

به کارگیری روش جنگل تصادفی در راستای تعیین میزان قند، اسیدیت و سفتی گیلاس با بهره‌مندی از روش طیف‌سنجی (مرئی-مادون قرمز)

غلامحسین شاهقلی^{۱*} و سینا فیض‌اله‌زاده اردبیلی^۲

چکیده

این مطالعه به بررسی کاربرد الگوریتم جنگل تصادفی در رابطه با روش‌های طیف‌سنجی، به‌ویژه طیف‌سنجی مادون قرمز نزدیک مرئی (Vis-NIR)، برای پیش‌بینی ویژگی‌های کیفیت کلیدی گیلاس می‌پردازد. هدف این مقاله ارائه یک روش قابل اعتماد و کارآمد برای ارزیابی غیر مخرب محتوای قند، اسیدیت و سفتی، عوامل مهم تعیین‌کننده کیفیت گیلاس و پذیرش مصرف‌کننده، است. در این مطالعه ۴۰۰ نمونه تصادفی از محصول گیلاس تهیه شد. داده‌های طیفی به دست آمدند. در ادامه داده‌های طیفی به دست‌آمده از طیف‌سنجی Vis-NIR توسط روش Relief مورد ارزیابی قرار گرفته و برای هر کدام از مقادیر متغیرهای وابسته (مقادیر قند، اسیدیت و سفتی)، پنج طول موج مؤثر به طور جداگانه انتخاب شدند. در ادامه مقادیر شدت بازتاب مربوط به طول موج‌های مؤثر انتخاب شده به‌عنوان ویژگی‌های ورودی برای مدل جنگل تصادفی انتخاب شدند (برای سفتی، محتوای جامد محلول و اسیدیت گیلاس، به ترتیب طول موج‌های بین ۶۱۵ تا ۶۵۰، ۶۵۰ تا ۶۷۵ و ۸۱۸ تا ۹۴۵ نانومتر). عملکرد مدل از نظر دقت پیش‌بینی از طریق اعتبارسنجی متقابل و آزمایش مستقل ارزیابی، با به کارگیری میانگین مربعات خطا (RMSE) و ضریب همبستگی (CC) مورد ارزیابی قرار گرفت. نتایج نشان‌دهنده اثربخشی الگوریتم جنگل تصادفی در پیش‌بینی دقیق محتوای قند (با ضریب همبستگی حدود ۰/۹۳ و میانگین مربعات خطای حدود ۳/۴)، اسیدیت (با ضریب همبستگی حدود ۰/۹۵ و میانگین مربعات خطای حدود ۰/۸۷) و سفتی گیلاس (با ضریب همبستگی حدود ۰/۹۲ و میانگین مربعات خطای حدود ۱/۲) بر اساس اطلاعات طیفی است. این رویکرد، یک راه حل سریع، غیر مخرب و مقرون به صرفه برای ارزیابی کیفیت محصول در صنعت گیلاس را ارائه می‌کند. همچنین با پیشرفت در این حوزه می‌توان تولیدکنندگان و ذینفعان را قادر ساخت تا تصمیم‌گیری آگاهانه در مورد برداشت، مرتب‌سازی و فرآیندهای پس از برداشت بگیرند.

واژه‌های کلیدی: مادون قرمز، طیف‌سنجی، گیلاس، سنجش غیرمخرب، انتخاب ویژگی، جنگل تصادفی.

ارجاع: شاهقلی غ. ح. و فیض‌اله‌زاده اردبیلی س. ۱۴۰۳. به کارگیری روش جنگل تصادفی در راستای تعیین میزان قند، اسیدیت و سفتی گیلاس با بهره‌مندی از روش طیف‌سنجی (مرئی-مادون قرمز). نشریه پژوهش‌های مکانیک ماشین‌های کشاورزی. ۳۳: ۳۵-۴۳.
<https://dx.doi.org/10.22034/JRMAM.2024.14719.700>

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

۲- استادیار گروه مهندسی نوین، دانشکده فناوری‌های نوین، دانشگاه محقق اردبیلی، اردبیل، ایران.

* نویسنده مسئول: gshahgholi@yahoo.com

تاریخ پذیرش: ۱۴۰۳/۰۵/۱۴

تاریخ دریافت: ۱۴۰۲/۰۲/۰۲

مقدمه

گیلاس، با طعم منحصر به فرد، جایگاه خود را به عنوان یکی از محبوب‌ترین میوه‌ها در سراسر جهان تثبیت کرده است. گیلاس سرشار از قندها، مواد معدنی، ویتامین‌ها و مجموعه‌ای از مواد مغذی ضروری دیگر (Cao et al., 2015; Shao et al., 2019)، یک میوه لذیذ است.

با این حال، گیلاس، با بافت گوشتی آبدار خود، در معرض آسیب‌هایی به ویژه ناشی از ضربه‌های مکانیکی است. در طول فرآیند برداشت، جابجایی، نگهداری، حمل و نقل و توزیع، گیلاس‌ها در صورت برخورد با یکدیگر یا تماس با سطوح سخت در معرض خطر آسیب دیدگی قرار دارند. بافت آسیب‌پذیر این میوه‌ها نیازمند توجه دقیق و رسیدگی در هر مرحله رشد محصول است تا اطمینان حاصل شود که کیفیت بکر آن‌ها به طور دست نخورده به دست مصرف‌کنندگان می‌رسد. از باغ تا مصرف، حفظ یکپارچگی گیلاس نیاز به مراقبت و دقت دقیق دارد تا کیفیت آنها حفظ شود (Díaz-Mula et al., 2009).

