بالاتر از ۲/۶) بهره برده، به عنوان یکی از دقیق‌ترین رویکردها در زمینه تشخیص غیرمخرب فعالیت آنزیمی، تثبیت می‌کند.

نوآوری متمایز تحقیق کنونی، در ارزیابی جامع شش الگوریتم فراابتکاری مختلف (شامل WCC، PSO، HTS، FOA، DSOS و CUK) در ترکیب با مدل SVM، برای کاهش ابعاد داده‌ها در میوه آلو است. در حالی که عمده تحقیقات تنها بر روی چند الگوریتم خاص تمرکز دارند، ارزیابی مقایسه‌ای گسترده ما، منجر به شناسایی الگوریتم بهینه PSO برای استخراج ۱۵ طول موج مؤثر شد. این کاهش ابعاد چشمگیر، نه تنها کارایی و سرعت محاسباتی مدل را افزایش می‌دهد، بلکه مهم‌تر از آن، پایه‌ای عملیاتی برای طراحی نسل بعدی ابزارهای طیفسنجی کوچک و ارزان‌قیمت ایجاد می‌کند که تنها بر روی این طول موج‌های کلیدی تمرکز می‌کنند و برای استفاده در محیط‌های صنعتی خطوط فرآوری و انبارداری، بسیار مناسب هستند. این رویکرد، مشابه با مطالعاتی است که از کاهش ابعاد برای توسعه سنسورهای خاص استفاده می‌کنند؛ برای مثال، لطیفی عموقین و همکاران (۲۰۲۵) در آنالیز محتوای نیترات در فلفل دلمه‌ای [۲۰]، تقی نژاد و همکاران (۲۰۲۵) در تشخیص ژلاتینه شدن نشاسته برنج [۱۵] و کونگ و همکاران (۲۰۱۴) در تشخیص سریع فعالیت POD در برگ‌های گوجه فرنگی [۹]، با موفقیت از الگوریتم‌های فراابتکاری برای دستیابی به مدل‌های غیرمخرب و قابل ساده‌سازی استفاده کرده‌اند. نتایج ما نیز با این روند پژوهشی جهانی همخوانی دارد که هدف آن حرکت از ابزارهای آزمایشگاهی گران‌قیمت به تجهیزات تخصصی و قابل حمل برای پایش کیفیت در لحظه است.

۳-۶- محدودیت‌های تحقیق

هر پژوهشی با محدودیت‌هایی روبه‌رو است که باید برای درک بهتر نتایج و قابلیت تعمیم‌پذیری آن‌ها مورد توجه قرار گیرد. پژوهش حاضر نیز از این قاعده مستثنی نیست. یکی از مهم‌ترین محدودیت‌های این مطالعه، محدود بودن تعداد ارقام مورد بررسی است. این تحقیق تنها بر روی دو رقم آلو شامل خرمایی و خونی انجام شده است، و نتایج به دست آمده ممکن است به طور کامل قابل تعمیم به سایر ارقام آلو نباشد که دارای ویژگی‌های فیزیکی و شیمیایی متفاوتی هستند.

علاوه بر این، تعداد نمونه‌ها (۱۶۰ نمونه) اگرچه برای این مطالعه کافی بود، اما افزایش آن می‌توانست به تقویت مدل‌ها و افزایش دقت پیش‌بینی‌ها کمک کند. همچنین، این پژوهش به تأثیر شرایط محیطی و فیزیولوژیکی مختلف در طول زمان نگهداری بر فعالیت POD نپرداخته است، که می‌تواند در تحقیقات آینده مورد بررسی قرار گیرد. در نهایت، با توجه به هدف پژوهش که توسعه یک روش غیرمخرب بود، این مطالعه تنها بر روی فعالیت POD تمرکز کرد و سایر پارامترهای کیفی مهم در میوه آلو مانند میزان قند، اسیدیته یا ترکیبات فنلی را به طور مستقیم مورد تحلیل قرار نداد.

۴- نتیجه‌گیری

در این پژوهش، یک روش غیرمخرب و کارآمد با استفاده از طیف‌سنجی Vis/NIR و الگوریتم‌های یادگیری ماشین PLSR و SVM، برای پیش‌بینی فعالیت آنزیم پراکسیداز (POD) در دو رقم آلو خرمایی و خونی توسعه داده شد. در بررسی رقم خرمایی، نتایج طیف کامل نشان داد که مدل‌های غیرخطی SVM در مقایسه با مدل‌های خطی مانند PLSR برتری قابل توجهی دارند. بهترین عملکرد در حالت طیف کامل با مدل SVM، هسته rbf و پیش‌پردازش Median Filter به $R^2=0.849$ و $RPD=2/656$ دست یافت که در محدوده طبقه‌بندی عالی قرار می‌گیرد. با این حال، با اعمال استراتژی انتخاب طول موج‌های مؤثر با استفاده از الگوریتم SVM-PSO، دقت مدل حتی بیشتر ارتقا یافت و به $R^2=0.852$ و $RPD=2/687$ رسید. اگرچه افزایش در این رقم جزئی بود، این دستاورد با حذف حجم زیادی از داده‌های طیفی و کاهش پیچیدگی محاسباتی همراه بوده است. در مورد رقم خونی، مدل‌ها در حالت طیف کامل نیز عملکرد بسیار قوی را نشان دادند. بهترین نتیجه مدل PLSR با پیش‌پردازش نرمال‌سازی به $R^2=0.7722$ و $RPD=2/16$ دست یافت که در محدوده طبقه‌بندی بسیار خوب قرار می‌گیرد. با این حال، استفاده از مدل غیرخطی SVM با هسته چندجمله‌ای و Mean Centering در حالت طیف کامل عملکرد را به $R^2=0.92$ و $RPD=3/642$ ارتقا داد. این نتیجه بهترین عملکرد مطلق در کل پژوهش برای رقم خونی را نشان می‌دهد. در حالت کاهش ابعاد، مدل PLSR با پیش‌پردازش نرمال‌سازی به $R^2=0.796$ و $RPD=2/29$ رسید که نسبت به حالت طیف کامل این مدل بهبود محسوس را نشان می‌دهد که بر توانایی الگوریتم‌های فراابتکاری در حذف نویز و بهبود عملکرد حتی در مدل‌های خطی تأکید می‌کند. در نهایت، این مطالعه ثابت کرد که طیف‌سنجی Vis/NIR ابزاری قدرتمند برای پایش غیرمخرب فعالیت آنزیمی در میوه‌های هسته‌دار است و با ترکیب آن با مدل و پیش‌پردازش مناسب، می‌توان به پیش‌بینی‌های دقیق و قابل اتکا دست یافت، که این امر می‌تواند راهگشای توسعه سیستم‌های هوشمند برای مدیریت پس‌از برداشت و تضمین کیفیت محصولات کشاورزی باشد.

