



پیش بینی شاخص بازداری ترکیبات تشکیل دهنده گیاه *Ammoides Atlantica* با استفاده از روش مطالعه ارتباط کمی ساختار - بازداری (QSRR)

سعید نکویی^{۱*}، سیدعباس طاهری^۲

^۱دانشکده شیمی، دانشگاه صنعتی شاهرود، شاهرود، ایران

^۲گروه شیمی، دانشکده علوم پایه، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۹/۰۶/۱۴، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۹/۰۸/۲۵، تاریخ پذیرش قطعی: ۱۳۹۹/۰۹/۰۴

چکیده

روغن های ضروری گیاهی و ترکیبات مرتبط، کاربرد گسترده ای در طب سنتی، صنایع غذایی، صنایع عطرسازی و داروسازی دارند. در این تحقیق، مطالعه ارتباط کمی ساختار - ویژگی بر روی شاخص بازداری ۳۲ ترکیب از روغن ضروری گیاه *Ammoides Atlantica* انجام پذیرفت. از الگوریتم ژنتیک برای انتخاب توصیف کننده ها استفاده شد. برای مدل سازی اندیس بازداری این ترکیبات از روش رگرسیون خطی چندگانه (MLR) به عنوان روش خطی استفاده گردید. برای بررسی اعتبار این مدل ها علاوه بر به کارگیری سری تست، از روش های دیگری همانند حذف مرحله ای تک تک داده ها و Y تصادفی استفاده گردید. ضرایب همبستگی برای سری تست توسط مدل MLR برابر ۰/۸۹۴ و درصد خطای نسبی برابر ۵/۵۷ بدست آمد. داده های آماری نشان می دهد که روش GA-MLR پیش بینی های قابل قبولی را ارائه نموده است.

واژه های کلیدی: ارتباط کمی ساختار - بازداری، الگوریتم ژنتیک، رگرسیون خطی چندگانه، گیاه *Ammoides Atlantica*.

۱. مقدمه

روغنهای ضروری که روغنهای فرار یا اتری نیز نامیده میشوند، مایعات روغنی معطری هستند که به روشهای تخمیر، استخراج و همینطور تقطیر با بخار آب که در تولید تجاری از همه رایجتر است (از مواد خام گیاهی) گلها، جوانهها، دانهها، برگها، چوب، میوه و ریشهها به دست می آیند. مشخصه مشترک اکثر این ترکیبات این است که دارای بوی تند و زننده هستند [۱]. یک روغن

*عهده دار مکاتبات: سعید نکویی

نشانی: دانشکده شیمی، دانشگاه صنعتی شاهرود، شاهرود، ایران

پست الکترونیک: E-mail: S_nekoei68@yahoo.com

تلفن: ۰۲۳۲۳۹۴۲۸۹

از این جهت اتری نام می‌گیرد که حاوی یک رایحه مشخص یا "اسانس" از یک گیاه معین است. خواص آنتی اکسیدانی، ضد میکروبی، ضد ویروسی و همینطور ضدانگلی این ترکیبات شناخته شده است. این روغن‌ها به عنوان ادویه در خورده‌نیها و آشامیدنیها، در مصارف بهداشتی نظیر صابون‌ها و شوینده‌ها، در صنعت داروسازی و همچنین مواد آرایشی و عطری کاربرد دارند. امروزه با پیشرفت بهداشت عمومی و چالش صنایع غذایی برای سالم‌سازی هرچه بیشتر فرآورده‌های غذایی، استفاده از این ترکیبات به عنوان جایگزینی برای مواد افزودنی سنتزی و همینطور نمک‌ها که میتوانند احتمال بیماریهای قلبی را افزایش دهند پیشنهاد شده است [۵-۲].

