



روش‌های محاسباتی و کمومتریکس ساده و قابل اعتماد جهت پیشگویی شاخص‌های بازداری کوتاس *Pistacia atlantica* Desf. ترکیبات طبیعی تشکیل دهنده‌ی روغن اسانسی فرار

مجید محمدحسینی*، مهدی نکویی

دانشگاه آزاد اسلامی، واحد شاهرود، دانشکده‌ی علوم پایه، گروه شیمی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۳/۴/۱۱، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۳/۶/۲۵، تاریخ پذیرش قطعی: ۱۳۹۳/۷/۴

چکیده

در این مقاله، مدل‌های معتبر مبتنی بر روابط کمی ساختار-بازداری جهت پیش‌گویی ضرائب بازداری کوتاس (RI) مجموعه‌ی وسیعی از ترکیبات طبیعی یافت‌شده در پروفایل‌های شیمیایی روغن‌های اسانسی توسعه یافته‌اند. این مدل‌ها، براساس رگرسیون خطی چندگانه‌ی مرحله‌ای و الگوریتم ژنتیک-رگرسیون خطی چندگانه‌ی می‌باشند. بر این اساس، روش‌های انتخاب متغیر مرحله‌ای و الگوریتم ژنتیک جهت گزینش مؤثرترین توصیف‌کننده‌های مولکولی از مجموعه‌ی توصیف‌کننده‌های محاسبه‌شده با نرم‌افزار دراگون مورد استفاده قرار گرفتند. بعلاوه، الگوی آنالیز جزء اصلی، حوزه‌ی کاربرد، نمودار ویلیام، فاکتور تورم تغییر (VIF)، اثر متوسط (MF_j)، ماتریکس هم‌بستگی، نمودارهای پیش‌گویی و باقی‌مانده همراه با آنالیز آماری مبسوط نیز ارزیابی شدند. مدل‌های خطی QSRR، درجات بالایی از هم‌بستگی را بین مقادیر مشاهده‌شده و محاسبه‌شده‌ی اندیس‌های کوتاس نشان‌داده که مؤید کاربرد موفقیت‌آمیز آنها در پیش‌گویی مقادیر RI ترکیبات طبیعی ناشناخته است. نهایتاً، پایداری و توانایی پیش‌گویی مدل‌ها با استفاده از تکنیک‌های حذف تکی-اعتبارسنجی تقاطعی (LOO-CV)، حذف گروهی-اعتبارسنجی تقاطعی (LGO-CV)، اعتبارسنجی خارجی روی سری آزمون و آزمون Y تصادفی مورد تأیید قرار گرفت.

واژه‌های کلیدی: روغن‌های اسانسی، کروماتوگرافی گازی-طیف‌سنجی جرمی، آویشن، فعالیت آنتی‌اکسیدان، فعالیت آنتی‌باکتریال.

۱. مقدمه

شیمی تجزیه که کمی‌ترین شاخه در میان شاخه‌های متعدد شیمی است، بیش از هر شاخه‌ی دیگر علم شیمی به استفاده از مفاهیم آماری نیازمند است. هدف نهایی شیمی تجزیه و روش‌های کلاسیک و دستگاهی به کاررفته در آن، دست‌یافتن به تشخیص یا اندازه‌گیری انتخابی مواد و ترکیبات مختلفی است که هر روز بر تعداد و تنوع آنها افزوده می‌شود. هر تجزیه‌گر برای دستیابی به نتایج قابل قبول در تشخیص یا تعیین مقدار، ناگزیر از انتخاب روش‌هایی است که در اجرای آنها، تأثیر نامطلوب عوامل مزاحم و مداخله‌گر محدود شده و به اصطلاح «شرایط بهینه» برای

*عهده‌دار مکاتبات: مجید محمدحسینی

نشانی: دانشگاه آزاد اسلامی واحد شاهرود، دانشکده‌ی علوم پایه، گروه شیمی

پست الکترونیک: majidmohammadhosseini@yahoo.com

تلفن: ۰۲۳-۳۲۳۹۴۵۳۰

آزمایش فراهم شده باشد. با این حال، بکارگیری مکرر چنین روش‌هایی نیز به نتایج کاملاً یکسان منتهی نمی‌شود و بدین ترتیب تصمیم‌گیری درباره‌ی ارائه‌ی جواب معینی برای یک اندازه‌گیری و یا قابل قبول بودن نتایج به دست آمده از روش‌های متفاوت برای نمونه‌های واحد و نظایر اینها، مستلزم بکارگیری مفاهیم آماری است [۱]. توانایی حل مسائل مختلف در زمینه‌ی درک سازوکار انجام فرآیندهای شیمیایی مختلف، کشف و توسعه‌ی مواد جدید و زمینه‌های دیگر شیمی هنوز به طور کامل محقق نشده است. روش‌های محاسباتی کمومتریکس توانسته‌اند در جهت حل این مشکلات به میزان زیادی مفید فایده واقع شوند. روش‌های نظری فنون کمومتریکس از چند دهه‌ی پیش شناخته شده‌اند، اما اخیراً با ظهور کامپیوترهای قدرتمند، کمومتریکس ماهیت واقعی خود را پیدا کرده است. بر طبق تعریف جامعی که انجمن بین‌المللی کمومتریکس از این روش دارد، کمومتریکس علم برقراری ارتباط بین سنجش‌های انجام شده بر روی یک سیستم یا فرآیند شیمیایی با استفاده از روش‌های آماری و ریاضی می‌باشد [۲]. علاوه بر مکانیک کوانتم، روش‌های مدل‌سازی نیز که ارزشی معادل یا حتی تأثیری بیشتر در کاربردهای صنعتی دارند در مراجع علمی ارائه شده‌اند. برخلاف مدل‌های مبتنی بر مکانیک کوانتم که انرژی را ارزیابی و براساس آن فرآیندهای ترمودینامیک را پیش‌بینی می‌کنند، این مدل‌های نوین بر پایه‌ی غیر از انرژی استوارند. گرافیک کامپیوتری، بزرگ‌ترین مؤلفه‌ی فن‌آوری انفورماتیک در گستره‌ی بسیار وسیعی از شبیه‌سازی‌های مولکولی است. زمینه‌ی دیگر که توسعه‌ی آن اثر مهمی در شبیه‌سازی و مدل‌سازی‌های مولکولی دارد، هندسه‌ی بعد است. سومین زمینه‌ی مدل‌سازی شامل ارتباط میان داده‌های آزمایشگاهی با خواص فیزیکوشیمیایی در قالب سیستم مدل‌سازی است. این روش‌ها، نقش مؤثر و غیر قابل انکاری در نتایج حاصل از مدل‌سازی در شیمی دارند [۳].

برای درک سازوکار فرآیندهای مختلف شیمی، کشف و تعمیم راهکارهای نوین، حفظ محیط زیست و زمینه‌های دیگر شیمی، هنوز توانایی حل مسائل به‌طور کامل وجود ندارد و عملی کردن بعضی از مسائل نیاز به سیستم‌های بسیار پیچیده دارد که استفاده از آنها مستلزم صرف هزینه‌های بسیار و انجام مطالعات گسترده است. روش‌های محاسباتی کمومتریکس جهت حل این مشکلات می‌توانند گزینه‌ی مفیدی باشند [۴]. شاخه‌های مختلف شیمی اعم از شیمی فیزیک، شیمی آلی، شیمی معدنی، بیوشیمی و به ویژه شیمی تجزیه به نحوی از این تکنیک‌ها بهره می‌جویند. یکی از شاخه‌های بسیار مهم کمومتریکس که مورد بحث این تحقیق است، ارتباط کمی ساختار-بازداری (QSRR) است. از جمله کاربردهای این تکنیک، پیش‌بینی اندیس کوواتس (شاخص بازداری) (RI) برخی از اسانس‌های گیاهان دارویی در شرایط مختلف است. در بین روش‌های به کاررفته جهت مطالعه‌ی ارتباط کمی ساختار-بازداری، می‌توان به روش رگرسیون خطی چندگانه (MLR) اشاره نمود. در این تحقیق، بر روی شاخص بازداری یک نمونه از اسانس‌های گیاهی، مدل‌سازی MLR انجام و توصیف‌کننده‌های مولکولی مؤثر شناسایی و بررسی شده است. در بررسی اسانس‌های بسیاری از گیاهان دارویی، مسائلی مانند سمیت بسیار زیاد آنها، گران قیمت بودن، تماس احتمالی آنها با پوست و مسائل زیست محیطی را نباید از نظر دور داشت. این محدودیت‌ها موجب می‌شوند که در کنار روش‌های آزمایشگاهی و تجربی، روش‌های نظری نیز جهت استفاده‌ی بهینه از نتایج تجربی توسعه یابند. تدبیر اساسی QSRR، به دست آوردن رابطه‌ی کمی بهینه‌ای از ساختار مولکولی برای پیش‌بینی خواص می‌باشد. در صورت حصول رابطه‌ی معتبر، امکان پیش‌بینی اندیس بازداری برای ساختارهای مشابه یا ساختارهای دیگر که هنوز اندازه‌گیری و یا تهیه نشده‌اند، وجود خواهد داشت [۷-۲].