- [1] Borji, M., & Rezaei, M. (2024). Characterization of frost-tolerant plum genotypes (*Prunus domestica* L.) in Nishabur region, Iran: a morphological and phenological assessment following natural spring frost event. *BMC Plant Biol.*, 24(1), 1098.
- [2] Jaiswal, R., Karaköse, H., Rühmann, S., Goldner, K., Neumüller, M., Treutter, D., & Kuhnert, N. (2013). Identification of phenolic compounds in plum fruits (*Prunus salicina* L. and *Prunus domestica* L.) by high-performance liquid chromatography/tandem mass spectrometry and characterization of varieties by quantitative phenolic fingerprints. *J. Agric. Food Chem.*, 61(49), 12020-12031.
- [3] Moscatello, S., Frioni, T., Blasi, F., Proietti, S., Pollini, L., Verducci, G., & Cossignani, L. (2019). Changes in absolute contents of compounds affecting the taste and nutritional properties of the flesh of three plum species throughout development. *Foods*, 8(10), 486.
- [4] Golmohammadi, A., Razavi, M. S., & Tahmasebi, M. (2022). Prediction of peroxidase activity using near infrared hyperspectral imaging in red delicious apple fruit during storage time. *J. Food Compos. Anal.*
- [5] 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. *Chemom. Intell. Lab. Syst.*, 105137.
- [6] Shrestha, L., Kulig, B., Moscetti, R., Massantini, R., Pawelzik, E., Hensel, O., & Sturm, B. (2020). Comparison between hyperspectral imaging and chemical analysis of polyphenol oxidase activity on fresh-cut apple slices. *J. Spectrosc.*, 2020(1), 7012525.
- [7] Serrano-Martínez, A., Fortea, M., Del Amor, F., & Núñez-Delicado, E. (2008). Kinetic characterisation and thermal inactivation study of partially purified red pepper (*Capsicum annuum* L.) peroxidase. *Food Chem.*, 107(1), 193-199.
- [8] Zhao, Z., Wang, Y., Gong, D., Niu, X., Cheng, W., & Gu, Y. (2016). Discrimination of plum browning with near infrared spectroscopy. *Guang pu xue yu Guang pu fen xi= Guang pu*, 36(7), 2089-2093.
- [9] Kong, W., Liu, F., Zhang, C., Bao, Y., Yu, J., & He, Y. (2014). Fast detection of peroxidase (POD) activity in tomato leaves which infected with *Botrytis cinerea* using hyperspectral imaging. *Spectrochim. Acta, Part A: Mol. Biomol. Spectrosc.*, 118, 498-502.
- [10] Posom, J., Klprachan, 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), 27909-27921.
- [11] Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Cruz-Gómez, 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.
- [12] Gardesh, A. S. K., Badii, F., Hashemi, M., Ardakani, A. Y., Maftoonazad, N., & Gorji, A. M. (2016). Effect of nanochitosan based coating on climacteric behavior and postharvest shelf-life extension of apple cv. Golab Kohanz. *LWT*, 70, 33-40.
- [13] Razavi, M. S., Sharabiani, V. R., Tahmasebi, M., & Szymanek, M. (2025). Unveiling the Fingerprint of Apple Browning: A Vis/NIR-Metaheuristic Approach for Rapid Polyphenol oxidase and Peroxidases Activities Detection in Red Delicious Apples. *J. Food Compos. Anal.*, 107499.
- [14] Terefe, N. S., Matthies, K., Simons, L., & Versteeg, C. (2009). Combined high pressure-mild temperature processing for optimal retention of physical and nutritional quality of strawberries (*Fragaria* × *ananassa*). *Innovative Food Sci. Emerg. Technol.*, 10(3), 297-307.

- [15] 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. *J. Food Compos. Anal.*, 107443.
- [16] 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 Sci. Soc. Am. J.*, 65(2), 480-490.
- [17] 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.
- [18] Rajkumar, D., Künnemeyer, R., Kaur, H., Longdell, J., & McGlone, A. (2022). Interactions of linearly polarized and unpolarized light on kiwifruit using aquaphotomics. *Molecules*, 27(2), 494.
- [19] Currà, A., Gasbarrone, R., Gattabria, R., Bonifazi, G., Serranti, S., Greco, D., & 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*.
- [20] 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.



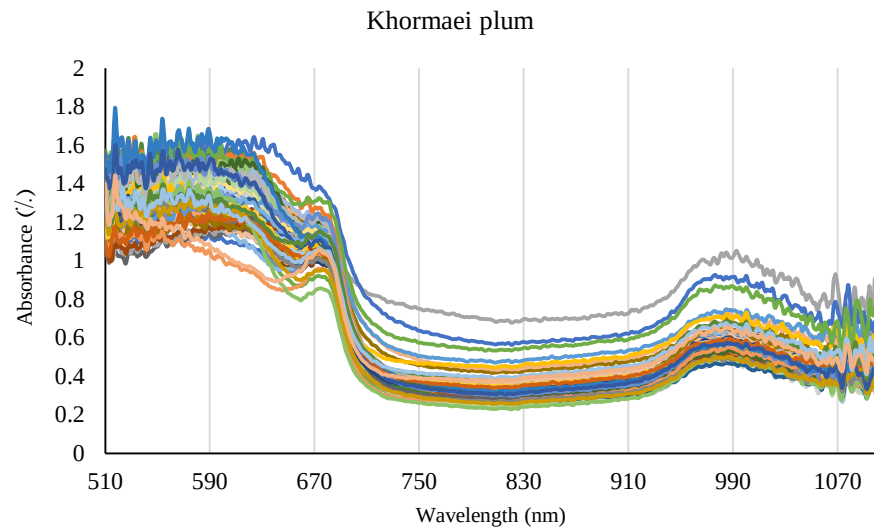
الف (a) ب (b)
شکل ۱. نمونه‌های مورد استفاده (الف) آلو خونی و (ب) آلو خرمايي

