

## مقایسه برخی روش‌های پیش‌پردازش طیف مرئی/فروسرخ نزدیک موج کوتاه بر عملکرد مدل حداقل مربعات جزئی (مطالعه موردی: ارزیابی رسیدگی میوه انجیر)

رضا محمدی گل<sup>۱\*</sup>، بابک ولی‌زاده کاجی<sup>۲</sup>، مجید لشگری<sup>۳</sup> و رضا صیاد حق‌شمار<sup>۴</sup>

۱. استادیار گروه مهندسی مکانیک بیوسیستم دانشگاه اراک (r-mohammadigol@araku.ac.ir)
۲. دانشیار گروه مهندسی باغبانی دانشگاه اراک (valizadehkaji@yahoo.com)
۳. دانشیار گروه مهندسی مکانیک بیوسیستم دانشگاه اراک (m-lashgari@araku.ac.ir)
۴. دانشجوی کارشناسی ارشد گروه مهندسی باغبانی دانشگاه اراک (rezasayad4747@gmail.com)

### چکیده

طیف‌های مرئی و فروسرخ نزدیک اخذ شده از مواد مختلف علاوه بر اطلاعات ناشی از ساختار مولکولی نمونه، حامل داده‌های ناخواسته از قبیل نوفه و نور پس‌زمینه است. لذا برای رسیدن به مدل واسنجی دقیق و قابل اعتماد معمولاً پیش‌پردازش داده‌های طیفی قبل از مدل‌سازی ضرورت دارد. در این پژوهش تأثیر روش‌های هموارسازی، نرمال‌سازی، Detrending، توزیع نرمال استاندارد (SNV) و تصحیح پراکنش افزایشنده (MSC) به صورت انفرادی یا ترکیبی در دقت پیش‌بینی مدل حداقل مربعات جزئی (PLS) مورد بررسی قرار گرفتند. تعداد ۱۶۷ میوه انجیر برای تدوین و اعتبارسنجی مدل‌ها انتخاب شدند. از روش آنالیز مولفه‌های اصلی (PCA) برای استخراج مولفه‌های اصلی طیف‌ها استفاده شد. عملکرد مدل مربعات حداقل جزئی (PLS) و روش‌های رایج پیش‌پردازش داده‌های طیفی با شاخص‌های: انحراف پیش‌بینی باقیمانده (RPD)، ضریب همبستگی تخمین (rp) و جذر میانگین مربعات خطای پیش‌بینی (RMSEP) مورد ارزیابی قرار گرفت. بیشترین شاخص RPD در پیش‌بینی سفتی بافت برابر ۱/۷۹ (RMSEP = ۱/۶۴ و rp = ۰/۸۴۵) در پیش‌پردازش De-trending + هموارسازی بدست آمد.

### کلمات کلیدی:

پیش‌پردازش، رسیدگی انجیر، حداقل مربعات جزئی، طیف‌سنجی فروسرخ نزدیک موج کوتاه/مرئی، ارزیابی غیرمخرب.

\*نویسنده مسئول

## مقایسه برخی روش‌های پیش‌پردازش طیف مرئی/فروسرخ نزدیک موج کوتاه بر عملکرد مدل حداقل مربعات جزئی (مطالعه موردی: ارزیابی رسیدگی میوه انجیر)

### مقدمه

به طور کلی تحقیقات فزاینده‌ای درباره استفاده از روش‌های طیف‌سنجی نوری به همراه الگوریتم‌ها و روش‌های تحلیل اطلاعات طیفی در مقوله‌های مختلف علمی موجود است. اما سهم کشاورزی در این تحقیقات نسبت به حوزه‌های دیگر مانند شیمی، فیزیک، علوم زیستی و غیره کمتر و البته رو به افزایش است که از دلایل اصلی آن می‌توان به توسعه قابل توجه و روزافزون ابزار مورد نیاز طیف‌سنجی رامان از قبیل تکنولوژی لیزر، دکتورها، میکروسکوپ‌ها، فیبرهای نوری و فیلترها اشاره نمود که باعث ارتقاء اطلاعات حاصله از طیف مواد مختلف می‌شود [۱۸].

میوه انجیر ۱ بدلیل طعم جذاب، رنگ، ارزش‌های تغذیه‌ای و دارویی آن شناخته شده است. سفتی بافت انجیر بعنوان یکی از شاخص‌های کیفی آن برشمرده شده است [۲۴] و استفاده از این شاخص به تنهایی بعنوان معیار رسیدگی، در درجه بندی غیر مخرب برخی از میوه‌ها تحقیق شده است [۱۲ و ۲۳].

