

Journal Pre-proofs

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

Fataneh Hashempour-Baltork, Somaye Vali Zade, Adel Mirza Alizadeh, Behrooz Jannat

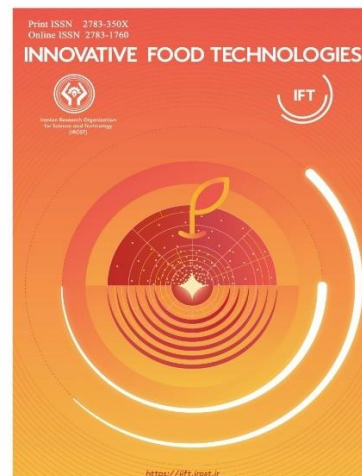
DOI: <https://doi.org/10.22104/ift.2025.7892.2240>

To appear in: Innovative Food Technologies (IFT)

Received Date: 24 September 2025

Revised Date: 24 October 2025

Accepted Date: 16 November 2025



Please cite this article as: Fataneh Hashempour-Baltork, Somaye Vali Zade, Adel Mirza Alizadeh, Behrooz Jannat, Chemometrics modeling for rapid prediction of physicochemical properties and hydroxymethylfurfural of honey based on hyperspectral Raman spectroscopy, *Innovative Food Technologies* (2025),

doi: <https://doi.org/10.22104/ift.2025.7892.2240>

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

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

مدل سازی کمومتریکس برای پیش بینی سریع خواص فیزیکوشیمیایی و هیدروکسی متیل فورفورال عسل مبتنی بر طیف سنجی فراطیفی رامان

فتانه هاشم پور بلترک^۱، سمیه ولی زاده^۱، عادل میرزاعلیزاده^{۲*}، بهروز جنت^۱، حسین رستگار^۱

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
 * نویسنده مسئول: a.alizadeh@zums.ac.ir

چکیده

هدف از این پژوهش، بررسی امکان سنجی پیش بینی سریع ویژگی های فیزیکوشیمیایی در عسل با استفاده از طیف سنجی رامان و مدل های کمومتریکس است. در این مطالعه، ۵۱ نمونه عسل از مناطق مختلف ایران جمع آوری شد و هر نمونه با طیف سنجی رامان (در محدوده ۱۰۰ تا ۳۵۰۰ cm^{-1}) آنالیز گردید. ویژگی های مورد بررسی شامل قندهای احیاکننده پیش از هیدرولیز، ساکاروز، رطوبت و هیدروکسی متیل فورفورال (HMF) بودند که با روش های استاندارد آزمایشگاهی اندازه گیری شدند. داده های طیفی پس از پیش پردازش میانگین-مرکزی با مقادیر آزمایشگاهی ویژگی های فیزیکوشیمیایی همبسته سازی شده و چهار مدل مستقل مبتنی بر رگرسیون حداقل مربعات جزئی (PLS) توسعه یافت. مدل ها با استفاده از الگوریتم Kennard-Stone برای تقسیم داده ها به مجموعه های کالیبراسیون و تست، اعتبارسنجی متقاطع با روش Leave-One-Out برای بهینه سازی تعداد متغیرهای پنهان، و شاخص هایی مانند RMSEP، %REP، و RMSECV ارزیابی شدند. نتایج نشان داد که مدل های پیش بینی قندهای احیاکننده پیش از هیدرولیز و رطوبت با خطای نسبی پایین (به ترتیب ۱.۶۴٪ و ۳.۰٪/۵) 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. به طور کلی، ترکیب طیف سنجی رامان با مدل های PLS روشی سریع، غیرمخرب، کم هزینه و سازگار با محیط زیست برای کنترل کیفیت عسل ارائه می دهد که می تواند در پایش اصالت، توزیع و توسعه اقتصادی این محصول نقش مؤثری ایفا کند.

کلمات کلیدی: کنترل کیفیت، ویژگی های فیزیکوشیمیایی، طیف سنجی رامان، عسل، مدل کمومتریکس، رگرسیون PLS

۱- مقدمه

طبق تعریف کمسیون کدکس مواد غذایی، عسل ماده غذایی شیرین طبیعی ساخته شده از شهد گلهای یا ترشحات بخشهای زنده گیاهان یا مواد دفعی حشرات مکند گیاه بر روی بخشهای زنده گیاه است که توسط زنبور عسل جمع آوری، تبدیل و با مواد خاص زنبور عسل 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 تحت تاثیر نوع گلهایی که زنبور عسل استفاده کرده است و همینطور شرایط منطقه ای و اقلیمی قرار میگیرد. عسل حاوی ۶۰-۸۰٪ گلوکز و فروکتوز به عنوان مونوساکارید، ۱۰٪ مالتوز و ساکارز به عنوان مهمترین دی ساکاریدها، مالزیتوز تری ساکارید اصلی و همچنین حاوی سایر اولیگو ساکاریدهایی با وزن مولکولی پایین میباشد. علاوه بر قندها، آنتی اکسیدانها (پنیوبانکسین، کرایسیتین و گالاجین)، ویتامین ها، اسیدها عمدتاً گلوکونیک اسید، پروتئینها، مواد معدنی، فلاونوئیدها، ویتامینها و آنزیمها نیز در عسل یافت میشوند (۱).

برای طبقه‌بندی و تعیین کیفیت عسل، استانداردها و روش‌هایی در کدکس آلیمانتاروس (۲)، کمیسیون بین المللی 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 می کنند. روش سنتی برای انجام آزمایشات کیفی عسل شامل آنالیز دانه های گرده (Melissopalynology) موجود در رسوبات آن توسط میکروسکوپ نوری است (۴). روش های دیگر گزارش شده در متون علمی شامل تکنیک های کروماتوگرافی، آنالیز رادیویی ایزوتوپ کربن پایدار و رزونانس مغناطیسی هسته ای می باشد (۵، ۶). اشکالات اصلی همه روش ها هزینه زیاد، صرف زمان زیاد، نیاز به تخصص همچنین تخریب نمونه است. این امر منجر به توسعه روش های تجزیه ای برای تشخیص اصالت عسل شده است که از جمله آن می توان به فناوری طیف‌سنجی در ترکیب با روش های کمومتریکس اشاره کرد که یک جایگزین خوب برای کنترل کیفیت سریع، قابل اعتماد و سازگار با محیط زیست است (۷). رایج ترین روش های کمومتری مورد استفاده در واکاوی عسل عبارتند از آنالیز اجزای اصلی (PCA)، آنالیز خوشه بندی سلسله مراتبی (HCA)، آنالیز تفکیک خطی (LDA)، حداقل مربع جزئی (PLS) و رگرسیون مؤلفه اصلی (PCR) (۸).

