

طیف‌سنجی مرئی و مادون قرمز نزدیک برای تعیین حجم آب میوه انار

محمد جمالی فردا، سید مهدی نصیری^۱، علی محمد شیرزادی فر^۲، علیرضا شمسوار^۳

^۱ دانشجوی کارشناسی ارشد، بخش مهندسی بیوسیستم، دانشگاه شیراز (mohammad.jamali.fard1@gmail.com)

^۲ دانشیار بخش مهندسی بیوسیستم، دانشگاه شیراز (nasiri@shirazu.ac.ir)

^۳ استادیار بخش مهندسی بیوسیستم، دانشگاه شیراز (shirzadifar@yahoo.com)

^۴ دانشیار بخش علوم باغبانی، دانشگاه شیراز (shahsava@shirazu.ac.ir)

چکیده

میوه انار یکی از محصولات باغی مهم در ایران است که علاوه بر ارزش بالای اقتصادی-تجاری از فواید غذایی فراوانی نیز برخوردار است. به دلیل عدم توسعه مناسب صنعت در بخش فرآوری این محصول، کارخانه‌ها از روش‌های کاملاً سنتی و محدود در سنجش کیفیت انار استفاده می‌کنند که باعث کاهش مرغوبیت کالای تولیدی می‌شود. در این پژوهش استفاده از فناوری غیرمخرب طیف‌سنجی مرئی-مادون قرمز نزدیک برای تعیین حجم آب میوه انار که از ویژگی‌های اساسی کیفیت این محصول است مورد بررسی قرار گرفت. داده‌های طیفی نمونه‌ها حاصل از اعمال امواج در دامنه ۴۰۰ تا ۲۵۰۰ نانومتر علاوه بر حالت پردازش نشده به پنج روش تصحیح پراکندگی ضربی (MSC)، متغیر نرمال استلندارد (SNV)، نرمال‌سازی بردار (VN)، مشتق اول (FD) و مشتق دوم (SD) پیش‌پردازش شدند و مورد ارزیابی قرار گرفتند. سپس به منظور تخمین حجم آب از رگرسیون حداقل مربعات جزئی (PLSR) استفاده شد. پیاده‌سازی این فرآیندها در قالب الگوریتم‌های یادگیری ماشین و با استفاده از نرم افزار PYTHON 3.8 صورت گرفت. نتایج پیش‌بینی حجم آب ضریب تعیین ۹۸ درصد را نشان داد که از ترکیب پیش‌پردازش SNV با رگرسیون PLSR حاصل شد. وابستگی نتایج به روش واکاوی داده‌های طیف‌سنجی مورد تأیید قرار گرفت.

کلمات کلیدی:

انار- طیف‌سنجی- یادگیری ماشین- روش غیرمخرب

*نویسنده مسئول: nasiri@shirazu.ac.ir

طیف‌سنجی مرئی و مادون قرمز نزدیک برای تعیین حجم آب میوه انار

مقدمه

انار (*Punica granatum* L) میوه‌ای از خانواده (Punicaceae) است که در مناطق گرمسیری و نیمه گرمسیری به عمل می‌آید و به دلیل مرغوبیت بالا، از نظر صادرات، محصولی بی‌همتا بوده و از نظر اقتصادی ارزش بالایی دارد. در آسیای مرکزی درخت انار به صورت وحشی در مناطق ایران، ترکمنستان، و شمال هند رشد می‌کند و به طور کلی مناطقی که دارای آب و هوای مدیترانه‌ای با نور زیاد خورشید، زمستان‌های ملایم با حداقل دمای ۱۲- درجه سلسیوس، تابستان‌های گرم و خشک و بارندگی در مراحل آخر نمو میوه هستند، بهترین اقلیم برای پرورش میوه انار به حساب می‌آیند. علاوه بر کشورهای یاد شده میوه انار در کشورهای زیادی نظیر پاکستان، ارمنستان، افغانستان، ازبکستان، تاجیکستان، آذربایجان، سوریه، لبنان، لیبی، سودان، تونس، بنگلادش، مصر، مراکش، اسپانیا، ایتالیا، آلمان، فرانسه، یونان، چین، روسیه، آمریکا و استرالیا تولید می‌شود که بیانگر پراکندگی بالای این محصول در سطح جهان و در نتیجه اهمیت اقتصادی فراوان آن است [۱]. طبق گزارش مرکز آمار وزارت جهاد کشاورزی (۱۳۹۸)، ایران با تولید سالانه ۹۱۷۵۲۸ تن انار در سال ۱۳۹۷ سومین تولید کننده برتر این میوه در جهان بوده است که این آمار بیانگر اهمیت ویژه تجاری این محصول در کشور ایران است.

علاوه بر پراکندگی و فراوانی انار در ایران و جهان، این میوه از لحاظ خواص و فواید نیز جایگاه ویژه‌ای در بین میوه‌ها دارد که این خود دلیلی بر اهمیت آن است. یافته‌های اخیر محققان نشان می‌دهد که از انار می‌توان به عنوان یک گیاه دارویی بهره برد. بافت، گل، پوست و برگ میوه انار دارای مواد شیمیایی زیست‌فعال است که ضد میکروب بوده و فشار خون را کاهش می‌دهد و همچنین در درمان بیماری‌هایی نظیر دیابت و سرطان مؤثر است [۱۴]. به علاوه میوه انار با دارا بودن مقادیر قابل توجهی از مواد جامد محلول، اسیدهای آلی، ویتامین‌ها، پلی‌ساکاریدها، عناصر معدنی، اسیدهای چرب و ویتامین C، ارزش غذایی بسیار بالایی دارد [۳].

مصرف تازه‌خوری انار به دلیل عمر ماندگاری پایین آن در تمام ماه‌های سال امکان‌پذیر نیست. بر همین اساس تولید فرآورده‌هایی که قابلیت نگهداری طولانی مدت داشته‌باشند اهمیت ویژه‌ای دارد. همچنین از آن‌جا که بخش قابل توجهی از محصول انار برداشت شده علاوه بر داشتن کیفیت مناسب آب و دانه به دلیل آفتاب‌سوختگی، اندازه کوچک و یا رنگ نامطلوب، قابلیت فروش و عرضه مستقیم به بازار را ندارد و ارزش تجاری آن پایین است، اهمیت تولید محصولات فرآوری شده این میوه را دوچندان می‌کند. افزون بر آن، با وجود محدودیت‌هایی نظیر هزینه بالای انبارداری و سردخانه، فاسد شدن و کاهش حجم انار در گذر زمان، خشکی پوست و سایر موارد، ارزش صادراتی این محصول همواره در خطر است که لزوم توجه به فرآوری آن را ضروری می‌سازد. بنابر دلایل ذکر شده امروزه فرآوری میوه انار به شدت مورد توجه کارخانه‌های فرآوری قرار گرفته و بستری بسیار مناسب برای انجام فرآیندهای متفاوت بر روی آن به وجود آمده است. انار از جمله میوه‌هایی است که قابلیت فرآوری و تبدیل بالایی دارد. محصولات بسیار متنوع و زیادی از جمله آبمیوه، رب، معجون، مارمالاد، مربا، ژله، انواع سس و چاشنی، پودر شربت، ترشی، بستنی، نوشابه، پکتین، روغن، اسانس، انار دانه، لواشک، سرکه و خوراکی دام، برخی از فرآورده‌های انار هستند.