ناحیه ۱۱۰۰-۳۵۰ نانومتر از طیف الکترومغناطیسی نور بنام مرئی/فروسرخ نزدیک موج کوتاه شناخته می‌شود [۸]. به علت محتوای بالای آب میوه و سبزی‌ها (۹۰-۸۰ درصد)، باندهای جذبی مربوط به آب در طیف NIR آن‌ها غالب است [۱۰]. مرکز تقریبی باندهای جذبی مربوط به آب (ارتعاشات پایه و اورتون‌های O-H) در طول موج‌های ۹۷۰، ۱۲۰۰، ۱۴۵۰، ۱۹۵۰ و ۲۲۵۰ نانومتر گزارش شده است [۱۰ و ۲۶] لذا انتخاب ناحیه طیفی فروسرخ نزدیک موج کوتاه/مرئی در موضوعات متکی بر آنالیزهای کمی ترکیباتی از میوه‌ها که علائم طیفی آن‌ها در این باند قرار می‌گیرد از نظر کاهش اثر آب درطیف نمونه، انتخاب مناسبی به نظر می‌رسد.

اسپکتروسکوپی فروسرخ نزدیک/مرئی این امکان را فراهم می‌سازد که ارزیابی کیفی بصورت آسان، سریع، دقیق، غیرمخرب، ارزان و بدون نیاز به آماده‌سازی قبلی نمونه انجام شده و اندازه‌گیری‌های برخط بدلیل وجود پتانسیل توسعه ابزار اندازه‌گیری مورد نیاز نیز ممکن شود [۴]. پیشرفت فعلی و رو به رشد فناوری تجهیزات نوری در دنیا باعث کوچک‌تر شدن ابعاد و اجزای اسپکترومترها شده و کاربردهای عملی انواع اسپکترومترهای قابل حمل که با تعداد طول موج‌های کمتری کار می‌کنند توسط پژوهشگران بررسی شده است [۱۱، ۱۹ و ۲۷].

مونرا و همکاران [۱۴] در تحقیق خود کیفیت داخلی میوه خرمالو رقم Rojo Brillanta را با استفاده از مدل‌سازی داده‌های حاصل از تصویربرداری فراطیفی در محدوده مرئی و فروسرخ نزدیک (۱۰۲۰-۴۵۰ نانومتر) مورد بررسی قرار دادند. آنها نمونه‌های مورد بررسی را در سطوح مختلف دی اکسید کربن مورد تیمار قرار دادند تا نمونه‌هایی با درجه رسیدگی مختلف و سطوح متفاوت از گسی (تانن محلول) داشته باشند. از داده‌های طیفی در تدوین مدل

<sup>1</sup> *Ficus carica L*

های طبقه بندی برای تشخیص سطوح مختلف رسیدگی و گسی استفاده کردند. آنها همچنین از مدل رگرسیونی مربعات حداقل جزئی برای پیش بینی سفتی گوشت میوه استفاده کردند. طول موج های بهینه توسط روش آنالیز مولفه های اصلی ۵۸۰، ۶۸۰ و ۱۰۵۰ نانومتر بدست آورده شدند. دقت طبقه بندی مدل ها بالای ۹۲ درصد بدست آمد. مدل QDA در تفکیک سطح گسی دقت ۹۵ درصد را نتیجه داد. در طول موج های ذکر شده در تخمین درجه سفتی،  $R^2=0/8$  و RPD برابر ۱/۸۶ بدست آمدند. آنها جمع بندی نمودند که تصویربرداری فراطیفی بعنوان یک ابزار غیرمخرب در ارزیابی سفتی، رسیدگی و مقدار گسی میوه خرما لپو پتانسیل لازم را دارد.

در مورد تخمین مقدار سفتی میوه هلو با استفاده از اسپکتروسکوپی فرو سرخ نزدیک (۲۵۰۰-۸۰۰ نانومتر) به کمک روش مدلسازی مربعات حداقل جزئی بهترین نتیجه با پیش پردازش MSC برابر  $rcv=0/864$  و  $0/42=RMSEP$  گزارش شده است (Fu et al., 2008). در سنجش کمی سفتی میوه کیوی بیشترین عملکرد مدل شبکه های عصبی مصنوعی (ورودی: نسبت های غلظت عناصر تغذیه ای) بدست آمده  $rp=0/85$  و  $RMSEP=0/539$  گزارش شده است [۲۳].