ضریب (اندیس یا شاخص) بازداری، به سال ۱۹۵۸ میلادی توسط دانشمندی به نام کوواتس به واژگان علم کروماتوگرافی افزوده شد و امروزه به عنوان یک مقیاس و یک معیار توانمند علی‌الخصوص در مقوله‌ی کیفی و شناسایی عرض اندام می‌کند. کوواتس به فراست دریافت که اندیس بازداری هر آلکان ۱۰۰ برابر تعداد اتم‌های کربن تشکیل‌دهنده‌ی آن است. او هم‌چنین دریافت که اگر لگاریتم زمان بازداری یک‌سری آلکان هومولوگ (همراه) بر حسب تعداد اتم‌های کربن آنها رسم شود می‌توان به یک نمودار خطی رسید. طبق نظریه‌ی کوواتس، اندیس بازداری هر

گونه را می توان به سادگی و با توجه به کروماتوگرام مخلوطی از آن با دو پارافین متوالی تعیین نمود مشروط به آن که زمان های بازداری دو آلکان یکی n کربنه و دیگری $n+1$ کربنه در دوطرف زمان بازداری گونه ی مورد نظر باشد. تحت شرایطی که ستون کروماتوگرافی مفروض، متحمل برنامه ریزی حرارتی نشود، اندیس کوواتس را می توان به توسط رابطه ی ۱ تعیین نمود:

$$I_x = 100 \frac{[\log(t_R - t_M) - \log(t_n - t_M)]}{[\log(t_{n+1} - t_M) - \log(t_n - t_M)]} + 100 n \quad (1)$$

در صورتی که ستون کروماتوگرافی تحت برنامه ریزی حرارتی قرار گیرد، رابطه ی لگاریتمی به صورت خطی درآمده و اندیس کوواتس با استناد به رابطه ی ۲ تعیین می شود:

$$I_x = 100 \frac{(t_R - t_n)}{(t_{n+1} - t_n)} + 100 n \quad (2)$$

در این روابط:

I_x : اندیس بازداری، t_R : زمان بازداری یا زمان مورد نیاز برای شویش و جداسازی گونه ی بدام افتاده در منافذ فاز ساکن به وسیله ی فاز متحرک در ستون کروماتوگرافی، t_M : زمان مرده یا زمان لازم برای طی شدن طول ستون کروماتوگرافی به وسیله ی فاز متحرک فاقد نمونه، n : تعداد اتم های کربن، t_n : زمان بازداری آلکان n کربنه و t_{n+1} : زمان بازداری آلکان $n+1$ کربنه می باشد. [۸]. تجزیه و تحلیل آماری و ریاضی داده های شیمیایی، کموتریکس خوانده می شود [۹]. به عبارت دیگر، کموتریکس یک روش کارآمد برای خلاصه کردن اطلاعات مفید از یک سری داده ی مشخص و پیش بینی سری دیگر داده هاست. هدف کموتریکس، بهبود بخشیدن فرآیندهای اندازه گیری و استخراج اطلاعات شیمیایی مفیدتر از داده های اندازه گیری شده ی فیزیکی و شیمیایی می باشد [۱۰]. کموتریکس، نخستین بار توسط دانشمند سوئدی به نام ولد [۱۱] در سال ۱۹۷۲ استفاده و کمی بعد توسط کووالسکی توسعه داده شد. پیشتر، عباراتی مانند بیومتریس و اقتصادسنجی نیز در زمینه های علوم زیست و اقتصاد معرفی شده اند. پس از آن، انجمن بین المللی کموتریکس تأسیس شد. از آن زمان به بعد، کموتریکس پیشرفت شگرفی داشته و امروزه در شاخه های مختلف شیمی، به ویژه شیمی تجزیه به میزان فرآیندهای مورد استفاده قرار می گیرد. اولین مرحله ی رشد کموتریکس قبل از سال ۱۹۷۰ میلادی است که در آن تعدادی از روش های ریاضی توسعه یافته و در زمینه های مختلف ریاضی، علم رفتار و مهندسی استفاده شدند. در این محدوده ی زمانی، شیمیدانان حوزه ی پژوهش خود را به آنالیز داده ها، از جمله محاسبه ی پارامترهای آماری مانند میانگین، انحراف استاندارد و حدود اطمینان محدود کرده بودند. دومین مرحله ی تکامل کموتریکس در سال ۱۹۷۰ میلادی صورت گرفت. در این سال، کموتریکس توجه شیمیدانان، به خصوص شیمی دانان تجزیه را هم برای پردازش و تحلیل داده ها و هم اتخاذ راهکارهای نوین جهت آنالیز نمونه های پیچیده، جلب کرد. تکامل بعدی کموتریکس که توسط هاوری و هیرش [۱۲] و سپس براون [۱۳] پیش بینی شده بود، از اوایل ۱۹۸۰ میلادی آغاز شد. هاوری و هیرش، تحقیق خود را به طور ویژه به هم بستگی تعداد زیادی از داده های شیمیایی به ویژگی های مولکولی مربوطه، اختصاص دادند. کموتریکس در دانشگاه های اروپا و آمریکا در گرایش های شیمی برای دانش آموزان و دانشجویان دوره های تحصیلات تکمیلی ارائه شد. بعلاوه، کموتریکس به عنوان ابزاری توانمند و یک شاخه ی وابسته ی شیمی در بسیاری از دانشگاه های آمریکا و اروپا، بعضی از دانشگاه ها در چین و سایر کشورها به شمار آمد. پیشرفت ها در علوم کامپیوتر، اطلاعات، آمار و ریاضی کاربردی عناصر جدیدی را در کموتریکس معرفی کرده است. شبکه ی عصبی مصنوعی [۱۴، ۱۵]، تکنیکی ریاضی که ارسال سیگنال از سیستم های بیولوژیکی را شبیه سازی می کند و الگوریتم ژنتیکی [۱۶، ۱۷-۲۳] نمونه هایی از این روش ها هستند. در تحقیق اخیر، مدل سازی های خطی توانمند مبتنی بر روش های انتخاب متغیر مرحله ای و الگوریتم ژنتیک جهت پیش بینی شاخص های بازداری کوواتس ترکیبات طبیعی موجود در روغن اسانسی فرار حاصل از گیاه *Pistacia atlantica* Desf. ارائه شده است.

۲. بخش مدل سازی

۱-۲. نرم افزارهای مورد استفاده

برای رسم و بهینه کردن شکل مولکولها از نرم افزار هایپرکم (Hyperchem) (نسخه ۷/۰) استفاده گردید. با توجه به موضوع مورد بحث (بررسی ارتباط کمی ساختار-بازداری)، ابتدا ساختار مولکولها به وسیله نرم افزار Hyperchem رسم و ساختار هندسی آنها با احتساب اتمهای هیدروژن و با استفاده از مدل هامیلتونی AM1 و الگوریتم Polak-Ribiere که از روشهای نیمه تجربی می باشند، بهینه شد. سپس، ساختارهای رسم شده، توسط نرم افزار مذکور به عنوان ورودی به نرم افزار دیگری به نام Dragon، معرفی شد. محاسبه‌ی توصیف کننده‌ها، توسط این نرم افزار صورت گرفت. بالغ بر ۱۴۰۰ توصیف کننده از هجده گروه مختلف برای هر ترکیب با این بسته نرم افزاری، قابل محاسبه است. از نرم افزار SPSS برای انتخاب بهترین توصیف کننده‌ها و از نرم افزار MATLAB برای به دست آوردن سایر پارامترهای آماری استفاده گردید. وظیفه‌ی خطیر مدل سازی، بر دوش ترکیبات سری آموزش قرار داشته و ترکیبات سری آزمون در این امر دخالت ندارند. در عوض، قدرت پیشگویی مدل و توانمندی آن با اعمال مدل به سری آزمون مورد بررسی قرار می گیرد.


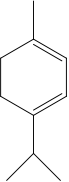
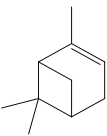
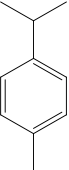
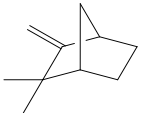
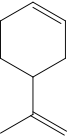
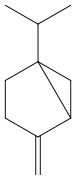
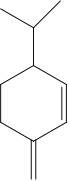
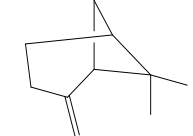
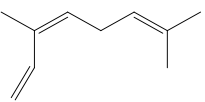
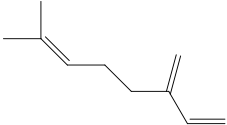
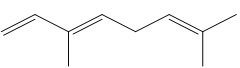
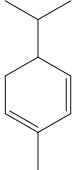
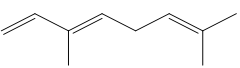
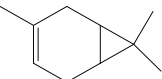
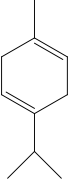
۲-۲. سری داده‌ها

در این بخش، اندیس کوواتس ۶۲ ترکیب از اسانس گیاه دارویی *Pistacia atlantica* Desf. (شکل ۱)، با استفاده از روش رگرسیون خطی چندگانه مورد بررسی قرار گرفته است. سری داده‌ی مربوطه، شامل مجموعه‌ای از مونوترپن‌ها، سزکوئی‌ترین‌ها و سایر ترکیبات طبیعی، از تحقیق یک گروه از محققین الجزایری به سال ۲۰۰۸ میلادی گرفته شده است [۲۴]. مراحل متوالی انجام این پژوهش، شامل انتخاب سری داده‌ها، تعیین توصیف کننده‌ها، رگرسیون خطی چندگانه و مدل سازی می باشد. نام، ساختار و مقادیر عددی اندیس‌های بازداری (RI) آنها در جدول ۱، نشان داده شده است. مطابق روال معمول، برای مدل سازی، ابتدا اندیس بازداری ترکیبات، به عنوان پارامتر مورد بحث، انتخاب شد. در مرحله‌ی دوم، ساختار تمام گونه‌های مورد نظر رسم و بهینه گردید.

