

### مقاله پژوهشی

## مدل سازی کمومتریکس برای پیش‌بینی سریع خواص فیزیکوشیمیایی و هیدروکسی متیل فورفورال عسل مبتنی بر طیف‌سنجی فراطیفی رامان عنوان کوتاه: پیش‌بینی خواص عسل با مدل‌سازی کمومتریکس و طیف‌سنجی رامان

فاتانه هاشم‌پور بلترک<sup>۱</sup>، سمیه ولی‌زاده<sup>۱</sup>، عادل میرزاعلیزاده<sup>۲\*</sup>، بهروز جنت<sup>۳</sup>، حسین رستگار<sup>۴</sup>

۱. مرکز تحقیقات حلال جمهوری اسلامی ایران، سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران
۲. دپارتمان بهداشت و ایمنی مواد غذایی، دانشکده بهداشت، دانشگاه علوم پزشکی زنجان، زنجان، ایران
۳. سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران.
۴. مرکز تحقیقات فرآورده های آرایشی و بهداشتی سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران.

(تاریخ ارسال: ۱۴۰۴/۰۷/۰۲، تاریخ آخرین بازنگری: ۱۴۰۴/۰۸/۰۲، تاریخ پذیرش: ۱۴۰۴/۰۸/۲۵)

### چکیده

هدف از این پژوهش، بررسی امکان‌سنجی پیش‌بینی سریع ویژگی‌های فیزیکوشیمیایی در عسل با استفاده از طیف‌سنجی رامان و مدل‌های کمومتریکس است. در این مطالعه، ۵۱ نمونه عسل از مناطق مختلف ایران جمع‌آوری شد و هر نمونه با طیف‌سنجی رامان (در محدوده ۱۰۰ تا  $3500\text{ cm}^{-1}$ ) آنالیز گردید. ویژگی‌های مورد بررسی شامل فندهای احیاکننده پیش از هیدرولیز، ساکاروز، رطوبت و هیدروکسی متیل فورفورال (HMF) بودند که با روش‌های استاندارد آزمایشگاهی اندازه‌گیری شدند. داده‌های طیفی پس از پیش‌پردازش میانگین-مرکزی با مقادیر آزمایشگاهی ویژگی‌های فیزیکوشیمیایی همبسته‌سازی شده و چهار مدل مستقل مبتنی بر رگرسیون حداقل مربعات جزئی (PLS) توسعه یافت. مدل‌ها با استفاده از الگوریتم Kennard-Stone برای تقسیم داده‌ها به مجموعه‌های کالیبراسیون و تست، اعتبارسنجی متقاطع با روش Leave-One-Out برای بهینه‌سازی تعداد متغیرهای پنهان، و شاخص‌هایی مانند RMSEP، REP% و RMSECV ارزیابی شدند. نتایج نشان داد که مدل‌های پیش‌بینی فندهای احیاکننده پیش از هیدرولیز و رطوبت با خطای نسبی پایین (به ترتیب ۱.۶۴٪ و ۳۰٪/۱۵) عملکرد عالی داشتند، مدل ساکاروز با خطای نسبی حدود ۱۱/۹۵٪ قابل قبول بود، در حالی که مدل HMF با خطای نسبی ۲۵/۸٪ دقت کمتری نشان داد. به‌طور کلی، ترکیب طیف‌سنجی رامان با مدل‌های PLS روشی سریع، غیرمخرب، کم‌هزینه و سازگار با محیط‌زیست برای کنترل کیفیت عسل ارائه می‌دهد که می‌تواند در پایش اصالت، توزیع و توسعه اقتصادی این محصول نقش مؤثری ایفا کند.

**کلمات کلیدی:** کنترل کیفیت، ویژگی‌های فیزیکوشیمیایی، طیف‌سنجی رامان، عسل، مدل کمومتریکس، رگرسیون PLS

## ۱. مقدمه

طبق تعریف کمسیون کدکس موادغذایی، عسل ماده‌غذایی شیرین طبیعی ساخته‌شده از شهد گلها یا ترشحات بخش‌های زنده گیاهان یا مواددفعی حشرات مکنده گیاه بر روی بخش‌های زنده گیاه است که توسط زنبورعسل جمع‌آوری، تبدیل و با مواد خاص زنبورعسل ترکیب و در کندوی عسل انبار می‌شود تا دوره رسیدن را طی کند. عسل باید حاوی حداقل ۶۰٪ قندهای احیا بوده (استاندارد ملی ایران کمینه ۶۵٪) و رطوبت آن نباید بیشتر از ۲۱٪ (استاندارد ملی ایران بیشینه ۲۰٪) باشد. ترکیب عسل بیشتر تحت‌تأثیر نوع گل‌هایی که زنبورعسل استفاده کرده است و همین‌طور شرایط منطقه‌ای و اقلیمی قرار می‌گیرد. عسل حاوی ۶۰-۸۰٪ گلوکز و فروکتوز به‌عنوان مونوساکارید، ۱۰٪ مالتوز و ساکارز به‌عنوان مهم‌ترین دی‌ساکاریدها، مالتیتوز تری‌ساکارید اصلی و همچنین حاوی سایر اولیگو ساکاریدهایی با وزن مولکولی پایین می‌باشد. علاوه‌بر قندها، آنتی‌اکسیدانها (پینوبانکسین، کرایسیتین و گالاجین)، ویتامین‌ها، اسیدها عمدتاً گلوکونیک اسید، پروتئین‌ها، موادمعدنی، فلاونوئیدها، ویتامین‌ها و آنزیم‌ها نیز در عسل یافت می‌شوند [۱].

برای طبقه‌بندی و تعیین کیفیت عسل، استانداردها و روش‌هایی در کدکس آلیمانتاروس [۲]، کمیسیون بین‌المللی عسل و انجمن رسمی شیمیدانان تجزیه (AOAC) [۳]، ارائه شده است. این استانداردها خواص فیزیکی و شیمیایی عسل را که باید برای تعیین کیفیت عسل ارزیابی شود، مشخص می‌کنند. روش سنتی برای انجام آزمایشات کیفی عسل شامل آنالیز دانه‌های گرده (Melissopalynology) موجود در رسوبات آن توسط میکروسکوپ نوری است [۴]. روش‌های دیگر گزارش‌شده در متون علمی شامل تکنیک‌های کروماتوگرافی، آنالیز رادیویی ایزوتوپ کربن پایدار و رزونانس مغناطیسی هسته‌ای می‌باشد [۵، ۶]. اشکالات اصلی همه روش‌ها هزینه زیاد، صرف زمان زیاد، نیاز به تخصص همچنین تخریب نمونه است. این امر منجر به توسعه روش‌های تجزیه‌ای برای تشخیص اصالت عسل شده است که از جمله آن می‌توان به فناوری طیف‌سنجی در ترکیب با روش‌های کمومتریکس اشاره کرد که یک جایگزین خوب

برای کنترل کیفیت سریع، قابل اعتماد و سازگار با محیط‌زیست است [۷]. رایج‌ترین روش‌های کمومتری مورد‌استفاده در واکاوی عسل عبارتند از: آنالیز اجزای اصلی (PCA)، آنالیز خوشه‌بندی سلسله‌مراتبی (HCA)، آنالیز تفکیک خطی (LDA)، حداقل مربع جزئی (PLS) و رگرسیون مؤلفه اصلی (PCR) [۸].

از تکنیک‌های طیف‌سنجی موجود، طیف‌سنجی رامان جهت واکاوی موادغذایی ویژگی‌های مناسبی دارد که شامل عدم تداخل آب موجود در نمونه با اندازه‌گیری رامان، سهولت نمونه‌برداری و اندازه‌گیری و حداقل تداخل فلورسانس تغییر ماتریس نمونه می‌باشد. در سال‌های اخیر، روش‌های واکاوی مبتنی‌بر طیف‌سنجی رامان به‌عنوان یک روش اقتصادی و سریع برای تعیین منشاء عسل مورد‌استفاده قرار گرفته است [۹، ۱۰].

کوروچی و همکاران [۹]، توانایی تشخیص منشاء گیاهی عسل را با استفاده از تکنیک melissopalynology و با طیف‌سنجی رامان در ترکیب با آنالیزهای چندمتغیره (PCA) مقایسه کرد. نمونه‌های عسل از ایتالیا، اروپای شرقی و اسپانیا جمع‌آوری شد. با توجه به نتایج، تمایز منشاء عسل‌های مورد‌مطالعه از ۸۵٪ به ۹۹٪ بهبود یافته است. روش تحلیلی فراستو-ریس و همکاران [۱۱]، جهت تعیین منشا گل و گیاه، در عسل تولیدشده توسط گونه خاصی از زنبور (*Apis Mellifera*)، از طیف‌سنجی رامان همراه با PCA استفاده کردند. در این مطالعه از ۶۶ نمونه عسل تک گل و چند گل جمع‌آوری‌شده از چندین منطقه مکزیک با انواع شرایط آب و هوایی استفاده شد. نتایج حاکی از آن بود که استفاده از روش کمومتری برای طبقه‌بندی عسل‌ها براساس منشاء و خلوص با صحت ۹۰٪ مناسب بود. ژاندریک و همکاران [۱۲] از ترکیب روش‌های طیف‌سنجی رامان، طیف‌سنجی فرسرخ نزدیک و فرسرخ تبدیل فوریه<sup>۱</sup> جهت تشخیص منشا نمونه‌های عسل به‌دست آمده از نیوزلند، در طیف بین  $\text{cm}^{-1}$  ۲۰۰ تا ۱۲۰۰۰ استفاده کردند. این روش با استفاده از PLS تکمیل شد. نتایج برازش مدل (R<sup>2</sup>)، خطای استاندارد کالیبراسیون (SEC) و خطای استاندارد پیش‌بینی (SEP) را به‌ترتیب ۰/۸۵، ۰/۲۱۹ و ۰/۳۱۵ نشان داد. ارویان و

1. Fourier-transform infrared spectroscopy

مطالعه طاهر و همکاران [۱۷]، طیف‌سنجی رامان همراه با PLS برای پیش‌بینی ترکیبات فنلی و فعالیت آنتی‌اکسیدانی عسل استفاده شد. مشخص شد که مدل‌های توسعه‌یافته مبتنی بر رامان نسبت به مدل‌های موجود برتری دارند. طیف‌سنجی رامان و همچنین مدل‌سازی PLS-LDA برای تعیین تقلب عسل چینی با شربت ذرت استفاده شده است [۱۸]. واکاوی اختلاط و تقلب نمونه‌ها را در محدوده ۱۰، ۲۰ و ۴۰٪ در نظر گرفت. صحت پیش‌بینی ۸۴/۴٪ به دست آمد که نشان می‌دهد ترکیب PLS-LDA با طیف رامان یک روش بالقوه در تشخیص ناخالصی در عسل است.