از تکنیک های طیف سنجی موجود، طیف سنجی رامان جهت واکاوی مواد غذایی ویژگی های مناسبی دارد که شامل عدم تداخل آب موجود در نمونه با اندازه گیری رامان، سهولت نمونه برداری و اندازه گیری و حداقل تداخل فلورسانس تغییر ماتریس نمونه می باشد. در سال های اخیر، روش های واکاوی مبتنی بر طیف سنجی رامان به عنوان یک روش اقتصادی و سریع برای تعیین منشا عسل مورد استفاده قرار گرفته است (۹، ۱۰).

کورووچی و همکاران (۹)، توانایی تشخیص منشاء گیاهی عسل را با استفاده از تکنیک melissopalynology و با طیف سنجی رامان در ترکیب با آنالیزهای چند متغیره (PCA) مقایسه کرد. نمونه های عسل از ایتالیا، اروپای شرقی و اسپانیا جمع آوری شد. با توجه به نتایج، تمایز منشاء عسل های مورد مطالعه از ۸۵٪ به ۹۹٪ بهبود یافته است. روش تحلیلی فراستو-ریس و همکاران (۱۱)، جهت تعیین منشا گل و گیاه، در عسل تولید شده توسط گونه خاصی از زنبور (*Apis Mellifera*)، از طیف سنجی رامان همراه با PCA استفاده کردند. در این مطالعه از ۶۶ نمونه عسل تک گل و چند گل جمع آوری شده از چندین منطقه مکزیک با انواع شرایط آب و هوایی استفاده شد. نتایج حاکی از آن بود که استفاده از روش کمومتری برای طبقه بندی عسل ها بر اساس منشاء و خلوص با صحت ۹۰٪ مناسب بود. ژاندریک و همکاران (۱۲) از ترکیب روشهای طیف سنجی رامان، طیف سنجی فروسرخ نزدیک و فروسرخ تبدیل فوریه^۱ جهت تشخیص منشا نمونه های عسل به دست آمده از نیوزلند، در طیف بین 200 cm^{-1} تا ۱۲۰۰۰ استفاده کردند. این روش با استفاده از PLS تکمیل شد. نتایج برازش مدل (R^2)، خطای استاندارد کالبراسیون (SEC) و خطای استاندارد پیش بینی (SEP) را به ترتیب ۰/۸۵، ۰/۲۱۹ و ۰/۳۱۵ نشان داد. اروپیان و Ropciuc (۱۰)، از واکاوی طیف رامان

¹ Fourier-transform infrared spectroscopy

برای تشخیص منشأ 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 درصد صحت اعتبارسنجی، بسیار عالی عمل نموده است.

همچنین از طیف سنجی رامان همراه با روش ها و مدل های آماری چند متغیره برای تشخیص عوامل خارجی موثر بر کیفیت عسل استفاده شده است (۱۳، ۱۴). طیف سنجی رامان و مدل های کمومتریکس برای پیش بینی غلظت گلوکز، فروکتوز، ساکارز و مالتوز موجود در نمونه های عسل ترکیه و یونان استفاده شده است (۱۵). همبستگی بین سطوح کمی قند و طیف رامان هم با استفاده از PLS و هم شبکه های 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، ۰/۹۵۶ و ۰/۹۱۳ برای شبکه عصبی به دست آمد که نشان می دهد هر دو روش کمومتریکس برای آنالیز سریع مقدار قند مناسب است. ارویان و همکاران (۱۰)، از طیف سنجی رامان برای تشخیص عسل تقلبی با قند (گلوکز، فروکتوز، قند اینورتر، شربت اینولین هیدرولیز شده، و مالت) استفاده کردند. در این مطالعه از ۹۰۰ نمونه با سطوح تقلب ۵، ۱۰، ۲۰، ۳۰، ۴۰ و ۵۰ درصد استفاده شد. تشخیص خلوص عسل با استفاده از PLS و PCR انجام شد. مدل های توسعه یافته کمومتریکس تناسب خوبی برای کالیبراسیون ($R^2_{cal} = ۰/۹۸۳$) و اعتبارسنجی ($R^2_{val} = ۰/۹۸۱$)، با خطاهای آماری کم ($SEC = ۰/۰۰۹$ و $SEP = ۰/۱۰۳$) نشان دادند. آنجوس و همکاران (۱۶)، پتانسیل طیف سنجی رامان را در پیش بینی ترکیب فیزیوشیمیایی عسل تک گل اسطوخودوس spp ارزیابی کرد. مدل های PLS برای تخمین کمی استفاده شد و نتایج حاصله با مقادیر به دست آمده از روش های مرجع مقایسه شد. همچنین از مدل های کمومتریکس برای pH، قندهای احیا، هدایت الکتریکی ساکارز، محتوای فنل کل، محتوای فلاونوئید کل، پرولین و کل اسیدها، استفاده شد. R^2 کالیبراسیون به دست آمده، در محدوده ۰/۹۷۳-۰/۹۹ و R^2 اعتبارسنجی در محدوده ۰/۸۳۳-۰/۹۹، SEC در محدوده ۰/۱-۲/۰۳ و SEP در محدوده ۱/۷۱-۰/۰۱ به دست آمد. در مطالعه طاهر و همکاران (۱۷)، طیف سنجی رامان همراه با PLS برای پیش بینی ترکیبات فنلی و فعالیت آنتی اکسیدانی عسل استفاده شد. مشخص شد که مدل های توسعه یافته مبتنی بر رامان نسبت به مدل های موجود برتری دارند. طیف سنجی رامان و همچنین مدل سازی PLS-LDA برای تعیین تقلب عسل چینی با شربت ذرت استفاده شده است (۱۸). واکاوی اختلاط و تقلب نمونه ها را در محدوده ۱۰، ۲۰ و ۴۰ درصد در نظر گرفت. صحت پیش بینی ۸۴/۴ درصد به دست آمد که نشان می دهد ترکیب PLS-LDA با طیف رامان یک روش بالقوه در تشخیص ناخالصی در عسل است.