به طور کلی، مشتریان، کیفیت گیلاس را بر اساس ویژگی‌های بیرونی آن مانند شکل، رنگ و اندازه قضاوت می‌کنند (Timm et al., 1995).

در صنعت کشاورزی امروز، نیاز به اطلاعات عددی برای استفاده توسط سامانه‌های برداشت محصول، دسته-

بندی و درجه‌بندی محصول پس از برداشت و در طول دوره انبارداری وجود دارد (Andersen et al., 2023).

در سال‌های اخیر ادغام روش‌های یادگیری ماشین و ابزارهای تشخیصی غیرمخرب، راه‌گشای ایجاد سامانه‌های تشخیص هوشمند در به دست آوردن اطلاعات فیزیکی و شیمیایی انواع محصول‌های کشاورزی شده‌اند.

طیف‌سنجی مرئی-مادون قرمز (Visible-NIR) یک روش غیرمخرب و برای اندازه‌گیری ویژگی‌های کیفی در میوه‌ها بسیار مناسب است. این روش سریع است، به آماده‌سازی نمونه کم‌تری نیاز دارد و تعیین هم‌زمان چندین ویژگی را با استفاده از یک اندازه‌گیری امکان‌پذیر می‌سازد (Lu, 2001; Ruiz-Altisent et al., 2010). استفاده از این روش هم توسط تولیدکنندگان در مزرعه و هم توسط مصرف‌کنندگان میوه در خط تولید و بسته بندی، یکنواختی دسته‌های میوه را، هم از نظر ظاهر خارجی و هم از نظر ویژگی‌های کیفی داخلی مانند سفتی و محتوای جامد محلول و اسیدپتیه، تضمین می‌کند که هر دو بر طعم میوه حاکم هستند.

در این راستا مطالعات متعددی انجام شده است. جدول ۱، مشخصاتی از مطالعات انجام شده با ادغام روش‌های یادگیری ماشین در انجام پروژه‌های تشخیصی خواص محصولات کشاورزی در حضور روش‌های غیرمخرب طیف‌سنجی را ارائه می‌دهد.

جدول ۱- مطالعات انجام شده در زمینه تشخیص ویژگی‌های محصولات کشاورزی

منبع	نوع روش یادگیری ماشین	نوع	بازه طیفی	نوع ویژگی	محصول
Walsh et al. (2024)	CNN	جذبی	۳۵۰ تا ۷۵۰	محتوای ماده خشک	انبه
Kalopesa et al. (2023)	CNN	بازتابی	۳۵۰ تا ۲۵۰۰	SSC, pH	انگور
Kim et al. (2023)	یادگیری جمعی	بازتابی	۴۵۰ تا ۱۱۰۰	SSC	مرکبات
Pourdarbani et al. (2021)	ANN	بازتابی	۴۵۰ تا ۱۰۰۰	سفتی، pH و SSC	سیب
Fazari et al. (2021)	CNN	بازتابی	۴۵۰ تا ۱۰۵۰	آنتراکتوز	زیتون
Pourdarbani et al. (2020)	ترکیبی ANN	بازتابی	۴۰۰ تا ۱۰۰۰	سفتی، pH و SSC	سیب
Rong et al. (2020)	CNN	بازتابی	۴۰۰ تا ۸۰۰	وارپته محصول	هلو
Guo et al. (2016)	ANN, SVM	جذبی	۲۵۰۰ تا ۸۰۰	محتوای شیمیایی	جوجوبا

طوری که می‌توان ادعا کرد در مطالعات پیشین هیچ اتفاق نظری در مورد طول موج‌های مؤثر بر تشخیص SSC، pH و سفتی محصول گیلاس وجود ندارد که نیاز به انجام مطالعه‌ای جامع و تکمیلی برای تشخیص این

بر اساس مطالعات انجام شده می‌توان عنوان کرد، پژوهش برای تشخیص سه ویژگی اصلی و مهم گیلاس شامل میزان اسیدپتیه (pH)، محتوای جامد محلول (SSC) و سفتی محصول به صورت محدود صورت گرفته است. به

استان اردبیل، ایران تهیه شدند. در مرحله اول، زمان تقریبی برداشت گیلاس توسط متخصصان انسانی تعیین شد. در مجموع ۴۰۰ نمونه، در هر چهار مرحله رشد انتخاب و توسط افراد متخصص طبقه‌بندی شدند:

محصول از واریته تکدانه مشهد انتخاب گردید. تاریخ برداشت در مرداد ماه سال ۱۴۰۲، ۲۰ روز قبل از تاریخ برداشت بهینه، ۱۰ روز قبل از تاریخ برداشت بهینه، در تاریخ برداشت بهینه و ۱۰ روز پس از تاریخ برداشت بهینه انجام گرفت. در هر مرحله ۱۰۰ نمونه برداشت شد. بلافاصله پس از جمع‌آوری، نمونه‌ها طیف‌نگاری و داده‌های طیفی جمع‌آوری شدند. این روش جمع‌آوری از رویکرد به کار رفته در مطالعه پوردربانی و همکاران تبعیت می‌کند (Pourdarbani *et al.*, 2020).