طیف‌سنجی رامان به‌عنوان روشی سریع، غیرمخرب و دقیق برای پیش‌بینی ویژگی‌های فیزیکوشیمیایی مواد غذایی مشابه عسل، جایگاه ویژه‌ای در پژوهش‌های اخیر یافته است. در این نوع محصولات که دارای ماهیتی ویسکوز و غنی از قند هستند، ویژگی‌هایی مانند ترکیب قندی (گلوکز، فروکتوز، ساکارز و مالتوز)، مواد جامد محلول کل (Brix°)، رطوبت، فعالیت آبی، اسیدیته، pH، اسیدهای آلی، شاخص‌های حرارت‌دیدگی نظیر ۵-هیدروکسی‌متیل‌فورفورال (HMF)، ترکیبات فنولی، هدایت الکتریکی، وضعیت فیزیکی (تبلور) و تشخیص تقلب یا تعیین منشاء گیاهی و جغرافیایی، با استفاده از طیف‌سنجی رامان و روش‌های شیمی‌سنجی به‌طور مؤثری ارزیابی شده‌اند [۱۹]. این روش بر پایه‌ی ثبت اثر انگشت ارتعاشی مولکول‌ها استوار است، به‌گونه‌ای که باندهای مربوط به پیوندهای C-O، C-C و C-H در نواحی ۷۰۰ تا ۱۲۰۰  $\text{cm}^{-1}$  و ناحیه‌ی کششی O-H اطلاعات دقیقی از نوع قند و میزان آب فراهم می‌کند [۲۰]. پس از جمع‌آوری طیف‌های خام، داده‌ها با حذف خط پایه و فلورسانس، هموارسازی و نرمال‌سازی پردازش می‌شوند و سپس مدل‌های آماری مانند PLSR، PCR یا مدل‌های غیرخطی نظیر ANN، FFNN، SVM و CNN برای پیش‌بینی کمی یا طبقه‌بندی اصالت نمونه‌ها به‌کار می‌روند [۱۹]. مطالعات متعدد نشان داده‌اند که این روش قادر است مقادیر گلوکز و فروکتوز را با ضریب تعیین بالای ۰/۹ و دقتی هم‌تراز با روش‌های مرجع کروماتوگرافی پیش‌بینی کند، و در تشخیص تقلب‌های رایج با شربت ذرت، مالتوز یا شکر معکوس عملکردی بالای ۰/۹۵ دارد [۲۰]. برای مقادیر جزئی

Ropciuc [۱۰]، از واکاوی طیف رامان برای تشخیص منشا گیاهی ۷۶ نمونه عسل رومانی استفاده کردند. نتایج حاکی از آن بود که استفاده از این روش تجزیه همراه با LDA، با دستیابی به ۸۳/۳٪ صحت اعتبارسنجی، بسیار عالی عمل نموده است.

همچنین از طیف‌سنجی رامان همراه با روش‌ها و مدل‌های آماری چندمتغیره برای تشخیص عوامل خارجی مؤثر بر کیفیت عسل استفاده شده است [۱۳، ۱۴]. طیف‌سنجی رامان و مدل‌های کمومتریکس برای پیش‌بینی غلظت گلوکز، فروکتوز، ساکارز و مالتوز موجود در نمونه‌های عسل ترکیه و یونان استفاده شده است [۱۵]. همبستگی بین سطوح کمی قند و طیف رامان هم با استفاده از PLS و هم شبکه‌های عصبی مصنوعی به دست آمد.  $R^2$  آماری برای گلوکز، فروکتوز، ساکارز و مالتوز به ترتیب با ۰/۹۲۹، ۰/۹۳۰، ۰/۹۳۷ و ۰/۸۹۳ برای PLS و ۰/۹۳۱، ۰/۹۳۰، ۰/۹۵۶ و ۰/۹۱۳ برای شبکه عصبی به دست آمد که نشان می‌دهد هر دو روش کمومتریکس برای آنالیز سریع مقدار قند مناسب است. ارویان و همکاران [۱۰]، از طیف‌سنجی رامان برای تشخیص عسل تقلبی با قند (گلوکز، فروکتوز، قند اینورت، شربت اینولین هیدرولیزشده، و مالت) استفاده کردند. در این مطالعه از ۹۰۰ نمونه با سطوح تقلب ۵، ۱۰، ۲۰، ۳۰، ۴۰ و ۵۰٪ استفاده شد. تشخیص خلوص عسل با استفاده از PLS و PCR انجام شد. مدل‌های توسعه‌یافته کمومتریکس تناسب خوبی برای کالیراسیون ( $R^2_{\text{cal}} = ۰/۹۸۳$ ) و اعتبارسنجی ( $R^2_{\text{val}} = ۰/۹۸۱$ )، با خطاهای آماری کم ( $\text{SEC} = ۰/۰۰۹$ ) و ( $\text{SEP} = ۰/۱۰۳$ ) نشان دادند. آنجوس و همکاران [۱۶]، پتانسیل طیف‌سنجی رامان را در پیش‌بینی ترکیب فیزیکوشیمیایی عسل تک گل اسطوخودوس spp ارزیابی کرد. مدل‌های PLS برای تخمین کمی استفاده شد و نتایج حاصله با مقادیر به دست آمده از روش‌های مرجع مقایسه شد. همچنین از مدل‌های کمومتریکس برای pH، قندهای احیا، هدایت الکتریکی ساکارز، محتوای فنل کل، محتوای فلاونوئید کل، پرولین و کل اسیدها، استفاده شد.  $R^2$  کالیراسیون به دست آمده، در محدوده ۰/۹۹-۰/۹۷۳ و اعتبارسنجی در محدوده ۰/۹۹-۰/۸۳۳، SEC در محدوده ۰/۰۱-۱/۷۱ و SEP در محدوده ۰/۰۱-۲/۰۳ و ۰/۰۱-۱/۷۱ به دست آمد. در

## ۲.۰۱. ویژگی‌های فیزیکوشیمیایی

موادشیمیایی مورد استفاده برای انجام آزمون‌های قند احیاکننده قبل از هیدرولیز، رطوبت، ساکارز و هیدروکسی‌متیل‌فورفورال شامل محلول محلول فهلینگ A و B، الکل اتیلیک، بی‌سولفیت سدیم، محلول استاندارد اسید هیدروکلریک (HCl)، معرف فنل فتالین، محلول استاندارد هیدروکسید سدیم (NaOH)، آب دی‌یونیزه، بی‌سولفیت سدیم و محلول‌های کاریزیک و دو بودند که همگی در گرید آزمایشگاهی تهیه شدند. توصیف دقیق مراحل و روش‌های به‌دست‌آوردن هر یک از ویژگی‌های فیزیکوشیمیایی در ادامه ارائه شده است:

### ۱.۱.۲. آزمون قند احیاکننده قبل از هیدرولیز

یک گرم از نمونه عسل در آب مقطر حل و به بالن ژوژه ۲۵۰ ml به حجم رسانده شد و از محلول تهیه‌شده جهت انجام تیتراسیون محلول فهلینگ A و B (هر کدام ۵ ml) استفاده شد [۲۲]. درصد قند احیا مطابق فرمول یک محاسبه شد:

$$S = \frac{F \times 250 \times 100}{V \times W \times 1000} \quad (1)$$

که در آن:

S قندهای احیاکننده بر حسب گرم

F عیار فهلینگ

V حجم مصرفی محلول بر حسب میلی لیتر

W وزن نمونه عسل بر حسب گرم

۱۰۰۰ تبدیل میلی گرم به گرم

### ۲.۰۱.۲. قندهای احیاکننده بعد از هیدرولیز

۵۰ ml از محلول نمونه تهیه‌شده را به یک بالن حجمی ۱۰۰ منتقل کرده و به آن ۲ ml اسید کلریدریک غلیظ اضافه شد و به مدت ۱۰ دقیقه در حمام آب گرم در دمای ۷۰°C قرار داده شد. پس از سرد شدن، در حضور معرف فنل فتالین با محلول سود تا ظهور رنگ ارغوانی ضعیف خنثی‌سازی انجام شد. سپس با آب مقطر به حجم ۱۰۰ رسانده شد. مطابق قند قبل از هیدرولیز، از این محلول برای تیتراسیون محلول‌های فهلینگ A و B در حالت جوش استفاده شد. درصد قند بعد از هیدرولیز طبق فرمول ۲ محاسبه شد [۲۲]:

ترکیباتی مانند HMF یا آلاینده‌ها، از رامن تقویت‌شده سطحی (SERS) جهت افزایش حساسیت بهره گرفته می‌شود و فناوری‌های جدیدی مانند رامن با جابه‌جایی فضایی و دستگاه‌های دستی، امکان غربالگری غیرمخرب در محل تولید یا بازار را فراهم کرده‌اند [۲۱]. مهم‌ترین مزایای این روش، سرعت بالا، عدم نیاز به آماده‌سازی نمونه و قابلیت حمل آن است، درحالی‌که چالش‌هایی مانند تداخل فلورسانس، تغییرپذیری ترکیب نمونه‌ها و ضرورت اعتبارسنجی مدل‌ها نسبت به روش‌های مرجع نیز باید مدیریت شوند. پژوهش‌های منتشرشده، شواهد مستندی از دقت، قابلیت اطمینان و پتانسیل صنعتی طیف‌سنجی رامن در پیش‌بینی سریع ویژگی‌های فیزیکوشیمیایی و ارزیابی اصالت محصولات غذایی مشابه عسل ارائه کرده‌اند [۱۹، ۲۰]. در این مطالعه، از مدل‌سازی چندمتغیره حداقل مربعات جزئی (PLS) برای توسعه یک سیستم یادگیری هوشمند استفاده شده است که قادر به برقراری ارتباط بین طیف‌های رامن عسل و چهار ویژگی فیزیکوشیمیایی آن باشد. این ویژگی‌ها شامل غلظت قندهای احیاکننده پیش از هیدرولیز، غلظت ساکاروز، میزان رطوبت و مقدار هیدروکسی‌متیل‌فورفورال (HMF) در نمونه‌های عسل است. مدل‌های توسعه‌یافته به‌منظور پیش‌بینی دقیق و کارآمد این ویژگی‌ها براساس طیف‌های رامن ثبت‌شده از نمونه‌های عسل طراحی شده‌اند و ابزاری مؤثر برای کنترل کیفی سریع و اقتصادی عسل تولیدشده در مناطق مختلف ایران فراهم می‌کنند.