طیف سنجی رامان به عنوان روشی سریع، غیرمخرب و دقیق برای پیش بینی ویژگی های فیزیوشیمیایی مواد غذایی مشابه عسل، جایگاه ویژه ای در پژوهش های اخیر یافته است. در این نوع محصولات که دارای ماهیتی ویسکوز و غنی از قند هستند، ویژگی هایی مانند ترکیب قندی (گلوکز، فروکتوز، ساکارز و مالتوز)، مواد جامد محلول کل ($Brix^\circ$)، رطوبت، فعالیت آبی، اسیدیته، pH، اسیدهای آلی، شاخص های حرارت دیدگی نظیر ۵-هیدروکسی متیل فورفورال (HMF)، ترکیبات فنولی، هدایت الکتریکی، وضعیت فیزیکی (تبلور) و تشخیص تقلب یا تعیین منشأ گیاهی و جغرافیایی، با استفاده از طیف سنجی رامان و روش های شیمی سنجی به طور مؤثری ارزیابی شده اند (۱۹). این روش بر پایه ی ثبت اثر انگشت ارتعاشی مولکول ها استوار است، به گونه ای که باندهای مربوط به پیوندهای C-O، C-C و C-H در نواحی ۷۰۰ تا ۱۲۰۰ cm^{-1} و ناحیه ی کششی O-H اطلاعات دقیقی از نوع قند و میزان آب فراهم می کنند (۲۰). پس از جمع آوری طیف های خام، داده ها با حذف خط پایه و فلورسانس، هموارسازی و نرمال سازی پردازش می شوند و سپس مدل های آماری مانند PLSR، PCR یا مدل های غیرخطی نظیر ANN، FFNN، SVM و CNN برای پیش بینی کمی یا طبقه بندی اصالت نمونه ها به کار می روند (۱۹). مطالعات متعدد نشان داده اند که این روش قادر است مقادیر گلوکز و فروکتوز را با ضریب تعیین بالای ۰.۹ و دقتی هم تراز با روش های مرجع کروماتوگرافی پیش بینی کند، و در تشخیص تقلب های رایج با شربت ذرت، مالتوز یا شکر معکوس عملکردی بالای ۹۵ درصد دارد (۲۰). برای مقادیر جزئی ترکیباتی مانند HMF یا آلانیندها، از رامان تقویت شده سطحی (SERS) جهت افزایش حساسیت بهره گرفته می شود و فناوری های جدیدی مانند رامان با جابه جایی فضایی و دستگاه های دستی، امکان غربالگری غیرمخرب در محل تولید یا بازار را فراهم کرده اند (۲۱). مهم ترین مزایای این روش، سرعت بالا، عدم نیاز به آماده سازی نمونه و قابلیت حمل آن است، در حالی که چالش هایی 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 رامان 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 رامان 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 بر اساس طیف‌های رامان 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 نمونه برداری جمع‌آوری و کدگذاری شد. آزمون‌های فیزیکوشیمیایی مطابق استاندارد ملی ایران شماره ۹۲ (۲۲) و همچنین طیف‌سنجی هر نمونه با دستگاه رامان انجام شد. روش‌های یادگیری ماشین (ML) به منظور مدل‌کردن ارتباط بین طیف رامان و ویژگی‌های فیزیکوشیمیایی عسل توسعه داده شد. جزئیات روش آزمون‌های فیزیکوشیمیایی، طیف‌گیری با اسپکتروسکوپی رامان و مدل‌سازی به شرح زیر می‌باشد:

۲-۱- ویژگی‌های فیزیکوشیمیایی

مواد شیمیایی مورد استفاده برای انجام آزمون‌های قند احیا کننده 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 فتالئین، محلول استاندارد هیدروکسید سدیم (NaOH)، آب دی‌یونیزه، بی‌سولفیت سدیم و محلول‌های کاریز یک 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:

۲-۱-۱- آزمون قند احیا کننده قبل از هیدرولیز

یک گرم از نمونه عسل در آب مقطر حل و به بالن ژوژه ۲۵۰ میلی‌لیتر به حجم رسانده شد و از محلول تهیه شده جهت انجام تیتراسیون محلول فهلینگ A و B (هر کدام ۵ میلی‌لیتر) استفاده شد (۲۲). درصد قند احیا مطابق فرمول یک محاسبه شد: فرمول ۱:

$$S = \frac{F \times 250 \times 100}{V \times W \times 1000}$$

که در آن:

S قندهای احیاکننده بر حسب گرم

F عیار فهلینگ

V حجم مصرفی محلول بر حسب میلی‌لیتر

W وزن نمونه عسل بر حسب گرم

۱۰۰۰ تبدیل میلی‌گرم به گرم

۲-۱-۲- قندهای احیا کننده بعد از هیدرولیز

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

فرمول ۲:

$$S_1 = \frac{F \times 250 \times 100 \times 100}{W \times V \times 50 \times 1000}$$

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

۳-۱-۲- درصد ساکارز

درصد ساکارز مطابق با فرمول ۳ محاسبه شد:

فرمول ۳:

$$(N) = (S_1 - S) \times 0.95 \text{ درصد ساکارز}$$

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 رفاکتومتری اندازه گیری شد. به این منظور، یک گرم عسل در دمای ۲۵ درجه سانتی گراد توسط رفاکتومتر (Refractometr BPTR-100, BenchTop)، آنالیز گردید. درصد رطوبت (گرم در ۱۰۰ گرم عسل) بر اساس جدول رابطه میران آب و ضریب شکست ارائه شده 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 گرم عسل در ۲۵ میلی لیتر آب دی یونیزه حل و به بالن ژوزه ۵۰ میلی لیتر منتقل گردید. محلول با افزودن ۰/۵ میلی لیتر معرف کاریز ۱ و کاریز ۲ شفاف شد و حجم آن به ۵۰ میلی لیتر رسانده شد. سپس محلول با استفاده از کاغذ صافی واتمن شماره ۴۲ صاف و به دو لوله آزمایش هر یک ۵ میلی لیتر از محلول صاف شده اضافه گردید. به یکی از دو لوله ۵ میلی لیتر اب مقطر و به لوله دیگر ۵ میلی لیتر بی سولیفیت سدیم (لوله شاهد) جهت غیر فعال شدن جذب هیدروکسی متیل فورفورال اضافه شد. جذب نوری نمونه در دستگاه طیف سنج UV-Vis مدل (Photonix Ar5- UV/Vis Spectrophotometer) در طول موج های ۲۸۴ و ۳۳۶ نانومتر در سل کوارتز اندازه گیری شد. غلظت هیدروکسی متیل فورفورال بر حسب میلی گرم بر کیلوگرم با استفاده از فرمول های ۴ و ۵ بیان گردید (۲۲):

فرمول ۴:

$$149.7 = \frac{126 \times 1000 \times 1000}{16830 \times 10 \times W}$$

فرمول ۵:

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

که در آن:

۱۴۹/۷ ضریب تبدیل جذب به غلظت

۱۶۸۳۰ ضریب جذب مولی هیدروکسی متیل فورفورال در ۲۸۴ نانومتر

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 طیف سنجی رامان مدل TDLG100 تکسان ساخت ایران با قابلیت گردآوری طیف و دیتا توسط نرم افزار Tansu مورد استفاده قرار گرفت. هر اندازه گیری در محدوده طیف ۱۰۰ cm⁻¹ تا ۳۵۰۰ انجام گردید. به منظور طیف گیری از نمونه ها از لیزر ۷۸۵ نانومتری YAG و همچنین برای تشخیص نور پراکنده رامان از آشکارساز Hamamatsu استفاده گردید. سه پارامتر اساسی برای بهینه سازی شدت پیک رامان انتخاب شد: قدرت لیزر، زمان ادغام و تعداد میانگین. این پارامترها به ترتیب ترتیب ۳۰۰ میلی وات، ۳ ثانیه و ۵ انتخاب شدند. برای انجام اسکن، دو گرم از هر یک از نمونه ها، در سه چاهک (سه تکرار) از پلیت ۲۰ خانه ریخته شده و سپس در نگهدارنده نمونه زیر میکروسکوپ رامان قرار گرفت. نور مستقیما بر روی نمونه ها متمرکز گردید و میانگین طیف آن ها پس از پنج اسکن برای سه تکرار اندازه گیری گردید.