Figure 1. Samples used (a) Khouni plum and (b) Khormaei plum

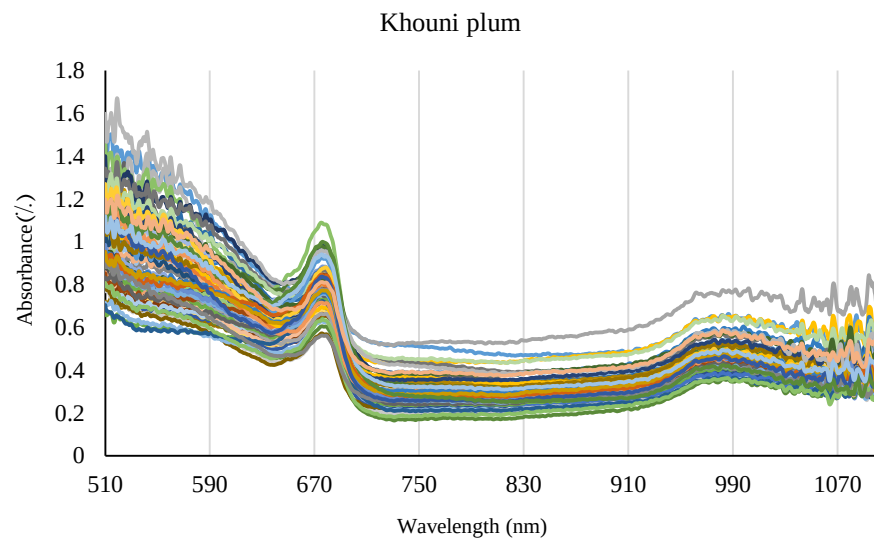


شکل ۲. نمایی از سیستم طیف‌سنجی VIS/NIR مورد استفاده برای جمع‌آوری داده‌ها

Figure 2. View of the VIS/NIR spectroscopy system used for data collection



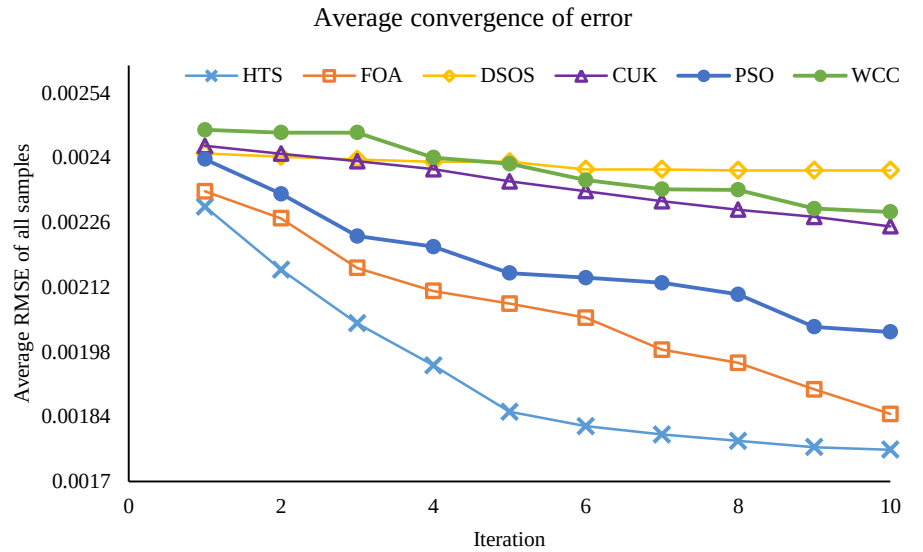
(a) الف



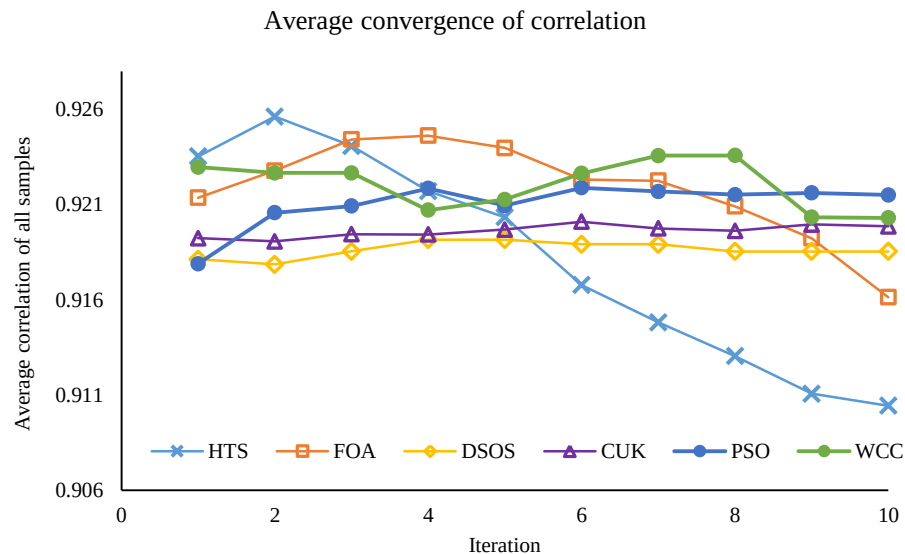
(b) ب

شکل ۳. طیف‌های جذبی نمونه‌های آلو خرمایی (الف) و آلو خونی (ب)