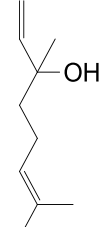
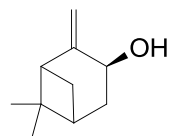
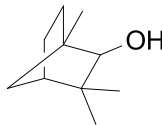
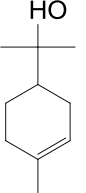
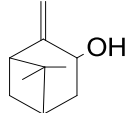
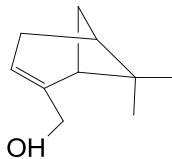
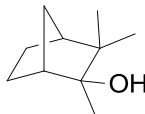
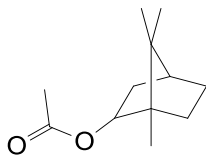
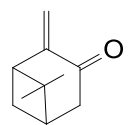
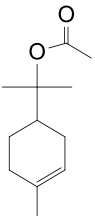
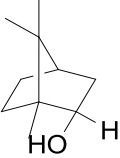
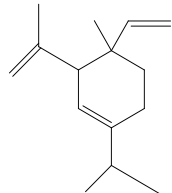
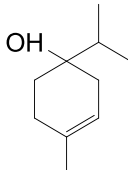
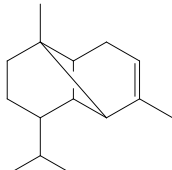


شکل ۱. تصویر گیاه *Pistacia atlantica* Desf.

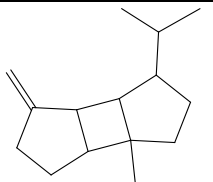
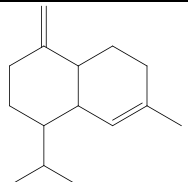
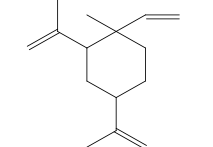
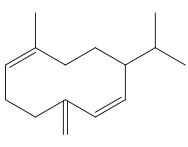
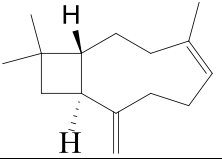
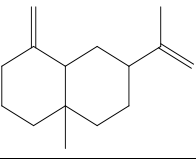
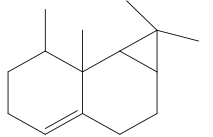
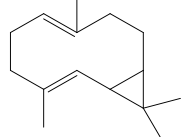
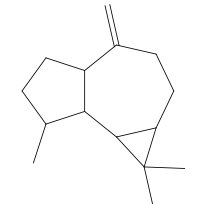
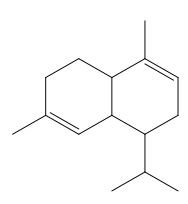
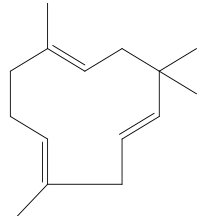
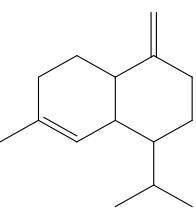
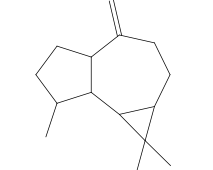
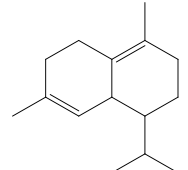
جدول ۱. نام، ساختار و اندیس کووانس ترکیبات اساسی گیاه *Pistacia atlantica* Desf.

شماره	نام ترکیب	RI	ساختار ترکیب	شماره	نام ترکیب	RI	ساختار ترکیب
۱	Tricyclene	920		۹	α -Terpinene	1009	
۲	α -Pinene	932		۱۰	p-Cymene	1011	
۳	Camphene	944		۱۱	Limonene	1021	
۴	Sabinene	965		۱۲	β -Phellandrene	1021	
۵	β -Pinene	972		۱۳	(Z)- β -Ocimene	1024	
۶	Myrcene	980		۱۴	(E)- β -Ocimene	1036	
۷	α -Phellandrene	997		۱۵	γ -Terpinene	1048	
۸	δ -3-Carene	1004		۱۶	Terpinolene	1078	

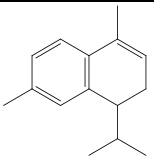
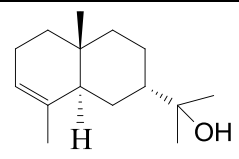
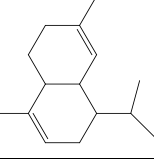
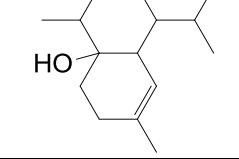
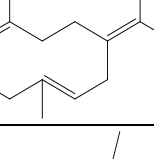
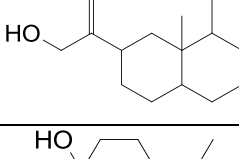
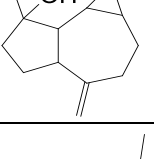
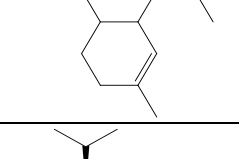
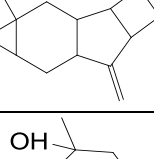
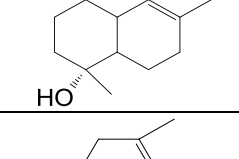
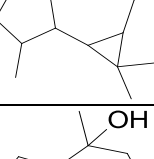
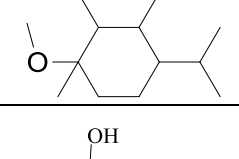
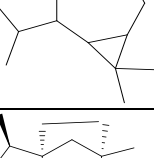
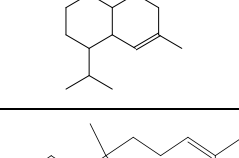
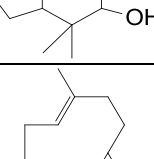
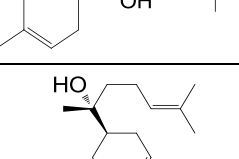
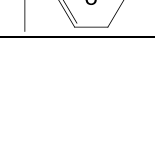
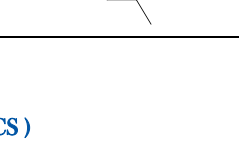
ادامه‌ی جدول ۱.

شماره	نام ترکیب	RI	ساختار ترکیب	شماره	نام ترکیب	RI	ساختار ترکیب
۱۷	Linalool	1082		۲۴	cis-Pinocarveol	1168	
۱۸	α -Fenchol	1098		۲۵	α -Terpineol	1172	
۱۹	trans-Pinocarveol	1124		۲۶	Myrtenol	1178	
۲۰	Camphene hydrate	1132		۲۷	Bornyl acetate	1267	
۲۱	Pinocarvone	1137		۲۸	α -Terpinyl acetate	1331	
۲۲	Borneol	1148		۲۹	δ -Elemene	1334	
۲۳	Terpinene-4-ol	1162		۳۰	α -Copaene	1377	

ادامه‌ی جدول ۱.

شماره	نام ترکیب	RI	ساختار ترکیب	شماره	نام ترکیب	RI	ساختار ترکیب
۳۱	β -Bourbonene	1384		۳۸	γ -Muurolene	1471	
۳۲	β -Elemene	1388		۳۹	Germacrene-D	1477	
۳۳	β -Caryophyllene	1420		۴۰	β -Selinene	1487	
۳۴	β -Gurjunene	1428		۴۱	Bicyclogermacrene	1492	
۳۵	Aromadendrene	1439		۴۲	α -Muurolene	1494	
۳۶	α -Humulene	1452		۴۳	γ -Cadinene	1507	
۳۷	Alloaromadendrene	1459		۴۴	δ -Cadinene	1515	

ادامه‌ی جدول ۱.