۲-۳- توسعه و ارزیابی مدل های کمومتریکس

برای توسعه مدل های کمومتریکس جهت پیش بینی ویژگی های فیزیکوشیمیایی عسل، از داده های طیف سنجی رامان به عنوان متغیرهای مستقل و چهار ویژگی فیزیکوشیمیایی (میزان قند پیش از هیدرولیز، مقدار ساکاروز، میزان رطوبت و درصد هیدروکسی متیل فورفورال (HMF)) به عنوان متغیرهای وابسته استفاده شد. طیف های رامان در محدوده ۱۰۰ تا ۳۵۰۰ cm⁻¹ برای تمامی نمونه های عسل ثبت شدند. این داده های طیفی به صورت ماتریس های عددی سازمان دهی شده و به عنوان ورودی های مدل سازی 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 داده های طیفی و مقادیر آزمایشگاهی این ویژگی ها به صورت دقیق مدل سازی شود.

۲-۳-۱- رگرسیون حداقل مربعات جزئی (PLS)

رگرسیون حداقل مربعات جزئی (PLS) یک روش کمومتریکس چندمتغیره است که برای مدل سازی روابط پیچیده بین مجموعه ای از متغیرهای مستقل (در اینجا داده های طیف رامان) و متغیرهای وابسته (ویژگی های فیزیکوشیمیایی عسل) به کار می رود. این روش به ویژه در مواردی که تعداد متغیرهای مستقل زیاد بوده و ممکن است هم خطی (collinearity) بین آن ها وجود داشته باشد، بسیار مؤثر است. در PLS، داده ها به متغیرهای پنهان (latent variables) تبدیل می شوند که ترکیبی خطی از متغیرهای اصلی هستند و بیشترین واریانس در داده های مستقل و وابسته را توضیح می دهند. این روش با کاهش ابعاد داده ها 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 فیزیکی‌شیمیایی 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 به کار رفت. تقسیم‌بندی داده‌ها به دو مجموعه کالیبراسیون و تست با استفاده از الگوریتم Kennard-Stone انجام شد تا اطمینان حاصل شود که نمونه‌های انتخاب‌شده برای هر مجموعه نماینده تنوع کامل داده‌ها هستند.

۲-۳-۲- Kennard-Stone الگوریتم

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 فیزیکی‌شیمیایی را پوشش می‌دهند، و در نتیجه، کارایی پیش‌بینی مدل‌ها بهبود یافت.

۲-۳-۴- 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 به درستی 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.

۲-۳-۵- نرم‌افزار و ابزارهای محاسباتی

تمامی مراحل مدل‌سازی در محیط MATLAB (MathWorks, Natick, MA, USA) انجام شد. برای پیاده‌سازی مدل‌های PLS، از نرم‌افزار MVC1 (Multivariate Calibration 1) استفاده شد که به‌طور خاص برای تحلیل‌های کمومتریکی طراحی شده است. داده‌های طیف 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- یافته ها و بحث

۳-۱- آنالیزهای فیزیکوشیمیایی

همان گونه که در بخش مواد و روش ها اشاره شد، در این پژوهش تعداد ۵۱ نمونه عسل جمع آوری و با استفاده از طیفسنجی رامان در سه تکرار مستقل آنالیز گردیدند. داده های طیفی استخراج شده به منظور توسعه چهار مدل کمومتریک بر پایه روش PLS با مقادیر آزمایشگاهی شاخص های فیزیکوشیمیایی منتخب شامل قندهای احیاکننده پیش از هیدرولیز، رطوبت، ساکاروز و HMF همبسته شدند. هدف از این مرحله، ایجاد مدلی غیرمخرب برای پیش بینی ویژگی های کیفی عسل و بررسی میزان دقت آن ها در قیاس با روش های مرجع آزمایشگاهی بود. در ادامه، نتایج به دست آمده از اندازه گیری این ویژگی ها و مقایسه آن ها با حدود استاندارد های ملی (شماره ۹۲) ارائه و مورد بحث قرار می گیرند.

جدول (۱) مقادیر به دست آمده از آزمون های 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

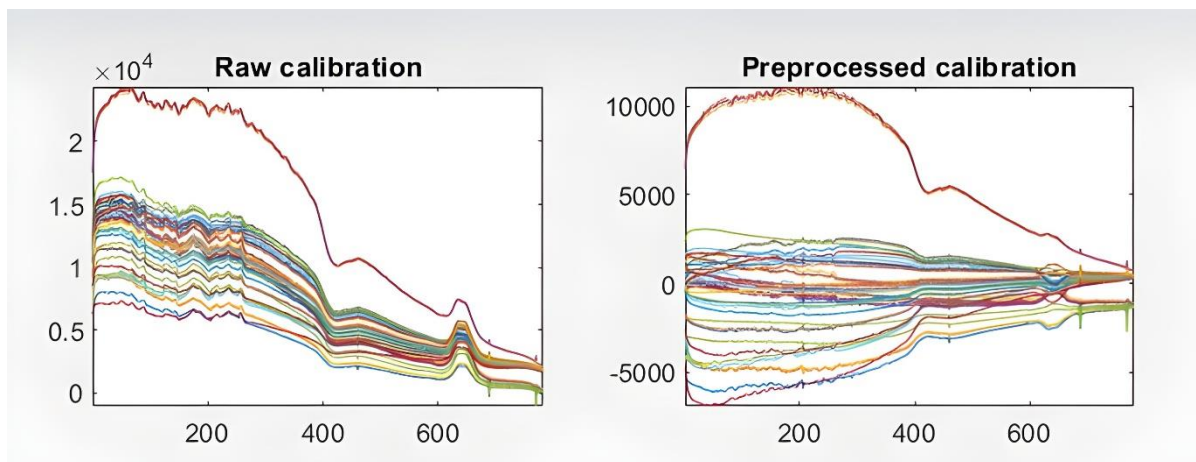
بر اساس نتایج حاصل از آنالیز نمونه های عسل، مهم ترین یافته های این پژوهش به شرح زیر است. مقدار ساکارز با میانگین $4/67 \pm 3/87$ گرم بر ۱۰۰ گرم، در محدوده قابل قبول استانداردهای ملی و بین المللی (کمتر از ۵ گرم بر ۱۰۰ گرم) قرار گرفت، هرچند در یک نمونه مقدار غیرمعمول $20/3$ گرم بر ۱۰۰ گرم مشاهده شد که احتمالاً نشان دهنده تغذیه زنبورها با شکر است. میزان HMF با میانگین $3/93 \pm 7/99$ میلی گرم بر کیلوگرم، در اکثر نمونه ها پایین بود که نشان دهنده تازگی و شرایط نگهداری مناسب عسل است. رطوبت نمونه ها بین $13/6$ تا 18 درصد متغیر بود و میانگین آن $15/56$ درصد به دست آمد که مطابق با استانداردهای جهانی (کمتر از ۲۰ درصد) است. این نتایج حاکی از کیفیت مطلوب و اصالت اکثر نمونه های مورد بررسی است، اگرچه نیاز به نظارت بیشتر بر برخی نمونه ها احساس می شود.