الهمدان و همکاران [۱] با استفاده از طیف سنجی مرئی/فروسرخ نزدیک (۱۲۰۰-۲۸۵ نانومتر) ویژگی های بافتی و حسی خرماي Khalal Barhi را در شرایط انبار شده ارزیابی نمودند. آنها نمونه های خود را در سه سطح رسیدگی (سبز، زرد و قهوه ای) فراهم نمودند. نمونه ها را به دو قسمت تقسیم کردند گروه اول برای بررسی در شرایط میوه تازه و گروه دوم سه شرایط محیطی مختلف تا ۱۲۰ روز انبار شدند. نتایج گزارش شده حاکی از عملکرد بسیار خوب مدل رگرسیون مولفه های اصلی در پیش بینی سفتی بافت و چسبندگی (پارامترهای بافتی) بوده است اما مدل مذکور در پیش بینی خاصیت جوندگی عملکرد ضعیفی داشته است. لذا عنوان نمودند که این نتایج در فرآیند درجه بندی برخط تجاری بر پایه ویژگی های کیفی با بکارگیری ابزار اندازه گیری مبتنی بر طیف های فروسرخ نزدیک کاربردی است.

کاربرد اسپکتروسکوپی فروسرخ نزدیک/مرئی در سنجش کمی و کیفی میوه ها و سبزیجات مختلف، به تفکیک محصول و جزئیات سامانه طیف سنجی بکار رفته، در مقالات [۲، ۴، ۹ و ۲۵] ارائه شده است.

هدف از انجام تحقیق حاضر مقایسه روش های پیش پردازش طیف های بازتابی فروسرخ نزدیک موج کوتاه/مرئی (۴۲۵-۹۵۰ نانومتر) در دقت پیش بینی مقدار سفتی بصورت غیرمخرب، بعنوان شاخص رسیدگی میوه انجیر بوده است.

#### مواد و روش ها

##### نمونه ها

نمونه های میوه انجیر سیاه (رقم محلی) در سه دسته ۶۰ عددی در سه مرحله زمانی رسیدگی کامل، نیمه رسیده و نرسیده و به تعداد کل ۱۸۰ نمونه از یک درخت در باغی واقع در منطقه هزاوه (استان مرکزی) تهیه شدند. با این

روش نمونه برداری، تنوع مورد نیاز نمونه ها در مقدار شاخص های سفتی و مزه ایجاد شد.

### سامانه طیف سنجی و طیف برداری

اجزای سامانه طیف سنجی به کار رفته شامل: اسپکترومتر از نوع قابل حمل (Flame سری Ocean optics)، دارای گذرگاه USB، ابعاد  $31/9 \times 63/5 \times 88/9$  میلیمتر، وزن ۲۶۵ گرم، محدوده کاری ۱۰۰۰-۳۵۰ نانومتر، تعداد پیکسل ۲۰۴۸، نسبت سیگنال به نویز ۲۵۰ به ۱ و تفکیک نوری ۰/۱ تا ۱۰ نانومتر، مجهز به دکتور CCD (Sony) و ILX511B linear silicon)، لامپ هالوژن (۱۲ ولت) به عنوان منبع نور، فیبرهای نوری (مدل QP400) و نگهدارنده نمونه و پراب برهمکنشی (Intractance) بوده است. طیف برداری از میوه ها در حالت های متفاوت موقعیت منبع نور و دکتور نسبت به یکدیگر از قبیل بازتابی، عبوری و برهم کنشی<sup>۲</sup> (دکتور و منبع نور بصورت موازی با هم قرار دارند) قابل انجام است [۸ و ۱۵]. در این تحقیق از مد طیف برداری برهمکنشی و با استفاده از رابطه ۱ [۳] درصد بازتاب نور در هر طول موج اخذ و ثبت شد.

$$R (\%) = 100(R_S - R_D) / (R_R - R_D) \quad (1)$$

که در آن R درصد بازتاب نور، RS شدت نور حاصل از نمونه، RD شدت نور محیط و RR شدت نور ماده مرجع سفید است.