با توجه به کاربرد گسترده‌ی روغنهای ضروری، اندازه‌گیری آنها امری اجتناب ناپذیر است؛ ولی در عین حال، به دلیل فرار بودن و وجود مقادیر کمی از هر یک از آنها در نمونه‌های گیاهی این اندازه‌گیری‌ها با مشکلاتی نیز همراه می‌باشد. یکی از راه‌های کسب اطلاعات کیفی و کمی راجع به این ترکیبات که اکثراً ایزومر بوده و دارای خواص فیزیکی و شیمیایی مشابهی هستند، اندازه‌گیری اندیس بازداری آنها است که با روشهای کروماتوگرافی قابل اندازه‌گیری می‌باشد [۱۰-۶]. این تکنیک‌ها به دلیل وجود تنوع زیاد ستونهای کروماتوگرافی و پیشرفتهای پدید آمده در این زمینه، توانایی بسیاری در حل مسائل مربوط به جداسازی و شناسایی ترکیبات دارند. اما با وجود این مزایا، در مورد اندازه‌گیری ترکیباتی که تهیه یا آماده‌سازی آنها مشکل بوده و یا شامل مواد ناپایدار هستند دارای محدودیت می‌باشند. وجود این محدودیت‌ها در روشهای آزمایشگاهی باعث شد که در کنار روشهای عملی، شیوه‌های تئوری نیز برای استفاده از نتایج تجربی توسعه یابند. مطالعات کموتریکس با استفاده از انجام محاسبات آماری و ریاضی، راهی برای استخراج حداکثر اطلاعات از نتایج تجربی می‌باشد [۱۵-۱۱]. مطالعات کمی روابط ساختار-خاصیت (QSPR) یکی از روشهای کموتریکس است که با استفاده از آن می‌توان ارتباط خطی یا غیرخطی بین ویژگیهای ساختاری و خواص ترکیبات (مانند اندیس بازداری) برقرار نموده و خاصیت مورد نظر را برای آن ترکیبات پیش‌بینی کرد [۲۰-۱۶]. شاخص بازداری معیاری کمی است که میزان بازداری نسبی هر یک از اجزای نمونه را نسبت به آلکانهای نرمال (هیدروکربنها) بر روی یک فاز ساکن معین و در یک دمای مشخص نشان می‌دهد. شاخص بازداری، متغیرهای مختلف سیستم کروماتوگرافی را نرمالیزه می‌کند، به گونه‌ای که امکان مقایسه نتایج سیستمهای مختلف فراهم میشود. شاخص بازداری زمانی مفید است که پیش‌بینی ترتیب خروج نسبی اجزاء نمونه از ستون مورد نظر باشد. اگر از شاخص بازداری با چنین منظوری استفاده شود، باید در نظر داشت که شاخص بازداری در یک شرایط دمایی معین و برای یک ستون خاص قابل کاربرد است. از شاخص بازداری، برای مقایسه بازداری و گزینش‌پذیری ستونهای مشابه (ستونهایی که از لحاظ جنس و ابعاد لوله موئین، نوع و ضخامت فاز ساکن و سایر موارد کاملاً مشابه اند) نیز می‌توان استفاده کرد [۲۱]. هدف از این تحقیق، ایجاد رابطه‌ی کمی بین ساختارهای مولکولی و شاخص بازداری ترکیب روغنهای ضروری گیاه *Ammoides Atlantica* با به‌کارگیری روشهای QSRR است. سپس با استفاده از این مدل، امکان پیش‌بینی شاخص بازداری برای ترکیباتی با ساختار مشابه وجود خواهد داشت.

۲. روش‌های محاسباتی

۱-۲. انتخاب سری داده‌ها

سری داده‌ها که برای مطالعه ارتباط کمی ساختار-بازداری انتخاب گردید، مربوط به اندیس کوتاس ۳۲ ترکیب از روغن‌های استحصال شده از گیاه *Ammoides Atlantica* می‌باشد [۲۲]. در این کار این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری تست تقسیم شدند (جدول ۱)، سری آموزش شامل ۲۲ مولکول و سری تست شامل ۱۰ مولکول می‌باشد. مقادیر شاخص بازداری (اندیس کوتاس) به عنوان متغیر وابسته و توصیف کننده‌ها به عنوان متغیر مستقل انتخاب شد. سری آموزش جهت ایجاد یک مدل مناسب و سری تست جهت ارزیابی مدل مورد استفاده قرار گرفت.

۲-۲. محاسبه توصیف کننده‌ها

برای محاسبه توصیف گرهای نظری ابتدا ساختارهای مولکولی به کمک نرم افزار Hyper Chem 7 رسم شدند [۲۳]. شکل و پیکربندی مولکولی نقش بسیار مهمی در پیش بینی و توصیف خاصیت یا فعالیت های بیولوژیکی بازی می کنند. بنابراین ابتدا ساختارهای مولکولی به وسیله الگوریتم AM1 بهینه شدند. سپس خروجی نرم افزار Hyper Chem برای هر ترکیب به برنامه Dragon منتقل شده و توصیف گرهای محاسبه شدند [۲۴-۲۷]. به این ترتیب تعداد ۱۴۹۷ توصیف گر مولکولی برای هر ترکیب محاسبه شدند.