۳-۲- تحلیل طیف رامان

طیف های رامان به دست آمده از نمونه های عسل دارای باندهای طیفی در بازه های مختلف هستند که هر یک بیان کننده ارتعاشات و ساختار مولکولی خاصی می باشد. مطالعات مختلف طیف های رامان به دست آمده از نمونه های عسل را مورد بررسی قرار داده اند و به طور کلی بیان داشته اند که ناحیه طیفی بین $230-510 \text{ cm}^{-1}$ مربوط به ارتعاشات کششی و خمشی گروه های C-O، C-C-O و C-C-C است که ساختار مولکولی قندها را تشکیل می دهند (۱۷). ناحیه $595-691 \text{ cm}^{-1}$ به ارتعاشات کششی حلقه های غیراشباع موجود در HMF، کاروتن ها، فلاون ها، فلاونوئیدها و پلی فنول ها نسبت داده می شود (۲۳). قله مشاهده شده در $691-752 \text{ cm}^{-1}$ مربوط به ارتعاشات کششی C-O و C-C-O و ارتعاشات خمشی O-C-O است. در مقابل، باند $770-917 \text{ cm}^{-1}$ حاصل ارتعاشات کششی گروه های C-H و C-H موجود در گلوکز است (۲۴). باندهای بین $820-1024 \text{ cm}^{-1}$ به ارتعاشات تغییر شکل C-H و پیوندهای متیلن $\text{-CH}_2\text{-}$ 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 پروتئین ها نسبت داده می شود (۲۶). باند $1094-1191 \text{ cm}^{-1}$ به ارتعاشات کششی پیوندهای C-O و C-O-C قندها و همچنین پیوندهای C-N در پروتئین ها 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، در حالی که باندهای $1300-1460 \text{ cm}^{-1}$ ناشی از ارتعاشات خمشی و نوسانی گروه های عاملی CH و OH هستند (۲۶).

۳-۳- توسعه و ارزیابی مدل‌های کمومتریکس

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، 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 ۳۵۰۰ 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 کمومتریکس 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.



شکل (۱) طیف‌های رامان ثبت‌شده از نمونه‌های عسل در محدوده ۱۰۰ تا ۳۵۰۰ 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.

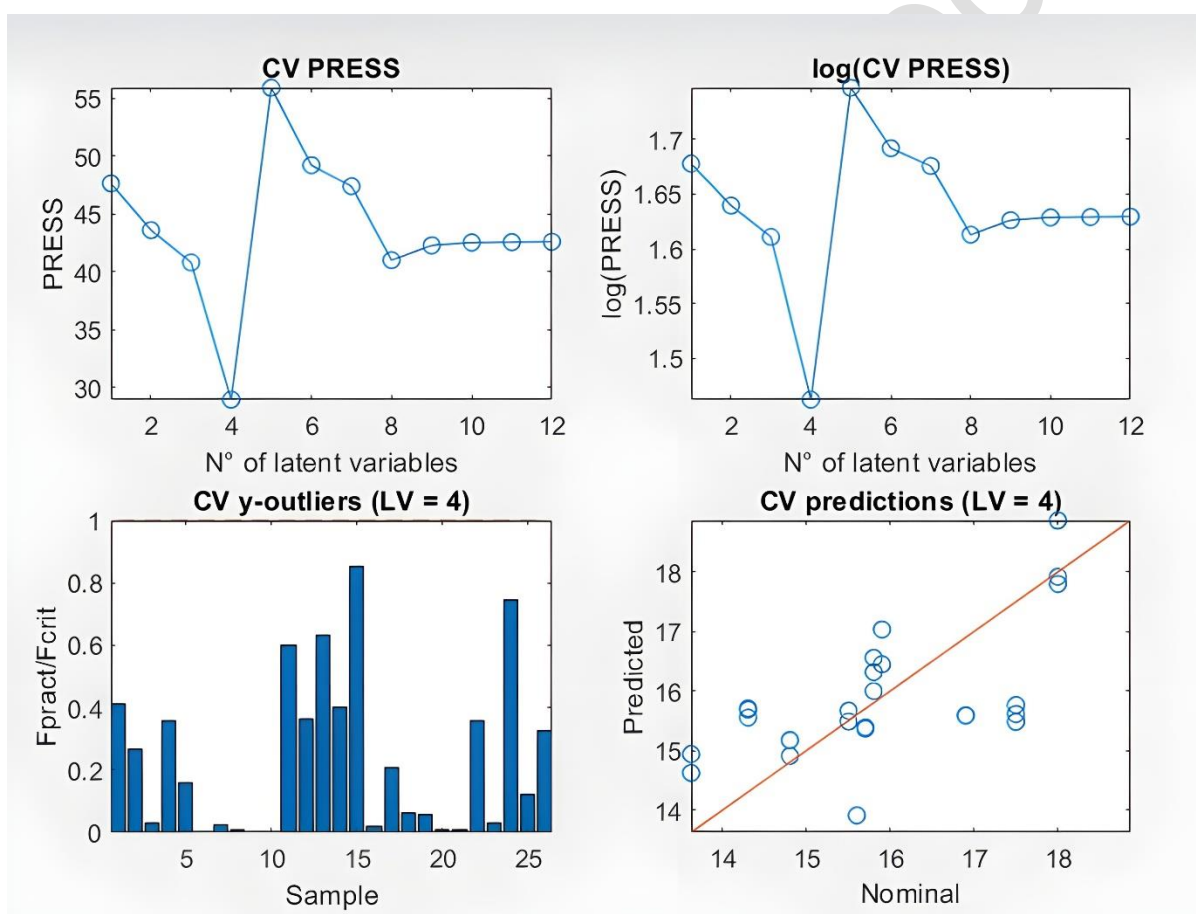
Figure 1. Raman spectra recorded from honey samples in the range of 100 to 3500 cm^{-1} . The left graph shows the raw spectra, and the right graph shows the spectra preprocessed with the mean-centering method. This preprocessing helps simplify calibration models and reduce noise.

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 بر اساس کمینه کردن 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 انتخاب شده 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.