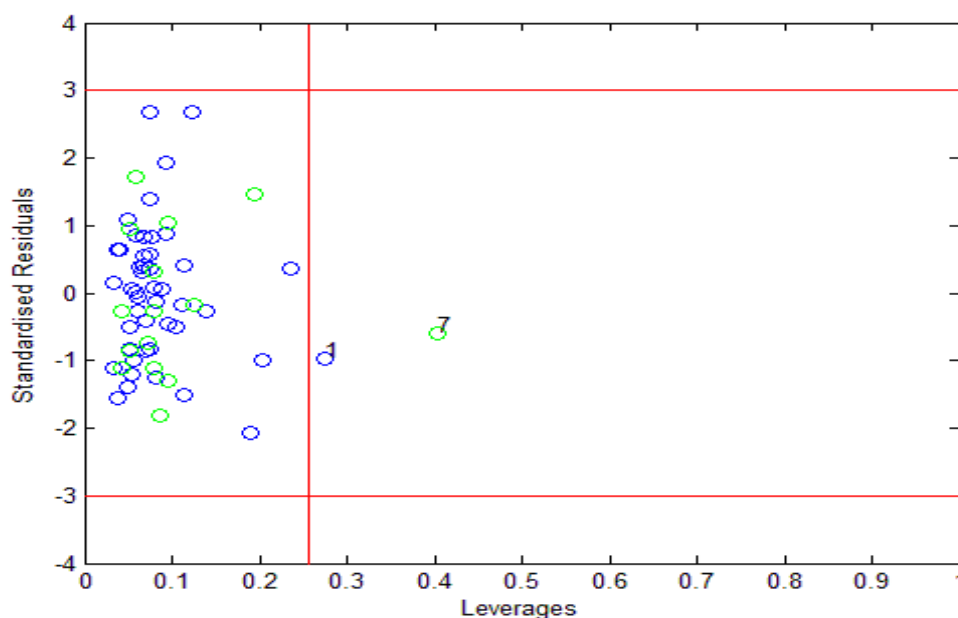
شماره	نام ترکیب	RI	ساختار ترکیب	شماره	نام ترکیب	RI	ساختار ترکیب
۴۵	α -Calacorene	1530		۵۴	7-Epi-eudesmol	1611	
۴۶	α -Cadinene	1532		۵۵	1-Epi-cubenol	1616	
۴۷	Germacrene-B	1542		۵۶	δ -9,10-Eremophil en-11-ol	1621	
۴۸	Spathulenol	1567		۵۷	T-Muurolol	1627	
۴۹	Caryophyllene oxide	1571		۵۸	α -Muurolol	1629	
۵۰	Globulol	1575		۵۹	T-Cadinol	1632	
۵۱	Viridiflorol	1583		۶۰	α -Cadinol	1639	
۵۲	Sesquithuriferol	1591		۶۱	α -Bisabolol	1667	
۵۳	Humulene epoxide II	1595		۶۲	Epi- α -bisabolol	1669	

۳-۲. محاسبه، انتخاب، تجزیه و تحلیل آماری توصیف‌کننده‌ها

پس از ترسیم و بهینه‌سازی ساختار مولکول‌ها به وسیله نرم‌افزار Hyperchem، ساختارهای رسم‌شده، به عنوان ورودی به نرم‌افزار Dragon، معرفی و محاسبه‌ی توصیف‌کننده‌ها توسط این نرم‌افزار صورت گرفت. برای ایجاد مدلی که بیان‌گر ارتباط ساختاری ترکیبات مورد بررسی با اندیس بازدارایی آنها باشد، از روش رگرسیون خطی چندگانه (MLR) استفاده شد. در این مدل‌سازی، هم تعدادی از این توصیف‌کننده‌ها که تمام مقادیر آنها صفر و یا دارای ۹۰٪ مقادیر یکسان بودند، حذف شدند. نهایتاً، پس از مرحله‌ی حذف مقدماتی اولیه، توصیف‌کننده‌های باقیمانده در یک ماتریس داده‌ی $n \times m$ جای گرفتند. در این ماتریس، جملات n و m ، به ترتیب بیان‌گر تعداد ترکیبات و تعداد توصیف‌کننده‌ها هستند. مقادیر عددی این دو جمله به ترتیب برابر ۶۲ و ۱۹۳ و شکل ماتریس مربوطه به صورت 62×192 می‌باشد. در پایان، بردار ستونی (Y) برای محاسبه‌ی متغیر وابسته (اندیس کوآتس)، ساخته شد.

۴-۲. الگوی پراکنش آنالیز جزء اصلی (PCA) و حوزه‌ی کاربرد (AD)

این الگو، با اعمال دستور متنی مربوطه متشکل از سه کلاستر به موازات یکدیگر می‌باشد. نهایتاً، با در نظر گرفتن الگوی پراکنش نقطه‌ای آنالیز مؤلفه‌ی اصلی، جهت تقسیم هوشمند سری داده به مجموعه‌های آموزش و آزمون ترکیباتی در سری آموزش جای گرفتند که نماینده‌ی مناسبی از کل ترکیبات در سری داده‌ی اصلی باشند. وظیفه‌ی مهم مدل‌سازی، بر عهده‌ی ترکیبات سری آموزش قرار داشته و ترکیبات سری آزمون در این امر دخالت ندارند. از طرف دیگر، قدرت پیش‌گوئی مدل و توانمندی آن با اعمال به سری آزمون مورد بررسی قرار می‌گیرد. در مدل‌سازی مبتنی بر رگرسیون خطی چندگانه روی اسانس گیاه داروئی *Pistacia atlantica Desf.*، سری اطلاعات اولیه‌ی متشکل از ۶۲ مولکول را به مجموعه‌های آموزشی شامل ۴۷ ترکیب و آزمایشی شامل ۱۵ ترکیب تقسیم می‌کنیم. این تقسیم‌بندی با نسبت تقریبی ۳ به ۱ صورت می‌پذیرد که در نهایت حدود ۷۵٪ از سری اطلاعات در مجموعه‌ی آموزشی و ۲۵٪ از آنها در مجموعه‌ی آزمایشی قرار می‌گیرند. بر اساس روال متعارف در اینگونه تحقیقات، نمودار ویلیام مربوطه ترسیم و نتایج مربوطه در شکل ۲، نمایش داده شده است. در این نمودار، ترکیب لورج و باقی‌مانده‌های استاندارد شده برای مشخص نمودن حوزه‌ی کاربرد مورد استفاده قرار گرفته است.



شکل ۲. نمودار ویلیام برای ترکیبات سری‌های آموزش و آزمون سری داده‌ی *Pistacia atlantica Desf.*

بررسی شکل ۲، خاطرنشان می‌سازد که دو ترکیب شماره ۱ و ۷ در سری آموزش و آزمون دارای مقادیر لورج بیشتر از مقدار هشدار (h^*) هستند. در این شکل، مقدار لورج هشدار برابر با $h^* = 0.2941$ است. بنابراین، می‌توانند جزو ترکیبات انحرافی باشند. خوشبختانه، در دو مورد اخیر مقادیر اطلاعات عددی پیشگویی شده مناسب و قابل قبول هستند. لذا، این دو ترکیب را لورج خوب نامیده و از حوزه‌ی کاربرد و سری داده‌ی اولیه حذف نمی‌کنیم. در حقیقت، نمودار حوزه‌ی کاربرد (نمودار ویلیام یا نمودار AD) هم مؤید مناسب بودن مدل ساخته‌شده و هم بیان‌گر تقسیم مناسب سری داده‌ی اولیه به مجموعه‌های آموزش و آزمون است.

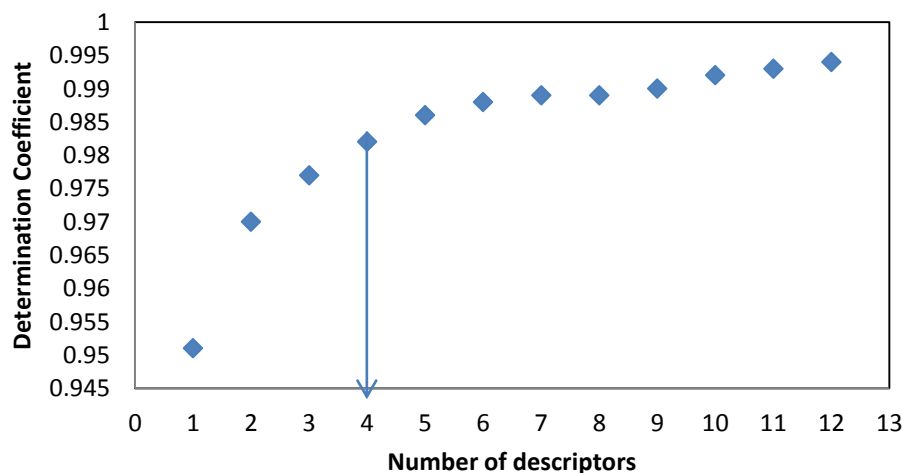
۲-۵. تحلیل مدل‌های آماری و انتخاب مدل مناسب

برای به‌دست آوردن مدلی مناسب نتایج عددی خلاصه‌ی مدل‌های ارائه شده در محیط SPSS با استفاده از روش انتخاب متغیر مرحله‌ای، در جدول ۲ نمایش داده شده است.

جدول ۲. خلاصه‌ی ۱۲ مدل منتخب پس از محاسبات رگرسیون در محیط نرم‌افزار SPSS سری داده‌ی *Pistacia atlantica* Desf.

شماره‌ی مدل	ضریب هم‌بستگی (r)	ضریب اندازه‌گیری (r^2)	ضریب اندازه‌گیری تنظیم شده ($Adj. r^2$)	خطای استاندارد تخمین
1	0.975	0.951	0.950	56.4
2	0.985	0.970	0.968	44.9
3	0.988	0.977	0.975	39.5
4	0.991	0.982	0.981	34.8
5	0.993	0.986	0.984	31.5
6	0.994	0.988	0.986	29.8
7	0.995	0.989	0.987	28.2
8	0.994	0.989	0.987	28.7
9	0.995	0.990	0.988	27.1
10	0.996	0.992	0.990	25.3
11	0.996	0.993	0.991	23.5
12	0.997	0.994	0.993	21.4

نهایتاً، با رسم روند تغییرات مقادیر ضرایب تعیین (R^2) بر حسب تعداد توصیف‌کننده‌ها (شکل ۳)، و تعیین نقطه‌ی شکست نمودار، تعداد سه توصیف‌کننده که نوید یک مدل بسیار پایدار و قابل تفسیر را می‌دهد، حاصل شد.