۳-۲. کاهش تعداد توصیف کننده‌ها

با توجه به این که بعضی از متغیرهای مستقل (توصیف کننده‌ها) ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می دهند، لذا به روش زیر بعضی از متغیرها حذف شدند.

۱) توصیف کننده هایی که مقادیر ثابت و یا تقریباً ثابت دارند (بیش از ۹۰٪ داده های ثابت دارند)، حذف می شوند. در این مرحله تعداد ۴۱۲ توصیف کننده کنار گذاشته شدند. بدین ترتیب ۱۰۸۵ توصیف کننده باقی ماند.

۲) با توجه به اینکه در برخی از موارد بعضی از متغیرهای مستقل با یکدیگر همبستگی بالایی دارند، و وجود تنها یکی از این متغیرها در مدل کردن کافی بوده و نیازی به حضور بقیه متغیرها نمی باشد. لذا داده ها از این نظر مورد بررسی قرار گرفتند. به منظور حذف متغیرهایی که همبستگی خطی بزرگتر از ۰/۹ با یکدیگر دارند، ماتریس 1085×1085 بین همه متغیرهای مستقل و نیز متغیر وابسته تشکیل شد. متغیرهای مستقلی که همبستگی بیشتر از ۰/۹ با متغیرهای دیگر دارند، بایستی حذف شوند. اما بایستی دقت داشت که هنگام حذف متغیر اضافی، متغیری کنار گذاشته شود که همبستگی کمتری با متغیر وابسته داشته باشد. بدین ترتیب تعداد ۶۵۶ توصیف کننده اضافی کنار گذاشته شدند و تعداد ۴۲۹ توصیف کننده باقی ماندند.

جدول ۱. مقادیر تجربی و محاسبه شده اندیس کواتس برای ترکیبات مختلف برای مجموعه های آموزش و پیش بینی با استفاده از GA-MLR همراه با

مقادیر درصد خطا

No.	Compound	RI (Exp) ^a	RI (GA-MLR) ^b	E % ^c
Training set				
1	α -Thujene	931	993.31	6.69
2	α -Pinene	941	967.82	2.85
3	Sabinene	975	976.75	0.18
4	β -Pinene	980	968.41	-1.18
5	α -Phellandrene	1007	1051.05	4.37
6	δ -3-Carene	1013	998.54	-3.38
7	<i>p</i> -Cymene	1026	1112.58	8.43
8	Limonene	1030	995.12	-3.38
9	γ -Terpinene	1064	1010.16	-5.05
10	Cis-Sabinene hydrate	1069	989.20	-7.46
11	linalool	1099	1178.61	7.24
12	Cis-p-menth-2-en 1-ol	1123	1193.25	6.25
13	α -Terpineol	1186	1118.92	-5.65
14	Methyl thymol	1236	1290.56	4.41
15	Thymol	1292	1265.62	-2.04
16	Carvacrol	1297	1257.97	-3.00
17	β -Caryophyllene	1420	1394.21	-1.81
18	α -Amorphene	1486	1490.01	0.26
19	γ -Cadinene	1516	1399.89	-7.65
20	δ -Cadinene	1527	1517.06	-0.65
21	Germacrene B	1560	1572.98	0.83
22	Caryophyllene oxide	1582	1617.89	2.26
Prediction set				
1	Camphene	954	1039.2	8.93
2	Myrcene	993	997.8	0.48
3	α -Terpinene	1020	1012.2	-0.76
4	1,8-Cineole	1032	1082.3	4.88
5	Terpinolene	1088	985	-9.45
6	Terpinen-4-ol	1177	1119.5	-4.87
7	Methyl carvacrol	1244	1272.5	2.29
8	α -Copaene	1379	1382.8	0.28
9	α -Murolene	1501	1482.9	-1.19
10	Selina-3,7(11)-diene	1545	1406.2	-8.98