عدم توسعه و بروزرسانی دستگاه‌های فرآوری میوه انار و همچنین حساسیت میوه انار به آسیب‌های مکانیکی در حین فرآوری، موجب شده است که مراحل گزینش، خرید و ارزیابی کیفیت این محصول توسط کارخانه‌های

فرآوری به صورت عینی و در سطح کاملاً محدود و با استفاده از روش‌های سنتی صورت گیرد. اگرچه ارزش تجاری انار وابسته به ویژگی‌های ظاهری آن نظیر اندازه، شکل، و رنگ پوست است اما این ویژگی‌های ظاهری همواره بیانگر کیفیت درونی این میوه نیستند [۱۳]. این موضوع باعث می‌شود تا میوه‌های نامناسب به راحتی وارد چرخه تولید شوند و در نتیجه کیفیت محصول نهایی به میزان قابل توجهی کاهش یابد. همچنین کارخانه‌ها نیز بابت خرید محصول بی کیفیت و یا کم کیفیت متحمل ضررهای اقتصادی می‌شوند، از این رو صنعت فرآوری انار نیازمند یک سیستم ارزیابی کیفیت با سرعت و دقت بالا، هزینه کم و غیرمخرب است.

امروزه روش‌های متنوعی برای کنترل کیفیت میوه‌ها در دنیا وجود دارد که در این بین روش‌های غیرمخرب به دلیل توانایی در تشخیص ویژگی‌های مختلف محصول بدون آسیب رساندن به آن‌ها از اهمیت ویژه‌ای برخوردار هستند. برخی از این روش‌ها شامل طیف سنجی مرئی-مادون قرمز نزدیک^۱ [۱۵]، تصویربرداری چند طیفی و فراطیفی^۲ [۱۹]، تصویربرداری رزونانس مغناطیسی هسته‌ای^۳ [۵]، توموگرافی محاسباتی با استفاده از اشعه ایکس^۴ [۱۶] و طیف سنجی دی‌الکتریک^۵ [۸] هستند. روش‌های طیف سنجی مانند مرئی، مادون قرمز نزدیک، مادون قرمز میانه^۶ و طیف سنجی فراطیفی^۷ غیر مخرب‌تر از روش‌های مکانیکی و شیمیایی هستند و داشتن ابزارهای ساده، سریع، قابل حمل و فراگیر، از دیگر فواید استفاده از این روش‌ها هستند. بنابر همین دلایل امروزه از این فناوری به شکل گسترده‌ای، در صنایع کشاورزی، غذا، دارو و پتروشیمی استفاده می‌شود [۱۸].

طیف سنجی مرئی-مادون قرمز نزدیک ثابت کرده است که قادر به ارائه اطلاعات طیفی بسیار گسترده‌ای در مورد ویژگی‌ها و خواص فیزیکی و شیمیایی از درون و بیرون طیف وسیعی از مواد غذایی از جمله میوه‌ها است چرا که پرتوهای ساطع شده مادون قرمز نزدیک می‌توانند با پیوندهای ملکولی اساسی و مرتعش اکسیژن-هیدروژن، کربن-هیدروژن و نیتروژن-هیدروژن فعل و انفعال کرده و اطلاعاتی را در مورد پراکندگی ریز ساختارهای ماده گزارش دهند که به طور غیر مستقیم بیانگر مشخصات فیزیکی و شیمیایی آن ماده هستند [۴]. در پژوهشی با استفاده از طیف‌های بازتابی و عبوری در بازه مرئی و مادون قرمز نزدیک با طول موج ۴۰۰ تا ۱۱۰۰ نانومتر سه شاخص مقدار مواد جامد محلول، اسیدیته و سفتی میوه انار بررسی شد. نتایج نشان داد که استفاده از هر دو حالت طیف‌های بازتابی و عبوری برای ارزیابی شاخص‌های بیان شده امکان پذیر است. هر چند که طیف‌های بازتابی نتیجه دقیق‌تری را به دنبال داشت. به علاوه مدل رگرسیون حداقل مربعات جزئی به عنوان روش مناسب برای تخمین شاخص‌های مد نظر معرفی شد [۱۱].

بررسی‌های میدانی انجام شده از کارخانه‌های فرآوری میوه انار نشان داد این مراکز از روش‌های سنتی و تجربی همراه با خطا برای تشخیص محتوای آب نمونه‌های خریداری شده استفاده می‌کنند. بنابراین در پژوهش حاضر امکان سنجی استفاده از فناوری غیرمخرب طیف سنجی مرئی-مادون قرمز نزدیک برای تخمین محتوای آب میوه انار به عنوان هدف اصلی مورد بررسی قرار گرفت. نتیجه این پژوهش ارائه الگوریتم مناسب بر پایه طیف سنجی مرئی-مادون قرمز نزدیک برای تخمین سریع توام با دقت و غیرمخرب بود.

¹ (Visible/Near Infrared) Spectroscopy

² Multispectral/Hyperspectral Imaging

³ Nuclear Magnetic Resonance (NMR) Imaging

⁴ X-ray Computed Tomography

⁵ Dielectric Spectroscopy

⁶ Mid Spectroscopy

⁷ Hyperspectral Spectroscopy

سلمانی زاده و همکاران با استفاده از فناوری غیرمخرب اشعه ایکس سعی کردند تا برخی مؤلفه‌های بیوفیزیک دو رقم میوه انار مانند حجم و وزن میوه، آب میوه، آریل، پوست و هسته و نسبت‌هایی نظیر نسبت حجمی آریل به میوه، نسبت حجمی آب به میوه و نسبت حجمی هسته به میوه را اندازه‌گیری کنند. نتایج نشان داد که عامل رقم اثر معناداری بر مؤلفه‌های بیان شده به جز وزن و حجم میوه دارد. همچنین عامل اندازه اثر معناداری بر برخی از مؤلفه‌ها داشته است. بهترین ضرائب تبیین روابط تخمین برای حجم کل میوه بدست آمد که به ترتیب برای ارقام رباب ملس و رباب ترش ۹۷/۸ و ۹۷/۳ درصد محاسبه گردید و به طور کل نتایج تجزیه و تحلیل رگرسیون و آزمون چاو نشان داد که روش غیرمخرب سی تی اسکن به خوبی می‌توان حجم انار و اجزاء درونی آن را تخمین بزند [۱۶].