شکل ۳. منحنی تغییرات مقادیر ضرایب تعیین (R^2) بر حسب تعداد توصیف‌کننده‌ها با استفاده از نرم‌افزار SPSS سری داده‌ی *Pistacia atlantica* Desf.

در این مرحله، مدل انتخابی را برای مجموعه‌ی آزمایشی نیز اعمال می‌کنیم. از آنجایی که توصیف‌کننده‌ها باید متغیرهایی مستقل باشند، لذا برای اطمینان از عدم هم‌بستگی بین آنها، در نرم‌افزار SPSS، با استفاده از منوی آنالیز، گزینه‌ی هم‌بستگی دو متغیره را انتخاب می‌کنیم. چنانچه هم‌بستگی بین توصیف‌کننده‌ها بیشتر از ۰/۸۵ باشد، توصیف‌کننده‌های وابسته را از سری اطلاعات حذف کرده و مراحل مذکور را تا جایکه هم‌بستگی بین توصیف‌کننده‌ها کمتر از این مقدار باشد، مجدداً انجام می‌دهیم.

۲-۶. بررسی مدل خطی بهینه‌ی ساخته‌شده

در مجموعه‌ی توصیف‌کننده‌های انتخاب‌شده‌ی بهینه با روش‌های انتخاب متغیر مختلف، توصیف‌کننده‌ی Mor01m با بیشترین مقدار MF_j بیشترین تأثیر را نسبت به سایر متغیرها در مدل ارائه‌شده دارد. مقادیر عددی محاسبه‌شده‌ی اثر متوسط (MF_j)، در جدول ۳ آورده شده‌اند. بعلاوه، بازه‌ی تغییرات فاکتور VIF (بین ۱/۰۰۶ تا ۱/۰۴۴) هم کاملاً مطلوب و از نظر آماری قابل قبول می‌باشد. در یک جمع‌بندی، معادله‌ی به دست آمده برای داده‌های اندیس بازداري این سری از ترکیبات به صورت معادله‌ی ریاضی ۳، قابل بیان می‌باشد:

$$RI = 682.543(\pm 18.8) + 5.755(\pm 0.14) \text{ Mor01m} - 135.143(\pm 26.7) \text{ Mor18m} - 410.370(\pm 110.7) \text{ MATS2e} \quad (3)$$

اعداد داخل پرانتز، بیان‌گر خطای استاندارد (انحراف استاندارد) ضرایب و عرض از مبدأ واردشده در مدل خطی بر اساس رگرسیون خطی چندگانه هستند. طبق جدول ۳، هر کدام از توصیف‌کننده‌های وارد شده در مدل، یک خصوصیت ساختاری از مولکول را بیان می‌کنند که نشان‌دهنده‌ی تأثیر بسزای ساختار مولکول‌های مورد بررسی بر مقادیر اندیس کوآتس مربوطه هستند.

علامت مثبت نسبت داده‌شده به هر توصیف‌کننده که در جدول ۳ آورده شده است، دلالت بر تقویت مدل به وسیله‌ی آنها داشته در حالیکه علامت منفی بیان‌گر رابطه‌ی معکوس بین متغیر وابسته (اندیس کوآتس) و متغیرهای مستقل (توصیف‌کننده‌های مولکولی) دارد. بر این اساس، توصیف‌کننده‌ی Mor01m از گروه تری دی موریس دارای تأثیر سازنده و افزایشی بر مدل خطی ارائه‌شده می‌باشد. بدین ترتیب، با افزایش مقدار این توصیف‌کننده مقدار اندیس کوآتس به عنوان یک متغیر وابسته نیز افزایش می‌یابد. در عین حال، توصیف‌گرهای Mor18m و MATS2e به ترتیب از گروه‌های تری دی موریس و تودی اتوکرولیشن (خود ارتباطی دوبعدی) دارای علامت منفی هستند.

جدول ۳. مشخصات مدل منتخب توسط روش رگرسیون خطی چندگانه سری داده‌ی *Pistacia atlantica* Desf.

No.	Symbol	Descriptor description	Descriptor group	Coefficient	Mean effect	VIF
1	Intercept	-	-	682.543	-	-
2	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	5.755	1.022	1.006
3	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE	-135.143	0.031	1.039
4	MATS2e	Moran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities	2D autocorrelations	-410.370	-0.053	1.044

همچنین، جدول ۴ ماتریس هم‌بستگی توصیف‌کننده‌های وارد شده در مدل نهایی برای سری داده‌ها را نمایش می‌دهد. مقادیر عددی مندرج در این جدول، بیان‌گر ضرایب هم‌بستگی بین هر جفت توصیف‌کننده‌ی انتخاب شده در مدل خطی نهایی هستند. بر اساس نتایج عددی مندرج در جدول ۴، بیشترین مقدار ضریب هم‌بستگی در این ماتریس برابر با ۰.۱۹۰ و بین توصیف‌کننده‌های Mor18m و MATS2e می‌باشد. بازه‌ی عددی بسیار کم توصیف‌کننده‌های انتخابی دلالت بر اعتبار مدل ارائه شده و شاهده‌ی بر رفتار مستقل این متغیرها می‌باشد.

جدول ۴. ماتریس هم‌بستگی ۳ توصیف‌کننده‌ی وارد شده در مدل منتخب سری داده‌ی *Pistacia atlantica* Desf. به روش SW-MLR

	Mor01m	Mor18m	MATS2e
Mor01m	1		
Mor18m	-0.027	1	
MATS2e	0.069	0.190	1

در ادامه، هنگامیکه از عدم وجود هم‌بستگی بین توصیف‌کننده‌ها اطمینان حاصل شد، با استفاده از معادله‌ی ریاضی ۳ مربوط به مدل ارائه شده، ضرایب رگرسیون موجود در جدول ۳ و نیز مقادیر وارد شده‌ی توصیف‌کننده‌ها در مدل، اندیس‌های بازداری ترکیبات را توسط نرم‌افزار قدرتمند MATLAB محاسبه و پیش‌بینی می‌کنیم. نتایج حاصل شده در این محاسبه در جدول ۵ نشان داده شده است.

۲-۷. نمودارهای پیش‌گوئی و باقی‌مانده

یکی از مواردی که می‌توان با استناد به آن از معتبر بودن مدل انتخابی اطمینان خاطر پیدا کرد، به دست آوردن ضریب اندازه‌گیری (R^2) بین مقادیر تجربی یک کمیت و مقادیر نظیر پیش‌گوئی شده برای آن می‌باشد. بر اساس یک قاعده‌ی کلی، هر قدر مقدار ضریب اندازه‌گیری در نمودار پیش‌گوئی به یک نزدیک‌تر باشد، مدل ساخته شده معتبرتر خواهد بود. جهت ارزیابی شمائی اعتبار مدل‌های انتخاب شده، مقادیر پیش‌بینی شده‌ی اندیس‌های بازداری را برحسب مقادیر تجربی مربوطه (شکل ۴) رسم می‌کنیم. بر اساس این شکل، یک هم‌گرایی مطلوب بین مقادیر تجربی و مقادیر پیش‌گوئی شده به چشم می‌خورد. در این شکل، مقدار ضریب اندازه‌گیری برای مجموعه‌های آموزش و آزمون به ترتیب معادل ۰/۹۶۹۴ و ۰/۹۴۶۱ است که دلالت بر اعتبار مدل و توانمندی بالای آن جهت پیش‌گوئی مقادیر اندیس کوتاس گستره‌ی وسیعی از ترکیبات دارد. مورد دیگری که می‌توان برای اعتبار مدل به آن استناد کرد، رسم مقادیر باقی‌مانده حاصل از اختلاف مقادیر پیش‌بینی شده‌ی اندیس‌های بازداری برحسب مقادیر تجربی آن می‌باشد. این نمودار، موسوم به نمودار باقی‌مانده در شکل ۵ آورده شده است. در نمودار باقی‌مانده، می‌توان پراکنش نرمال نقاط را در دو طرف محور X، معیاری از عدم وجود خطای معین در روش و اعتبار بالای مدل دانست.

۲-۸. اعتبارسنجی تقاطعی و آزمون Y تصادفی مدل خطی ارائه شده

در مدل‌سازی انجام شده به روش انتخاب مرحله‌ای با استفاده از روش اعتبارسنجی تقاطعی (CV)، مقدار Q^2 برای روش‌های حذف تکی (LOO) و حذف گروهی (LGO)، به ترتیب برابر ۰/۹۷۲ و ۰/۹۶۴ است. این مقادیر، نشان‌گر اعتبار تقاطعی مدل ارائه شده هستند. همچنین، نتایج آزمون Y تصادفی و عدم وجود هرگونه رابطه‌ی منطقی بین توصیف‌کننده‌ها و متغیر وابسته در جدول ۶، پس از ۱۰ بار تکرار آورده شده‌اند. بنابراین، انتظار می‌رود که با استفاده از این رویکرد پس از چندین بار تکرار، مدل‌های ساخته شده دارای مقادیر کم و قابل اغماض R^2 و Q^2 باشند.