(a) مقادیر تجربی (b) مقادیر پیش بینی (c) درصد خطا

۲-۴. انتخاب توصیف کننده‌های مناسب

در این کار، از الگوریتم ژنتیک^۱ برای انتخاب توصیف کننده های مناسب استفاده گردید. نرم افزار مورد استفاده برای این کار تحت MATLAB نوشته شده است. ابتدا یک سری از توصیف کننده ها به صورت شانسی به عنوان جمعیت یا کروموزوم اولیه در نظر گرفته می شود و به صورت یک رشته با کد گذاری دوتایی (۰ و ۱) مشخص می گردد. تعداد کروموزوم های اولیه توسط کاربر تعیین می شود. با افزایش تعداد اعضاء جمعیت، زمان محاسبه نیز افزایش می یابد. برای یافتن مناسب ترین تعداد جمعیت اولیه عمل بهینه سازی با ۱۰ بار تکرار برای هر جمعیت اولیه صورت گرفته و کمترین ارزش محاسبه شده برای میانگین ۱۰ بار تکرار، بدست آمد. پس از ایجاد جمعیت اولیه کلیه کروموزوم ها (راه حل ها) ارزیابی می گردد. برای این کار هر راه حل اولیه برای ساخت مدل به کار گرفته می شود. راه حل های بهتر، راه حل هایی هستند که مدل بهتری ساخته و خطای کمتری در پیش بینی سری حذف شده ایجاد کنند [۲۸-۳۰].

۳. نتایج و بحث

همانطور که گفته شد برای انتخاب منسب ترین توصیف کننده ها از الگوریتم ژنتیک استفاده گردید. جدول ۲ پارامترهای بهینه شده الگوریتم ژنتیک را نشان می دهد. با این روش از بین ۴۲۹ توصیف کننده باقی مانده تعداد ۳ توصیف کننده به عنوان مناسب ترین آنها انتخاب و توسط روش MLR مدل سازی شد و مورد ارزیابی قرار گرفت.

جدول ۲. پارامترهای بهینه شده برای الگوریتم ژنتیک

مشخصات الگوریتم ژنتیک	
اندازه جمعیت ^۲	۱۰۰
(%) جمعیت اولیه ^۳	۱۰
حداکثر تعداد نسل ^۴	۱۰۰
نرخ جهش ^۵	۰/۰۵
همگذری	دوتایی

لیست توصیف کننده های انتخاب شده توسط الگوریتم ژنتیک به همراه توصیف مختصری از آنها در جدول ۳ آورده شده است.

¹ Genetic Algorithm

² Population size

³ Initial terms

⁴ Max Generation

⁵ Mutation rate

جدول ۳. توصیف کننده های انتخاب شده با الگوریتم ژنتیک و توصیف آنها

شماره	علامت توصیف کننده	مختصری در مورد توصیف کننده
1	<i>RDF035m</i>	Radial distribution function – 3.5/weighted by atomic mass
2	<i>PCD</i>	Difference of multiple path counts to path counts
3	<i>Q2</i>	Total squared charge

۳-۱. ارزیابی توصیف کننده های انتخاب شده

به منظور ارزیابی توصیف کننده های انتخاب شده مبنی بر مستقل بودن از همدیگر در جدول ۴ ضرایب همبستگی توصیف کننده های انتخاب شده نسبت به یکدیگر آورده شده است. همانطور که در این جدول مشاهده می شود بیشترین ضریب همبستگی بین توصیف کننده *Q2* و توصیف کننده *RDF035m* با مقدار ضریب همبستگی ۰/۵۳۲ می باشد این نتایج نشان می دهد که بین توصیف کننده های انتخاب شده همبستگی بالایی وجود نداشته و توصیف کننده ها تقریباً مستقل از هم هستند و نتایج بدست آمده از مدلسازی دال بر وابستگی توصیف کننده ها نمی باشد.

جدول ۴. ماتریس ضرایب همبستگی توصیف کننده های انتخاب شده

	<i>PCD</i>	<i>Q2</i>	<i>RDF035m</i>
<i>PCD</i>	1		
<i>Q2</i>	0.259	1	
<i>RDF035m</i>	0.108	0.532	1