## ۲. مواد و روش

حدود ۵۱ نمونه عسل‌های موجود در بازار مناطق مختلف ایران با توجه به نکات نمونه‌برداری جمع‌آوری و کدگذاری شد. آزمون‌های فیزیکوشیمیایی مطابق استاندارد ملی ایران شماره ۹۲ [۲۲] و همچنین طیف‌سنجی هر نمونه با دستگاه رامن انجام شد. روش‌های یادگیری ماشین (ML) به‌منظور مدل‌کردن ارتباط بین طیف رامن و ویژگی‌های فیزیکوشیمیایی عسل توسعه داده شد. جزئیات روش آزمون‌های فیزیکوشیمیایی، طیف‌گیری با اسپکتروسکوپی رامن و مدل‌سازی به شرح زیر می‌باشد:

شد. غلظت هیدروکسی متیل فورفورال بر حسب mg/kg با استفاده از فرمول‌های ۴ و ۵ بیان گردید [۲۲]:

$$149.7 = \frac{126 \times 1000 \times 1000}{16830 \times 10 \times W} \quad (4)$$

$$HMF = (A_{284} - A_{336}) \times 149.7 \times 5 \times \frac{D}{W} \quad (5)$$

که در آن:

۱۴۹/۷ ضریب تبدیل جذب به غلظت

۱۶۸۳۰ ضریب جذب مولی هیدروکسی متیل فورفورال

در ۲۸۴ nm

۱۲۶ وزن مولکولی هیدروکسی متیل فورفورال

۱۰۰۰ تبدیل Kg/g به g/mg

D ضریب رقت (حجم محلول نهایی تقسیم بر ده)

## ۲.۲. تحلیل طیف سنجی رامان

دستگاه طیف‌سنجی رامان مدل TDLG100 تکسان ساخت ایران با قابلیت گردآوری طیف و دیتا توسط نرم‌افزار Tansu مورد استفاده قرار گرفت. هر اندازه‌گیری در محدوده طیف ۱۰۰ تا ۳۵۰۰ cm<sup>-1</sup> انجام گردید. به‌منظور طیف‌گیری از نمونه‌ها از لیزر ۷۸۵ نانومتری YAG و همچنین برای تشخیص نور پراکنده رامان از آشکارساز Hamamatsu استفاده گردید. سه پارامتر اساسی برای بهینه‌سازی شدت پیک رامان انتخاب شد: قدرت لیزر، زمان ادغام و تعداد میانگین. این پارامترها به ترتیب ۳۰۰ mW، ۳ ثانیه و ۵ انتخاب شدند. برای انجام اسکن، دو گرم از هر یک از نمونه‌ها، در سه چاهک (سه تکرار) از پلیت ۲۰ خانه ریخته‌شده و سپس در نگهدارنده نمونه زیر میکروسکوپ رامان قرار گرفت. نور مستقیماً بر روی نمونه‌ها متمرکز گردید و میانگین طیف آن‌ها پس از پنج اسکن برای سه تکرار اندازه‌گیری گردید.

## ۳.۲. توسعه و ارزیابی مدل‌های کمومتریکس

برای توسعه مدل‌های کمومتریکس جهت پیش‌بینی ویژگی‌های فیزیکوشیمیایی عسل، از داده‌های طیف‌سنجی رامان به‌عنوان متغیرهای مستقل و چهار ویژگی فیزیکوشیمیایی (میزان قند پیش از هیدرولیز، مقدار ساکاروز، میزان رطوبت و درصد هیدروکسی‌متیل‌فورفورال

$$S_1 = \frac{F \times 250 \times 100 \times 100}{W \times V \times 50 \times 1000} \quad (2)$$

که در آن:

F عیار فهلینگ

V حجم مصرفی محلول بر حسب میلی لیتر

W وزن نمونه عسل بر حسب گرم

۱۰۰۰ تبدیل میلی گرم به گرم

## ۲.۱.۳. درصد ساکارز

درصد ساکارز مطابق با فرمول ۳ محاسبه شد:

$$(\bar{N}) = (S_1 - S) \times 0.95 \quad (3)$$

که در آن:

S درصد قند قبل از هیدرولیز

S1 درصد قند بعد از هیدرولیز

## ۲.۱.۴. آزمون رطوبت

میزان رطوبت با استفاده از روش رفراکتومتری اندازه‌گیری شد. به این منظور، یک گرم عسل در دمای °C ۲۵ توسط رفراکتومتر (BPTR-100, BenchTop Refractometer)، آنالیز گردید. درصد رطوبت (g/100g عسل) بر اساس جدول رابطه میران آب و ضریب شکست ارائه‌شده در استاندارد ملی ایران [۲۲] گزارش گردید.

## ۲.۱.۵. آزمون هیدروکسی متیل فورفورال

جهت تعیین میزان هیدروکسی متیل فورفورال موجود در نمونه عسل، ۵ g عسل در ۲۵ ml آب دی‌یونیزه حل و به بالن ژوزه ۵۰ ml منتقل گردید. محلول با افزودن ۰/۵ ml معرف کاریز ۱ و کاریز ۲ شفاف شد و حجم آن به ۵۰ ml رسانده شد. سپس محلول با استفاده از کاغذ صافی واتمن شماره ۴۲ صاف و به دو لوله آزمایش هر یک ۵ ml از محلول صاف‌شده اضافه گردید. به یکی از دو لوله ۵ ml آب مقطر و به لوله دیگر ۵ ml بی سولفیت سدیم (لوله شاهد) جهت غیرفعال‌شدن جذب هیدروکسی متیل فورفورال اضافه شد. جذب نوری نمونه در دستگاه طیف‌سنج UV-Vis مدل (Photonix Ar5- UV/Vis Spectrophotometer) در طول‌موج‌های ۲۸۴ nm و ۳۳۶ در سل کوارتز اندازه‌گیری



انجام شد تا اطمینان حاصل شود که نمونه‌های انتخاب‌شده برای هر مجموعه نماینده تنوع کامل داده‌ها هستند.

### ۲.۳.۲. الگوریتم Kennard-Stone

الگوریتم Kennard-Stone یک روش سیستماتیک برای تقسیم داده‌ها به مجموعه‌های کالیبراسیون و تست است که با هدف حداکثر کردن تنوع داده‌ها در هر مجموعه طراحی شده است. این الگوریتم با انتخاب نمونه‌هایی که بیشترین فاصله اقلیدسی را در فضای داده‌ها دارند، شروع می‌کند و به تدریج نمونه‌های دیگر را بر اساس توزیع یکنواخت انتخاب می‌کند. این روش تضمین می‌کند که مجموعه کالیبراسیون شامل نمونه‌هایی با تنوع کافی برای آموزش مدل باشد و مجموعه تست نیز نماینده کل داده‌ها باشد. در این مطالعه، استفاده از این الگوریتم باعث شد که مدل‌های PLS با داده‌هایی آموزش ببینند که توزیع گسترده‌ای از ویژگی‌های طیفی و فیزیکوشیمیایی را پوشش می‌دهند، و در نتیجه، کارایی پیش‌بینی مدل‌ها بهبود یافت.

### ۴.۳.۲. اعتبارسنجی متقاطع و الگوریتم Leave-One-Out

برای بهینه‌سازی تعداد متغیرهای پنهان در مدل‌های PLS و ارزیابی اولیه کارایی آن‌ها، از تکنیک اعتبارسنجی متقاطع (Cross-Validation) با الگوریتم Leave-One-Out استفاده شد. در این روش، هر بار یک نمونه از مجموعه کالیبراسیون کنار گذاشته می‌شود و مدل با استفاده از سایر نمونه‌ها آموزش داده می‌شود. سپس، نمونه کنار گذاشته‌شده برای پیش‌بینی و ارزیابی خطای مدل استفاده می‌شود. این فرآیند برای تمامی نمونه‌های مجموعه کالیبراسیون تکرار می‌شود تا میانگین خطای پیش‌بینی محاسبه شود. این روش به زبان ساده، مانند امتحان کردن مدل با حذف تک‌تک نمونه‌ها و بررسی اینکه مدل چقدر می‌تواند ویژگی‌های نمونه حذف‌شده را به‌درستی پیش‌بینی کند، عمل می‌کند. این تکنیک به‌ویژه برای مجموعه‌های داده کوچک، مانند مورد حاضر، بسیار مناسب است، زیرا از تمام داده‌ها به‌طور مؤثر برای ارزیابی مدل استفاده می‌کند. در این مطالعه، تعداد متغیرهای پنهان (LV) برای هر مدل با استفاده از این روش بهینه شد تا بهترین تعادل بین دقت و پیچیدگی مدل به دست آید.

(HMF) به‌عنوان متغیرهای وابسته استفاده شد. طیف‌های رامان در محدوده ۱۰۰ تا  $3500\text{ cm}^{-1}$  برای تمامی نمونه‌های عسل ثبت شدند. این داده‌های طیفی به‌صورت ماتریس‌های عددی سازمان‌دهی شده و به‌عنوان ورودی‌های مدل‌سازی مورد استفاده قرار گرفتند. برای هر یک از چهار ویژگی فیزیکوشیمیایی، یک مدل مستقل مبتنی بر رگرسیون حداقل مربعات جزئی (PLS) طراحی و توسعه داده شد تا ارتباط بین داده‌های طیفی و مقادیر آزمایشگاهی این ویژگی‌ها به‌صورت دقیق مدل‌سازی شود.

### ۱.۳.۲. رگرسیون حداقل مربعات جزئی (PLS)

رگرسیون حداقل مربعات جزئی (PLS) یک روش کمومتریکس چندمتغیره است که برای مدل‌سازی روابط پیچیده بین مجموعه‌ای از متغیرهای مستقل (در اینجا داده‌های طیف رامان) و متغیرهای وابسته (ویژگی‌های فیزیکوشیمیایی عسل) به‌کار می‌رود. این روش به‌ویژه در مواردی که تعداد متغیرهای مستقل زیاد بوده و ممکن است هم‌خطی (collinearity) بین آن‌ها وجود داشته باشد، بسیار مؤثر است. در PLS، داده‌ها به متغیرهای پنهان (latent variables) تبدیل می‌شوند که ترکیبی خطی از متغیرهای اصلی هستند و بیشترین واریانس در داده‌های مستقل و وابسته را توضیح می‌دهند. این روش با کاهش ابعاد داده‌ها و کاهش نویز، دقت پیش‌بینی را بهبود می‌بخشد. در این مطالعه، از الگوریتم PLS1 استفاده شد که برای پیش‌بینی یک متغیر وابسته در هر مدل بهینه‌سازی شده است. این الگوریتم با انتخاب تعداد مناسب متغیرهای پنهان، مدل‌هایی با دقت بالا و خطای پیش‌بینی کم را مورد هدف قرار می‌دهد.