جدول ۵. مقادیر تجربی و پیش‌بینی‌شده‌ی اندیس‌های بازداری (RI)، اسانس گیاه *Pistacia atlantica* Desf. توسط روش رگرسین خطی چندگانه‌ی مرحله‌ای

No.	Compound name	K.I. (Lit.) ^a	K.I. (Cal.) ^b	Dif. ^c	R.E. ^d %
Training set					
1	Tricyclene	920	887.3	-32.7	-3.6
2	α -Pinene	932	965.3	33.3	3.6
3	Camphene	944	934.2	-9.8	-1
4	β -Pinene	972	984.1	12.1	1.2
5	Myrcene	980	1052.2	72.2	7.4
6	α -Phellandrene	997	1049.7	52.7	5.3
7	α -Terpinene	1009	1040.5	31.5	3.1
8	p-Cymene	1011	1026.1	15.1	1.5
9	Limonene	1021	1042.1	21.1	2.1
10	(Z)- β -Ocimene	1024	1040	16	1.6
11	(E)- β -Ocimene	1036	1050	14	1.4
12	γ -Terpinene	1048	1043.2	-4.8	-0.5
13	Linalool	1082	1106.6	24.6	2.3
14	α -Fenchol	1098	1091.8	-6.2	-0.6
15	trans-Pinocarveol	1124	1107.3	-16.7	-1.5
16	Pinocarpone	1137	1083.5	-53.5	-4.7
17	Borneol	1148	1091.5	-56.5	-4.9
18	Terpinene-4-ol	1162	1130.4	-31.6	-2.7
19	α -Terpineol	1172	1111.4	-60.6	-5.2
20	Myrtenol	1178	1143	-35	-3
21	Bornyl acetate	1267	1279.9	12.9	1
22	δ -Elemene	1334	1435.7	101.7	7.6
23	α -Copaene	1377	1418.7	41.7	3
24	β -Bourbonene	1384	1406	22	1.6
25	β -Caryophyllene	1420	1445.3	25.3	1.8

26	β -Gurjunene	1428	1412.3	-15.7	-1.1
27	Aromadendrene	1439	1444.6	5.6	0.4
28	Alloaromadendrene	1459	1411.7	-47.3	-3.2
29	γ -Muurolene	1471	1503.4	32.4	2.2
30	Germacrene-D	1477	1479.4	2.4	0.2
31	Bicyclogermacrene	1492	1448.9	-43.1	-2.9
32	α -Muurolene	1494	1494.5	0.5	0
33	γ -Cadinene	1507	1504.8	-2.2	-0.1
34	α -Calacorene	1530	1456.4	-73.6	-4.8
35	α -Cadinene	1532	1493.8	-38.2	-2.5
36	Germacrene-B	1542	1496.1	-45.9	-3
37	Caryophyllene oxide	1571	1670.3	99.3	6.3
38	Globulol	1575	1577.7	2.7	0.2
39	Viridiflorol	1583	1564.2	-18.8	-1.2
40	Humulene epoxide II	1595	1627.1	32.1	2
41	7-Epi-eudesmol	1611	1613.4	2.4	0.1
42	1-Epi-cubenol	1616	1630.4	14.4	0.9
43	T-Muurolol	1627	1616.8	-10.2	-0.6
44	α -Muurolol	1629	1609.6	-19.4	-1.2
45	T-Cadinol	1632	1630	-2	-0.1
46	α -Bisabolol	1667	1634.5	-32.5	-1.9
47	Epi- α -bisabolol	1669	1637.4	-31.6	-1.9
Test set					
1	Sabinene	965	977.1	12.1	1.3
2	δ -3-Carene	1004	976	-28	-2.8
3	β -Phellandrene	1021	1060	39	3.8
4	Terpinolene	1078	1029.6	-48.4	-4.5

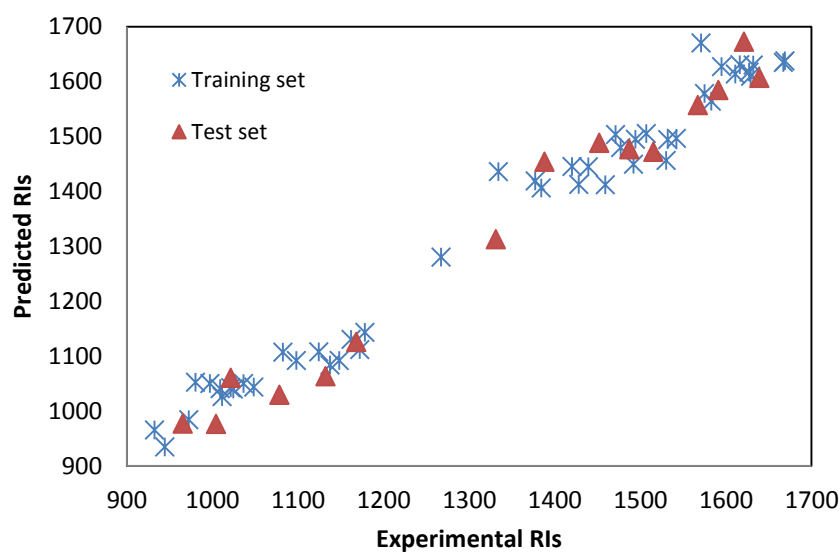
5	Camphene hydrate	1132	1063.1	-68.9	-6.1
6	cis-Pinocarveol	1168	1125.7	-42.3	-3.6
7	α -Terpinyl acetate	1331	1312.6	-18.4	-1.4
8	β -Elemene	1388	1453.7	65.7	4.7
9	α -Humulene	1452	1488.4	36.4	2.5
10	β -Selinene	1487	1476.7	-10.3	-0.7
11	δ -Cadinene	1515	1471.7	-43.3	-2.9
12	Spathulenol	1567	1556.8	-10.2	-0.7
13	Sesquithuriferol	1591	1584.6	-6.4	-0.4
14	δ -9,10-Eremophilen-11-ol	1621	1672.6	51.6	3.2
15	α -Cadinol	1639	1606.5	-32.5	-2

A : Retention index in literature: اندیس کواتس در منابع علمی:

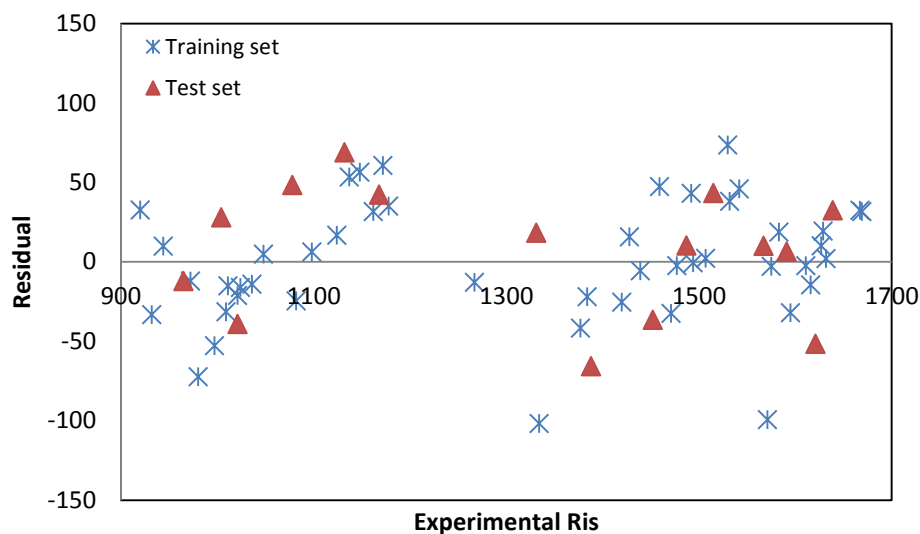
B: Retention index calculated: اندیس کواتس محاسبه شده:

C: Difference : تفاوت:

D: Relative Error: خطای نسبی:



شکل ۴. مقادیر پیش‌بینی‌شده‌ی اندیس‌های بازداری ترکیبات اسانس گیاه *Pistacia atlantica Desf.* برحسب مقادیر تجربی آن به روش SW-MLR



شکل ۵. مقادیر باقی مانده حاصل از اختلاف مقادیر پیش‌بینی شده‌ی اندیس‌های بازداری ترکیبات اسانس گیاه *Pistacia atlantica* Desf. بر حسب مقادیر تجربی آن به روش SW-MLR

جدول ۶. پارامترهای آماری شاخص حاصل از انجام آزمون Y تصادفی به روش SW-MLR

الف) PRESS	ب) SPRESS	ج) SST	د) R^2CV	ه) PRESS/SST	و) REP
3075162.9863	267.4237	401378.2342	0.0124	7.6615	233407.0706
3086581.6825	267.9197	352262.4416	0.0078	8.7622	234273.7579
3165668.4909	271.3304	259132.4216	0.0000	12.2164	240276.5032
3518810.0873	286.0643	214481.2522	0.0598	16.4061	267080.2030
3333643.8189	278.4360	194541.1884	0.0224	17.1359	253025.9508
3212520.6613	273.3309	256072.2700	0.0007	12.5454	243832.6164
3026969.3826	265.3199	166429.9676	0.0015	18.1876	229749.1416
4135318.6858	310.1132	244800.2829	0.3352	16.8926	313873.6467
3880612.7821	300.4111	180082.7306	0.2966	21.5491	294541.2864
3562451.9719	287.8328	190071.9903	0.0960	18.7426	270392.6533

الف) مجموع مربعات باقی مانده‌ی پیش‌بینی (ب) عدم قطعیت مجموع مربعات باقی مانده‌ی پیش‌بینی (ج) مجموع مربعات کل (د) مجذور ضریب هم‌بستگی اعتبار تقاطعی (ه) مجموع مربعات باقی مانده‌ی پیش‌بینی تقسیم بر مجموع مربعات کل (و) خطای نسبی پیش‌بینی

۳. نتایج و بحث

۳-۱. نتایج محاسبات و مدل‌های خطی حاصل از سری داده‌ی *Pistacia atlantica* Desf. به روش SW-MLR

در این بخش، مقادیر تجربی و پیش‌بینی شده‌ی اندیس‌های بازداری را به‌طور جداگانه در یک فایل متنی قرار داده و با استفاده از یک برنامه‌ی جامع در نرم‌افزار MATLAB به همراه نتایج حاصل از نرم‌افزار SPSS، پارامترهای آماری شاخص مربوطه را محاسبه و نتایج حاصل در جدول ۷ آمده است.