در هر نمونه در سه موقعیت مرکزی آن به فاصله تقریبی ۱۲۰ درجه از یکدیگر و ۱۲ اسکن در هر موقعیت به وسیله نرم افزار Spectrasuite طیف برداری شد. برای کاهش نویزهای ناشی از دمای دکتور معمولاً در فرآیند طیف برداری، میانگین اسکن ها در نظر گرفته می شود [۱۵] لذا در مراحل بعدی میانگین ۳۶ اسکن به عنوان طیف هر نمونه انجیر در نظر گرفته شد.

### اندازه گیری سفتی

برای اندازه گیری سفتی بافت میوه انجیر از دستگاه سفتی سنج (STEP SYSTEM, Germany) با پیستون با قطر ۸ میلی متر استفاده شد. پس از برداشتن اپیدرم در دو ناحیه در مرکز میوه، از نوک پیستون ۵ میلی متری برای اندازه گیری سفتی میوه ها استفاده شد و قرائت ها به صورت کیلوگرم بر سانتی متر مربع ثبت شدند.

### شیمی سنجی<sup>۳</sup>

<sup>2</sup> Intractance

<sup>3</sup> Chemometrics

## پیش‌پردازش

پیش‌پردازش‌های متفاوتی برای داده‌های طیفی ابداع شده و در حال توسعه‌اند که عمدتاً هر کدام کاربری خاصی دارند و معمولاً هدف خاصی را دنبال می‌کنند. با این وجود نمی‌توان یک روش پیش‌پردازش مشخص را به‌عنوان بهترین روش معرفی کرد چرا که در تحقیقات مختلف و متنوع شرایط نمونه و آزمایش تغییر می‌نمایند. بر این اساس، در این تحقیق پس از اخذ طیف‌ها، از روش‌های معمول پیش‌پردازش استفاده شد [۱۵]. میانگین‌گیری متحرک از معمول‌ترین روش‌های هموارسازی است که در تحقیقات مشابه در مورد میوه‌های دیگر بصورت ترکیبی با روش‌های پیش‌پردازش مختلف مورد استفاده قرار گرفته است [۲۵]. از گروه پیش‌پردازش‌های نرمال‌سازی، توزیع نرمال استاندارد  $\epsilon$  و تصحیح پراکندگی افزایشنده بطور کلی برای تصحیح اثر پراکندگی نور در طیف‌های اخذ شده استفاده شد [۲۲].

## آنالیز مولفه‌های اصلی<sup>۴</sup>

آنالیز مولفه‌های اصلی روشی است که باعث می‌شود اطلاعات مفید از داده‌ها استخراج شود، داده‌ها، متغیرها و روابط بین آنها لکاو شده و همبستگی کلی بین متغیرها فهمیده شود [۲]. در این روش متغیرهای جدیدی ایجاد می‌شود که هر کدام یک ترکیب خطی از متغیرهای اصلی هستند. این متغیرهای جدید مولفه‌های اصلی نامیده می‌شوند و بیشترین اطلاعات (واریانس) موجود در ماتریس متغیرهای اصلی را در خود دارند. اولین مولفه اصلی بیشترین واریانس را دربر دارد در حالی که مولفه اصلی دوم اطلاعاتی را در خود دارد که توسط مولفه اصلی اول ارائه نشده است و این رویه به همین شکل در مولفه‌های بعدی نیز وجود دارد. مولفه‌های اصلی اول تا چهارم بعنوان ورودی مدل پیش‌بینی در نظر گرفته شدند.

## مدلسازی

در آنالیز رگرسیون مولفه‌های اصلی<sup>۶</sup> ترتیب مولفه‌های اصلی (نهفته) بر مبنای کاهش مقدار واریانس پوششی توسط آن‌ها در ماتریس داده‌های طیفی است. اشکال موجود اینست که اولین مولفه‌های اصلی لزوماً موثرترین (پر اطلاعات‌ترین) مولفه‌ها در ارتباط با متغیر هدف نیستند. تکنیک رگرسیون حداقل مربعات جزئی (PLS) برای غلبه بر این مشکل ارائه شده است. در این روش متغیرهای نهفته که نسبت به هم متعامد هستند به گونه‌ای قرار می‌گیرند که کوواریانس بین ماتریس ورودی و بردار خروجی بیشینه گردد، در این روش اطمینان حاصل می‌شود که چیدمان متغیرهای نهفته (یا همان مولفه‌های اصلی) مطابق با اهمیت آنها در پیش‌بینی متغیر خروجی است [۱۵]. برای پیش‌بینی مقدار سفتی بافت از مدل PLS استفاده شد.