Figure 3. Absorption spectra of Khormaei plum (a) and Khouni plum (b) samples



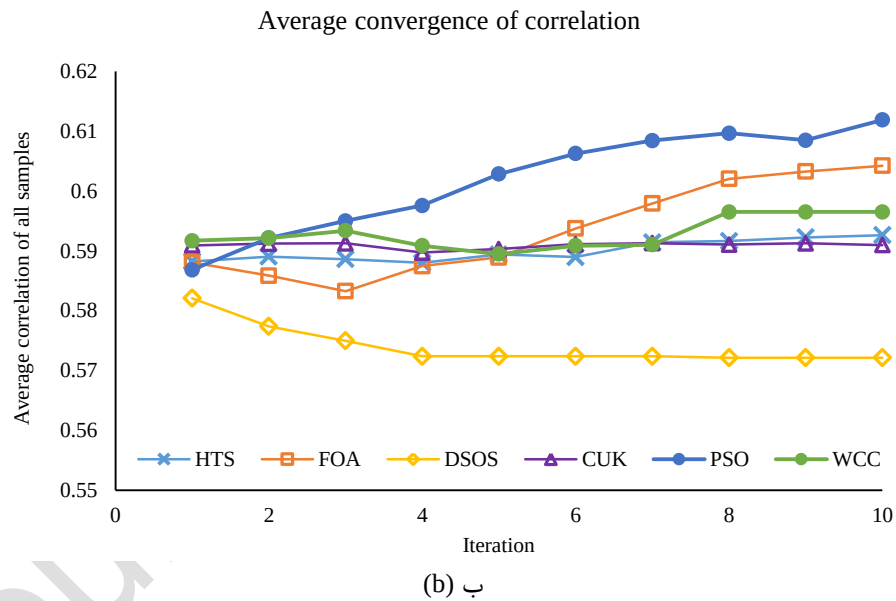
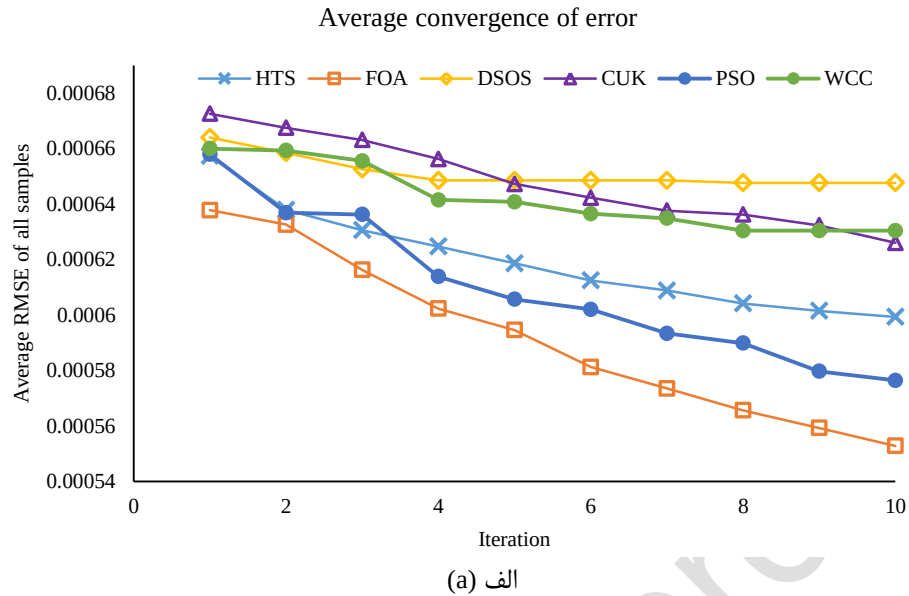
(a) الف



(b) ب

شکل ۴. عملکرد مدل SVM همراه با الگوریتم‌های فراابتکاری مختلف شامل مسابقه رقابتی جهانی (WCC)، بهینه‌سازی ازدحام ذرات (PSO)، الگوریتم بهینه‌سازی انتقال حرارت (HTS)، الگوریتم بهینه‌سازی جنگل (FOA)، جستجوی ارگانیسم‌های همزیست گسسته (DSOS) و بهینه‌سازی فاخته (CUK) برای انتخاب طول موج‌های مؤثر در نمونه‌های رقم خرمایی الف (میانگین RMSE برای همه نمونه‌ها ب) میانگین همبستگی همه نمونه‌ها

Figure 4. Performance of the SVM model with various meta-heuristic algorithms, including World Competitive Contest (WCC), Particle Swarm Optimization (PSO), Heat Transfer Search (HTS) algorithm, Forest Optimization Algorithm (FOA), Discrete Symbiotic Organisms Search (DSOS), and Cuckoo Optimization (CUK), for the selection of effective wavelengths in Khormaei cultivar samples: (a) Average RMSE for all samples, (b) Average correlation for all samples.



شکل ۵. عملکرد مدل SVM همراه با الگوریتم‌های فراابتکاری مختلف شامل مسابقه رقابتی جهانی (WCC)، بهینه‌سازی ازدحام ذرات (PSO)، الگوریتم بهینه‌سازی انتقال حرارت (HTS)، الگوریتم بهینه‌سازی جنگل (FOA)، جستجوی ارگانیسم‌های همزیست گسسته (DSOS) و بهینه‌سازی فاخته (CUK) برای انتخاب طول موج‌های مؤثر در نمونه‌های رقم خونی الف) میانگین RMSE برای همه نمونه‌ها ب) میانگین همبستگی همه نمونه‌ها

Figure 5. Performance of the SVM model with various meta-heuristic algorithms, including World Competitive Contest (WCC), Particle Swarm Optimization (PSO), Heat Transfer Search (HTS) algorithm, Forest Optimization Algorithm (FOA), Discrete Symbiotic Organisms Search (DSOS), and Cuckoo Optimization (CUK), for the selection of effective wavelengths in Khouni cultivar samples: (a) Average RMSE for all samples, (b) Average correlation for all samples

جدول ۱. مشخصات پارامترهای تنظیمی هر یک از الگوریتم‌های انتخاب طول موج مؤثر

Table 1. Specifications of the tuning parameters for each effective wavelength selection algorithms