مواد و روش‌ها

تهیه نمونه

میوه‌های انار شامل رقم‌های رباب، اتابکی، شیرین شهوار و خفری استفاده شده در این پژوهش از مجموعه باغ‌های قصردشت واقع در شیراز، استان فارس تهیه شدند. رقم‌های تهیه شده از رایج‌ترین رقم‌های تجاری موجود در استان فارس هستند که رقم رباب دارای پوست قرمز با دانه‌های قرمز تیره رنگ، رقم اتابکی دارای پوست قرمز رنگ با دانه‌هایی قرمز، رقم خفری دارای پوست سبز مایل به زرد با دانه‌هایی قرمز رنگ و رقم شیرین شهوار با پوستی سبز مایل به زرد با دانه‌هایی صورتی رنگ هستند. تعداد میوه‌های تهیه شده در مجموع ۵۰۹ عدد بود که سهم رقم رباب ۲۰۵ عدد، رقم اتابکی ۱۰۵ عدد، رقم شیرین شهوار ۱۰۰ عدد و رقم خفری ۹۹ عدد بود. نمونه‌ها به منظور جلوگیری از فساد بلافاصله پس از خریداری در یخچال با دمای ۴ درجه سلسیوس انبار شدند و برای جلوگیری از تغییر در ویژگی‌های فیزیکی مانند کاهش ضخامت پوست، با پلاستیک ضخیم پوشش دهی شد. لازم به ذکر است که تمامی آزمایش‌ها و بررسی‌های لازم در کوتاه‌ترین زمان ممکن پس از خریداری نمونه‌ها صورت گرفت. ارقام خریداری شده مطابق استاندارد سازمان خواروبار جهانی در چهار گروه اندازه A، B، C و D دسته‌بندی شدند (جدول ۱) [۷].

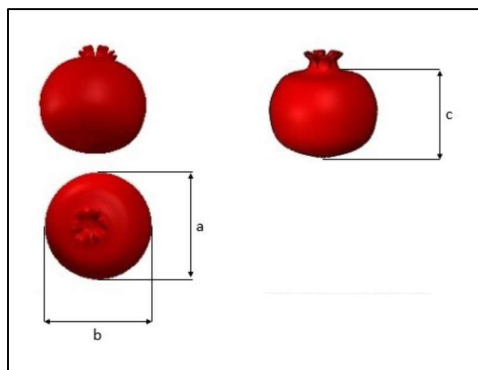
جدول ۱- استاندارد FAO برای طبقه‌بندی میوه انار بر اساس قطر میوه

اندازه	قطر (mm)
A	≥ 81
B	۷۱-۸۰
C	۶۱-۷۰
D	۵۱-۶۰
E	۴۰-۵۰

برای اندازه‌گیری قطر نمونه‌ها از قطر معادل هندسی مطابق رابطه (۱) استفاده شد [۱۲]:

$$D = (abc)^{\frac{1}{3}} \quad (1)$$

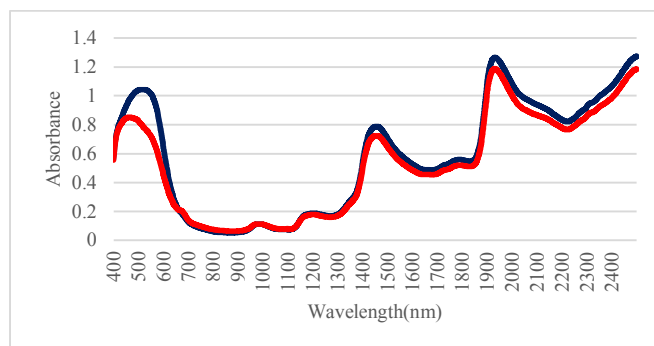
در رابطه بالا a قطر اول، b قطر دوم و c طول میوه است که a و b در یک صفحه و عمود بر یکدیگر و c عمود بر آن دو است (شکل ۱). اندازه‌گیری ابعاد با کولیس دیجیتال (Guanglu ساخت کشور چین) با دقت ± 0.01 میلی‌متر انجام شد.



شکل ۱- ابعاد مشخص شده میوه انار برای اندازه‌گیری قطر معادل هندسی

طیف‌سنجی نمونه‌ها و تعیین حجم آب

برای سنجش پاسخ طیفی نمونه‌های انار، از دستگاه اسپکتروفوتومتر مرئی-مادون قرمز نزدیک (مدل NIR XDS ساخت کمپانی Metrohm در کشور سوئیس) استفاده شد. آزمایش‌ها در طیف جذبی بازه طول موج ۴۰۰ تا ۲۵۰۰ نانومتر با فواصل ۰/۵ نانومتر انجام شد. تمامی نمونه‌ها کاملاً سالم و فاقد هرگونه فساد و آسیب دیدگی بودند. قبل از طیف‌سنجی، دستگاه مطابق روش توصیه شده بدون وجود نمونه واسنجی (کالیبره) شد. برای این که پاسخ‌های طیفی از اعتبار بیشتری برخوردار باشند از هر نمونه دوبار و در دو جهت مختلف طیف‌برداری شد. سپس از طیف‌های برداشت شده برای هر نمونه میانگین‌گیری شد تا به ازای هر نمونه یک دسته پاسخ طیفی به همراه میزان جذب نور حاصل شود (شکل ۲).



شکل ۲- نمونه‌ای از امضاء طیفی ثبت شده از انار توسط دستگاه طیف‌سنج

در طیف‌سنجی، غلظت نمونه با واحد مول بر لیتر و طول مسیر نور با واحد سانتی‌متر سنجیده می‌شود. بنابر این واحدهای جذب مولی لیتر بر مول بر سانتی‌متر هستند. از آنجا که واحد جذب مولی همواره به همین شکل است به طور معمول آن را بدون واحد اعلام می‌شود [۹]. محتوای آب نمونه‌های طیف‌سنجی شده با استفاده از دستگاه

آبمیوه‌گیری دستی استخراج شد و حجم آن با استوانه مدرج با دقت ± 1 میلی‌لیتر اندازه‌گیری شد. دستگاه آب میوه‌گیر دارای مکانیزم فشاری برای خارج کردن کامل آب با فشار یکسان برای همه نمونه‌ها بود.