جدول ۷. پارامترهای آماری جامع با اعمال روش رگرسیون خطی چندگانه‌ی مرحله‌ای برای مجموعه‌های آموزش و آزمون سری داده‌ی *Pistacia atlantica* Desf. به روش

SW-MLR			
شماره	سری منتخب	پارامتر آماری	مقدار عددی
۱	سری آموزش	R^2 (الف)	0.977
	سری آزمون		0.976
۲	سری آموزش	Adjusted R^2 (ب)	0.975
۳	سری آموزش	REP (%) (ج)	2.9
	سری آزمون		3.0
۴	سری آموزش	RMSEP (د)	37.8
	سری آزمون		39.3
۵	سری آموزش	SEP (ه)	38.2
	سری آزمون		40.7
۶	سری آموزش	RSEP (و)	2.8
	سری آزمون		2.9
۷	سری آموزش	MAE (ز)	78.9
	سری آزمون		151.1
۸	سری آموزش	F (ح)	606.8
	سری آزمون		145.5

الف) ضریب اندازه‌گیری، ب) ضریب اندازه‌گیری تنظیم‌شده، ج) درصد خطای نسبی پیش‌گویی، د) جذر میانگین مجذور خطای پیش‌گویی، ه) خطای استاندارد پیش‌گویی، و) خطای استاندارد نسبی پیش‌گویی، ز) خطای مطلق میانگین، ح) آماره‌ی F

۲-۳. نتایج محاسبات و مدل‌های خطی حاصل از سری داده‌ی *Pistacia atlantica* Desf. به روش GA-MLR

پس از انجام محاسبات با استفاده از الگوریتم توانمند GA-MLR، با تغییر متوالی در پارامترهای مربوطه در دستورهای متنی در محیط نرم‌افزار متلب، مجموعه‌های شامل ۳ تا ۷ توصیف‌کننده‌ی مولکولی حاصل شد. در جداول متوالی ۸ تا ۱۰، مشخصات و جزئیات کامل این مدل‌ها و در جداول ۱۱ تا ۱۵ فهرست کاملی از پارامترهای آماری مربوطه آورده شده است.

جدول ۸. فهرست توصیف‌کننده‌های منتخب و برخی از مشخصات آماری آنها در مدل‌های خطی سری داده‌ی *Pistacia atlantica* Desf. با استفاده از روش GA-MLR

Model number	Descriptor Name(s)	R^2		Q^2	
		Training set	Test set	Q^2_{LOO}	Q^2_{LGO}
1	-	-	-	-	-
2	-	-	-	-	-
3	MATS1p; Mor01m; Mor18m	0.9723	0.9704	0.967	0.958
4	MATS6e; MATS1p; Mor01m; Mor29m	0.9671	0.9682	0.959	0.950
5	SRW08; ATS1v; MATS1p; Mor01m; Mor18m	0.973	0.9686	0.964	0.953
6	MATS1p; Mor13u; Mor01m; Mor18m; H6p; R5e	0.9739	0.9699	0.963	0.952
7	MATS1p; Mor01m; Mor18m; Mor21m; Mor25p; Vm; HATS2v	0.9776	0.9641	0.968	0.953

(JQCS)

جدول ۹. مشخصات مدل‌های منتخب سری داده *Pistacia atlantica* Desf. به روش GA-MLR

Linear Model 1 (3 molecular descriptors)						
No.	Symbol	Descriptor description	Descriptor group	Coefficient	Mean effect	VIF
1	Intercept	-	-	645.8892	-	-
2	MATS1p	Moran autocorrelation - lag 1 / weighted by atomic polarizabilities	2D autocorrelations	225.7802	0.014	1.820
3	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	5.6266	0.945	1.131
4	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE	-193.3676	0.042	1.710
Linear Model 2 (4 molecular descriptors)						
1	Intercept	-	-	701.9639	-	-
2	MATS6e	Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities	2D autocorrelations	86.6420	-0.005	1.102
3	MATS1p	Moran autocorrelation - lag 1 / weighted by atomic polarizabilities	2D autocorrelations	-100.5831	-0.007	1.183
4	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	5.7891	1.061	1.178
5	Mor29m	3D-MoRSE - signal 29 / weighted by atomic masses	3D-MoRSE	220.9469	-0.049	1.167
Linear Model 3 (5 molecular descriptors)						
1	Intercept	-	-	745.8663	-	-
2	SRW08	self-returning walk count of order 08	molecular walk counts	-0.0009	-0.009	2.066
3	ATS1v	Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic van der Waals volumes	2D autocorrelations	-162.8574	-0.169	1.114
4	MATS1p	Moran autocorrelation - lag 1 / weighted by atomic polarizabilities	2D autocorrelations	261.2837	0.018	2.109
5	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	5.6431	1.113	1.181
6	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE	-183.9368	0.047	2.039
Linear Model 4 (6 molecular descriptors)						
1	Intercept	-	-	582.5359	-	-
2	MATS1p	Moran autocorrelation - lag 1 / weighted by atomic polarizabilities	2D autocorrelations	109.1306	0.006	2.459
3	Mor13u	3D-MoRSE - signal 13 / unweighted	3D-MoRSE	-7.7792	-0.015	2.503
4	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	5.5793	0.856	6.099
5	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE	-208.4041	0.041	3.234
6	H6p	H autocorrelation of lag 6 / weighted by atomic polarizabilities	GETAWAY	-80.9091	-0.007	2.880
7	R5e	R autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities	GETAWAY	45.6813	0.120	3.690
Linear Model 5 (7 molecular descriptors)						
1	Intercept	-	-	673.3378	-	-
2	MATS1p	Moran autocorrelation - lag 1 / weighted by atomic polarizabilities	2D autocorrelations	189.6975	0.012	2.644
3	Mor01m	3D-MoRSE - signal 01 / weighted by atomic masses	3D-MoRSE	4.9306	0.863	13.621
4	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses	3D-MoRSE	-172.9809	0.039	1.960
5	Mor21m	3D-MoRSE - signal 21 / weighted by atomic masses	3D-MoRSE	-184.0501	0.097	2.115
6	Mor25p	3D-MoRSE - signal 25 / weighted by atomic polarizabilities	3D-MoRSE	-12.3479	-0.004	1.220
7	Vm	V total size index / weighted by atomic masses	WHIM	1.5574	0.058	18.732
8	HATS2v	leverage-weighted autocorrelation of lag 2 / weighted by atomic van der Waals volumes	GETAWAY	-412.9801	-0.065	7.359

جدول ۱۰. فهرست مدل‌های خطی نهائی در سری داده‌ی *Pistacia atlantica* Desf. به روش GA-MLR

Model number	Linear Model Form
1	$RI = 645.8892 (\pm 18.6) + 225.7802 (\pm 140.5) \times MATS1p + 5.6266 (\pm 0.2) \times Mor01m - 193.3676 (\pm 36.8) \times Mor18m$
2	$RI = 701.9639 (\pm 21.5) - 86.6420 (\pm 34.7) \times MATS6e - 100.5831 (\pm 124.8) \times MATS1p + 5.7891 (\pm 0.2) \times Mor01m + 220.9469 (\pm 65.1) \times Mor29m$
3	$RI = 745.8663 (\pm 159.9) - 0.0009 (\pm 0.002) \times SRW08 - 162.8574 (\pm 277.0) \times ATS1v + 261.2837 (\pm 153.4) \times MATS1p + 5.6431 (\pm 0.2) \times Mor01m - 183.9368 (\pm 40.7) \times Mor18m$
4	$RI = 582.5359 (\pm 61.3) + 109.1306 (\pm 164.1) \times MATS1p - 7.7792 (\pm 25.8) \times Mor13u + 5.5793 (\pm 0.4) \times Mor01m - 208.4041 (\pm 50.8) \times Mor18m - 80.9091 (\pm 183.6) \times H6p + 45.6813 (\pm 33.4) \times R5e$
5	$RI = 673.3378 (\pm 112.8) + 189.6975 (\pm 159.9) \times MATS1p + 4.9306 (\pm 0.5) \times Mor01m - 172.9809 (\pm 37.2) \times Mor18m - 184.0501 (\pm 64.5) \times Mor21m - 12.3479 (\pm 62.7) \times Mor25p + 1.5574 (\pm 2.0) \times Vm - 412.9801 (\pm 1122.3) \times HATS2v$