No	Algorithm	Parameters
1	WCC	Match Time: 2 ,Seasons: 10 ,Groups: 3 ,Teams: 6
2	PSO	P Bests: 8 ,Episodes: 10 ,Particles: 40
3	HTS	RDF: 2 ,COF: 10 ,CDF: 2 ,Iterations: 10 ,Molecules: 15
4	FOA	Iterations: 10 ,Trees: 20
5	DSOS	Iterations: 10 ,Organisms: 20
6	CUK	Iterations: 10 ,Cuckoos: 40 ,Eggs: 10

جدول ۲. آمار توصیفی میزان فعالیت آنزیم پراکسیداز (absorbance/min.g) ارقام مختلف آلو

Table 2. Descriptive statistics of peroxidase enzyme activity (absorbance/min.g) for different plum varieties

Variety	Mean	Standard Deviation	Maximum	Minimum
Khouni	0.4236	0.1861	0.734	0.0808
Khormaei	0.4334	0.2029	0.7314	0.0787

جدول ۳. نتایج مدل سازی PLSR رقم خرمایی با انواع پیش پردازش های طیفی (۵ مؤلفه)

Table 3. PLSR modeling results for the Khormaei plum variety with various spectral preprocessing methods (5 Component)

No	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	No Preprocessing	0.9506	0.0419	4.55	0.8719	0.0715	2.89	0.3533	0.2797	0.89
2	SNV	0.8917	0.0621	3.07	0.5358	0.2329	0.89	0.7815	0.1567	1.6
3	MSC	0.8921	0.0619	3.08	0.5378	0.2346	0.88	0.6939	0.1723	1.45
4	Normalization	0.9385	0.0467	4.08	0.8666	0.073	2.83	0.6405	0.2907	0.86
5	Moving Average	0.9437	0.0447	4.26	0.8632	0.0739	2.79	0.4909	0.2752	0.91
6	Gaussian Filter	0.9354	0.0479	3.98	0.8676	0.0727	2.84	0.4262	0.2721	0.92
7	Median Filter	0.949	0.0426	4.47	0.873	0.0712	2.9	0.5328	0.2796	0.89
8	Detrending	0.8213	0.0797	2.39	0.4438	0.2229	0.93	0.7043	0.1524	1.64
9	Mean Centering	0.9506	0.0419	4.55	0.8719	0.0715	2.89	0.5334	0.2797	0.89

جدول ۴. نتایج مدل سازی PLSR رقم خونی با انواع پیش پردازش های طیفی (۸ مؤلفه)

Table 4. PLSR modeling results for the Khouni plum variety with various spectral preprocessing methods (8 Component)

No	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	No Preprocessing	0.9814	0.0248	7.41	0.968	0.0379	5.77	0.7406	0.0745	2.03
2	SNV	0.9252	0.0498	3.7	0.5282	0.1454	1.5	0.0482	0.1428	1.06
3	MSC	0.9264	0.0494	3.73	0.5477	0.1424	1.54	0.0156	0.1475	1.02
4	Normalization	0.9799	0.0259	7.12	0.9467	0.0488	4.48	0.7722	0.0698	2.16
5	Moving Average	0.9778	0.0272	6.78	0.9681	0.0378	5.79	0.7439	0.0741	2.04
6	Gaussian Filter	0.9761	0.0282	6.54	0.9649	0.0396	5.52	0.7554	0.0724	2.09
7	Median Filter	0.9804	0.0255	7.21	0.9655	0.0393	5.56	0.7411	0.0745	2.03
8	Detrending	0.9773	0.0275	6.7	0.8502	0.0819	2.67	0.2249	0.1289	1.17
9	Mean Centering	0.9814	0.0248	7.41	0.968	0.0379	5.77	0.7406	0.0745	2.03

جدول ۵. مقایسه عملکرد مدل SVM-R رقم خرمایی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 5. Performance comparison of the SVM-R model for the Khormaei plum variety with different kernels and preprocessing methods

Kernel	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
Linear	No Preprocessing	0.974	0.031	6.267	0.874	0.083	2.91	0.557	0.326	0.646
	SNV	0.985	0.024	8.21	0.568	0.153	1.572	0.578	0.231	0.822
	MSC	0.801	0.09	2.268	0.517	0.153	1.485	0.212	0.203	0.938
	Normalization	0.988	0.021	9.055	0.835	0.096	2.545	0.069	0.179	1.07
	Moving Average	0.916	0.057	3.488	0.868	0.076	2.845	0.646	0.125	1.735
	Gaussian Filter	0.893	0.07	3.084	0.831	0.079	2.513	0.737	0.091	2.016
	Median Filter	0.88	0.07	2.913	0.897	0.06	3.22	0.243	0.193	1.187
	Detrending	0.791	0.089	2.209	0.586	0.135	1.605	0.172	0.199	1.135
	Mean Centering	0.966	0.036	5.47	0.927	0.059	3.836	0.503	0.15	1.465
rbf	No Preprocessing	0.983	0.028	7.807	0.801	0.063	2.317	0.789	0.083	2.247
	SNV	0.879	0.068	2.908	0.473	0.149	1.422	0.843	0.22	0.761
	MSC	0.917	0.057	3.508	0.368	0.165	1.299	0.041	0.201	1.012
	Normalization	0.982	0.027	7.572	0.815	0.085	2.4	0.648	0.107	1.74
	Moving Average	0.983	0.029	7.65	0.85	0.058	2.669	0.143	0.155	1.116
	Gaussian Filter	0.984	0.027	7.89	0.809	0.085	2.364	0.562	0.117	1.561
	Median Filter	0.983	0.028	7.845	0.887	0.056	3.076	0.849	0.072	2.656
	Detrending	0.869	0.078	2.792	0.214	0.163	1.165	0.156	0.168	1.124
	Mean Centering	0.981	0.028	7.41	0.823	0.082	2.453	0.365	0.166	1.296
Polynomial	No Preprocessing	0.983	0.027	7.733	0.832	0.08	2.52	0.062	0.211	1.002
	SNV	0.985	0.026	8.134	0.491	0.124	1.447	0.199	0.178	1.154
	MSC	0.949	0.046	4.473	0.431	0.155	1.369	0.174	0.177	1.137
	Normalization	0.986	0.024	8.56	0.939	0.044	4.186	0.667	0.127	1.789
	Moving Average	0.985	0.024	8.264	0.846	0.081	2.628	0.19	0.183	1.148
	Gaussian Filter	0.986	0.024	8.505	0.845	0.086	2.624	0.245	0.289	0.689
	Median Filter	0.985	0.025	8.326	0.807	0.078	2.349	0.681	0.122	1.827
	Detrending	0.987	0.023	8.732	0.549	0.135	1.538	0.049	0.196	1.059
	Mean Centering	0.984	0.027	8.032	0.949	0.041	4.583	0.508	0.11	1.473