ساخت مدل ریاضی برای برآورد حجم آب

برای ساخت مدل ریاضی برآورد حجم آب اندازه‌گیری شده بر اساس پاسخ‌های طیفی، الگوریتم‌های یادگیری ماشین و در نرم افزار PYTHON 3.8 پیاده‌سازی شد. ابتدا پیش‌پردازش داده‌ها صورت گرفت و داده‌های پرت با استفاده از روش دامنه میان چارکی^۸ حذف گردید (روابط ۲ تا ۴). برای این منظور دامنه اعداد مشخص و مقادیر چارک اول (Q_1) و سوم (Q_3) برای حجم آب به طور جداگانه محاسبه گردید. سپس محدوده بالا (UL) و محدوده پایین (LL) محاسبه و اعداد خارج این دامنه به عنوان داده پرت حذف شد.

$$IQR = Q_3 - Q_1 \quad (2)$$

$$\text{Upper Limit (UL)} = Q_3 + 1.5(IQR) \quad (3)$$

$$\text{Lower Limit (LL)} = Q_1 - 1.5(IQR) \quad (4)$$

از آنجا که داده‌های ثبت شده زیاد علاوه بر داشتن اطلاعات ارزشمند [۱۰]، دارای اطلاعات نامربوط و نوفه‌هایی نیز هستند که به دلایلی نظیر پراکندگی نور، بروز تفاوت در سطح تماس نمونه و شناساگر دستگاه طیف‌نگار هنگام نمونه‌برداری‌های متعدد و ویژگی‌های فیزیکی نمونه مثل سطح آن رخ می‌دهند [۱۷]، پنج روش پیش‌پردازش نوفه‌ها بر اساس بررسی منابع و سعی و خطا انتخاب شد. این روش‌ها عبارت از تصحیح پراکندگی ضربی^۹، متغیر نرمال استاندارد^{۱۰}، نرمال‌سازی بردار^{۱۱}، مشتق اول^{۱۲} و مشتق دوم^{۱۳} بود.

همچنین در این تحقیق سعی شد با آزمون و خطاهای متعدد، نوفه‌های موجود در داده‌های طیف‌برداری که روش‌های ذکر شده قادر به شناسایی و حذف آن‌ها نشده‌اند، با استفاده از سه روش شناسایی ناهنجاری^{۱۴} شامل جنگل انزوا^{۱۵}، تعیین کوواریانس حداقل^{۱۶} و ماشین بردار پشتیبان تک کلاسه^{۱۷}، شناسایی و حذف شوند.

کاهش ابعاد و یافتن مؤلفه‌های اصلی

هر کدام از طول موج‌هایی که به نمونه‌ها ساطع شده‌اند یک بعد یا یک ویژگی محسوب می‌شوند. نکته حائز اهمیت این است که همه این ویژگی‌ها در ساخت مدل کالیبراسیون به یک اندازه دارای اهمیت نیستند و این اهمیت را می‌توان از روی پراکندگی هر ویژگی از نمونه‌ای به نمونه دیگر تعیین کرد. به این مفهوم که اگر با تغییر

⁸ Interquartile Range (IQR)

⁹ Multiplicative Scatter Correction (MSC)

¹⁰ Standard Normal Variate (SNV)

¹¹ Vector Normalization (VN)

¹² First Derivative (FD)

¹³ Second Derivative (SD)

¹⁴ Anomaly Detection

¹⁵ Isolation Forest

¹⁶ Minimum Covariance Determinant

¹⁷ One-Class Support Vector Machine

در نمونه‌ها مقادیر جذب در یک طول موج خاص، دستخوش تغییر نشوند و یا به عبارت دیگر پراکندگی پائینی داشته باشند، می‌توان نتیجه گرفت که آن طول موج خاص اهمیت بالایی ندارد و قادر به نشان دادن اختلاف بین نمونه‌ها نیست به همین رو می‌توان به منظور کاهش حجم محاسبات، و ساده‌تر کردن مدل آن را حذف کرد. با بررسی منابع مختلف و همچنین با طی چند مرحله آزمون و خطا در استفاده از روش‌های مختلف، نهایتاً به منظور کاهش ابعاد و تعیین مؤلفه‌های اصلی، در مرحله تخمین حجم آب از روش حداقل مربعات جزئی استفاده شد. الگوریتم کاهش ابعاد PLS امکان کار با مجموعه داده‌هایی که در آن‌ها تعداد مشاهدات از تعداد ویژگی‌ها کمتر هستند یا به عبارت دیگر چندخطی^{۱۸} بوده و تعداد متغیرهای بالایی دارند را به خوبی فراهم می‌کند. ویژگی دیگر این روش این است که نظارت شده^{۱۹} است، یعنی به هنگام یافتن مؤلفه‌های اصلی هم از مقادیر X (ویژگی‌ها) و هم از مقادیر Y (برچسب‌ها) بهره می‌برد و بارگذاری آن توسط تجزیه مقدار منفرد S که همان ماتریس حاصل شده از برهمکنش مقادیر X و Y است، انجام می‌شود [۶]. نکته حائز اهمیت دیگر این است که در هنگام استفاده از تمام مدل‌ها، از ۱ الی ۴۰ مؤلفه اساسی در ساخت مدل استفاده شد و نهایتاً تعداد مؤلفه اساسی که منجر به کمترین خطای ریشه میانگین مربع شد، در ساخت مدل نهایی مورد استفاده قرار گرفت [۱۸].

پس از اتمام پیش‌پردازش و آماده‌سازی داده‌ها مرحله ساخت مدل اجرا شد. سایر پژوهشگران نیز از روش‌های متنوعی برای تخمین زدن یک شاخص استفاده کرده‌اند که به منظور دستیابی به مدل‌هایی با بازدهی و دقت بالاتر برخی از این پژوهش‌ها مورد بررسی قرار گرفت. بنابر پیشنهاد اغلب محققان در مطالعات مشابه به منظور تخمین حجم آب از الگوریتم رگرسیون حداقل مربعات جزئی استفاده شد [۱۱]، [۱۵]، [۱۸]. روش PLSR یک روش چند متغیره است که در تجزیه و تحلیل طیف‌سنجی مرئی-مادون قرمز نزدیک به میزان وسیعی به کار گرفته می‌شود [۱۰] و نیز به منظور ایجاد ارتباط بین مجموعه داده طیف‌ها (X) و شاخص‌های فیزیکی یا شیمیایی اندازه‌گیری شده (Y) کاربرد فراوان دارد [۳].