### ۲.۳.۲. تقسیم‌بندی داده‌ها

داده‌های جمع‌آوری شده برای هر ویژگی فیزیکوشیمیایی به دو مجموعه کالیبراسیون (آموزش) و تست (اعتبارسنجی) تقسیم شدند تا کارایی مدل‌های PLS ارزیابی شود. مجموعه کالیبراسیون برای آموزش مدل و تنظیم پارامترهای آن استفاده شد، در حالی که مجموعه تست، که شامل نمونه‌های مستقل بود، برای ارزیابی توانایی مدل در پیش‌بینی داده‌های جدید به‌کار رفت. تقسیم‌بندی داده‌ها به دو مجموعه کالیبراسیون و تست با استفاده از الگوریتم Kennard-Stone

## ۵.۳.۲. نرم‌افزار و ابزارهای محاسباتی

تمامی مراحل مدل‌سازی در محیط MATLAB (MathWorks, Natick, MA, USA) انجام شد. برای پیاده‌سازی مدل‌های PLS، از نرم‌افزار (Multivariate Calibration 1) MVC1 استفاده شد که به‌طور خاص برای تحلیل‌های کمومتریکس طراحی شده است. داده‌های طیفی رامان ابتدا در نرم‌افزار Microsoft Excel 2013 سازمان‌دهی و به ماتریس‌های داده تبدیل شدند. سپس، این داده‌ها به MATLAB منتقل شده و برای مدل‌سازی و تحلیل‌های آماری پردازش شدند. برای کاهش نویز و بهبود کیفیت داده‌ها، تنها از پیش‌پردازش mean centering طیف‌ها استفاده شد.

## ۳. یافته‌ها و بحث

## ۳.۱. آنالیزهای فیزیکوشیمیایی

همان‌گونه که در بخش مواد و روش‌ها اشاره شد، در این پژوهش تعداد ۵۱ نمونه عسل جمع‌آوری و با استفاده از طیف‌سنجی رامان در سه تکرار مستقل آنالیز گردیدند. داده‌های طیفی استخراج‌شده به‌منظور توسعه چهار مدل کمومتریک بر پایه روش PLS با مقادیر آزمایشگاهی شاخص‌های فیزیکوشیمیایی منتخب شامل قندهای احیاکننده پیش از هیدرولیز، رطوبت، ساکاروز و HMF همبسته شدند. هدف از این مرحله، ایجاد مدلی غیرمخرب برای پیش‌بینی ویژگی‌های کیفی عسل و بررسی میزان دقت آن‌ها در قیاس با روش‌های مرجع آزمایشگاهی بود. در ادامه، نتایج به‌دست‌آمده از اندازه‌گیری این ویژگی‌ها و مقایسه آن‌ها با حدود استانداردهای ملی (شماره ۹۲) ارائه و مورد بحث قرار می‌گیرند.

جدول (۱) مقادیر به دست آمده از آزمون‌های فیزیکوشیمیایی

Table 1. Values obtained from physicochemical tests

Parameter	Unit	National Standard	Mean $\pm$ Standard Deviation	Minimum	Median	Maximum
Sugar before hydrolysis	g/100g	Minimum 65	72.49 $\pm$ 1.93	68.62	72.00	76.00
Sugar after hydrolysis	g/100g	-	76.08 $\pm$ 1.45	74.00	75.65	79.05
Sucrose	g/100g	Maximum 5	4.67 $\pm$ 3.87	2.00	3.88	20.30
HMF	mg/kg	Maximum 40	7.99 $\pm$ 3.93	2.82	6.60	15.00
Moisture	%	Maximum 20	15.56 $\pm$ 1.32	13.60	15.60	18.00

## ۳.۲. تحلیل طیف رامان

طیف‌های رامان به‌دست‌آمده از نمونه‌های عسل دارای باندهای طیفی در بازه‌های مختلف هستند که هر یک بیان‌کننده ارتعاشات و ساختار مولکولی خاصی می‌باشد. مطالعات مختلف طیف‌های رامان به‌دست آمده از نمونه‌های عسل را مورد بررسی قرار داده‌اند و به‌طور کلی بیان داشته‌اند که ناحیه طیفی بین ۲۳۰-۵۱۰  $\text{cm}^{-1}$  مربوط به ارتعاشات کششی و خمشی گروه‌های C-O، C-C-O و C-C-C است که ساختار مولکولی قندها را تشکیل می‌دهند (۱۷). ناحیه ۵۹۵-۶۹۱  $\text{cm}^{-1}$  به ارتعاشات کششی حلقه‌های غیراشباع موجود در HMF، کاروتن‌ها، فلاون‌ها، فلاونوئیدها و پلی‌فنول‌ها نسبت داده می‌شود (۲۳). قله مشاهده‌شده در ۶۹۱-۷۵۲  $\text{cm}^{-1}$  مربوط به ارتعاشات کششی C-O و C-C و O و ارتعاشات خمشی O-C-O است. در مقابل، باند ۷۷۰-

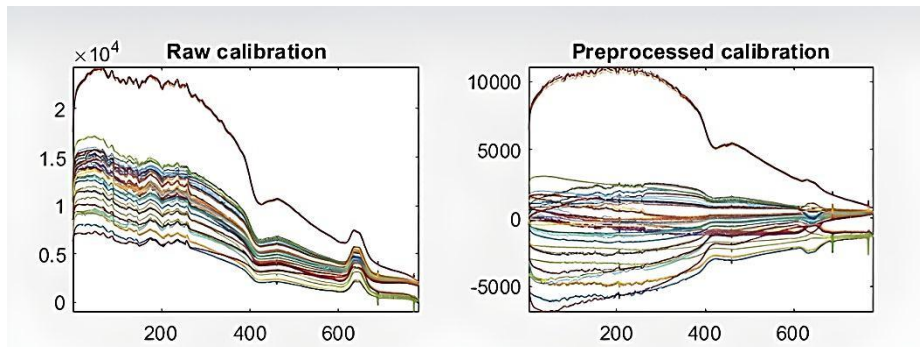
براساس نتایج حاصل از آنالیز نمونه‌های عسل، مهم‌ترین یافته‌های این پژوهش به‌شرح زیر است. مقدار ساکارز با میانگین  $4/67 \pm 3/87$  g/100g، در محدوده قابل‌قبول استانداردهای ملی و بین‌المللی (کمتر از ۵ g/100g) قرار گرفت، هرچند در یک نمونه مقدار غیرمعمول ۲۰/۳ g/100g مشاهده شد که احتمالاً نشان‌دهنده تغذیه زنبورها با شکر است. میزان HMF با میانگین  $7/99 \pm 3/93$  mg/kg، در اکثر نمونه‌ها پایین بود که نشان‌دهنده تازگی و شرایط نگهداری مناسب عسل است. رطوبت نمونه‌ها بین ۱۳/۶ تا ۱۸٪ متغیر بود و میانگین آن ۱۵/۵۶٪ به دست آمد که مطابق با استانداردهای جهانی (کمتر از ۲۰٪) است. این نتایج حاکی از کیفیت مطلوب و اصالت اکثر نمونه‌های مورد بررسی است، اگرچه نیاز به نظارت بیشتر بر برخی نمونه‌ها احساس می‌شود.

استفاده از این مدل‌ها برای پیش‌بینی ویژگی‌هایی مانند قندهای احیاکننده، ساکاروز، رطوبت و هیدروکسی‌متیل‌فورفورال (HMF) می‌تواند به‌طور قابل‌توجهی فرآیندهای آنالیز را تسریع کرده و هزینه‌ها را کاهش دهد. این روش‌ها با بهره‌گیری از داده‌های طیف‌سنجی رامان و الگوریتم‌های یادگیری ماشین، مانند رگرسیون حداقل مربعات جزئی (PLS)، امکان توسعه سیستم‌های هوشمند و غیرمخرب را فراهم می‌کنند که می‌توانند به‌صورت سریع و دقیق ویژگی‌های کیفی عسل را تخمین بزنند. شکل (۱) طیف‌های رامان ثبت‌شده از مجموعه نمونه‌های عسل مورد بررسی در این مطالعه را نشان می‌دهد. این طیف‌ها در محدوده ۱۰۰ تا  $3500\text{ cm}^{-1}$  ثبت شده‌اند و پس از اعمال پیش‌پردازش میانگین-مرکزی نمایش داده شده‌اند. پیش‌پردازش میانگین-مرکزی با کسر میانگین داده‌ها از هر نقطه، به کاهش نویز و ساده‌سازی مدل‌های کالیبراسیون کمک می‌کند. این فرآیند باعث می‌شود مدل‌های کومتریکس با پیچیدگی کمتری توسعه یابند و دقت پیش‌بینی آن‌ها بهبود یابد، زیرا تغییرات غیرمرتبط با متغیرهای هدف حذف شده و تمرکز مدل بر الگوهای معنادار داده‌ها افزایش می‌یابد.

$917\text{ cm}^{-1}$  حاصل ارتعاشات کششی گروه‌های C-H و C-C موجود در گلوکز است [۲۴]. باندهای بین  $820-1024\text{ cm}^{-1}$  به ارتعاشات تغییرشکل C-H و پیوندهای متیلن- $\text{CH}_2$  و همچنین ارتعاشات خمشی C-O-H مربوط هستند [۲۵]. قله موجود در  $1024-1094\text{ cm}^{-1}$  به ارتعاشات خمشی پیوندهای C-H و C-O-H قندها و همچنین ارتعاشات خمشی پیوند C-N در اسیدهای آمینه و پروتئین‌ها نسبت داده می‌شود [۲۶]. باند  $1094-1191\text{ cm}^{-1}$  به ارتعاشات کششی پیوندهای C-O و C-O-C قندها و همچنین پیوندهای C-N در پروتئین‌ها و اسیدهای آمینه اختصاص دارد [۱۵]. در نهایت، ناحیه طیفی  $1262-1300\text{ cm}^{-1}$  مربوط به ارتعاشات C-H و O-C-H بوده، درحالی‌که باندهای  $1300-1460\text{ cm}^{-1}$  ناشی از ارتعاشات خمشی و نوسانی گروه‌های عاملی CH و OH هستند [۲۶].

### ۳.۳. توسعه و ارزیابی مدل‌های کومتریکس

مدل‌سازی مبتنی بر یادگیری ماشین به‌عنوان ابزاری قدرتمند در تحلیل داده‌های پیچیده، امکان پیش‌بینی ویژگی‌های فیزیکوشیمیایی را فراهم می‌کند که اندازه‌گیری آن‌ها با روش‌های سنتی معمولاً زمان‌بر، پرهزینه و نیازمند تجهیزات تخصصی است. در حوزه کنترل کیفیت عسل،



شکل (۱). طیف‌های رامان ثبت‌شده از نمونه‌های عسل در محدوده ۱۰۰ تا  $3500\text{ cm}^{-1}$  نمودار سمت چپ طیف‌های خام و نمودار سمت راست طیف‌های پیش‌پردازش‌شده با روش میانگین-مرکزی را نشان می‌دهد. این پیش‌پردازش به ساده‌سازی مدل‌های کالیبراسیون و کاهش نویز کمک می‌کند.

**Fig. 1.** Raman spectra recorded from honey samples in the range of 100 to  $3500\text{ cm}^{-1}$ . The left graph shows the raw spectra, and the right graph shows the spectra preprocessed with the mean-centering method. This preprocessing helps simplify calibration models and reduce noise.