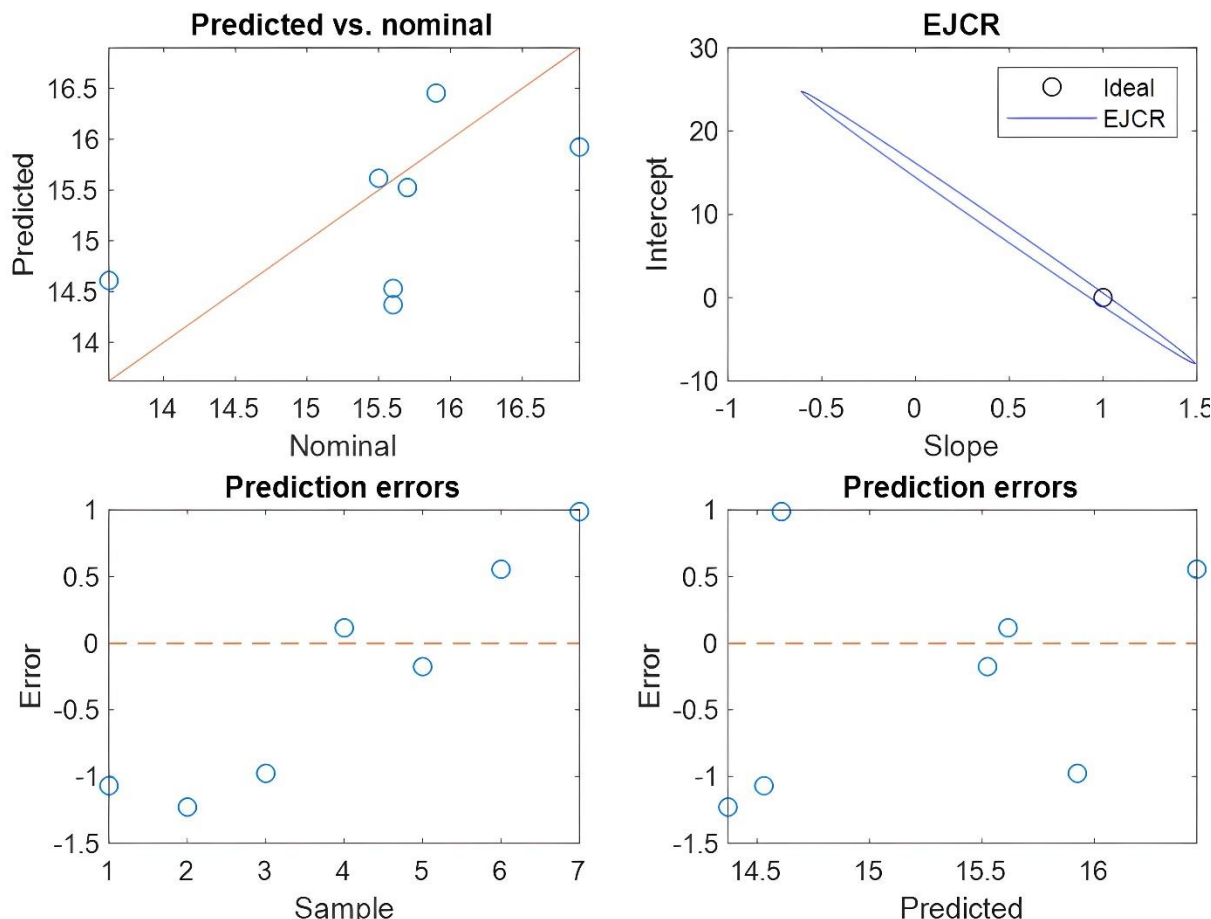
شکل ۲ نتایج اعمال تکنیک اعتبارسنجی متقاطع (Cross-Validation) برای مدل رطوبت عسل را به عنوان یک مثال نمایش می‌دهد. در ردیف بالای این شکل، نمودارهای مربوط به تعیین تعداد متغیرهای پنهان در مدل PLS برای داده‌های رطوبت نشان داده شده‌اند. این نمودارها به وضوح نشان می‌دهند که خطای پیش‌بینی برای نمونه‌های حذف شده از مجموعه کالیبراسیون در تعداد متغیرهای پنهان برابر با ۴ به حداقل می‌رسد، که تعداد بهینه متغیرهای پنهان برای این مدل را تأیید می‌کند. در ردیف پایین (چپ)، نمودار میله‌ای تست پرت (outlier) برای نمونه‌های مجموعه کالیبراسیون بر اساس مقایسه با خطای RMSECV ارائه شده است. این تست نشان می‌دهد که هیچ کدام از نمونه‌ها به عنوان پرت شناسایی نشده‌اند، زیرا مقدار تست F نسبی برای همه نمونه‌ها کمتر از مقدار مرزی ۱ است. همچنین، نمودار پایین سمت راست، مقادیر پیش‌بینی شده در مقابل مقادیر واقعی رطوبت در فرآیند اعتبارسنجی متقاطع را برای مدل PLS با ۴ متغیر پنهان نشان می‌دهد، که ارتباط خطی قوی و قابل قبولی را تأیید می‌کند.



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 میله‌ای 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.

Figure 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.

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 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.



شکل (۳) نتایج پیش‌بینی مدل کمومتریکس برای میزان رطوبت عسل در نمونه‌های مجموعه تست. ردیف بالا (چپ): نمودار مقادیر پیش‌بینی‌شده در مقابل مقادیر واقعی رطوبت. ردیف بالا (راست): نمودار منطقه اطمینان مشترک بیضی‌گون برای نمایش دامنه مجاز شیب و عرض از مبدأ نمودار مقادیر پیش‌بینی‌شده 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 پیش‌بینی‌شده رطوبت.

Figure 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.

معیارهای ارزیابی عملکرد مدل‌ها در پیش‌بینی مجموعه تست شامل ریشه میانگین مربعات پیش‌بینی (RMSEP) و درصد خطای نسبی (REP%) بودند. علاوه بر این، نمودارهای مقدار پیش‌بینی‌شده در مقابل مقدار واقعی برای مجموعه تست، نمودارهای EJCR (Ellipse Joint Confidence Region) و نمایش خطاهای مطلق برای هر نمونه در مجموعه تست نیز بررسی شدند. RMSEP نشان‌دهنده میزان انحراف پیش‌بینی‌های مدل از مقادیر واقعی است، در حالی که REP% خطای نسبی پیش‌بینی را

به صورت درصد بیان می کند و معیاری قابل فهم برای ارزیابی دقت مدل فراهم می آورد. نمودارهای مقدار پیش بینی شده در مقابل مقدار واقعی، انطباق بین مقادیر پیش بینی شده و واقعی را به صورت بصری نمایش می دهند. نمودارهای EJCR توصیفی آماری از توانایی پیش بینی مدل ارائه میکنند. بر اساس این آماره ناحیه مجاز انحراف شیب و عرض از مبدا خط مربوط به نمایش مقادیر پیش بینی شده نسبت به مقادیر واقعی (شکل 3) لازم است نقطه ای با مختصات یک و صفر را در بر بگیرد. همانطور که در نمودار EJCR در شکل 3 نشان داده است، مثالی از این تطابق در مدل توسعه یافته برای پیش بینی میزان رطوبت عسل مشاهده میشود. این معیارها در کنار یکدیگر، تصویری جامع از دقت و قابلیت اطمینان مدل های توسعه یافته ارائه می دهند. جدول ۱ مقادیر RMSEP و REP% را برای هر مدل نشان می دهد.