پیش‌بینی حجم آب

در این پژوهش برای تخمین مقدار حجم آب میوه انار ابتدا نمونه‌ها به شکل تصادفی به دو دسته ۳۰ درصد و ۷۰ درصد تقسیم شدند. از ۷۰ درصد نمونه‌ها برای ساخت مدل کالیبراسون بین مقادیر جذب در هر طول موج به عنوان مقادیر X و مقادیر حجم آب به عنوان مقادیر Y با استفاده از الگوریتم PLSR استفاده شد و ۳۰ درصد باقی مانده برای ارزیابی مدل به کار گرفته شد. همچنین برای بررسی تفاوت پاسخ‌ها در کل‌بازه، بازه مرئی و بازه مادون قرمز نزدیک سه‌بازه ۴۰۰ تا ۲۵۰۰ نانومتر، ۴۰۰ تا ۹۰۰ نانومتر و ۹۰۰ تا ۲۵۰۰ نانومتر به صورت جداگانه مورد تحلیل و بررسی قرار گرفتند. خطای ریشه میانگین مربع (RMSE)، انحراف پیش‌بینی باقی مانده (RPD) و ضریب تعیین (R^2) که رابطه‌ها به ترتیب در موارد (۵)، (۶) و (۷) بیان شده است، شاخص‌های آماری بودند که برای بررسی مدل در مرحله کالیبراسون استفاده شد.

$$RSME = \sqrt{\sum \frac{(y_{cal} - y_{act})^2}{n}} \quad (5)$$

¹⁸ Multicollinearity

¹⁹ Supervised

$$RPD = \frac{SD}{RMSE} \quad (6)$$

$$R^2 = 1 - \frac{\sum (y_{cal} - y_{act})^2}{\sum (y_{cal} - y_{mean})^2} \quad (7)$$

در روابط بالا y_{cal} ، مقادیر حجم پیش‌بینی شده در مرحله کالیبراسیون، y_{act} مقادیر واقعی حجم آب، y_{mean} مقدار میانگین حجم آب پیش‌بینی شده، SD انحراف معیار مقادیر حجم آب واقعی است و n تعداد نمونه‌ها است. ارزیابی مدل علاوه بر ضریب تعیین و خطای ریشه میانگین مربع، میانگین خطای اریبی، میانگین خطای مطلق و شاخص ویلموت^{۲۰} بودند که در روابط (۸)، (۹) و (۱۰) آمده است.

$$MBE = \frac{1}{n} \sum (y_{pred} - y_{act}) \quad (8)$$

$$MAE = \frac{1}{n} \sum |y_{pred} - y_{act}| \quad (9)$$

$$d = 1 - \frac{\sum (y_{act} - y_{pred})^2}{\sum (|y_{pred} - \bar{y}_{act}| + |y_{act} - \bar{y}_{act}|)^2}, \quad 0 \leq d \leq 1 \quad (10)$$

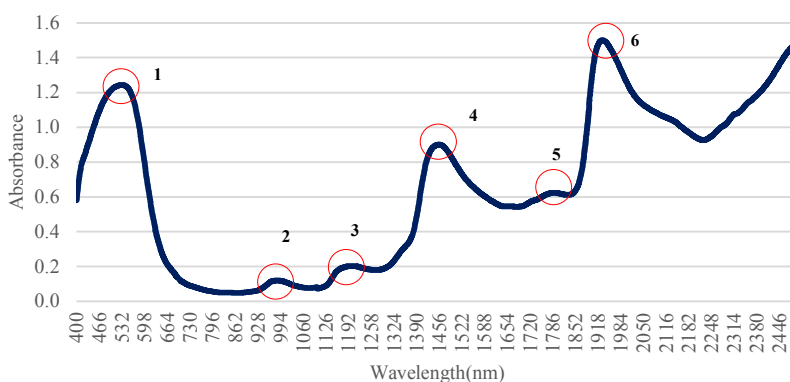
از آن جا که ارزیابی داخلی مدل بر ارزیابی خارجی آن مقدم است تمام مدل‌های ساخته شده در این تحقیق پیش از ارزیابی خارجی، از یک مرحله ارزیابی داخلی عبور کردند. عملکرد مدل‌های رگرسیون و طبقه‌بندی با استفاده از روش ارزیابی متقابل ۱۰ تایی در مرحله ساخت مورد بررسی قرار گرفت. در این روش ۷۰ درصد داده‌ها که به منظور ساخت مدل در نظر گرفته شده بودند خود به ۱۰ قسمت مساوی تقسیم شدند. سپس یکی از این ده قسمت برای امتحان مدل و ۹ قسمت باقی مانده برای ساخت مجدد آن به کار گرفته شدند. این عملیات به تعداد ۱۰ مرتبه صورت گرفت تا هر ده قسمت حداقل یکبار برای ارزیابی مدل مورد استفاده قرار گیرند. شاخص‌های آماری که برای ارزیابی داخلی مدل‌های رگرسیون مربوط به پیش‌بینی حجم آب مورد استفاده قرار گرفتند عبارتند از ضریب تعیین و خطای ریشه میانگین مربع.

تحلیل نتایج

بررسی ویژگی‌های امضاء طیفی میوه انار

امضاء طیفی میوه‌های انار که در دستگاه طیف‌سنج قرار گرفتند به صورت نموداری از میزان مقادیر جذب امواج بر حسب طول موج اعمال شده است. در این نمودار نقاط اوج متعددی وجود دارد که هر کدام از آن‌ها بیانگر یک ویژگی خاص از میوه انار هستند که در این جا به بررسی برخی از آن‌ها پرداخته خواهد شد (شکل ۳).

²⁰ Willmott Index (d)



شکل ۳- نمونه امضاء طیفی میوه انار (میزان جذب بر حسب طول موج)

نقطه اوج شماره ۱ در بازه ۴۵۰ تا ۶۵۰ نانومتر حاصل شد که بیانگر وجود رنگدانه‌های موجود در پوست انار است که این رنگدانه‌ها طول موج‌های بازه مذکور که در محدوده مرئی هستند را بیشتر جذب می‌کنند. نقطه اوج شماره ۲ در محدوده ۹۳۰ تا ۹۸۰ نانومتر رویداد که به دلیل وجود پیوندهای بین ملکولی اکسیژن-هیدروژن (O-H) موجود در دانه‌های انار است [۱۱]. نقاط اوج ۳، ۵ و ۶ موجود در محدوده ۱۱۰۰ تا ۱۲۰۰ نانومتر، ۱۷۵۰ تا ۱۸۰۰ نانومتر و ۱۸۵۰ تا ۱۹۵۰ نانومتر مربوط به پیوندهای ملکولی تغییر شکل یافته اکسیژن-هیدروژن و کربن-هیدروژن (C-H) موجود در قند محلول در آب انار است و نقطه اوج شماره ۴ ایجاد شده در محدوده ۱۳۵۰ تا ۱۴۵۰ نانومتر نیز مربوط به پیوندهای ملکولی اکسیژن-هیدروژن مرتبط با جذب آب انار است [۳].