برای توسعه ماشین‌های یادگیری هوشمند که بتوانند با دریافت طیف‌های رامان نمونه‌های عسل، چهار ویژگی فیزیکوشیمیایی کلیدی (قندهای احیاکننده پیش از

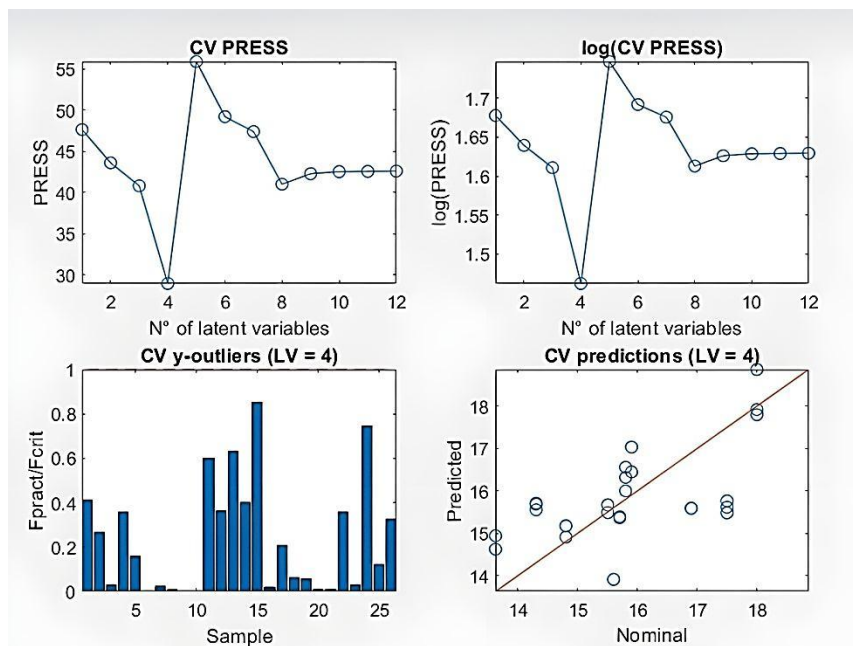
هیدرولیز، ساکاروز، رطوبت و HMF) را پیش‌بینی کنند، از مدل‌های کالیبراسیون مبتنی بر رگرسیون حداقل مربعات جزئی (PLS) استفاده شد. این مدل‌ها به‌گونه‌ای طراحی

شدند که با بهره‌گیری از داده‌های طیفی به‌عنوان متغیرهای مستقل، ارتباط دقیقی با مقادیر آزمایشگاهی ویژگی‌های هدف برقرار کنند. هر مدل به‌صورت مستقل برای یکی از ویژگی‌های ذکرشده توسعه یافت تا توانایی پیش‌بینی دقیق و قابل‌اعتماد برای هر ویژگی را بررسی کند.

در هر چهار مدل توسعه‌یافته، تعداد متغیرهای پنهان (Latent Variables) با استفاده از تکنیک اعتبارسنجی متقاطع (Cross-Validation) و براساس کمینه‌کردن خطای ریشه میانگین مربعات اعتبارسنجی (RMSECV) انتخاب شد. این روش با حذف متوالی نمونه‌ها از مجموعه کالیبراسیون و ارزیابی عملکرد مدل در پیش‌بینی آن‌ها، تعداد بهینه متغیرهای پنهان را تعیین کرد. جدول ۱ تعداد متغیرهای پنهان انتخاب‌شده برای هر مدل را نشان می‌دهد. این رویکرد تضمین کرد که مدل‌ها نه تنها از پیچیدگی غیرضروری جلوگیری کنند، بلکه توانایی تعمیم‌دهی مناسبی برای پیش‌بینی داده‌های جدید داشته باشند.

شکل (۲) نتایج اعمال تکنیک اعتبارسنجی متقاطع

شکل (۲). نتایج اعتبارسنجی متقاطع برای مدل پیش‌بینی میزان رطوبت عسل با استفاده از رگرسیون حداقل مربعات جزئی. ردیف بالا (چپ و راست): نمودارهای تعیین تعداد متغیرهای پنهان با کمینه‌سازی خطای پیش‌بینی، نشان‌دهنده بهینه بودن ۴ متغیر پنهان. ردیف پایین (چپ): نمودار تست پرت بر اساس تست F نسبی، ردیف پایین (راست): نمودار مقادیر پیش‌بینی شده در مقابل مقادیر واقعی در اعتبارسنجی متقاطع با ۴ متغیر پنهان.

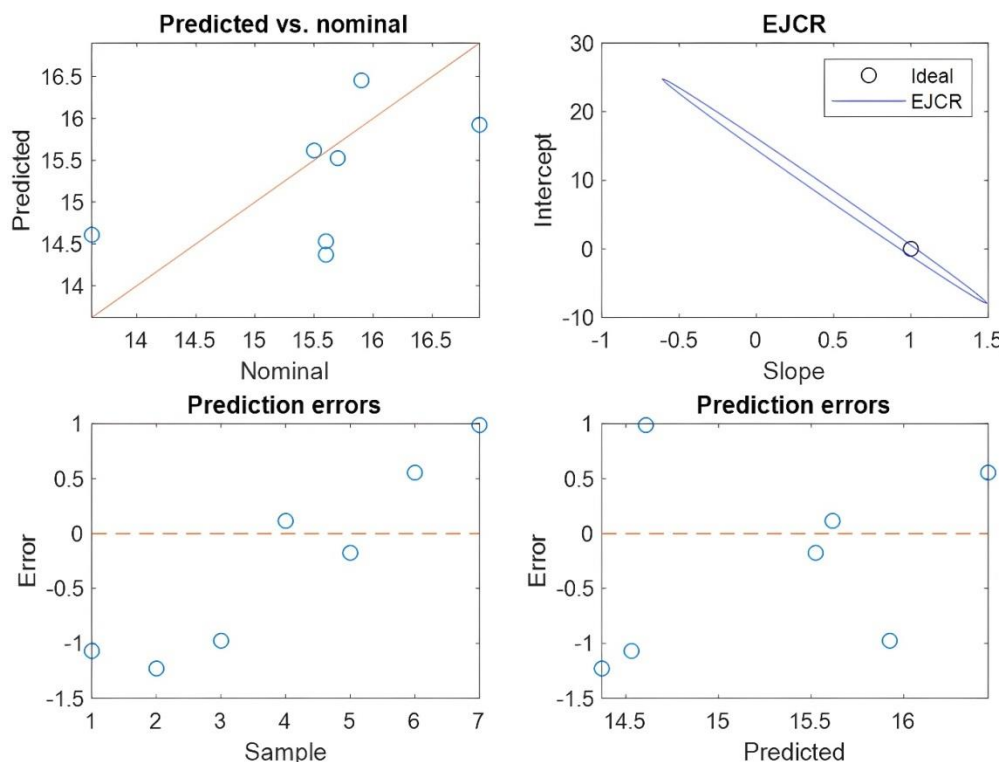


شکل (۲). نتایج اعتبارسنجی متقاطع برای مدل پیش‌بینی میزان رطوبت عسل با استفاده از رگرسیون حداقل مربعات جزئی. ردیف بالا (چپ و راست): نمودارهای تعیین تعداد متغیرهای پنهان با کمینه‌سازی خطای پیش‌بینی، نشان‌دهنده بهینه بودن ۴ متغیر پنهان. ردیف پایین (چپ): نمودار تست پرت بر اساس تست F نسبی، ردیف پایین (راست): نمودار مقادیر پیش‌بینی شده در مقابل مقادیر واقعی در اعتبارسنجی متقاطع با ۴ متغیر پنهان.

**Fig. 2.** Results of the cross-validation for the model predicting honey moisture content using Partial Least Squares (PLS) regression. Top row (left and right): Plots for determining the number of latent variables by minimizing the prediction error, indicating that 4 latent variables are optimal. Bottom row (left): Bar plot of outlier test based on relative F-test; Bottom row (right): Plot of predicted values versus actual values in cross-validation with 4 latent variables.

قابلیت تعمیم‌دهی مدل عمل می‌کند و اطمینان می‌دهد که مدل‌های توسعه‌یافته می‌توانند در شرایط واقعی، مانند کنترل کیفیت عسل در مقیاس صنعتی، عملکرد قابل‌قبولی داشته باشند. در این مطالعه، مجموعه‌های تست برای هر ویژگی شامل نمونه‌هایی بودند که با استفاده از الگوریتم Kennard-Stone انتخاب شده بودند تا نماینده تنوع کامل داده‌ها باشند.

ارزیابی مدل‌ها با استفاده از مجموعه تست مستقل، که شامل نمونه‌هایی بود که در فرآیند کالیبراسیون استفاده نشده بودند، از اهمیت بسزایی برخوردار است. اعتبارسنجی متقاطع توانایی مدل را برای پیش‌بینی داده‌های مجموعه کالیبراسیون ارزیابی می‌کند، اما مجموعه تست مستقل نشان‌دهنده عملکرد واقعی مدل در مواجهه با داده‌های جدید و نادیده است. این فرآیند به‌عنوان معیاری برای سنجش



شکل (۳). نتایج پیش‌بینی مدل کمومتریکس برای میزان رطوبت عسل در نمونه‌های مجموعه تست. ردیف بالا (چپ): نمودار مقادیر پیش‌بینی شده در مقابل مقادیر واقعی رطوبت. ردیف بالا (راست): نمودار منطقه اطمینان مشترک بیضی‌گون برای نمایش دامنه مجاز شیب و عرض از مبدأ نمودار مقادیر پیش‌بینی شده در مقابل مقادیر واقعی. ردیف پایین (چپ): نمایش خطای مطلق پیش‌بینی برای هر نمونه بر اساس شماره نمونه. ردیف پایین (راست): نمایش خطای مطلق پیش‌بینی نسبت به مقادیر پیش‌بینی شده رطوبت.

**Fig. 3.** Results of the chemometrics model prediction for honey moisture content in the test set samples. Top row (left): Plot of predicted values versus actual moisture values. Top row (right): Elliptical Joint Confidence Region (EJCR) plot to display the allowable range for the slope and intercept of the predicted versus actual values plot. Bottom row (left): Display of the absolute prediction error for each sample based on sample number. Bottom row (right): Display of the absolute prediction error relative to the predicted moisture values.

مجموعه تست نیز بررسی شدند. RMSEP نشان‌دهنده میزان انحراف پیش‌بینی‌های مدل از مقادیر واقعی است، درحالی‌که  $REP\%$  خطای نسبی پیش‌بینی را به‌صورت درصد بیان می‌کند و معیاری قابل‌فهم برای ارزیابی دقت مدل فراهم می‌آورد. نمودارهای مقدار پیش‌بینی‌شده در مقابل مقدار واقعی، انطباق بین مقادیر پیش‌بینی‌شده و واقعی را به‌صورت

معیارهای ارزیابی عملکرد مدل‌ها در پیش‌بینی مجموعه تست شامل ریشه میانگین مربعات پیش‌بینی (RMSEP) و درصد خطای نسبی ( $REP\%$ ) بودند. علاوه‌براین، نمودارهای مقدار پیش‌بینی‌شده در مقابل مقدار واقعی برای مجموعه تست، نمودارهای ( Ellipse Joint Confidence ) EJCR (Region و نمایش خطاهای مطلق برای هر نمونه در

مثالی از این تطابق در مدل توسعه‌یافته برای پیش‌بینی میزان رطوبت عسل مشاهده می‌شود. این معیارها در کنار یکدیگر، تصویری جامع از دقت و قابلیت اطمینان مدل‌های توسعه‌یافته ارائه می‌دهند. جدول (۱) مقادیر RMSEP و REP% را برای هر مدل نشان می‌دهد.