<sup>4</sup> Standard Normal Variate (SNV)

<sup>5</sup> Principle Components Analysis (PCA)

<sup>6</sup> Principle Components Regression (PCR)

### ارزیابی مدل

تدوین مدل بدون آزمون آن معنی ندارد لذا حدود ۲۷ درصد از کل طیف‌ها قبل از مدل‌سازی توسط نرم‌افزار بطور تصادفی برای اعتبارسنجی کنار گذاشته شدند سپس تدوین مدل (پیش‌بینی و طبقه‌بندی) با داده‌های باقیمانده انجام شد. توانایی پیش‌بینی مدل PLS، با استفاده از شاخص‌های: ریشه میانگین مربعات خطای پیش‌بینی (RMSEP)، ضریب همبستگی پیش‌بینی (rp) و انحراف پیش‌بینی باقیمانده (روابط ۲-۵) ارزیابی شد [۲، ۵، ۷، ۲۱ و ۲۵].

$$RMSEP = \sqrt{\sum_{i=1}^n (Y_i - X_i)^2 / n_p} \quad (2)$$

$$r_p = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / (\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}) \quad (3)$$

$$RPD = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / n_p - 1} / RMSEP = SD / RMSEP \quad (4)$$

که در آن  $X_i$  مقدار اندازه‌گیری شده متغیر نمونه  $i$ ،  $Y_i$  مقدار پیش‌بینی شده متغیر مورد نظر برای نمونه  $i$  و  $n_p$  تعداد نمونه‌های پیش‌بینی،  $\bar{X}$  میانگین مقدارهای اندازه‌گیری شده،  $SD$  انحراف استاندارد مقدارهای اندازه‌گیری شده در دسته نمونه‌های پیش‌بینی است. از نرم‌افزار اسپکتراسوییت<sup>۷</sup> در طیف برداری و برای آنالیز مولفه‌های اصلی، اعمال پیش‌پردازش‌های طیفی و مدل‌سازی از نرم‌افزار آن اسکرملر<sup>۸</sup> (نسخه ۱۰/۴) استفاده شد.

### تحلیل نتایج

جدول ۱ مشخصات آماری نمونه‌های انجیر استفاده شده در تحقیق پس از حذف داده‌های پرت را نشان می‌دهد.

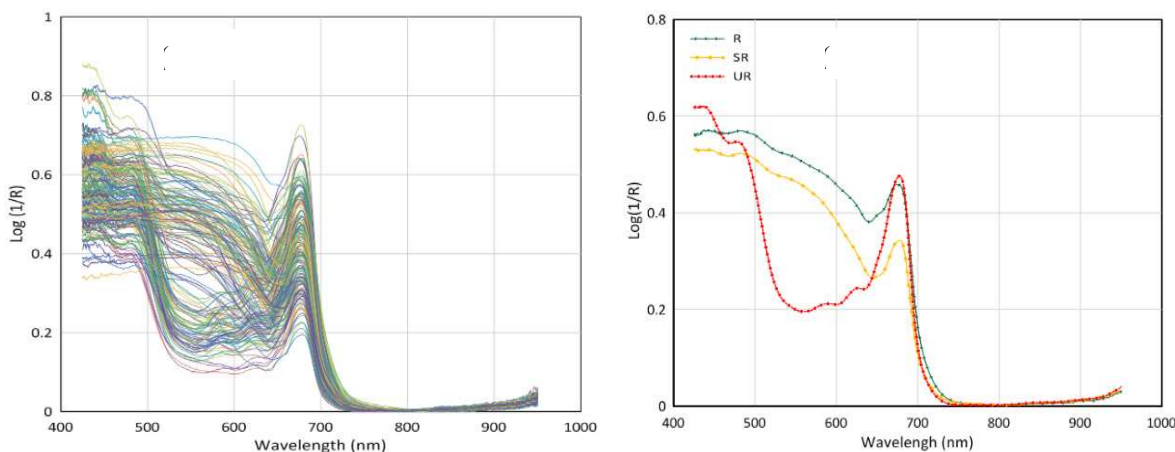
جدول ۱- مشخصات آماری دسته نمونه‌های آموزش و آزمون