تخمین حجم آب و نتایج حاصل شده

برای تخمین حجم آب میوه انار ابتدا داده‌های طیفی در شش حالت شامل داده‌های خام و داده‌های پیش پردازش شده با روش‌های تصحیح پراکندگی ضربی (MSC)، متغیر نرمال استاندارد (SNV)، نرمال‌سازی بردار (VN)، مشتق اول (FD) و مشتق دوم (SD) آماده شدند و سپس در سه بازه طول موج ۴۰۰ تا ۹۰۰، ۹۰۰ تا ۲۵۰۰ و ۴۰۰ تا ۲۵۰۰ نانومتر، با استفاده از روش رگرسیون حداقل مربعات جزئی تخمین زده شدند. بازه ۴۰۰ تا ۹۰۰ نانومتر به این علت بررسی شد که پاسخگو بودن این بازه در تخمین حجم آب، سرعت نمونه برداری را به میزان قابل توجهی کاهش می‌دهد. نتایج حاصل شده به شرح جدول‌های ۲، ۳ و ۴ است.

جدول ۲- نتایج مرحله کالیبراسیون مدل PLSR در تخمین حجم آب میوه انار

طول موج (nm)	روش پیش پردازش	RMSE	R ²	RPD
۴۰۰-۹۰۰	MSC	۸/۳۹	۰/۹۲	۰/۴۲
	SNV	۷/۳۶	۰/۹۳	۰/۵۳
	VN	۸/۴۴	۰/۹۳	۰/۴۳
۹۰۰-۲۵۰۰	MSC	۸/۱۷	۰/۹۳	۰/۴۵
	SNV	۷/۴۶	۰/۹۳	۰/۵۱
	VN	۶/۱۷	۰/۹۵	۰/۷۳
۴۰۰-۲۵۰۰	MSC	۷/۱۹	۰/۹۴	۰/۵۶
	SNV	۴/۳۱	۰/۹۸	۰/۵۵
	VN	۷/۶۴	۰/۹۳	۰/۵۰

جدول ۳- نتایج ارزیابی داخلی مدل PLSR در تخمین حجم آب میوه انار

R ²	RMSE	روش پیش پردازش	طول موج (nm)
۰/۸۶	۱۱/۰۲	MSC	۴۰۰-۹۰۰
۰/۹۰	۹/۱۶	SNV	
۰/۸۱	۱۳/۵۰	VN	
۰/۸۸	۱۰/۴۷	MSC	۹۰۰-۲۵۰۰
۰/۸۷	۱۰/۰۳	SNV	
۰/۹۲	۷/۷۱	VN	
۰/۹۰	۹/۰۸	MSC	۴۰۰-۲۵۰۰
۰/۹۴	۶/۹۴	SNV	
۰/۸۹	۹/۴۰	VN	

جدول ۴- نتایج حاصل از ارزیابی مدل PLSR در تخمین حجم آب میوه انار

MBE	MAE	d	R ²	RMSE	روش پیش پردازش	طول موج (nm)
-۶/۴۶	۱۳/۰۶	۰/۹۲	۰/۸۲	۲۱/۵۷	MSC	۴۰۰-۹۰۰
-۲/۲۴	۱۰/۵۰	۰/۹۵	۰/۸۸	۱۳/۵۳	SNV	
-۳/۶۰	۱۱/۷۶	۰/۹۴	۰/۸۳	۱۶/۸۱	VN	
-۶/۰۷	۱۱/۷۶	۰/۹۳	۰/۸۴	۱۸/۹۶	MSC	۹۰۰-۲۵۰۰
-۳/۸۲	۸/۴۲	۰/۹۷	۰/۸۹	۱۲/۳۸	SNV	
-۸/۴۶	۱۲/۹۸	۰/۹۲	۰/۸۵	۲۰/۸۶	VN	
-۸/۹۶	۱۴/۰۸	۰/۹۱	۰/۸۲	۲۴/۰۷	MSC	۴۰۰-۲۵۰۰
-۰/۳۳	۵/۳۰	۰/۹۸	۰/۹۴	۸/۴۱	SNV	
-۵/۵۸	۱۱/۲۳	۰/۹۴	۰/۸۵	۱۷/۱۲	VN	

نتایج پیش بینی حجم آب در جدول ۴ نشان می دهد که در هر سه بازه ۴۰۰ تا ۹۰۰، ۹۰۰ تا ۲۵۰۰ و ۲۵۰۰ تا ۴۰۰ نانومتر خروجی های آماری حاصل شده از روش پیش پردازش SNV وضعیت بهتری دارند که ممکن است به این علت باشد که سطح پوست میوه انار و اجزاء داخلی آن صاف و یکدست نیستند و این ناهمواری ها باعث ایجاد نوفه هایی در طیف های تاییده شده می شود. از آنجا که پیش پردازش SNV فرایند هموارسازی نوفه های طیفی را در سطح محدودی نسبت به سایر روش ها انجام می دهد و به نوفه های موجود در طیف های برداشته شده از انار حساس تر است، نتایج بهتری را نسبت به سایر روش های پیش پردازش به دست داده است [۱۷].

همچنین نتایج حاصل شده در دو بازه ۹۰۰ تا ۲۵۰۰ و ۴۰۰ تا ۲۵۰۰ نانومتر نسبت به بازه ۴۰۰ تا ۹۰۰ نانومتر بهتر هستند به این علت که پیوندهای ملکولی اکسیژن-هیدروژن موجود در آب دانه های انار طول موج هایی در بازه ۱۳۵۰ تا ۱۴۵۰ نانومتر را بیشتر به خود جذب می کنند، لذا این بازه مختص جذب آب است [۳]. از آنجا که بازه ۴۰۰ تا ۹۰۰ نانومتر فاقد اطلاعات مربوط به آب دانه های انار است در نتیجه شاخص های آماری ضعیف تری دارد.