استخراج داده‌های طیفی

از یک اسپکتروفوتومتر (StellarNet.USA, Tampa, EPP200NIR Vis-NIR) مجهز به آشکارساز عکس ایندیوم-گالیوم-آرسنید (InGaAs) برای به دست آوردن اطلاعات طیفی بازتابی استفاده شد. در این سامانه، روشنایی با استفاده از لامپ تنگستن ۱۰ وات انجام شد. برای پردازش بیشتر داده‌های به دست آمده، به یک رایانه همراه (Corei7, 1000M, 3.2 GHz, 8GB RAM) با سامانه نرم افزاری شامل Windows 10 و MatLab 2021 منتقل شد.

اندازه‌گیری خواص فیزیکی-شیمیایی

برای اندازه‌گیری سفتی از روش ارائه شده در مطالعه پوردربانی و همکاران استفاده شد (Pourdarbani *et al.*, 2020). در این روش از یک دستگاه نفوذسنج دستی با نفوذگر مخصوص (قطر ۵ و ارتفاع ۲۵ میلی متر) استفاده شد و نفوذگرها در دو طرف گیلاس قرار گرفتند. نفوذگرها به داخل نمونه‌ها فشار داده شدند تا دستگاه میزان نیروی اعمال شده، بر حسب واحد کیلوگرم بر سانتی‌متر را ثبت کند. میانگین نیروی اعمال شده در دو طرف گیلاس به عنوان معیار سفتی در نظر گرفته شد.

برای استخراج SSC از دستگاه رفاکتومتر (Optizen 2120 UVplus, China) استفاده شد.

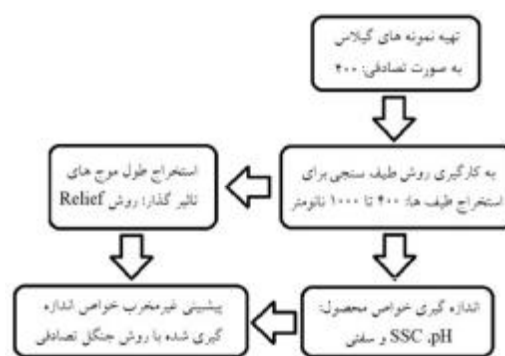
ویژگی‌ها را توجیه می‌کند. هم‌چنین مطالعات انجام شده، روش‌های یادگیری ماشین جمعی را به عنوان روش‌های دقیق در مدلسازی ویژگی‌های محصولات با حجم داده‌های پایین پیشنهاد کردند. بر این اساس در این مطالعه از روش جنگل تصادفی به عنوان یک روش یادگیری ماشین جمعی برای مدل‌سازی استفاده شد.

نوآوری اصلی این پژوهش استخراج طول موج‌های مؤثر بر تشخیص SSC، pH و سفتی محصول گیلاس هستند که در ادامه، پژوهش با همین طول موج‌ها به سمت تخمین pH و سفتی محصول گیلاس حرکت می‌کند.

هدف اصلی این مطالعه به کارگیری روش جنگل تصادفی به عنوان یکی از قوی‌ترین روش‌های یادگیری ماشین برای ادغام با روش طیف‌سنجی به عنوان یک روش غیرمخرب در تشخیص دقیق اسیدیته (pH)، SSC و سفتی محصول گیلاس است. این مطالعه از چهار گام کلی شامل آماده‌سازی نمونه‌های گیلاس، انجام طیف‌سنجی، استخراج ویژگی‌های اصلی و طول موج‌های مؤثر و مدل‌سازی با روش جنگل تصادفی تشکیل شده است.

مواد و روش‌ها

شکل ۱ سامانه پیش‌بینی پیشنهادی غیر مخرب را نشان می‌دهد که شامل پنج مرحله مورد نیاز برای تولید داده‌های ورودی، آموزش و ارزیابی سامانه است که در ادامه به تفصیل مورد بحث قرار خواهند گرفت.



شکل ۱- مراحل انجام روش کار

جمع‌آوری نمونه

جمع‌آوری نمونه‌های مختلف اولین گام برای ایجاد سامانه پیش‌بینی پیشنهادی بود. نمونه‌ها از شهرستان هیر در

نحوه عملکرد جنگل تصادفی به صورت خلاصه به قرار زیر است (Parmar *et al.*, 2019):