دسته آموزش (۲۲ نمونه)				دسته آزمون (۴۵ نمونه)			
کمینه	بیشینه	میانگین	انحراف	کمینه	بیشینه	میانگین	انحراف
Min	Max	Mean	استاندارد	Min	Max	Mean	استاندارد
2.3	15.03	5.85	2.94	2.17	14.9	6.24	2.94
سفتی بافت (کیلوگرم بر سانتیمتر مربع)							

<sup>7</sup> Spectrasweet

<sup>8</sup> Unscrambler

طول موج های ابتدا و انتهای طیف های حاصل از نمونه ها بدلیل وجود نویز قابل توجه حذف شدند و آنالیز داده های طیفی در محدوده ۴۲۵ تا ۹۵۰ نانومتر به انجام رسید. در شکل (۱) قسمت A کل طیف های جذبی اخذ شده و در قسمت B میانگین طیف های جذبی نمونه های انجیر رسیده، نیمه رسیده و نرسیده در محدوده ۴۲۵-۹۵۰ نانومتر دیده می شود (طول موج های ابتدایی و انتهایی به دلیل نویز زیاد حذف شدند). شکل نشان می دهد که در بیشتر طول موج ها میانگین جذب نور در نمونه انجیرهای نرسیده کمتر از بقیه نمونه ها دیده می شود در توضیح این مساله می توان بیان کرد که اصولاً با تکمیل مراحل رسیدگی میوه تجزیه پکتین ها و هموسلولوزهای داخل میوه با آنزیم های خاصی شروع می شود و این باعث تغییر دیواره سلولی، نرم شدن، شفافیت میوه، افزایش عمق نفوذ پرتوهای نور و به طبع آن کاهش درصد انعکاس نور (افزایش جذب) و افزایش تغییرات در پخش پرتوهای نوری می شود (Cavaco *et al.*, 2009; Pourdarbani *et al.*, 2020).



شکل ۱- (A) کل طیف های اخذ شده در محدوده ۴۲۵-۹۵۰ نانومتر و (B) طیف های میانگین انجیرهای رسیده (R)، نیمه رسیده (SR) و نرسیده (UR)

شکل نمودارها نشان می دهد که الگوی جذبی طیف ها در ۶۸۰-۴۵۰ نانومتر متفاوت بوده و در محدوده باقیمانده الگوی جذبی نمونه ها بسیار مشابه است در چنین مواردی برای استخراج اطلاعات مفید احتمالی طیف ها ضرورت بکارگیری پیش پردازش های مناسب نمایان می شود (Teye *et al.*, 2019). از طرفی برای حذف اطلاعات ناخواسته (نویز)، طیف های خام معمولاً با روش های پیش پردازش مناسب اصلاح می شوند.

مدل PLS

در تدوین یک مدل پیش بینی کننده اگر مقدار  $RPD < 1/5$  باشد مدل قابل استفاده نبوده و با  $RPD < 2$  تخمین مقدارهای کم و زیاد پارامتر مورد نظر امکان پذیر خواهد بود و چنانچه  $RPD < 2/5$  امکان

سنجش کمی توسط مدل وجود داشته و نهایتاً  $RPD = 2/5$  و  $RPD < 3$  به ترتیب مبین دقت پیش بینی خوب و عالی هستند (Beghi et al., 2017). در حدود پایین  $RPD$  (کوچکتر از ۲) تفاوت های جزئی در بیان شرایط مدل ارائه شده است بعنوان مثال بدست آمدن  $1/8 < RPD < 1/4$  نشانه نسبتاً خوب بودن مدل و با  $2 < RPD < 1/8$  سنجش کمی امکان پذیر دانسته شده است (Mouazen et al., 2010; Rossel et al., 2006).

جدول ۲- نتایج مدل PLS تدوین شده در پیش بینی مقدار سفتی بر پایه ترکیب روش های مختلف پیش پردازش

RPD	$r_p$	RMSEP	$r_c$	RMSEC	LVs	پیش پردازش
1.78	.846	1.64	.856	1.51	7	No Preprocessing
1.69	.825	1.73	.876	1.41	8	MA+SNV
1.71	.827	1.72	.870	1.44	7	MA+MSC
1.60	.821	1.74	.873	1.42	7	Normalize
1.79	.845	1.64	.853	1.52	7	MA+De-trending

در پیش بینی سفتی،  $RPD = 1/79$  و  $r_p = 0/845$  و  $RMSEP = 1/64$  و در پیش پردازش MA+De-trending بدست آمد.