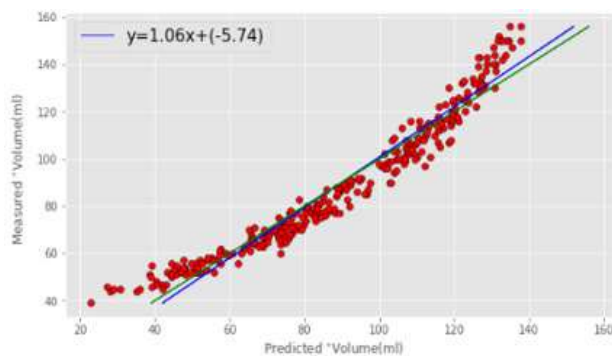
از طرفی می توان نتیجه گرفت که اگر چه بازه ۴۰۰ تا ۲۵۰۰ و ۲۵۰۰ تا ۹۰۰ نانومتر بهترین نتایج آماری را به خود اختصاص داده اند، اما استفاده از دستگاه طیف سنجی که این بازه وسیع را پوشش دهد زمان اسکن طولانی تری خواهد داشت و به دلیل استفاده از سخت افزارهای بیشتر و پیچیده تر هزینه بالاتری نیز در بر خواهد داشت. لذا در صورت

اقدام به ساخت دستگاه شناساگر حجم آب انار بر پایه طیف‌سنجی، استفاده از بازه ۴۰۰ تا ۹۰۰ نانومتر مقرون به صرفه‌تر بوده و از دقت خوبی نیز برخوردار خواهد بود.

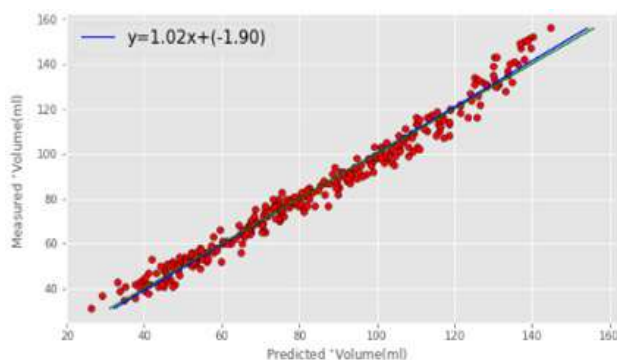
در پژوهشی سعی شد با استفاده از توموگرافی محاسباتی اشعه ایکس حجم آب انار به صورت غیرمخرب تعیین شود. حجم آب اندازه‌گیری شده در روش مخرب $16/3 \pm 14/1$ و در روش غیرمخرب $16/4 \pm 14/2$ گزارش شد. نتایج آزمون t نشان می‌دهد که اختلاف بین مقادیر مشاهده شده و پیش‌بینی شده در سطح ۰/۰۵ معنا دار نیست و روش سی‌تی‌اسکن به طور مناسب قادر به تخمین حجم آب انار است [۲]. در یک پژوهش دیگر نیز حجم آب انار با استفاده از پردازش تصاویر حاصل از روش سی‌تی‌اسکن اشعه ایکس تخمین زده شد. بهترین پاسخ در تخمین حجم آب در این روش با ضریب تعیین ۹۱/۹ درصد همراه بود و نتایج آزمون چاو نشان داد که اختلاف بین حجم آب مشاهده شده و پیش‌بینی شده معنادار نیست و در نتیجه با استفاده از پردازش تصاویر حاصل از سی‌تی‌اسکن می‌توان با دقت مناسبی حجم آب انار را به صورت غیرمخرب تخمین زد.

در پژوهش حاضر با استفاده از طیف‌سنجی حجم آب انار با ضریب تعیین ۹۸ درصد تخمین زده شد که از لحاظ دقت اندازه‌گیری فراتر از روش سی‌تی‌اسکن است. همچنین فناوری طیف‌سنجی مرئی-مادون قرمز نزدیک با داشتن ابزارهای ساده و قابل حمل شرایط به مراتب راحت‌تری را برای نمونه‌برداری نسبت به روش سی‌تی‌اسکن فراهم می‌کند. به علاوه بنا بر بررسی‌های میدانی صورت گرفته و نظر محققان پیشین هزینه نمونه‌برداری با دستگاه سی‌تی‌اسکن بسیار بیشتر از دستگاه طیف‌سنجی است که خود دلیلی بر برتری استفاده از این روش است [۳].

نمودارهای آموزش و ارزیابی حجم آب واقعی بر اساس حجم آب پیش‌بینی شده با مدل PLSR برای بهترین (SNV) در بازه ۴۰۰ تا ۲۵۰۰) و ضعیف‌ترین (MSC در بازه ۴۰۰ تا ۲۵۰۰) نتیجه در شکل‌های ۴ و ۵ نمایش داده شده‌اند (خط $y=x$ با رنگ سبز مشخص شده است).

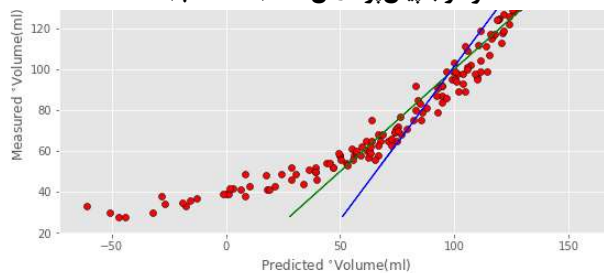


(الف)

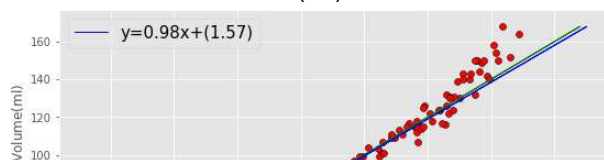


(ب)

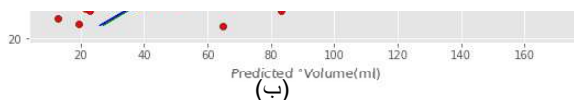
شکل ۴- نمودارهای آموزش حجم آب پیش‌بینی شده در بازه ۴۰۰ تا ۲۵۰۰ نانومتر با پیش‌پردازش: الف) MSC، ب) SNV



(الف)



شکل ۵- نمودارهای ارزیابی حجم آب پیش‌بینی شده در بازه ۴۰۰ تا ۲۵۰۰ نانومتر با پیش‌پردازش: الف) MSC، ب) SNV



(ب)

نتیجه‌گیری

در این پژوهش دو هدف اساسی دنبال شد. هدف اول امکان‌سنجی استفاده از فناوری غیرمخرب طیف‌سنجی مرئی-مادون قرمز نزدیک در تعیین حجم آب میوه انار، که نتایج این امکان‌سنجی نشان داد که می‌توان حجم آب را با دقت بسیار بالا تعیین کرد. هدف دوم این پژوهش تعیین الگوریتم‌های مناسب به منظور تعیین این ویژگی بود که می‌توان در یک نتیجه‌گیری کلی بیان کرد که ترکیب علم یادگیری ماشین با داده‌های طیف‌نگاری و به کارگیری مدل PLSR در تخمین حجم آب با ثبت ضریب تعیین ۹۸ درصد موفقیت‌آمیز بوده و راهی جدید را برای سایر پژوهشگران باز کرده‌است تا این مسیر را ادامه و گسترش دهند.