- نمونه‌برداری از نقاط داده با جایگزینی (نمونه‌برداری تصادفی)، که چندین درخت تصمیم می‌سازد. این بدان معنی است که هر درخت در جنگل بر روی یک زیر مجموعه تصادفی از داده‌های آموزشی، آموزش داده می‌شود.
 - در هر گره درخت تصمیم، یک زیرمجموعه تصادفی از ویژگی‌ها برای تقسیم در نظر گرفته می‌شود. تعداد ویژگی‌های در نظر گرفته شده در هر تقسیم معمولاً جذر تعداد کل ویژگی‌های مجموعه داده است.
 - هر درخت تصمیم با استفاده از زیرمجموعه ای از داده‌ها و ویژگی‌های آموزشی ساخته می‌شود. درختان در اعماق رشد می‌کنند، اغلب تا زمانی که تمام برگ‌ها خالص یا حاوی حداقل تعداد نمونه هستند.
 - برای کارهای طبقه‌بندی، (متداول‌ترین طبقه) حالت طبقه‌های پیش‌بینی شده توسط درختان، جداگانه به عنوان پیش‌بینی نهایی در نظر گرفته می‌شود. برای وظایف رگرسیون، میانگین پیش‌بینی همه درختان گرفته می‌شود.
- روش جنگل تصادفی چندین مزیت دارد (Zhang & Ma, 2012):
- در مقایسه با درختان تصمیم فردی، کم‌تر مستعد بیش‌برازش است.
 - می‌تواند مجموعه داده‌های بزرگ و با ابعاد بالا را مدیریت کند.
 - تخمین‌هایی از اهمیت ویژگی ارائه می‌دهد.
- با توجه به این مزایا و دقت آن، جنگل تصادفی به طور گسترده در روش‌های مختلف یادگیری ماشین از جمله طبقه‌بندی، رگرسیون و انتخاب ویژگی استفاده می‌شود.

در این مطالعه از روش جنگل تصادفی برای مدل‌سازی غیرمخرب تخمین مقادیر سفتی، pH و SSC محصول گیلان استفاده شد. به این ترتیب که طول موج‌های منتخب از مرحله انتخاب ویژگی به عنوان متغیرهای مستقل و مقادیر اندازه‌گیری شده به عنوان متغیرهای وابسته انتخاب شدند. در ادامه این داده‌ها به مجموعه‌های آموزش و آزمون شبکه (به صورت تصادفی با نسبت ۷۵ به ۲۵ درصد) تقسیم شدند، بدون این که

در ادامه اندازه‌گیری مقادیر pH نیز از یک دستگاه سنجش اسیدیته با دقت ۰/۰۱ (hi2211, Thailand) استفاده شد.

انتخاب طول موج مؤثر

برای انتخاب طول موج مؤثر از روش انتخاب ویژگی Relief استفاده شد. الگوریتم Relief یک روش انتخاب ویژگی است که معمولاً در یادگیری ماشین و داده‌کاوی استفاده می‌شود (Urbanowicz *et al.*, 2018). ارتباط هر ویژگی با متغیرهای وابسته را با در نظر گرفتن سهم آن‌ها در دقت طبقه‌بندی ارزیابی می‌کند. امتیاز Relief اهمیت ویژگی‌ها را با محاسبه تفاوت در مقادیر ویژگی بین نمونه‌هایی که در فضای ویژگی به یکدیگر نزدیک هستند، اندازه‌گیری می‌کند (Kira & Rendell, 1992). این الگوریتم امتیازهای بالاتری را به ویژگی‌هایی اختصاص می‌دهد که تفاوت‌های قابل توجهی در مقادیر برای نمونه‌های متعلق به طبقه‌های مختلف دارند که نشان دهنده ارتباط آن‌ها با طبقه‌بندی است.

امتیاز Relief اهمیت ویژگی‌ها را در طبقه‌بندی براساس توانایی آن‌ها در تمایز بین طبقات مختلف را کمی‌سازی می‌کند. ویژگی‌های با امتیاز Relief بالاتر برای طبقه‌بندی مرتبط‌تر در نظر گرفته و اغلب برای ساخت مدل‌های پیش‌بینی انتخاب می‌شوند.

برای اجرای این الگوریتم، ویژگی‌های به دست آمده از آزمایش تجربی به عنوان متغیرهای وابسته و طول موج‌ها به عنوان متغیرهای مستقل در کتابخانه نرم‌افزار MATLAB انتخاب گردیدند. بر این اساس یک مجموعه امتیازاتی برای طول موج‌ها به دست آمد و پنج طول موج مؤثر به عنوان متغیرهای مستقل تأثیرگذار انتخاب شدند.

مدل‌سازی با جنگل تصادفی

جنگل تصادفی یک روش یادگیری گروهی محبوب است که برای کارهای طبقه‌بندی و رگرسیون در یادگیری ماشین استفاده می‌شود و با ساخت تعداد زیادی درخت تصمیم در طول آموزش عمل می‌کند. در این مرحله طبقه‌ای را که به ارتباط با متغیر وابسته یا متغیر پیش‌بینی شونده نزدیک است خروجی می‌دهد (Genuer *et al.*, 2020).

متغیرهای وابسته در سطح احتمال پنج درصد معنی‌دار هستند.

نتایج انجام طیف‌سنجی

شکل ۲ نتایج شدت‌های بازتابی برای طول موج‌های ۴۰۰ تا ۱۰۵۰ نانومتر را برای نمونه‌های مختلف نشان می‌دهد. در این مرحله، طول موج‌های ۴۰۰ تا ۶۱۴ و طول موج‌های ۱۰۵۱ تا ۱۱۰۰ به دلیل ورود خطا از مسیر ارزیابی با روش بصری خارج شدند.