بصری نمایش می‌دهند. نمودارهای EJCR توصیفی آماری از توانایی پیش‌بینی مدل ارائه می‌کند. بر اساس این آماره ناحیه مجاز انحراف شیب و عرض از مبدا خط مربوط به نمایش مقادیر پیش‌بینی‌شده نسبت به مقادیر واقعی (شکل ۳) لازم است نقطه‌ای با مختصات یک و صفر را در بر بگیرد. همانطور که در نمودار EJCR در شکل (۳) نشان داده است،

جدول (۲). پارامترها و نتایج مدل‌های کمومتریک مبتنی بر رگرسیون حداقل مربعات جزئی برای پیش‌بینی چهار ویژگی فیزیکی شیمیایی عسل.

Table 2. Parameters and results of chemometrics models based on Partial Least Squares (PLS) regression for predicting four physicochemical properties of honey.

Physicochemical Property	Number of Latent Variables	RMSEP	%REP
Sugar before hydrolysis	5	10.18	1.64
Sucrose	7	0.46	11.95
Moisture	6	2.4	25.8
HMF	4	0.84	5.30

عملکرد ضعیف‌تری داشت ( $R^2p = 0.9068$ ). RMSEP = 10.54، که نشان‌دهنده چالش‌های موجود در کمی‌سازی جامع تقلب است. در مقابل، پژوهش حاضر که بر پارامترهای کیفی ذاتی عسل تمرکز دارد نه بر شناسایی تقلب، توانسته است پیش‌بینی‌های دقیقی برای مؤلفه‌های کلیدی عسل ارائه دهد بدون نیاز به دانستن نوع تقلب‌کننده احتمالی [۲۷].

دقت بالای پیش‌بینی برای قندهای احیاء‌کننده و رطوبت در این پژوهش با نتایج سایر مطالعاتی که از روش‌های طیف‌سنجی ارتعاشی استفاده کرده‌اند، هم‌راستا است. برای نمونه، مرور جامعی از El Haji & Estephan, 2025 نشان می‌دهد که طیف‌سنجی مادون قرمز میانی (MIR)، یکی از روش‌های مکمل در طیف‌سنجی ارتعاشی، با موفقیت بالا برای پیش‌بینی قند و رطوبت در عسل به‌کار رفته و معمولاً همراه با روش رگرسیون حداقل مربعات جزئی (PLS) استفاده می‌شود. هم‌خوانی نتایج ما با یافته‌های موفق MIR نشان می‌دهد که روش‌های طیف‌سنجی ارتعاشی، در صورت استفاده از ابزارهای شیمی‌سنجی مناسب، برای تحلیل ترکیبات اصلی عسل بسیار کارآمد و قابل اعتماد هستند [۲۸]. به‌طور مشابه، Hu et al., 2022 نیز از طیف‌سنجی رامان همراه با الگوریتم‌های یادگیری ماشین برای شناسایی عسل تقلبی استفاده کردند و به دقتی نزدیک به صد درصد دست یافتند. موفقیت آن‌ها در تشخیص عسل خالص از

این نتایج نشان داد که مدل‌های اول و چهارم، که به‌ترتیب برای پیش‌بینی قندهای احیاء‌کننده پیش از هیدرولیز و رطوبت توسعه یافته بودند، با خطاهای نسبی بسیار پایین (REP% به‌ترتیب ۱/۶۴ و ۵/۳۰) عملکرد بسیار مطلوبی داشتند. این امر نشان‌دهنده توانایی بالای طیف‌سنجی رامان در ترکیب با مدل‌های PLS برای پیش‌بینی دقیق این دو ویژگی در محدوده‌های نرمال آن‌ها است. مدل دوم، که برای پیش‌بینی ساکاروز توسعه یافت، با خطای نسبی حدود ۱۱/۹۵٪ عملکرد قابل‌قبولی داشت، در حالی که مدل سوم، که برای پیش‌بینی HMF طراحی شده بود، با خطای نسبی حدود ۲۵/۸٪ دقت کمتری نشان داد. این تفاوت در دقت می‌تواند به دلیل غلظت پایین HMF در نمونه‌های عسل و چالش‌های مرتبط با تشخیص آن باشد. به‌طور کلی، این نتایج تأیید می‌کنند که مدل‌های PLS مبتنی بر طیف‌سنجی رامان ابزار قدرتمندی برای پیش‌بینی ویژگی‌های فیزیکی شیمیایی عسل هستند و می‌توانند به‌عنوان روشی سریع و غیرمخرب در کنترل کیفیت عسل مورد استفاده قرار گیرند.

در مقایسه، وو و همکاران (۲۰۲۲) از رگرسیون PLS برای تعیین میزان تقلب در عسل با شربت‌های مجزا استفاده کرده و مقادیر  $R^2p$  بالاتر از ۰/۹۸ و RMSEP کمتر از ۳/۵۰ برای شربت ذرت با فروکتوز بالا (HFCS) و شربت برنج گزارش کرده‌اند. باین‌حال، مدل کلی آنان برای چند نوع شربت

کمی ترکیبات خاص، اما کاربرد کلی روش مشابه است. یکی از دلایل احتمالی تفاوت در دقت پیش‌بینی HMF می‌تواند ماهیت روش‌شناختی باشد؛ زیرا ویژگی‌های طیفی HMF در غلظت‌های پایین (همانند نمونه‌های عسل تازه در این پژوهش) ممکن است بسیار ضعیف بوده و تحت‌تأثیر سیگنال‌های غالب قندها قرار گیرند، در نتیجه مدل قادر به تشخیص دقیق آن‌ها نباشد [۲۳]. دیوید و همکاران (۲۰۲۵) با ترکیب داده‌های حاصل از طیف‌سنجی فرورسرخ (IR) و رامان و به‌کارگیری شبکه عصبی سه‌لایه، توانستند دقت بالایی (۹۳.۸٪ در مرحله آزمون) در تفکیک جغرافیایی نمونه‌های عسل به‌دست آورند. در حالی که مطالعه حاضر تنها بر داده‌های رامان و رگرسیون PLS تکیه دارد؛ روشی که اگرچه برای پیش‌بینی کمی پارامترهای فیزیکوشیمیایی خاص مؤثر است، اما ممکن است تنوع ترکیبی موردنیاز برای طبقه‌بندی‌های پیچیده‌تر مانند تشخیص دقیق منشاء جغرافیایی را در بر نگیرد. این تفاوت روش‌شناختی، یعنی ادغام داده‌ها و استفاده از مدل‌های غیرخطی پیشرفته در مقابل تکیه بر یک تکنیک طیف‌سنجی و مدل خطی، احتمالاً علت تفاوت در کاربرد و عملکرد است [۳۱].

سونگ و همکاران (۲۰۱۹) از طیف‌سنجی  $^1\text{H NMR}$  به همراه روش‌های کمومتریکس مانند PLS و OPLS-DA برای شناسایی و تعیین کمی تقلب در عسل اقلایا با عسل ارزان‌تر کلزا استفاده کردند. مدل PLS آن‌ها ضریب همبستگی بسیار بالایی ( $R^2 = 0.9996$ ) را در پیش‌بینی سطوح تقلب نشان داد که قدرت داده‌های طیفی همراه با کالیبراسیون چندمتغیره را در احراز اصالت عسل نمایان می‌کند؛ نتیجه‌ای که با عملکرد پیش‌بینی قوی مدل‌های ما در تعیین محتوای قند و رطوبت هم‌خوانی دارد [۳۲]. به‌طور مشابه، اوزبالچی و همکاران (۲۰۱۳) نیز کاربرد موفق طیف‌سنجی رامان و رگرسیون PLS را در تحلیل سریع فروکتوز و گلوکز در عسل گزارش کردند و به ضرایب تعیین بالا دست‌یافتند. موفقیت مطالعه آن‌ها در پیش‌بینی اجزای قندی، یافته‌های ما را در خصوص دقت بالای پیش‌بینی قندهای احیاکننده تأیید و استحکام چارچوب رامان PLS را در تعیین کمی ترکیبات کربوهیدراتی اصلی عسل تقویت می‌کند [۱۵]. به‌طور مشابه، مگداس و برگیان-گروسان (۲۰۲۳) نیز کارایی طیف‌سنجی

تقلبی بر اساس ترکیب کربوهیدراتی، با نتایج ما در اندازه‌گیری دقیق قندهای اصلی هم‌راستا است و حساسیت بالای روش رامان نسبت به ارتعاشات مولکولی قندها را در ترکیب پیچیده‌ای مانند عسل نشان می‌دهد [۲۹].

در مقابل، مدل ما برای پیش‌بینی HMF دقت پایین‌تری داشت. این نتیجه با برخی مطالعات دیگر تفاوت دارد؛ برای مثال، Stöbener et al., 2019 با استفاده از طیف‌سنجی ATR-FTIR مدلی با عملکرد مناسب برای اندازه‌گیری مقادیر کم HMF ارائه کردند [۳۰]. تفاوت عملکرد احتمالاً ناشی از تفاوت در روش‌هاست. طیف‌سنجی مادون قرمز میانی (MIR) ارتعاشات بنیادی مولکول‌ها را بررسی می‌کند و می‌تواند باندهای جذبی مشخصی برای گروه‌های عاملی خاص ایجاد کند. از آنجا که HMF دارای گروه‌های کربونیل ( $\text{C}=\text{O}$ ) و حلقه‌ی فورانی است، سیگنال آن در ناحیه‌ی MIR واضح‌تر و قابل‌اندازه‌گیری‌تر از طیف رامان است؛ در حالی که در طیف رامان، سیگنال آن ممکن است ضعیف‌تر بوده یا در میان نویز فلورسانس یا قله‌های قوی‌تر قندها پنهان شود. علاوه‌براین، غلظت پایین طبیعی HMF در نمونه‌های عسل تازه ما (میانگین  $7/99 \text{ mg/kg}$ ) نیز چالشی جدی برای هر روش طیف‌سنجی ایجاد می‌کند، زیرا مقدار آن در محدوده حد تشخیص دستگاه قرار دارد. مطالعات همچنین نشان داده است که طیف‌سنجی NIR نیز برای اندازه‌گیری HMF به‌کار رفته، اما طیف‌های NIR به‌دلیل داشتن باندهای هم‌پوشان، تحلیل ترکیبات جزئی را دشوار می‌سازند [۲۸]. به نظر می‌رسد طیف‌سنجی رامان نیز با چالش مشابهی روبه‌رو است؛ چرا که سیگنال‌های ضعیف HMF در میان قله‌های قوی فروکتوز، گلوکز و آب قرار می‌گیرند. عملکرد بهتر مدل‌های ما برای قندها و رطوبت نیز مؤید همین نکته است، زیرا این ترکیبات درصد بالایی از عسل را تشکیل می‌دهند و سیگنال‌های واضح‌تری در طیف ایجاد می‌کنند که مدل PLS می‌تواند به‌خوبی آن‌ها را با داده‌های مرجع مرتبط کند.