جدول (۲) پارامترها و نتایج مدل های کمومتریک مبتنی بر رگرسیون حداقل مربعات جزئی برای پیش بینی چهار ویژگی فیزیکوشیمیایی 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

این نتایج نشان داد که مدل های اول و چهارم، که به ترتیب برای پیش بینی قندهای احیاکننده پیش از هیدرولیز و رطوبت توسعه یافته بودند، با خطاهای نسبی بسیار پایین (REP% به ترتیب ۱/۶۴ و ۵/۳۰) عملکرد بسیار مطلوبی داشتند. این امر نشان دهنده توانایی بالای طیفسنجی رامان در ترکیب با مدل های PLS برای پیش بینی دقیق این دو ویژگی در محدوده های نرمال آنها است. مدل دوم، که برای پیش بینی ساکاروز توسعه یافت، با خطای نسبی حدود ۱۱/۹۵٪ عملکرد قابل قبولی داشت، در حالی که مدل سوم، که برای پیش بینی HMF طراحی شده بود، با خطای نسبی حدود ۲۵/۸٪ دقت کمتری نشان داد. این تفاوت در دقت می تواند به دلیل غلظت پایین HMF در نمونه های عسل و چالش های مرتبط با تشخیص آن باشد. به طور کلی، این نتایج تأیید می کنند که مدل های PLS مبتنی بر طیفسنجی رامان ابزار قدرتمندی برای پیش بینی ویژگی های فیزیکوشیمیایی عسل هستند و می توانند به عنوان 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 شربت های 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 کرده اند. با این حال، مدل کلی آنان برای چند نوع شربت عملکرد ضعیف تری داشت ($R^2p = 0.9068$, RMSEP = 10.54)، که نشان دهنده چالش های موجود در کمی سازی جامع تقلب است. در مقابل، پژوهش حاضر که بر پارامترهای کیفی ذاتی عسل تمرکز دارد نه بر شناسایی تقلب، توانسته است پیش بینی های دقیقی برای مؤلفه های کلیدی عسل ارائه دهد بدون نیاز به دانستن نوع تقلب کننده ای احتمالی (۲۷).

دقت بالای پیش بینی برای قندهای احیا کننده و رطوبت در این پژوهش با نتایج سایر مطالعاتی که از روش های طیفسنجی ارتعاشی استفاده کرده اند، هم راستا است. برای نمونه، مرور جامعی از El Haji & Estephan, 2025 نشان می دهد که طیفسنجی مادون قرمز میانی (MIR)، یکی از روش های مکمل در طیفسنجی ارتعاشی، با موفقیت بالا برای پیش بینی قند و رطوبت در عسل به کار رفته و معمولاً همراه با روش رگرسیون حداقل مربعات جزئی (PLS) استفاده می شود. هم خوانی نتایج ما با یافته های موفق MIR نشان می دهد که روش های طیفسنجی ارتعاشی، در صورت استفاده از ابزارهای شیمی سنجی مناسب، برای تحلیل ترکیبات اصلی عسل بسیار کارآمد و قابل اعتماد هستند (۲۸). به طور مشابه، Hu et al., 2022 نیز از طیفسنجی رامان همراه با الگوریتم های یادگیری ماشین برای شناسایی عسل تقلبی استفاده کردند و به دقتی نزدیک به صد درصد دست یافتند. موفقیت آنها در

تشخیص عسل خالص از 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 (C=O) و حلقه‌ی فورانی است، سیگنال آن در ناحیه‌ی MIR واضح‌تر و قابل اندازه‌گیری‌تر 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. علاوه بر این، غلظت پایین طبیعی HMF در نمونه‌های عسل تازه‌ی ما (میانگین ۷.۹۹ mg/kg) نیز چالشی جدی برای هر روش طیف‌سنجی ایجاد می‌کند، زیرا مقدار آن در محدوده‌ی حد تشخیص دستگاه قرار دارد. مطالعات همچنین نشان داده است که طیف‌سنجی NIR نیز برای اندازه‌گیری HMF به کار رفته، اما طیف‌های NIR به دلیل 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 به خوبی 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 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 تحت تأثیر 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 فروسرخ (IR) و رامان و به کارگیری شبکه عصبی سه‌لایه، توانستند دقت بالایی (۹۳.۸٪ در مرحله آزمون) در تفکیک جغرافیایی نمونه‌های عسل به دست آورند. در حالی که مطالعه حاضر تنها بر داده‌های رامان و رگرسیون PLS تکیه دارد؛ روشی که اگرچه برای پیش‌بینی کمی پارامترهای فیزیکی‌وشیمیایی خاص مؤثر است، اما ممکن است تنوع ترکیبی موردنیاز برای طبقه‌بندی‌های پیچیده‌تر مانند تشخیص دقیق منشأ 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 افاقیا 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 دست یافتند. 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 رامان همراه با الگوریتم‌های یادگیری ماشین را برای تشخیص منشأ 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 به‌راحتی آن‌ها را یاد می‌گیرد. در مقابل، دقت کمتر در پیش‌بینی HMF به دلیل 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. اندازه‌ی نمونه‌ها برای توسعه‌ی اولیه‌ی 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 به دقت داده‌های آزمایشگاهی مرجع نیز وابسته است و هرگونه خطا در آن داده‌ها مستقیماً به مدل منتقل می‌شود. در پژوهش‌های آینده، گسترش پایگاه داده‌ی 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 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 قابل حمل رامان با مدل‌های اعتبارسنجی شده می‌تواند زمینه‌ساز انجام آزمایش‌های میدانی شود و تحول بزرگی در پایش زنجیره‌ی تأمین و شناسایی تقلب در عسل ایجاد کند.

۴- نتیجه گیری

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 به دلیل 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 به دلیل 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 سازگاری با محیط زیست، ابزاری قدرتمند برای کنترل کیفیت عسل است. مدل‌های PLS توسعه‌یافته، با بهره‌گیری از داده‌های طیفی پیش‌پردازش شده 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 رامن تمرکز کنند تا این فناوری به‌طور گسترده‌تری در زنجیره تأمین عسل اعمال شود. در نهایت، این مطالعه نشان‌دهنده پتانسیل 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 171 و شناسه اخلاق IR.ZUMS.BLC.1404.060 انجام شده است. بدین وسیله مراتب قدردانی خود را از همکاری آن‌ها اعلام می‌داریم.