همان‌طور که از شکل ۲ مشخص است، روند حرکتی تمام طول موج‌ها تقریباً یکسان است، به طوری که در بازه تقریباً ۶۵۰ تا ۷۷۰ همگی دارای یک کمینه مقدار هستند. از بازه ۷۷۱ تا تقریباً ۹۲۰ روند رشد صعودی بوده و در بازه ۹۲۱ تا ۹۴۰ یک افت ناگهانی را دارند. از ۹۴۱ تا ۱۰۵۰ دوباره روند تغییرات به صورت صعودی در می‌آید. این روند برای تمام طول موج‌ها و نمونه‌ها تقریباً مشابه است.

جدول ۳- نتایج تحلیل واریانس (ANOVA)

متغیر وابسته	P-value	F-Value	متغیر مستقل
SSC	۰/۰۴۲*	۴/۳۵	۶۶۶
	۰/۰۳۷*	۴/۵۸	۶۶۶/۵
	۰/۰۳۲*	۴/۸۵	۶۶۷
	۰/۰۴۶*	۴/۲۵	۶۶۷/۵
	۰/۰۴۸*	۴/۰۲	۶۶۸
pH	۰/۰۲۳*	۵/۹۹	۸۹۵
	۰/۰۲۵*	۵/۷۳	۸۹۵/۵
	۰/۰۲۶*	۵/۶۵	۸۹۶
	۰/۰۲۴*	۵/۷۹	۸۹۶/۵
	۰/۰۲۸*	۵/۴۶	۸۹۷
سفتی	۰/۰۳۱*	۴/۸۹	۶۳۳/۵
	۰/۰۳*	۵/۰۲	۶۳۴
	*۰/۰۳۴	۴/۷۴	۶۳۴/۵
	۰/۰۳۹*	۴/۴۷	۶۳۵
	۰/۰۴۱*	۴/۳۲	۶۳۵/۵

* معنی‌دار در سطح احتمال پنج درصد

روش انتخاب ویژگی

شکل ۳ روند تغییرات امتیاز مربوط به طول موج‌های مؤثر را در هر کدام از متغیرهای وابسته نشان می‌دهد.

نمونه مشترک واحدی بین مجموعه‌های آموزش، و آزمون شبکه به اشتراک گذاشته شود. مجموعه‌های آموزش و آزمون شبکه برای معماری‌های مختلف مدل جنگل تصادفی استفاده شدند تا بهترین شبکه و مدل را تولید کنند.

در ادامه تنظیمات ساختار شبکه به نحوی انجام گرفت تا بهترین شبکه برای پیش‌بینی مقادیر اندازه‌گیری شده تحت آموزش قرار بگیرد. برای ارزیابی عملکرد شبکه نیز از مقادیر ریشه میانگین مربعات خطا (RMSE) و ضریب همبستگی (CC) استفاده شد.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (1)$$

$$CC = \frac{Cov(x, \hat{x})}{\sigma_x \sigma_{\hat{x}}} \quad (2)$$

که در این معادلات \hat{x} مقادیر پیش‌بینی شده و x مقادیر واقعی هستند. در این معادلات N تعداد نمونه‌ها است.

نتایج و بحث

محصولات به دست آمده تحت فرایند طیف‌سنجی قرار گرفتند تا شدت‌های مختلف بازتاب‌ها از طول موج ۴۰۰ تا ۱۰۰۰ برای نمونه‌ها استخراج شوند.

جدول ۲ خلاصه آماری خواص فیزیکی-شیمیایی (اندازه‌گیری شده) شامل سفتی، اسیدیته و SSC نمونه‌های گیلان را ارائه می‌دهد. طبق جدول ۲، انحراف استاندارد نسبی، که به عنوان ضریب تغییرات نیز شناخته می‌شود، محتوای SSC (۳/۳) از انحراف معیار اسیدیته و سفتی بالاتر است.

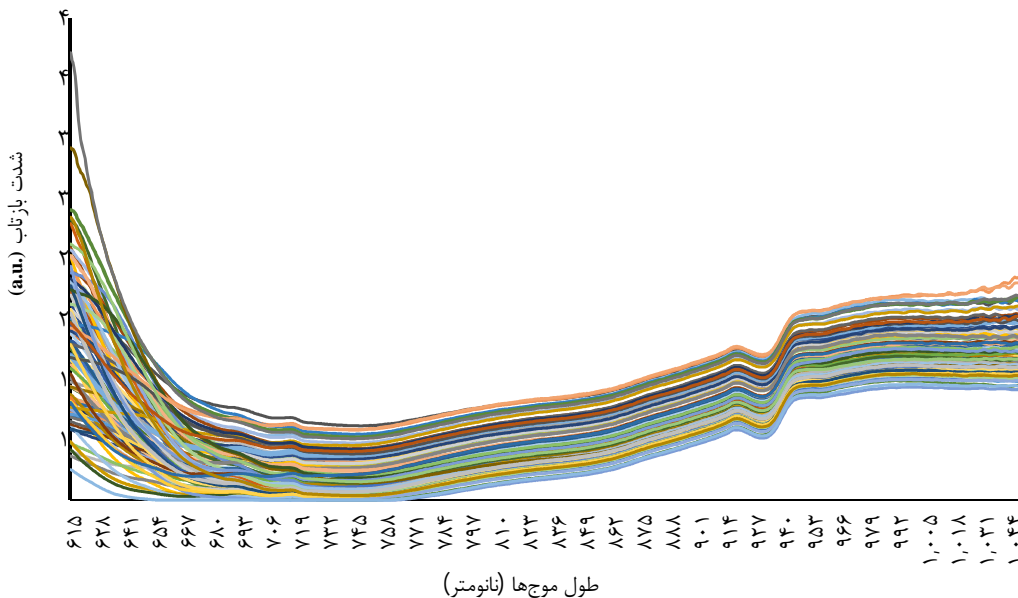
جدول ۲- مقادیر مربوط به خواص فیزیکی-شیمیایی