برخی مطالعات نتایجی متفاوت با پژوهش حاضر گزارش کرده‌اند، به‌ویژه در زمینه پیش‌بینی HMF، لی و همکاران (۲۰۱۲) از طیف‌سنجی رامان برای شناسایی تقلب در عسل با شربت ذرت با فروکتوز بالا و شربت مالتوز استفاده کردند. هرچند تمرکز آن‌ها بر شناسایی تقلب بود و نه پیش‌بینی

ویژگی‌های ضعیف طیفی مرتبط با HMF کمک کند [۲۹]. همچنین، ترکیب داده‌های حاصل از روش رامان با سایر روش‌های سریع مانند NIR یا MIR در قالب رویکرد هم‌جوشی داده‌ها (Data Fusion) می‌تواند نقاط قوت هر روش را با یکدیگر ترکیب کرده و امکان شناسایی جامع‌تر ترکیبات عسل را فراهم کند. در نهایت، توسعه‌ی دستگاه‌های قابل‌حمل رامان با مدل‌های اعتبارسنجی‌شده می‌تواند زمینه‌ساز انجام آزمایش‌های میدانی شود و تحول بزرگی در پایش زنجیره‌ی تأمین و شناسایی تقلب در عسل ایجاد کند.

#### ۴. نتیجه‌گیری

در این پژوهش، امکان‌سنجی استفاده از طیف‌سنجی رامان در ترکیب با مدل‌های کمومتریکس مبتنی بر رگرسیون حداقل مربعات جزئی (PLS) برای پیش‌بینی سریع و دقیق ویژگی‌های فیزیکوشیمیایی عسل بررسی شد. نتایج نشان داد که این رویکرد، به‌عنوان یک روش غیرمخرب و سریع، می‌تواند جایگزین مناسبی برای روش‌های سنتی آزمایشگاهی باشد که اغلب زمان‌بر، پرهزینه و نیازمند مواد شیمیایی هستند. مدل‌های توسعه‌یافته برای پیش‌بینی قندهای احیاکننده پیش از هیدرولیز و رطوبت با خطاهای نسبی بسیار پایین (۱/۶۴٪ و ۳۰/۵٪) عملکرد مطلوبی از خود نشان دادند، که این امر به‌دلیل حساسیت بالای طیف‌های رامان به ارتعاشات مولکولی مرتبط با قندها و آب در عسل است. مدل پیش‌بینی ساکاروز نیز با خطای نسبی حدود ۱۱/۹۵٪ قابل قبول بود و توانایی مناسبی در تشخیص سطوح ساکاروز را ارائه کرد، که می‌تواند در شناسایی عسل‌های تقلبی مفید باشد. با این حال، مدل HMF با خطای نسبی بالاتر (۲۵/۸٪) دقت کمتری داشت، که احتمالاً به‌دلیل غلظت پایین این ترکیب در نمونه‌های عسل طبیعی و تداخل‌های طیفی ناشی از سایر اجزای پیچیده عسل است. این محدودیت نشان‌دهنده نیاز به بهینه‌سازی بیشتر مدل‌ها، مانند استفاده از پیش‌پردازش‌های پیشرفته‌تر یا ترکیب با تکنیک‌های طیف‌سنجی دیگر، برای بهبود دقت در پیش‌بینی ترکیبات با غلظت کم است.

به‌طور کلی، یافته‌ها تأیید می‌کنند که فناوری رامان، با مزایایی همچون عدم نیاز به آماده‌سازی نمونه، سرعت بالا و

رامان همراه با الگوریتم‌های یادگیری ماشین را برای تشخیص منشاء گیاهی عسل و تحلیل کمی آن به اثبات رساندند. این شواهد، قدرت طیف‌سنجی رامان را به‌عنوان ابزاری کارآمد در شناسایی اصالت و پیش‌بینی ویژگی‌های عسل تقویت می‌کند. هم‌سویی نتایج این مطالعات با پژوهش حاضر نشان‌دهنده قابلیت تعمیم‌پذیری رویکردهای مبتنی بر رامان در ارزیابی ترکیبات اصلی عسل است [۳۳].

از دیدگاه اصول شیمی تجزیه و شیمی‌سنجی، این نتایج قابل انتظار هستند. دقت بالای پیش‌بینی در ترکیبات اصلی طبیعی است، زیرا حضور زیاد آن‌ها در عسل موجب ایجاد الگوهای طیفی قوی و پایدار می‌شود که مدل به راحتی آن‌ها را یاد می‌گیرد. در مقابل، دقت کمتر در پیش‌بینی HMF به‌دلیل غلظت پایین آن و سطح مقطع پراکندگی ضعیف‌تر رامان نسبت به جذب مادون قرمز است. اگرچه استفاده از میان‌مرکزسازی (mean-centering) به ساده‌سازی مدل و کاهش نویز کمک کرده، اما برای جداسازی کامل سیگنال‌های ضعیف HMF از زمینه‌ی قوی قندها کافی نبوده است. در غلظت‌های پایین، سیگنال رامان HMF ضعیف و مستعد تداخل با زمینه و نویز است، که توان پیش‌بینی مدل را محدود می‌کند. این مسئله در کمومتریکس امری رایج است، زیرا حد تشخیص و حد تعیین کمی روش بر عملکرد آن برای تحلیل ترکیبات جزئی اثر می‌گذارد. با این حال، این مطالعه محدودیت‌هایی نیز دارد. اندازه نمونه‌ها برای توسعه اولیه مدل کافی بود، اما افزایش آن می‌تواند به بهبود دقت و قابلیت تعمیم مدل برای انواع مختلف عسل از مناطق گوناگون کمک کند. همچنین، عملکرد مدل HMF نشان می‌دهد که تنظیمات و پارامترهای دستگاه رامان مورد استفاده ممکن است دارای حد تشخیص بالاتری برای این ترکیب خاص باشند. از سوی دیگر، دقت نهایی مدل به‌دقت داده‌های آزمایشگاهی مرجع نیز وابسته است و هرگونه خطا در آن داده‌ها مستقیماً به مدل منتقل می‌شود. در پژوهش‌های آینده، گسترش پایگاه داده طیفی با نمونه‌های متنوع‌تر عسل ضروری است. استفاده از روش‌های پیش‌پردازش پیشرفته‌تر و الگوریتم‌های یادگیری ماشین جدید مانند ماشین بردار پشتیبان (SVM) یا شبکه‌های عصبی پیچشی - (CNN) می‌تواند به استخراج بهتر

بالای ادغام طیف‌سنجی و کمومتریکس در ارزیابی مواد غذایی پیچیده است و می‌تواند الگویی برای بررسی سایر محصولات کشاورزی باشد.

### تعارض منافع

هیچ گونه تعارض منافع توسط نویسندگان بیان نشده است.

### تشکر و قدردانی

این مطالعه با حمایت مرکز ملی تحقیقات حلال جمهوری اسلامی ایران، سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران به شماره گرنت ۱۷۱ و شناسه اخلاق IR.ZUMS.BLC.1404.060 انجام شده است. بدین‌وسیله مراتب قدردانی خود را از همکاری آن‌ها اعلام می‌داریم.

سازگاری با محیط‌زیست، ابزاری قدرتمند برای کنترل کیفیت عسل است. مدل‌های PLS توسعه‌یافته، با بهره‌گیری از داده‌های طیفی پیش‌پردازش‌شده و اعتبارسنجی دقیق (از جمله RMSECV برای بهینه‌سازی متغیرهای پنهان و ارزیابی خارجی با مجموعه تست مستقل)، قابلیت تعمیم‌دهی مناسبی نشان دادند و می‌توانند در مقیاس کاربردی برای پایش سریع کیفیت عسل‌های تولیدشده در مناطق مختلف ایران مورد استفاده قرار گیرند. این روش نه تنها هزینه‌های آزمایشگاهی را کاهش می‌دهد، بلکه به ارتقای فرآیندهای توزیع و صادرات عسل کمک می‌کند. پیشنهاد می‌شود تحقیقات آینده بر روی گسترش پایگاه داده با نمونه‌های بیشتر، ترکیب با تکنیک‌های یادگیری عمیق مانند شبکه‌های عصبی و توسعه دستگاه‌های قابل حمل مبتنی بر رامان تمرکز کنند تا این فناوری به‌طور گسترده‌تری در زنجیره تأمین عسل اعمال شود. در نهایت، این مطالعه نشان‌دهنده پتانسیل

### منابع

- [1] Duca A, Sturza A, Moacă E-A, Negrea M, Lalescu V-D, Lungeanu D, et al. (2019). Identification of resveratrol as bioactive compound of propolis from western Romania and characterization of phenolic profile and antioxidant activity of ethanolic extracts. *Molecules*, 24(18), 3368. <https://doi.org/10.3390/molecules24183368>
- [2] Bogdanov, S. (2019). *Harmonised methods of the International Honey Commission*. Retrieved November 1, 2019, from <http://www.ihc-platform.net>
- [3] AOAC International. (2005). *Official methods of analysis of AOAC International* (18th ed.). Rockville, MD: Author. ISBN 0935584544.
- [4] Maione, C., Barbosa, F., Jr., & Barbosa, R. M. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Computers and Electronics in Agriculture*, 157, 436–446. <https://doi.org/10.1016/j.compag.2019.01.020>
- [5] Se, K. W., Wahab, R. A., Yaacob, S. N. S., & Ghoshal, S. K. (2019). Detection techniques for adulterants in honey: Challenges and recent trends. *J. Food Compos. Anal.*, 80, 16–32. <https://doi.org/10.1016/j.jfca.2019.04.001>
- [6] Das C, Chakraborty S, Acharya K, Bera NK, Chattopadhyay D, Karmakar A, et al. (2017). FT-MIR supported electrical impedance spectroscopy based study of sugar adulterated honeys from different floral origin. *Talanta*, 171, 327–334. <https://doi.org/10.1016/j.talanta.2017.05.016>
- [7] Brereton RG. (2000). Introduction to multivariate calibration in analytical chemistry: Electronic Supplementary Information available. *Analyst*, 125(11), 2125–2154. <https://doi.org/10.1039/B003805I>
- [8] Aliaño-González MJ, Ferreiro-González M, Espada-Bellido E, Palma M, & Barbero GF. (2019). A screening method based on visible-NIR spectroscopy for the identification and quantification of different adulterants in high-quality honey. *Talanta*, 203, 235–241. <https://doi.org/10.1016/j.talanta.2019.05.067>
- [9] Corvucci F, Nobili L, Melucci D, & Grillenzoni F-V. (2015). The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis. *Food Chem.*, 169, 297–304. <https://doi.org/10.1016/j.foodchem.2014.07.122>
- [10] Oroian, M., & Ropciuc, S. (2018). Botanical authentication of honeys based on Raman spectra. *J. Food Meas. Charact.*, 12(1), 545–554. <https://doi.org/10.1007/s11694-017-9755-8>
- [11] Frausto-Reyes C, Casillas-Peñuelas R, Quintanar-Stephano J, Macías-López E, Bujdud-Pérez J, & Medina-Ramírez, I. (2017). Spectroscopic study of honey from *Apis mellifera* from different regions in Mexico. *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, 178, 212–217. <https://doi.org/10.1016/j.saa.2017.02.009>
- [12] Jandrić Z, Haughey S, Frew R, McComb K, Galvin-King P, & Cannavan, A. (2015).