جدول ۶. مقایسه عملکرد مدل SVM-R رقم خونی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 6. Performance comparison of the SVM-R model for the Khouni plum variety with different kernels and preprocessing methods

Kernel	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
Linear	No Preprocessing	0.983	0.023	7.693	0.946	0.047	4.44	0.765	0.097	2.133
	SNV	0.99	0.017	9.988	0.703	0.125	1.894	0.122	0.186	0.975
	MSC	0.698	0.104	1.838	0.748	0.097	2.057	0.355	0.119	1.286
	Normalization	0.988	0.02	9.323	0.946	0.043	4.426	0.817	0.078	2.412
	Moving Average	0.96	0.036	5.049	0.933	0.055	3.981	0.918	0.052	3.611
	Gaussian Filter	0.964	0.034	5.349	0.953	0.034	4.758	0.802	0.094	2.323
	Median Filter	0.957	0.038	4.857	0.954	0.044	4.823	0.911	0.052	3.464
	Detrending	0.946	0.04	4.347	0.865	0.085	2.812	0.601	0.105	1.635
	Mean Centering	0.969	0.033	5.773	0.968	0.033	5.749	0.867	0.066	2.836
rbf	No Preprocessing	0.981	0.025	7.429	0.824	0.075	2.46	0.496	0.127	1.455
	SNV	0.96	0.036	5.024	0.499	0.147	1.459	0.31	0.157	1.244
	MSC	0.937	0.045	4.028	0.635	0.105	1.71	0.45	0.161	1.392
	Normalization	0.981	0.025	7.4	0.873	0.061	2.903	0.538	0.143	1.519
	Moving Average	0.985	0.024	8.184	0.86	0.071	2.764	0.862	0.058	2.779
	Gaussian Filter	0.98	0.023	7.124	0.813	0.083	2.388	0.443	0.168	1.384
	Median Filter	0.981	0.026	7.371	0.753	0.093	2.076	0.633	0.107	1.704
	Detrending	0.966	0.035	5.501	0.681	0.092	1.828	0.651	0.116	1.748
	Mean Centering	0.984	0.023	7.963	0.815	0.089	2.398	0.876	0.056	2.932
Polynomial	No Preprocessing	0.968	0.034	5.626	0.943	0.042	4.308	0.77	0.072	2.154
	SNV	0.986	0.021	8.586	0.655	0.121	1.758	0.557	0.211	0.828
	MSC	0.449	0.133	1.361	0.636	0.116	1.711	0.459	0.229	0.855
	Normalization	0.987	0.021	8.831	0.95	0.04	4.602	0.326	0.161	1.258
	Moving Average	0.988	0.02	9.137	0.874	0.06	2.908	0.869	0.068	2.858
	Gaussian Filter	0.921	0.055	3.603	0.865	0.068	2.81	0.358	0.115	1.289
	Median Filter	0.985	0.022	8.337	0.915	0.054	3.542	0.697	0.097	1.877
	Detrending	0.986	0.021	8.585	0.785	0.076	2.227	0.742	0.089	2.033
	Mean Centering	0.987	0.021	8.954	0.936	0.05	4.089	0.92	0.048	3.642

جدول ۷. طول موج های منتخب الگوریتم SVM-PSO

Table 7. Selected wavelengths of the SVM-PSO algorithm

Variety	Selected EWs (nm)
Khormaei	864, 1027, 932, 1045.5, 978, 1069.5, 777, 987.5, 1085.5, 1011, 1076.5, 992.5, 954.5, 962.5, 526.5
Khouni	789, 579.5, 1069, 1034.5, 1067, 606.5, 780.5, 614.5, 702.5, 618, 617, 1079, 837, 539.5, 780

جدول ۸. نتایج مدل سازی PLSR رقم خرمایی با انواع پیش پردازش های طیفی (۶ مؤلفه)

Table 8. PLSR modeling results for the Khormaei plum variety with various spectral preprocessing methods (6 components)

No	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	No Preprocessing	0.838	0.0833	2.51	0.9111	0.0533	3.46	0.0494	0.1856	1.01
2	SNV	0.3631	0.1651	1.27	0.314	0.2051	0.9	0.5503	0.2256	0.83
3	MSC	0.3709	0.1641	1.27	0.3097	0.2048	0.9	0.4926	0.2214	0.85
4	Normalization	0.8324	0.0847	2.47	0.902	0.056	3.3	0.0458	0.177	1.06
5	Moving Average	0.7985	0.0929	2.25	0.8722	0.064	2.89	0.0357	0.1779	1.05
6	Gaussian Filter	0.74	0.1055	1.98	0.8279	0.0742	2.49	0.1506	0.1944	0.96
7	Median Filter	0.7751	0.0981	2.13	0.7761	0.0847	2.18	0.5271	0.1246	1.5
8	Detrending	0.3628	0.1652	1.27	0.8429	0.2429	0.76	0.7546	0.24	0.78
9	Mean Centering	0.838	0.0833	2.51	0.9111	0.0533	3.46	0.0494	0.1856	1.01

جدول ۹. نتایج مدل سازی PLSR رقم خونی با انواع پیش پردازش های طیفی (۱۰ مؤلفه)

Table 9. PLSR modeling results for the Khouni plum variety with various spectral preprocessing methods (10 components)