ویژگی	واحد	حداقل	حداکثر	میانگین	انحراف معیار
سفتی	kg/cm	۲/۲۵	۷/۵۵	۴/۳۳	۰/۸۶
pH	-	۲/۸۲	۴	۳/۳۱	۰/۲۱
SSC	درصد	۱۲/۶	۲۸/۷	۱۷/۷۶	۳/۳۱

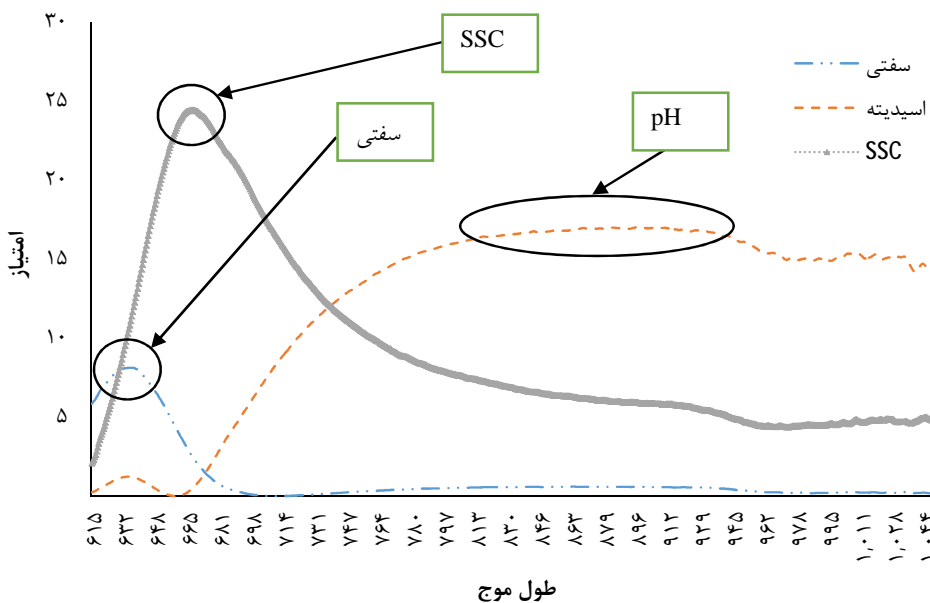
جدول ۳ نتایج حاصل از تحلیل واریانس (ANOVA) را برای بررسی معنی‌داری تأثیر متغیرهای مستقل بر وابسته را ارائه می‌دهد. مطابق با جدول ۳ می‌توان عنوان کرد که تأثیر تمامی متغیرهای مستقل بر

را به عنوان طول موج‌های مؤثر بر محتوای زیست توده شناسایی کردند (Mahmoodi-Eshkaftaki et al., 2024). همان‌طور که مشخص است این بازه‌های طیفی خواص تشخیصی بالایی دارند که می‌تواند به دلیل ماهیت طول موج‌های مرتبط باشد. پرداختن بیشتر به این بخش از مطالعه می‌تواند بینش‌های ارزشمندی را در علم تشخیص غیرمخرب آشکار کند.

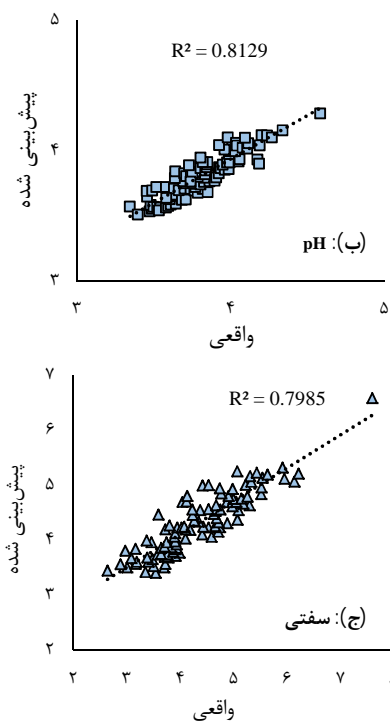
مطابق با شکل ۳ می‌توان عنوان کرد، برای سفتی گیلان، طول موج‌های بین ۶۱۵ تا ۶۵۰ بالاترین تأثیر را دارند. از طرفی دیگر برای SSC گیلان، طول موج‌های موجود در بازه ۶۵۰ تا ۶۷۵ بالاترین تأثیر را دارند. اما در مورد اسیدیته، بازه تأثیرگذاری طول موج‌ها کمی گسترده‌تر است به طوری که دامنه بین ۸۱۸ تا حدود ۹۴۵ را می‌توانند پوشش دهند. محمودی و همکاران در مطالعه‌ای طول موج‌های ۵۷۳، ۶۰۲، ۶۳۸، ۷۳۷ و ۸۸۲