- Discrimination of honey of different floral origins by a combination of various chemical parameters. *Food Chem.*, 189, 52–59.  
<https://doi.org/10.1016/j.foodchem.2014.11.165>
- [13] Oroian M, Ropciuc S, & Paduret, S. (2018). Honey adulteration detection using Raman spectroscopy. *Food Anal. Methods*, 11(4), 959–968.  
<https://doi.org/10.1007/s12161-017-1155-x>
- [14] Anjos O, Santos AJ, Paixão V, & Estevinho, L. M. (2018). Physicochemical characterization of *Lavandula* spp. honey with FT-Raman spectroscopy. *Talanta*, 178, 43–48.  
<https://doi.org/10.1016/j.talanta.2017.08.099>
- [15] Özbalci B, Boyacı İH, Topcu A, Kadılar C, & Tamer, U. (2013). Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chem.*, 136(3–4), 1444–1452.  
<https://doi.org/10.1016/j.foodchem.2012.10.159>
- [16] Tahir HE, Xiaobo Z, Zhihua L, Jiyong S, Zhai X, Wang S, et al. (2017). Rapid prediction of phenolic compounds and antioxidant activity of Sudanese honey using Raman and Fourier transform infrared (FT-IR) spectroscopy. *Food Chem.*, 226, 202–211.  
<https://doi.org/10.1016/j.foodchem.2017.01.024>
- [17] Li S, Shan Y, Zhu X, Zhang X, & Ling, G. (2012). Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy. *J. Food Compos. Anal.*, 28, 69–74.  
<https://doi.org/10.1016/j.jfca.2012.07.006>
- [18] Anguebes-Franceschi, F., Abatal, M., Pat, L., Flores, A., Córdova Quiroz, A. V., Ramírez-Elias, M. A., ... & Bassam, A. (2019). Raman spectroscopy and chemometric modeling to predict physical-chemical honey properties from Campeche, Mexico. *Molecules*, 24(22), 4091.  
<https://doi.org/10.3390/molecules24224091>
- [19] Batsoulis, A. N., Siatis, N. G., Kimbaris, A. C., Alissandrakis, E. K., Pappas, C. S., & Tarantilis, P. A., et al. (2005). FT-Raman spectroscopic simultaneous determination of fructose and glucose in honey. *J. Agric. Food Chem.*, 53(2), 207–210.  
<https://doi.org/10.1021/jf0486818>
- [20] Craig, A. P., Franca, A. S., & Irudayaraj, J. (2013). Surface-enhanced Raman spectroscopy applied to food safety. *Annu. Rev. Food Sci. Technol.*, 4, 369–380.  
<https://doi.org/10.1146/annurev-food-030212-182554>
- [21] Iranian National Standardization Organization (INSO). (2020). *Honey—Specifications and test methods*. Tehran, Iran: Author.
- [22] Li, S., Shan, Y., Zhu, X., Zhang, X., & Ling, G. (2012). Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy. *J. Food Compos. Anal.*, 28(1), 69–74.  
<https://doi.org/10.1016/j.jfca.2012.07.006>
- [23] White, J. W., Jr. (1969). Moisture in honey: Review of chemical and physical methods. *J. Assoc. Off. Anal. Chem.*, 52(4), 729–737.
- [24] Kek, S. P., Chin, N. L., Yusof, Y. A., Tan, S. W., & Chua, L. S. (2017). Classification of entomological origin of honey based on its physicochemical and antioxidant properties. *Int. J. Food Prop.*, 20(sup3), S2723–S2738.  
<https://doi.org/10.1080/10942912.2017.1382518>
- [25] Belay, A., Solomon, W., Bultossa, G., Adgaba, N., & Melaku, S. (2013). Physicochemical properties of the Harena forest honey, Bale, Ethiopia. *Food Chem.*, 141(4), 3386–3392.  
<https://doi.org/10.1016/j.foodchem.2013.06.035>
- [26] Wu, X., Xu, B., Ma, R., Niu, Y., Gao, S., Liu, H., et al. (2022). Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, 274, 121133.  
<https://doi.org/10.1016/j.saa.2022.121133>
- [27] El Hajj, R., & Estephan, N. (2024). Advances in infrared spectroscopy and chemometrics for honey analysis: A comprehensive review. *Crit. Rev. Food Sci. Nutr.*, 64, 1–14.  
<https://doi.org/10.1080/10408398.2024.2305036>
- [28] Hu, S., Li, H., Chen, C., Chen, C., Zhao, D., Dong, B., et al. (2022). Raman spectroscopy combined with machine learning algorithms to detect adulterated Suichang native honey. *Sci. Rep.*, 12(1), 3456. <https://doi.org/10.1038/s41598-022-07391-4>
- [29] Stöbener, A., Naefken, U., Kleber, J., & Liese, A. (2019). Determination of trace amounts with ATR FTIR spectroscopy and chemometrics: 5-(Hydroxymethyl)furfural in honey. *Talanta*, 204, 1–5.  
<https://doi.org/10.1016/j.talanta.2019.06.004>
- [30] David, M., Berghian-Grosan, C., & Magdas, D. A. (2025). Honey differentiation using infrared and Raman spectroscopy analysis and the employment of machine-learning-based authentication models. *Foods*, 14(6), 1032.  
<https://doi.org/10.3390/foods14061032>
- [31] Song, X., She, S., Xin, M., Chen, L., Li, Y., Vander Heyden, Y., et al. (2020). Detection of adulteration in Chinese monofloral honey using <sup>1</sup>H nuclear magnetic resonance and chemometrics. *J. Food Compos. Anal.*, 86, 103390.  
<https://doi.org/10.1016/j.jfca.2019.103390>
- [32] Magdas, D. A., Guyon, F., Berghian-Grosan, C., & Molnár, C. M. (2021). Challenges and a step forward in honey classification based on Raman spectroscopy. *Food Control*, 123, 107769.  
<https://doi.org/10.1016/j.foodcont.2020.107769>

*Research Article*

## Chemometrics modeling for rapid prediction of physicochemical properties and hydroxymethylfurfural of honey based on hyperspectral Raman spectroscopy

Fataneh Hashempour-Baltork <sup>1</sup>, Somaye Vali Zade <sup>1</sup>, Adel Mirza Alizadeh <sup>2\*</sup>, Behrooz Jannat <sup>3</sup>, Hossein Rastegar <sup>4</sup>

1. Halal Research Center of IRI, Iran Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran
2. Department of Food Safety and Hygiene, School of Public Health, Zanjan University of Medical Sciences, Zanjan, Iran
3. Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran.
4. Cosmetic Products Research Center, Iran Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran.

(Received 24 September 2025, Received in revised form 24 October 2025, Accepted 16 November 2025)

**Introduction:** According to the Codex Alimentarius, honey is a natural sweet substance produced by honey bees from plant nectar or secretions, containing at least 60% reducing sugars (65% per Iranian national standard) and not more than 21% moisture (20% per Iranian standard). Its composition varies based on botanical and geographical origins. Honey consists mainly of sugars (glucose, fructose, sucrose, maltose), along with antioxidants, acids, proteins, minerals, and enzymes. Traditional methods for quality assessment, such as melissopalynology and chromatographic techniques, are time-consuming, costly, and destructive. This has led to the adoption of spectroscopic techniques combined with chemometrics for rapid, non-destructive, and environmentally friendly honey authentication. Raman spectroscopy, in particular, is advantageous due to its minimal interference from water and fluorescence. Recent studies demonstrate the effectiveness of Raman spectroscopy with multivariate analysis (e.g., PCA, PLS, LDA) in determining botanical origin, detecting adulteration, and predicting physicochemical properties like sugar content, moisture, HMF, and antioxidant activity. For instance, research has achieved high accuracy (up to 99%) in botanical classification and excellent predictive performance ( $R^2$  up to 0.99) for quantitative analysis. This study utilizes PLS regression to develop a chemometrics model linking Raman spectra to four key honey properties: reducing sugars, sucrose, moisture, and HMF. The aim is to provide a rapid, reliable tool for quality control of Iranian honey.

**Materials and methods:** Approximately 51 honey samples from various regions of Iran were collected from the market following appropriate sampling guidelines and were coded accordingly. Physicochemical tests were conducted in accordance with Iranian National Standard No. 92, and Raman spectroscopy was performed for each sample. Machine learning (ML) methods were developed to model the relationship between the Raman spectra and the physicochemical properties of honey.

**Results and discussion:** Following mean-centering preprocessing, the spectral data was correlated with laboratory-measured physicochemical values. Four independent Partial Least Squares (PLS) regression models were developed. The data was split into calibration and test sets using the Kennard-Stone algorithm, and the optimal number of latent variables was determined via Leave-One-Out cross-validation. Model performance was evaluated using RMSEP, %REP, and RMSECV. The results demonstrated excellent predictive accuracy for

---

\* Corresponding author: [a.alizadeh@zums.ac.ir](mailto:a.alizadeh@zums.ac.ir)



the models estimating reducing sugars (before hydrolysis) and moisture, which achieved low relative errors of 1.64% and 3.50%, respectively. The sucrose model performed acceptably with a relative error of 11.95%. In contrast, the HMF prediction model showed lower accuracy, with a higher relative error of 25.8%.

**Conclusions:** This research demonstrated that Raman spectroscopy combined with PLS regression is an effective, non-destructive method for predicting key honey quality parameters. Models excelled at predicting reducing sugars (1.64% error) and moisture (3.50% error), performed acceptably for sucrose (11.95% error), but were less accurate for HMF (25.8% error) due to its low natural concentration. The technique offers a rapid, eco-friendly alternative to traditional lab methods, requiring minimal sample preparation. The validated models show strong potential for large-scale quality control in Iran's honey industry, reducing costs and improving efficiency for distribution and export. Future research should expand the sample database, explore deep learning integration, and develop portable devices for broader supply chain application. This approach sets a valuable precedent for analyzing complex agricultural products.

**Keywords:** *Quality control, physicochemical properties, raman spectroscopy, honey, chemometrics model, PLS regression*