1. Duca A, Sturza A, Moacă E-A, Negrea M, Lalescu V-D, Lungeanu D, et al. Identification of resveratrol as bioactive compound of propolis from western Romania and characterization of phenolic profile and antioxidant activity of ethanolic extracts. *Molecules*. 2019;24(18):3368.
2. Bogdanov S. Harmonised methods of the International Honey Commission. Available online: <http://www.iwhc-platform.net/> (accessed on 1 November 2019). 2019.
3. AOAC. Official Methods of Analysis of AOAC. International; Association of Official Analysis Chemists International: Rockville, MD, USA, 2005; ISBN 0935584544. 2005.
4. Maione C, Barbosa Jr F, Barbosa RM. Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Computers and Electronics in Agriculture*. 2019;157:436-46.
5. Se KW, Wahab RA, Yaacob SNS, Ghoshal SK. Detection techniques for adulterants in honey: Challenges and recent trends. *Journal of Food Composition and Analysis*. 2019;80:16-32.
6. Das C, Chakraborty S, Acharya K, Bera NK, Chattopadhyay D, Karmakar A, et al. FT-MIR supported Electrical Impedance Spectroscopy based study of sugar adulterated honeys from different floral origin. *Talanta*. 2017;171:327-34.
7. Brereton RG. Introduction to multivariate calibration in analytical chemistry. Electronic Supplementary Information available. *Analyst*. 2000;125(11):2125-54.
8. Aliaño-González MJ, Ferreiro-González M, Espada-Bellido E, Palma M, Barbero GF. A screening method based on Visible-NIR spectroscopy for the identification and quantification of different adulterants in high-quality honey. *Talanta*. 2019;203:235-41.
9. Corvucci F, Nobili L, Melucci D, Grillenzoni F-V. The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis. *Food chemistry*. 2015;169:297-304.
10. Oroian M, Ropciuc S. Botanical authentication of honeys based on Raman spectra. *Journal of Food Measurement and Characterization*. 2018;12(1):545-54.
11. Frausto-Reyes C, Casillas-Peñuelas R, Quintanar-Stephano J, Macías-López E, Bujdud-Pérez J, Medina-Ramírez I. Spectroscopic study of honey from *Apis mellifera* from different regions in Mexico. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2017;178:212-7.
12. Jandrić Z, Haughey S, Frew R, McComb K, Galvin-King P, Elliott C, et al. Discrimination of honey of different floral origins by a combination of various chemical parameters. *Food Chemistry*. 2015;189:52-9.
13. Oroian M, Ropciuc S, Paduret S. Honey adulteration detection using Raman spectroscopy. *Food analytical methods*. 2018;11(4):959-68.
14. Anjos O, Santos AJ, Paixão V, Estevinho LM. Physicochemical characterization of *Lavandula* spp. honey with FT-Raman spectroscopy. *Talanta*. 2018;178:43-8.
15. Özbacı B, Boyacı İH, Topcu A, Kadılar C, Tamer U. Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food chemistry*. 2013;136(3-4):1444-52.
16. Anjos O, Santos AJA, Paixão V, Estevinho LM. Physicochemical characterization of *Lavandula* spp. honey with FT-Raman spectroscopy. *Talanta*. 2018;178:43-8.
17. Tahir HE, Xiaobo Z, Zhihua L, Jiyong S, Zhai X, Wang S, et al. Rapid prediction of phenolic compounds and antioxidant activity of Sudanese honey using Raman and Fourier transform infrared (FT-IR) spectroscopy. *Food Chemistry*. 2017;226:202-11.
18. Li S, Shan Y, Zhu X, Zhang X, Ling G. Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy. *J Food Compos Anal* 2012;28, 69–74.
19. Anguebes-Franceschi F, Abatal M, Pat L, Flores A, Córdova Quiroz A, Ramírez-Elias M, et al. Raman spectroscopy and chemometric modeling to predict physical-chemical honey properties from Campeche, Mexico. *Molecules*. 2019;24(22):4091.
20. Batsoulis AN, Siatis NG, Kimbaris AC, Alissandrakis EK, Pappas CS, Tarantilis PA, et al. FT-Raman spectroscopic simultaneous determination of fructose and glucose in honey. *Journal of agricultural and food chemistry*. 2005;53(2):207-10.
21. Craig AP, Franca AS, Irudayaraj J. Surface-enhanced Raman spectroscopy applied to food safety. *Annual review of food science and technology*. 2013;4:369-80.
22. INSO. Honey – Specifications and test methods. Iranian National Standardization Organization; 2020.
23. Li S, Shan Y, Zhu X, Zhang X, Ling G. Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy. *Journal of Food Composition and Analysis*. 2012;28(1):69-74.

24. White Jr JW. Moisture in honey: Review of chemical and physical methods. *Journal of the Association of Official Analytical Chemists*. 1969;52(4):729-37.
25. Kek SP, Chin NL, Yusof YA, Tan SW, Chua LS. Classification of entomological origin of honey based on its physicochemical and antioxidant properties. *International journal of food properties*. 2017;20(sup3):S2723-S38.
26. Belay A, Solomon W, Bultossa G, Adgaba N, Melaku S. Physicochemical properties of the Harennna forest honey, Bale, Ethiopia. *Food chemistry*. 2013;141(4):3386-92.
27. Wu X, Xu B, Ma R, Niu Y, Gao S, Liu H, et al. Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2022;274:121133.
28. El Hajj R, Estephan N. Advances in infrared spectroscopy and chemometrics for honey analysis: a comprehensive review. *Critical Reviews in Food Science and Nutrition*. 2024:1-14.
29. Hu S, Li H, Chen C, Chen C, Zhao D, Dong B, et al. Raman spectroscopy combined with machine learning algorithms to detect adulterated Suichang native honey. *Scientific reports*. 2022;12(1):3456.
30. Stöbener A, Naefken U, Kleber J, Liese A. Determination of trace amounts with ATR FTIR spectroscopy and chemometrics: 5-(hydroxymethyl) furfural in honey. *Talanta*. 2019;204:1-5.
31. David M, Berghian-Grosan C, Magdas DA. Honey differentiation using infrared and raman spectroscopy analysis and the employment of machine-learning-based authentication models. *Foods*. 2025;14(6):1032.
32. Song X, She S, Xin M, Chen L, Li Y, Vander Heyden Y, et al. Detection of adulteration in Chinese monofloral honey using ^1H nuclear magnetic resonance and chemometrics. *Journal of Food Composition and Analysis*. 2020;86:103390.
33. Magdas DA, Guyon F, Berghian-Grosan C, Molnar CM. Challenges and a step forward in honey classification based on Raman spectroscopy. *Food Control*. 2021;123:107769.

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

Short title: Prediction of Honey Properties Using Chemometrics and Raman Hyperspectral Spectroscopy

Fataneh Hashempour-Baltork ¹, Somaye Vali Zade ¹, Adel Mirza Alizadeh ^{2*}, Behrooz Jannat ¹, Hossein Rastegar ¹

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.

* Corresponding author: A. Mirza Alizadeh (a.alizadeh@zums.ac.ir)

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 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 work 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