No	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
1	No Preprocessing	0.8803	0.0633	2.92	0.9484	0.0437	4.55	0.7913	0.0797	2.26
2	SNV	0.6792	0.1037	1.78	0.2016	0.172	1.16	0.0631	0.2506	0.72
3	MSC	0.688	0.1023	1.81	0.2098	0.1711	1.16	0.0975	0.1658	1.09
4	Normalization	0.8801	0.0634	2.92	0.9476	0.0441	4.51	0.796	0.0788	2.29
5	Moving Average	0.8802	0.0634	2.92	0.9302	0.0509	3.91	0.7464	0.0879	2.05
6	Gaussian Filter	0.8769	0.0642	2.88	0.9333	0.0497	4	0.7911	0.0797	2.26
7	Median Filter	0.8269	0.0762	2.43	0.827	0.0801	2.48	0.6429	0.1043	1.73
8	Detrending	0.838	0.0737	2.51	0.7836	0.0896	2.22	0.2843	0.1476	1.22
9	Mean Centering	0.8803	0.0633	2.92	0.9484	0.0437	4.55	0.7913	0.0797	2.26

جدول ۱۰. مقایسه عملکرد مدل SVM-R رقم خرمایی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 10. Performance comparison of the SVM-R model for the Khormaei plum variety with different kernels and preprocessing methods

Kernel	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
Linear	No Preprocessing	0.272	0.165	1.184	0.883	0.061	3.024	0.128	0.259	0.973
	SNV	0.242	0.175	1.161	0.497	0.132	1.456	0.025	0.215	1.046
	MSC	0.091	0.185	1.06	0.317	0.187	1.25	0.043	0.199	1.056
	Normalization	0.756	0.097	2.047	0.868	0.074	2.844	0.064	0.193	1.001
	Moving Average	0.526	0.13	1.468	0.782	0.101	2.21	0.31	0.182	1.243
	Gaussian Filter	0.423	0.157	1.33	0.852	0.075	2.687	0.136	0.17	1.111
	Median Filter	0.451	0.148	1.364	0.885	0.065	3.052	0.494	0.142	1.452
	Detrending	0.102	0.195	1.066	0.382	0.17	1.314	0.16	0.182	0.959
	Mean Centering	0.457	0.148	1.372	0.861	0.077	2.768	0.715	0.106	1.935
rbf	No Preprocessing	0.98	0.028	7.125	0.947	0.038	4.475	0.792	0.096	2.264
	SNV	0.739	0.106	1.98	0.614	0.12	1.663	0.28	0.166	1.218
	MSC	0.718	0.109	1.904	0.521	0.135	1.492	0.073	0.19	1.072
	Normalization	0.985	0.026	8.148	0.957	0.039	4.992	0.815	0.083	2.404
	Moving Average	0.972	0.034	6.035	0.895	0.063	3.193	0.754	0.108	2.081
	Gaussian Filter	0.97	0.035	5.826	0.873	0.074	2.897	0.673	0.114	1.806
	Median Filter	0.983	0.028	7.739	0.954	0.038	4.838	0.852	0.072	2.687
	Detrending	0.555	0.135	1.515	0.433	0.141	1.372	0.507	0.213	0.841
	Mean Centering	0.983	0.027	7.751	0.957	0.033	4.965	0.655	0.119	1.758
Polynomial	No Preprocessing	0.985	0.025	8.253	0.94	0.043	4.205	0.147	0.547	0.417
	SNV	0.985	0.024	8.14	0.107	0.227	0.981	0.7852	0.417	0.526
	MSC	0.154	0.767	0.269	0.125	0.219	0.974	0.549	0.339	0.548
	Normalization	0.986	0.022	8.488	0.978	0.028	6.956	0.514	1.292	0.176
	Moving Average	0	0.2	1.011	0.002	0.231	1.032	0	0.179	1.033
	Gaussian Filter	0	0.207	1.011	0	0.179	1.033	0.403	0.22	0.872
	Median Filter	0	0.199	1.011	0	0.215	1.033	0.034	0.202	1.016
	Detrending	0.983	0.025	7.643	0.114	0.227	0.978	0.082	0.368	0.588
	Mean Centering	0.986	0.024	8.435	0.952	0.036	4.733	0.354	2.31	0.107

جدول ۱۱. مقایسه عملکرد مدل SVM-R رقم خونی با هسته‌ها و پیش‌پردازش‌های مختلف

Table 11. Performance comparison of the SVM-R model for the Khouni plum variety with different kernels and preprocessing methods

Kernel	Preprocessing	Training			Validation			Test		
		R ²	RMSE	RPD	R ²	RMSE	RPD	R ²	RMSE	RPD
Linear	No Preprocessing	0.652	0.101	1.714	0.812	0.089	2.38	0.746	0.101	2.049
	SNV	0.419	0.135	1.326	0.664	0.11	1.781	0.48	0.148	1.432
	MSC	0.232	0.169	1.153	0.364	0.129	1.295	0.056	0.18	1.063
	Normalization	0.849	0.069	2.599	0.888	0.066	3.09	0.638	0.088	1.716
	Moving Average	0.631	0.111	1.663	0.843	0.074	2.603	0.635	0.116	1.71
	Gaussian Filter	0.669	0.106	1.756	0.783	0.09	2.22	0.536	0.119	1.516
	Median Filter	0.67	0.101	1.76	0.861	0.071	2.769	0.699	0.113	1.883
	Detrending	0.46	0.133	1.375	0.649	0.125	1.742	0.238	0.143	1.183
	Mean Centering	0.678	0.102	1.781	0.839	0.082	2.57	0.599	0.116	1.631
rbf	No Preprocessing	0.914	0.053	3.444	0.829	0.085	2.498	0.432	0.131	1.371
	SNV	0.772	0.088	2.116	0.619	0.122	1.672	0.166	0.149	1.131
	MSC	0.796	0.087	2.236	0.687	0.105	1.846	0.355	0.168	0.887
	Normalization	0.954	0.04	4.721	0.83	0.082	2.506	0.446	0.124	1.388
	Moving Average	0.932	0.048	3.862	0.683	0.101	1.834	0.34	0.162	1.271
	Gaussian Filter	0.881	0.068	2.928	0.706	0.1	1.904	0.362	0.124	1.293
	Median Filter	0.893	0.061	3.088	0.854	0.077	2.707	0.605	0.1	1.643
	Detrending	0.906	0.055	3.29	0.817	0.089	2.417	0.489	0.115	1.445
	Mean Centering	0.935	0.05	3.973	0.824	0.079	2.462	0.49	0.099	1.446
Polynomial	No Preprocessing	0.983	0.024	7.729	0.878	0.057	2.959	0.708	0.112	1.912
	SNV	0.985	0.023	8.368	0.082	0.174	1.078	0.368	0.313	0.563
	MSC	0.981	0.025	7.308	0.276	0.164	1.214	0.148	0.284	0.705
	Normalization	0.989	0.019	9.552	0.772	0.079	2.163	0.11	0.279	0.711
	Moving Average	0.938	0.047	4.064	0.849	0.064	2.659	0.24	0.162	1.184
	Gaussian Filter	0.925	0.049	3.685	0.788	0.079	2.242	0.366	0.164	1.298
	Median Filter	0.832	0.075	2.465	0.71	0.109	1.919	0.385	0.136	1.316
	Detrending	0.989	0.017	9.721	0.641	0.113	1.723	0.5	0.152	1.461
	Mean Centering	0.981	0.025	7.393	0.799	0.074	2.305	0.092	0.563	0.343