شکل ۲- طیف‌های بازتابی برای نمونه‌های مورد آزمایش



شکل ۳- روند تغییرات امتیاز مربوط به طول موج هر کدام از متغیرهای وابسته



ادامه شکل ۴ - مقادیر پیش‌بینی برای متغیرهای وابسته

نتیجه‌گیری

این مطالعه کارایی الگوریتم جنگل تصادفی را در پیش‌بینی ویژگی‌های کیفیت کلیدی گیلاس، یعنی محتوای قند، اسیدیته و سفتی، با استفاده از داده‌های طیفی به‌دست‌آمده از طیف‌سنجی مادون قرمز مرئی - نزدیک (Vis-NIR) نشان می‌دهد. با استفاده از روش‌های یادگیری ماشین و فناوری طیف‌سنجی، تولیدکنندگان و ذینفعان می‌توانند بینش‌های سریع و دقیقی در مورد کیفیت گیلاس به دست آورند و تصمیم‌گیری آگاهانه در مورد برداشت، مرتب‌سازی و فرآیندهای پس از برداشت را تسهیل کنند. علاوه بر این، ماهیت غیرمخرب طیف‌سنجی Vis-NIR حداقل آماده‌سازی نمونه را تضمین می‌کند و یکپارچگی میوه را حفظ می‌کند و به افزایش کارایی و کاهش ضایعات در زنجیره تولید کمک می‌کند. جهت‌های تحقیقاتی آینده ممکن است شامل اصلاح معماری مدل، کاوش ویژگی‌های طیفی اضافی و گسترش کاربرد جنگل تصادفی به سایر انواع میوه و ویژگی‌های کیفیت است. به‌طور کلی، ادغام جنگل تصادفی با طیف‌سنجی Vis-NIR نویدبخش پیشرفت شیوه‌های کنترل کیفیت در صنعت گیلاس است که در نهایت به نفع تولیدکنندگان و مصرف‌کنندگان است.

جدول ۴ طول موج‌های مؤثر با ضریب ارتباط بالا را برای تمام متغیرهای وابسته نشان می‌دهد.

جدول ۴ - مقادیر طول موج‌های مؤثر با ضریب ارتباط بالا برای متغیرهای وابسته

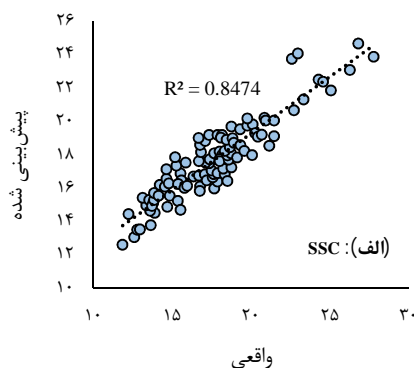
متغیر	طول موج‌های مؤثر				
اسیدیته	۸۹۶	۸۹۶/۵	۸۹۵	۸۹۵/۵	۸۹۷
SSC	۶۶۷	۶۶۶/۵	۶۶۷/۵	۶۶۶	۶۶۸
سفتی	۶۳۵	۶۳۴	۶۳۴/۵	۶۳۵/۵	۶۳۳/۵

مدل‌سازی

جدول ۵ به ارائه ویژگی‌های ساختاری شبکه در کنار داده‌های به دست آمده برای دقت شبکه می‌پردازد. شکل ۴ نقاط پیش‌بینی شده توسط روش جنگل تصادفی را در مقابل داده‌های واقعی نشان می‌دهد. مطابق با شکل ۴ می‌توان عنوان کرد همبستگی مناسبی در مرحله آزمون شبکه برای مقادیر SSC (حدود ۰/۹۲)، اسیدیته (حدود ۰/۹۳) و سفتی (حدود ۰/۹۱) وجود دارد. بر این اساس می‌توان عنوان کرد، روش غیرمخرب جنگل تصادفی می‌تواند به عنوان یک روش قابل اعتماد برای پیش‌بینی خواص مهم محصول گیلاس انتخاب شود.

جدول ۵ - مقادیر مربوط به تنظیمات شبکه

متغیر	تنظیمات شبکه		تعداد درخت‌ها	تعداد ویژگی‌ها
	CC	RMSE		
اسیدیته	۰/۹۵	۰/۸۷	۲۵۰	۴
SSC	۰/۹۳	۳/۴	۳۱۰	۴
سفتی	۰/۹۲	۱/۲	۲۷۰	۴