#### نتیجه گیری

در طیف های فرسرخ نزدیک موج کوتاه / مرئی (۴۲۵-۹۵۰ نانومتر) اخذ شده از نمونه های انجیر تعداد پیک های کمی مشاهده شد که علت آن جذب نور توسط اورتون ها و ترکیب ارتعاشات پیوندهای پایه ای مواد ارگانیک تشکیل دهنده میوه انجیر است. الگوی انعکاسی طیف های اخذ شده در طول موج های زیادی به یکدیگر شباهت داشته و لذا عملکرد پیش پردازش های MA+MSC، Normalize، MA+SNV و MA+De-trending برای استخراج اطلاعات مفید موجود احتمالی در طیف ها بررسی شد. نتایج ارزیابی دسته نمونه های آزمون نشان داد که می توان مدل PLS را برای سنجش مقادیر زیاد و کم سفتی میوه بکار برد ( $RPD = 1/79$ ).

#### منابع

1. Alhamdan, A.M., Fickak, A. and Atia, A.R. 2019. Evaluation of sensory and texture profile analysis properties of stored Khalal Barhi dates nondestructively using Vis/NIR spectroscopy. Journal of Food Process Engineering, 42(6): e13215.
2. Beghi, R., Buratti, S., Giovenzana, V., Benedetti, S. and Guidetti, R. 2017. Electronic nose and visible-near infrared spectroscopy in fruit and vegetable monitoring. Reviews in Analytical Chemistry, 36(4): 20160016.
3. Cavaco, A.M., Pinto, P., Antunes, M.D., da Silva, J.M. and Guerra, R. 2009. 'Rocha' pear firmness predicted by a Vis/NIR segmented model. Postharvest Biology and Technology, 51(3): 311-319.



4. Cortés, V., Blasco, J., Aleixos, N., Cubero, S. and Talens, P. 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends in Food Science & Technology*, 85: 138-148.
5. Forthofer, R.N., Hernandez, M. and Lee, E.S. 2007. *Biostatistics: a guide to design, analysis and discovery*. 349-386.
6. Fu, X.-p., Ying, Y.-b., Zhou, Y., Xie, L.-j. and Xu, H.-r. 2008. Application of NIR spectroscopy for firmness evaluation of peaches. *Journal of Zhejiang University SCIENCE B*, 9(7): 552-557.
7. Guo, Y., Ni, Y. and Kokot, S. 2016. Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 153: 79-86.
8. Jamshidi, B. 2020. Ability of near-infrared spectroscopy for non-destructive detection of internal insect infestation in fruits: Meta-analysis of spectral ranges and optical measurement modes. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 225: 117479.
9. Lin, H. and Ying, Y. 2009. Theory and application of near infrared spectroscopy in assessment of fruit quality: a review. *Sensing and instrumentation for food quality and safety*, 3(2): 130-141.
10. Magwaza, L.S., Opara, U.L., Nieuwoudt, H., Cronje, P.J., Saeys, W. and Nicolai, B. 2012. NIR spectroscopy applications for internal and external quality analysis of citrus fruit—a review. *Food and Bioprocess Technology*, 5(2): 425-444.
11. Maraphum, K., Saengprachatanarug, K., Wongpichet, S., Phuphaphud, A. and Posom, J. 2020. In-field measurement of starch content of cassava tubers using handheld vis-near infrared spectroscopy implemented for breeding programmes. *Computers and Electronics in Agriculture*, 175: 105607.
12. Matteoli, S., Diani, M., Massai, R., Corsini, G. and Remorini, D. 2015. A spectroscopy-based approach for automated nondestructive maturity grading of peach fruits. *IEEE Sensors Journal*, 15(10): 5455-5464.
13. Mouazen, A., Kuang, B., De Baerdemaeker, J. and Ramon, H. 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, 158(1-2): 23-31.
14. Munera, S., Besada, C., Aleixos, N., Talens, P., Salvador, A., Sun, D.-W., Cubero, S. and Blasco, J. 2017. Non-destructive assessment of the internal quality of intact persimmon using colour and VIS/NIR hyperspectral imaging. *LWT*, 77: 241-248.
15. Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. and Lammertyn, J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2): 99-118.
16. Pourdarbani, R., Sabzi, S., Kalantari, D. and Arribas, J.I. 2020. Non-destructive visible and short-wave near-infrared spectroscopic data estimation of various physicochemical properties of Fuji apple (*Malus pumila*) fruits at different maturation stages. *Chemometrics and Intelligent Laboratory Systems*, 206: 104147.
17. Rossel, R.V., McGlynn, R. and McBratney, A. 2006. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 137(1-2): 70-82.