(Extended Abstract)

Non-destructive detection of peroxidase enzyme activity in different plum varieties using VIS/NIR spectroscopy and machine learning methods

Meysam Latifi-Amoghin¹, Yousef Abbaspour-Gilandeh^{2*}

¹Ph.D Graduate., Department of Biosystems Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran.

²Professor, Department of Biosystems Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran.

Introduction: Plums (*Prunus domestica* L.) represent a commercially important group of stone fruits, valued globally for their nutritional composition and health-promoting properties. Despite their economic and dietary significance, postharvest quality deterioration remains a major challenge, primarily driven by enzymatic processes such as peroxidase (POD) activity, which contributes to internal browning and reduced marketability. Conventional approaches for quantifying POD activity are often labor-intensive, destructive, and influenced by multiple environmental and physiological variables, limiting their practical application in quality control workflows. In recent years, visible/near-infrared (Vis/NIR) spectroscopy has emerged as a promising non-destructive technique for rapid assessment of internal fruit attributes. This optical method enables real-time monitoring of biochemical changes without compromising sample integrity. The present study aimed to develop and validate a Vis/NIR spectroscopy-based predictive model for estimating POD enzyme activity in two Iranian plum cultivars, Khormaei and Khouni, by integrating spectral data with machine learning algorithms.

Materials and Methods: A total of 160 plum samples were employed in this study, comprising 80 specimens each from the Khormaei and Khouni cultivars. Non-destructive spectral measurements were initially acquired using a Vis/NIR spectrometer operating within the 350–1100 nm wavelength range. These absorption spectra provided preliminary insights into the internal biochemical properties of the fruits. Subsequently, POD enzyme activity was quantified through a conventional destructive colorimetric assay, serving as the reference method. To enhance the predictive performance of the models, spectral data underwent preprocessing using a suite of techniques, including normalization, standard normal variate (SNV), multiplicative scatter correction (MSC), median filtering, and mean centering. These procedures aimed to mitigate noise and variability inherent in raw spectral signals. Two machine learning algorithms—Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM)—were implemented to construct predictive models for POD activity. To address the high dimensionality of spectral data and identify the most informative wavelengths, the SVM framework was coupled with six meta-heuristic optimization algorithms. The selected wavelengths were then used to reconstruct both PLSR and SVM models, and their predictive accuracies were benchmarked against models developed using the full spectral range.

Results and Discussion: The findings of this study demonstrated that Vis/NIR spectroscopy is a reliable and efficient technique for predicting POD enzyme activity in plum fruits. However, model performance varied significantly between the two cultivars, underscoring the influence of varietal-specific biochemical and physical traits. In full-spectrum modeling, the SVM consistently outperformed the linear PLSR model across both varieties. For the Khormaei cultivar, the SVM model employing a radial basis function (RBF) kernel combined with Median Filter preprocessing achieved the highest predictive accuracy, classified as excellent (RPD = 2.656). In contrast, for the Khouni cultivar, although the PLSR model with normalization preprocessing yielded very good results (RPD = 2.16), the SVM model with a polynomial kernel and mean centering preprocessing attained superior performance (RPD = 3.642), indicating excellent predictive capability. These differences are attributed to the distinct physicochemical properties of the Khormaei and Khouni plums, which affect spectral response and enzyme activity patterns. To enhance model efficiency and reduce computational complexity, six meta-heuristic optimization algorithms were evaluated for effective wavelength selection. Among them, Particle Swarm Optimization (PSO) emerged as the most accurate and computationally efficient, successfully reducing the spectral dimensionality to 15 key wavelengths while preserving model integrity. Post-optimization modeling further confirmed that predictive

performance is highly dependent on cultivar-specific characteristics and the choice of modeling strategy. The SVM model, owing to its capacity to capture complex and non-linear relationships, generally exhibited superior performance compared to the linear PLSR approach. Specifically, for the Khormaei variety, the optimized SVM-R model with an RBF kernel achieved an RPD of 2.687, maintaining excellent classification. Conversely, for the Khoumi variety, the best result was obtained using the PLSR model with normalization preprocessing (RPD = 2.29), suggesting that linear modeling can be highly effective when paired with appropriate data treatment.

Conclusions: This study demonstrates the efficacy of Vis/NIR spectroscopy, in conjunction with advanced machine learning techniques, as a robust and non-destructive approach for the rapid assessment of POD enzyme activity in postharvest plum fruits. The results underscore the importance of cultivar-specific model selection, with non-linear algorithms such as SVM consistently outperforming linear models like PLSR in predictive accuracy. Moreover, the integration of meta-heuristic optimization algorithms—particularly Particle Swarm Optimization (PSO)—proved highly effective in dimensionality reduction, enabling the identification of key wavelengths that enhance model performance while improving computational efficiency. These selected wavelengths offer a practical foundation for the development of compact, dedicated devices tailored to non-invasive quality monitoring of plum fruits. Overall, the proposed methodology holds significant promise for advancing postharvest quality control practices in the fruit industry.

Keywords: Plum, Peroxidase enzyme, Spectroscopy, Non-destructive detection, Machine learning

Journal Pre-proofs