مراجع:

۱. سلمان‌زاده، ف. ۱۳۹۱. تعیین مقدار کمی پارامترهای بیوفیزیکی دو رقم میوه انار محلی با استفاده از فناوری غیرمخرب اشعه ایکس. پایان‌نامه کارشناسی ارشد، دانشگاه شیراز، شیراز، ایران.
2. Arendse, E., Fawole, O. A., Magwaza, L. S., & Opara, U. L. (2016). Non-destructive characterization and volume estimation of pomegranate fruit external and internal morphological fractions using X-ray computed tomography. *Journal of Food Engineering*, 186, 42-49.
3. Arendse, E., Fawole, O. A., Magwaza, L. S., Nieuwoudt, H. H., & Opara, U. L. 2017. Development of calibration models for the evaluation of pomegranate aril quality by Fourier-transform near infrared spectroscopy combined with chemometrics. *Biosystems Engineering*, 159, 22-32.
4. Arendse, E., Fawole, O. A., Magwaza, L. S., Nieuwoudt, H., and Opara, U. L. 2018b. Fourier transform near infrared diffuse reflectance spectroscopy and two spectral acquisition modes for evaluation of external and internal quality of intact pomegranate fruit. *Postharvest Biology and Technology*, 138, 91-98.
5. Baek, S., Lim, J., Lee, J. G., McCarthy, M. J., and Kim, S. M. 2020. Investigation of the Maturity Changes of Cherry Tomato Using Magnetic Resonance Imaging. *Applied Sciences*, 10(15), 5188.
6. Bulut, E., and Alma, Ö. G. 2010. Dimensionality Reduction Methods: PCR, PLSR, RRR and health application. *Physical Sciences*, 6(2), 36-47.
7. Food and Agriculture Organization of the United Nations (2013). Standard for Pomegranate. Retrieved from <http://www.fao.org/home/search/en/?q=standardforpomegranate>
8. Fazayeli, A., Kamgar, S., Nassiri, S. M., Fazayeli, H., and De La Guardia, M. 2019. Dielectric spectroscopy as a potential technique for prediction of kiwifruit quality indices during storage. *Information Processing in Agriculture*, 6(4), 479-486.
9. Hagen, C. L., and Sanders, S. T. 2007. Investigation of multi-species (H₂O₂ and H₂O) sensing and thermometry in an HCCI engine by wavelength-agile absorption spectroscopy. *Measurement Science and Technology*, 18(7), 1992.

10. Khodabakhshian, R., Emadi, B., Khojastehpour, M., Golzarian, M. R., and Sazgarnia, A. 2017. Non-destructive evaluation of maturity and quality parameters of pomegranate fruit by visible/near infrared spectroscopy. *International Journal of Food Properties*, 20(1), 41-52.
11. Khodabakhshian, R., Emadi, B., Khojastehpour, M., and Golzarian, M. R. 2019. A comparative study of reflectance and transmittance modes of Vis/NIR spectroscopy used in determining internal quality attributes in pomegranate fruits. *Journal of Food Measurement and Characterization*, 13(4), 3130-3139.
12. Mohsenin, N. N. (1996). Physical properties of plant and animal materials. Vol. 1. Structure, physical characteristics and mechanical properties. *Physical properties of plant and animal materials*. Vol. 1. Structure, physical characteristics and mechanical properties., 1.
13. Munera, S., Hernández, F., Aleixos, N., Cubero, S., and Blasco, J. 2019. Maturity monitoring of intact fruit and arils of pomegranate cv. 'Mollar de Elche' using machine vision and chemometrics. *Postharvest Biology and Technology*, 156, 110936.
14. Patel, M. B., Nath, A., and Mayani, J. M. 2018. A study on physical properties of pomegranate (*Punica granatum* L., Punicaceae) fruits. *International Journal of Communication Systems* 6(5), 1460-1463.
15. 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.
16. Salmanizadeh, F., Nassiri, S. M., Jafari, A., and Bagheri, M. H. 2015. Volume estimation of two local pomegranate fruit (*Punica granatum* L.) cultivars and their components using non-destructive X-ray computed tomography technique. *International Journal of Food Properties*, 18(2), 439-455.
17. Shirzadifar, A., Bajwa, S., Mireei, S. A., Howatt, K., and Nowatzki, J. 2018. Weed species discrimination based on SIMCA analysis of plant canopy spectral data. *Biosystems Engineering*, 171, 143-154.
18. Xiao, H., Feng, L., Song, D., Tu, K., Peng, J., and Pan, L. 2019. Grading and sorting of grape berries using visible-near infrared spectroscopy on the basis of multiple inner quality parameters. *Sensors*, 19(11), 2600.
19. Yang, L., Gao, H., Meng, L., Fu, X., Du, X., Wu, D., and Huang, L. 2020. Nondestructive measurement of pectin polysaccharides using hyperspectral imaging in mulberry fruit. *Food Chemistry*, 334, 127614.

Visible and Near Infrared Spectroscopy for Determining Pomegranate juice volume

Mohammad Jamalifard^{1*}, Seyed Mehdi Nassiri^{2*}, Alimohammad Shirzadifar³ and Alireza Shahsavari⁴

1. Biosystems Engineering Department, Shiraz University, Shiraz, Iran
2. Biosystems Engineering Department, Shiraz University, Shiraz, Iran
3. Biosystems Engineering Department, Shiraz University, Shiraz, Iran
4. Horticulture Science and Engineering Department, Shiraz University, Shiraz, Iran

Abstract

Pomegranate is one of the most important garden products in Iran. In addition to its high commercial-economic value, it has many nutritional benefits. Due to the lack of appropriate development of processing technologies in this sector, factories use completely traditional and limited methods for measuring the quality of pomegranate, which reduces the quality of the final product. In the present study, the feasibility of utilizing non-destructive technology of visible and near-infrared (Vis/NIR) spectroscopy was investigated to determine the pomegranate fruit juice volume. In addition to raw spectral data of the samples were subjected to a wavelength of 400 to 2500 nm, five preprocessed data series from multiplicative scatter correlation (MSC), standard normal variate (SNV), vector normalization (VN), first derivative (FD), second derivative (SD) were used for data analyzing. The prediction models of juice volume were developed by PLSR. The model was established according to machine learning algorithms using PYTHON 3.8 software. The results predicted the juice volume by 98% coefficient of determination which was obtained from the combination of SNV and PLSR model.

Key words: Pomegranate, Spectroscopy, Machine Learning, Non-Destructive Method

*Corresponding author
E-mail: nasiri@shirazu.ac.ir