شکل ۴ - مقادیر پیش‌بینی برای متغیرهای وابسته

- Genuer, R., Poggi, J.-M., Genuer, R., & Poggi, J.-M. (2020). *Random forests*. Springer.
- Guo, Y., Ni, Y., & 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.
- Kalopesa, E., Gkrimpizis, T., Samarinas, N., Tsakiridis, N. L., & Zalidis, G. C. (2023). Rapid Determination of Wine Grape Maturity Level from pH, Titratable Acidity, and Sugar Content Using Non-Destructive In Situ Infrared Spectroscopy and Multi-Head Attention Convolutional Neural Networks. *Sensors*, 23(23): 9536.
- Kim, S. Y., Hong, S. J., Kim, E., Lee, C. H., & Kim, G. (2023). Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy. *Journal of Food Engineering*, 338: 111254.
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992*. Elsevier. pp. 249-256
- Lu, R. (2001). Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Transactions of the ASAE*, 44(5): 1265.
- Mahmoodi-Eshkaftaki, M., Mahbod, M., & Ghenaatian, H. R. (2024). Non-destructive estimation of biomass characteristics: Combining hyperspectral imaging data with neural networks. *Renewable energy*, 224: 120137.
- Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. International conference on intelligent data communication technologies and internet of things (ICICI) 2018.
- Pourdarbani, R., Sabzi, S., & Arribas, J. I. (2021). Non-destructive estimation of three apple fruit properties at various ripening levels with optimal Vis-NIR spectral wavelength regression data. *Heliyon*, 7(9).
- Pourdarbani, R., Sabzi, S., Kalantari, D., & 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.
- Rong, D., Wang, H., Ying, Y., Zhang, Z., & Zhang, Y. (2020). Peach variety detection using VIS-NIR spectroscopy and deep learning. *Computers and Electronics in Agriculture*, 175: 105553.
- در این مطالعه نتایج زیر حاصل شدند:
- برای بررسی سفتی گیلاس، طول موج‌های بین ۶۱۵ تا ۶۵۰ بالاترین نقش را دارند.
 - برای بررسی SSC گیلاس، طول موج‌ها در بازه ۶۵۰ تا ۶۷۵ بالاترین نقش را دارند.
 - برای بررسی اسیددیده گیلاس، طول موج‌های بین ۸۱۸ تا حدود ۹۴۵ بالاترین نقش را دارند.
 - مطابق با یافته‌های مرحله آزمون واریانس، تأثیر تمام پنج طول موج مؤثر در مقادیر سفتی، SSC و اسیددیده در سطح احتمال پنج درصد معنی‌دار است.
 - مدل ارائه شده توانست با همبستگی مناسبی در مرحله آزمون شبکه مقادیر SSC (حدود ۰/۹۲)، اسیددیده (حدود ۰/۹۳) و سفتی (حدود ۰/۹۱) را تخمین بزند.
- با پرداختن به این یافته‌ها، می‌توانیم فرصت‌های جدیدی را برای بهینه‌سازی راهبردهای برداشت، به حداقل رساندن تلفات پس از برداشت، و ارائه گیلاس با کیفیت بالا به مصرف‌کنندگان باز کنیم و در عین حال از پایداری و کارایی در طول فرآیند تولید اطمینان حاصل کنیم.

منابع

- Andersen, P. V., Afseth, N. K., Aaby, K., Gaarder, M. Ø., Remberg, S. F., & Wold, J. P. (2023). Prediction of chemical and sensory properties in strawberries using Raman spectroscopy. *Post-harvest Biology and Technology*, 201: 112370.
- Cao, J., Li, X., Liu, Y., Leng, F., Li, X., Sun, C., & Chen, K. (2015). Bioassay-Based Isolation and Identification of Phenolics from Sweet Cherry That Promote Active Glucose Consumption by HepG2 Cells. *Journal of food science*, 80(2): C234-C240.
- Díaz-Mula, H. M., Castillo, S., Martínez-Romero, D., Valero, D., Zapata, P. J., Guillén, F., & Serrano, M. (2009). Sensory, nutritive and functional properties of sweet cherry as affected by cultivar and ripening stage. *Food Science and Technology International*, 15(6): 535-543.
- Fazari, A., Pellicer-Valero, O. J., Gómez-Sánchez, J., Bernardi, B., Cubero, S., Benalia, S., Zimbalatti, G., & Blasco, J. (2021). Application of deep convolutional neural networks for the detection of anthracnose in olives using VIS/NIR hyperspectral images. *Computers and Electronics in Agriculture*, 187: 106252.

- Ruiz-Altisent, M., Ruiz-Garcia, L., Moreda, G., Lu, R., Hernandez-Sanchez, N., Correa, E., Diezma, B., Nicolai, B., & Garcia-Ramos, J. (2010). Sensors for product characterization and quality of specialty crops—A review. *Computers and Electronics in agriculture*, 74(2): 176-194.
- Shao, Y., Xuan, G., Hu, Z., Gao, Z., & Liu, L. (2019). Determination of the bruise degree for cherry using Vis-NIR reflection spectroscopy coupled with multivariate analysis. *PLoS One*, 14(9): e0222633.
- Timm, E., Guyer, D., Brown, G., & Schulte, N. (1995). Michigan sweet cherry color measurement and prototype color chip development. *Applied Engineering in Agriculture*, 11(3): 403-407.
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85: 189-203.
- Walsh, J., Neupane, A., & Li, M. (2024). Evaluation of 1D convolutional neural network in estimation of mango dry matter content. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 124003.
- Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.