جدول ۱۱. پارامترهای آماری جامع برای مجموعه‌های آموزش و آزمون در سری داده *Pistacia atlantica* Desf. به روش GA-MLR در مدل شامل ۳ توصیف‌کننده

مقدار عددی	پارامتر آماری	سری منتخب	شماره
0.9723	R^2	سری آموزش	1
0.9704		سری آزمون	
0.970	Adjusted R^2	سری آموزش	2
3.1	REP (%)	سری آموزش	3
3.1		سری آزمون	
41.4	RMSEP	سری آموزش	4
41.7		سری آزمون	
41.9	SEP	سری آموزش	5
43.2		سری آزمون	
3.1	RSEP	سری آموزش	6
3.1		سری آزمون	
83.8	MAE	سری آموزش	7
150.6		سری آزمون	
502.5	F	سری آموزش	8
120.9		سری آزمون	

جدول ۱۲. پارامترهای آماری جامع برای مجموعه‌های آموزش و آزمون در سری داده *Pistacia atlantica* Desf. به روش GA-MLR در مدل شامل ۴ توصیف‌کننده

شماره	سری منتخب	پارامتر آماری	مقدار عددی
1	سری آموزش	R^2	0.9671
	سری آزمون		0.9682
2	سری آموزش	Adjusted R^2	0.964
3	سری آموزش	REP (%)	3.4
	سری آزمون		3.2
4	سری آموزش	RMSEP	45.1
	سری آزمون		43.2
5	سری آموزش	SEP	45.6
	سری آزمون		44.7
6	سری آموزش	RSEP	3.4
	سری آزمون		3.2
7	سری آموزش	MAE	85.4
	سری آزمون		154.8
8	سری آموزش	F	308.6
	سری آزمون		77.8

جدول ۱۳. پارامترهای آماری جامع برای مجموعه‌های آموزش و آزمون در سری داده *Pistacia atlantica* Desf. به روش GA-MLR در مدل شامل ۵ توصیف‌کننده

شماره	سری منتخب	پارامتر آماری	مقدار عددی
1	سری آموزش	R^2	0.973
	سری آزمون		0.9686
2	سری آموزش	Adjusted R^2	0.969
3	سری آموزش	REP (%)	2.8
	سری آزمون		3.2
4	سری آموزش	RMSEP	37.3
	سری آزمون		43.0
5	سری آموزش	SEP	37.7
	سری آزمون		44.5
6	سری آموزش	RSEP	2.8
	سری آزمون		3.2
7	سری آموزش	MAE	79.4
	سری آزمون		153.1
8	سری آموزش	F	292.2
	سری آزمون		56.3

جدول ۱۴. پارامترهای آماری جامع برای مجموعه‌های آموزش و آزمون در سری داده *Pistacia atlantica* Desf. به روش GA-MLR در مدل شامل ۶ توصیف کننده

شماره	سری منتخب	پارامتر آماری	مقدار عددی
1	سری آموزش	R^2	0.9739
	سری آزمون		0.9699
2	سری آموزش	Adjusted R^2	0.970
3	سری آموزش	REP (%)	3.0
	سری آزمون		3.2
4	سری آموزش	RMSEP	40.2
	سری آزمون		42.4
5	سری آموزش	SEP	40.6
	سری آزمون		43.9
6	سری آموزش	RSEP	3.0
	سری آزمون		3.1
7	سری آموزش	MAE	82.2
	سری آزمون		152.5
8	سری آموزش	F	249.2
	سری آزمون		44.1

جدول ۱۵. پارامترهای آماری جامع برای مجموعه‌های آموزش و آزمون در سری داده *Pistacia atlantica* Desf. به روش GA-MLR در مدل شامل ۷ توصیف کننده

شماره	سری منتخب	پارامتر آماری	مقدار عددی
1	سری آموزش	R^2	0.9776
	سری آزمون		0.9641
2	سری آموزش	Adjusted R^2	0.974
3	سری آموزش	REP (%)	2.8
	سری آزمون		3.4
4	سری آموزش	RMSEP	37.3
	سری آزمون		45.2
5	سری آموزش	SEP	37.7
	سری آزمون		46.8
6	سری آموزش	RSEP	2.8
	سری آزمون		3.3
7	سری آموزش	MAE	79.4
	سری آزمون		163.0
8	سری آموزش	F	242.7
	سری آزمون		27.4

۴. نتیجه گیری

در تحقیق اخیر، ارتباط کمی بین بازداري و ساختار ترکیباتی از روغن های اسانسی با استفاده از روش انتخاب متغیر به صورت مرحله ای- رگرسیون خطی چندگانه (SW-MLR) و الگوریتم ژنتیک- رگرسیون خطی چندگانه (GA-MLR) جهت پیش بینی شاخص بازداري کوتاس ترکیبات طبیعی موجود در روغن اسانسی گیاه *Pistacia atlantica* با موفقیت مورد مطالعه قرار گرفته است. در گام نخست، مجموعه ی داده ها شامل شاخص بازداري ۶۲ ترکیب از مولکول های روغن های اسانسی از مقاله ی ارائه شده توسط گروهی از محققین الجزایری انتخاب و سپس توصیف کننده های مولکولی با استفاده از روش های فوق الذکر انتخاب شدند. پیشگویی کمی با مدل QSRR برای ۱۵ ترکیب از ۶۲ ترکیب که به صورت تصادفی انتخاب شده بودند، مورد ارزیابی قرار گرفت. مدل SW-MLR با ۴ توصیف کننده و مدل GA-MLR در قالب مدل های ۳، ۴، ۵، ۶ و ۷ متغیره ساخته شد. پارامترهای آماری نشان می دهد که انواع مدل های ارائه شده، توانایی پیش بینی شاخص بازداري مولکول ها را با قدرت و توانایی بالا دارد. در عین حال، در این مطالعه انعطاف پذیری و صحت مدل های مبتنی بر روش GA-MLR در مجموع به طور قابل ملاحظه ای از مدل مبتنی بر روش SW-MLR بیشتر است.

۵. مراجع

- [1] <http://chemistryiran.parsiblog.com/1602908.htm>.
- [2] F. Keshvarzinejad, Department of Chemistry, Islamic Azad University, Shahrood Branch, Shahrood, Iran, (2009).
- [3] <http://www.nano.ir>.
- [4] P.C. Jurs *Computer Software Applications in Chemistry*, John Wiley & Sons New York, (1996).
- [5] M. Mohammadhosseini, O. Deeb, A. Alavi-Gharabagh and M. Nekoei, *Anal. Chem. Lett.*, 2 (2012) 80.
- [6] M. Mohammadhosseini, H.A. Zamani, H. Akhlaghi and M. Nekoei, *J. Essent. Oil-Bear. Plants*, 14 (2011) 559.
- [7] M. Nekoei and M. Mohammadhosseini, *Anal. Chem. Lett.*, 4 (2014) 93.
- [8] D. Skoog, F.J. Holler and T.A. Nieman *Principles of Instrumental Analysis, Fifth Ed.*, (1998).
- [9] R.A. Mayer *Encyclopedia of Analytical Chemistry*, WILEY, (2000).
- [10] D.L. Massart, B.G.M. Vandeginste, S.N. Deming and L. Kaufman *Chemometrics: A Textbook*, Elsevier, Amsterdam, (1998).
- [11] S. Wold, *Chemometr. Intell. Lab. Syst.*, 30 (1995) 109.
- [12] D.G. Howery and R.F. Hirsch, *J. Chem. Educ.*, 60 (1983) 656.
- [13] <http://www.emsl.pnl.gov:2080/docs/hncinic/homepage.html>.
- [14] G.G. Andersson and P. Kaufmann, *Chemometr. Intell. Lab. Syst.*, 50 (2000) 101.
- [15] J. Zupan and J. Gasteiger, *Anal. Chim. Acta*, 248 (1991) 1.
- [16] D.B. Hibert, *Chemometr. Intell. Lab. Syst.*, 19 (1993) 277.
- [17] N. Goudarzi, M. Goodarzi, M.M. Mohammadhosseini and M. Nekoei, *Mol. Phys.*, 107 (2009) 1739.
- [18] J. Hosseini, M. Nekoei, M. Mohammadhosseini and N. Goudarzi, *J. Appl. Res. Chem.*, 5 (2011) 5.
- [19] A. Jalali, M. Nekoei and M. Mohammadhosseini, *J. Chem. Health Risks*, 6 (2016) 49.
- [20] M. Mohammadhosseini, *Asian J. Chem.*, 24 (2012) 3814.
- [21] M. Nekoei, M. Mohammadhosseini and E. Pourbasheer, *Med. Chem. Res.*, 24 (2015) 3037.
- [22] M. Nekoei, M. Mohammadhosseini, E. Pourbasheer and M. Maham, *Semnan J. Appl. Chem.*, 6 (2012) 35.
- [23] M. Nekoei, M. Salimi, M. Dolatabadi and M. Mohammadhosseini, *Monatsh. Chem.*, 142 (2011) 943.
- [24] S. Mecherara-Idjeri, A. Hassani, V. Castola and J. Casanova, *J. Essent. Oil Res.*, 20 (2008) 215.