18. Slater, J.B., Tedesco, J.M. and Fairchild, R.C. 2001. Raman Spectrometry and Its Adaptation to the Industrial Environment. Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line: 41 .
19. Szuvandzsiev, P., Helyes, L., Lugasi, A., Szántó, C., Baranowski, P. and Pék, Z. 2014. Estimation of antioxidant components of tomato using VIS-NIR reflectance data by handheld portable spectrometer. International Agrophysics, 28(4). (
20. Teye, E., Amuah, C.L., McGrath, T. and Elliott, C. 2019. Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 217: 147-154.
21. Theanjumpol, P., Wongzeewasakun, K., Muenmanee, N., Wongsaipun, S., Krongchai, C., Changrue, V., Boonyakiat, D. and Kittiwachana, S. 2019. Non-destructive identification and estimation of granulation in 'Sai Num Pung' tangerine fruit using near infrared spectroscopy and chemometrics. Postharvest Biology and Technology, 153: 13-20.
22. Tiecher, T., Moura-Bueno, J.M., Caner, L., Minella, J.P., Evrard, O., Ramon, R., Naibo, G., Barros, C.A., Silva, Y.J. and Amorim, F.F. 2020. Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet-visible, near-and middle-infrared spectroscopy. Geoderma, 384: 114815.
23. Torkashvand, A.M., Ahmadi, A. and Nikraves, N.L. 2017. Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). Journal of integrative agriculture, 16(7): 1634-1644 .
24. Tuai, P., Tinjauan, S., JUSOH, N.A.M., DING, P. and YEAT, C.S. 2020. Extending Post-Harvest Quality of Fresh Fig (*Ficus carica* L.) Fruit Through Manipulation of Pre-and Post-Harvest Practices: A Review. Sains Malaysiana, 49(3): 553-560.
25. Wang, H., Peng, J., Xie, C., Bao, Y. and He, Y. 2015. Fruit quality evaluation using spectroscopy technology: a review. Sensors, 15(5): 11889-11927.
26. Williams, P.C. and Norris, K. 2001. Variables affecting near-infrared spectroscopic analysis. Near-infrared technology in the agricultural and food industries: 171-189.
27. Yan, H. and Siesler, H.W. 2018. Identification of textiles by handheld near infrared spectroscopy: Protecting customers against product counterfeiting. Journal of Near Infrared Spectroscopy, 26(5): 311-321 .

## Comparison of some Vis/SWNIR preprocessing methods on the performance of the partial least squares model (Case study: evaluation of fig fruit ripeness)

Reza Mohammadigol<sup>1\*</sup>, Babak Valizadehkaji<sup>2</sup>, Majid Lashgari<sup>3</sup> and Reza Sayad Haghshomar<sup>4</sup>

1. Department of Biosystems Engineering, Arak University, Arak, Iran
2. Department of Horticultural Engineering, Arak University, Arak, Iran
3. Department of Biosystems Engineering, Arak University, Arak, Iran
4. Department of Horticultural Engineering, Arak University, Arak, Iran

### Abstract

Vis/NIR spectra have included useless data such as noise and background light in addition to the sample molecular structure information. Therefore, to achieve accurate and reliable calibration models, before modeling, usually spectral data preprocessing is necessary. In this study, the effect of smoothing, normalization, Detrending, standard normal variation (SNV) and multiplicative scatter correction (MSC) individually or in combination to each other on the accuracy of Partial least squares (PLS) modeling techniques were studied. A total of 167 fig fruits were selected for the development and validation of the models. Principal component analysis (PCA) was employed to extract the principal components of the spectra. PLS performance and common spectral data pretreatment methods were evaluated using the residual prediction deviation (RPD), predictive correlation coefficient (rp), and root mean square error of prediction (RMSEP). Based on the combined pretreatment method of MA + De-trending, the highest mean value of RPDs, 1.79 (RMSEP = 0.845, rp = 1.64) was achieved.

**Key words:** Preprocessing, Fig fruit ripeness, Partial least squares, Vis/SWNIR spectroscopy, non-destructive evaluation.

\*Corresponding author

E-mail: r-mohammadigol@araku.ac.ir