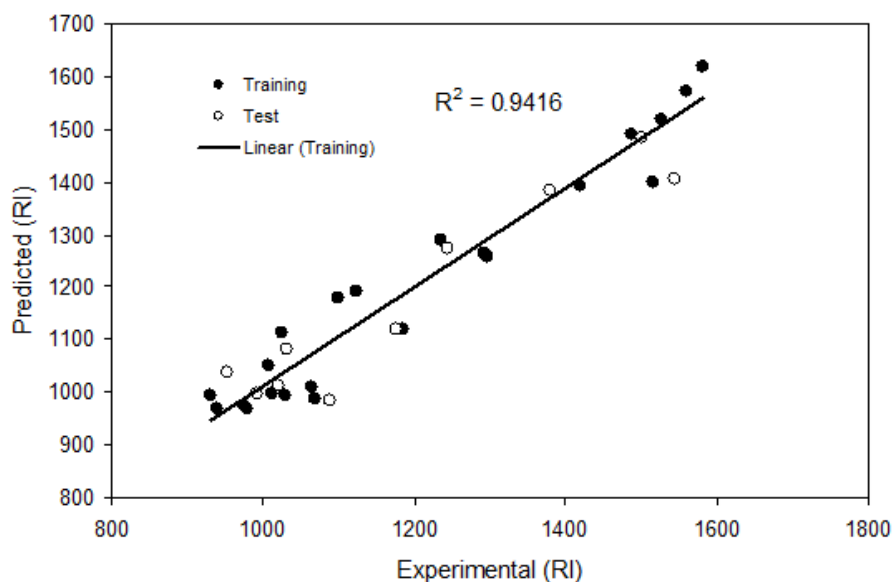
۳-۲. ایجاد مدل با استفاده از GA-MLR

پس از انتخاب مناسب ترین توصیف کننده ها توسط الگوریتم ژنتیک مرحله بعدی ایجاد مدل، میان توصیف کننده های انتخاب شده و اندیس کواتس ترکیبات می باشد. که با استفاده از روش MLR رابطه زیر بدست آمد:

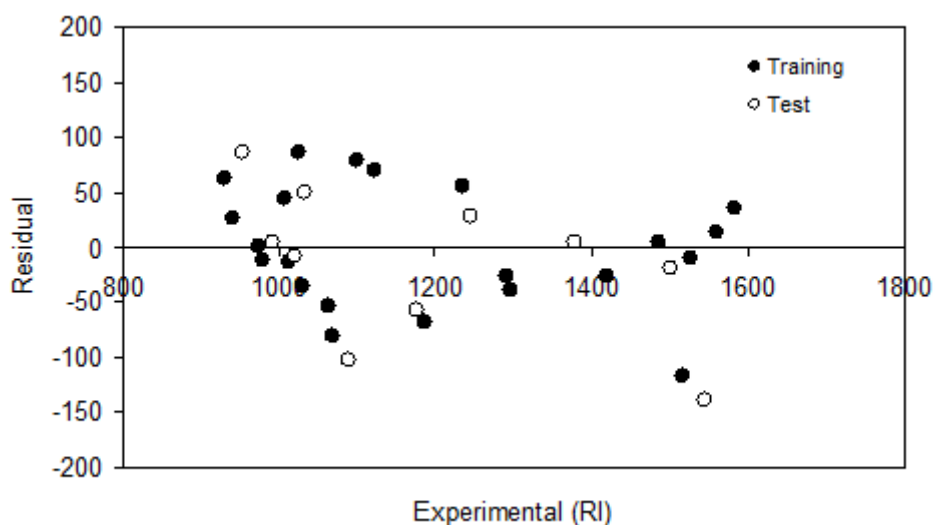
$$RI = 590.81 + 20.56 (RDF035m) + 107.41 (PCD) + 839.37 (Q2)$$

سپس از معادله بدست آمده برای پیش بینی اندیس کواتس سری تست استفاده گردید.

مقادیر واقعی و پیش بینی شده اندیس کواتس و همچنین خطاهای پیش بینی برای کلیه ترکیبات مجموعه آموزش و تست در جدول ۱ آورده شده است. همچنین مقادیر اندیس کواتس محاسبه شده و تجربی برای ترکیبات براساس مدل GA-MLR در دو مجموعه آموزشی و تست در شکل ۱ آورده شده است. در این شکل، میزان نزدیکی داده ها به خط راست قدرت پیش بینی مدل را نشان می دهد. شکل ۲ مقادیر باقیمانده خطاها را نسبت به مقادیر تجربی نشان می دهد. میزان پراکندگی خطاها در اطراف خط صفر نشان دهنده این است که خطای سیستماتیک در مدل وجود ندارد.



شکل ۱. نمودار مقادیر اندیس کوتاس محاسبه شده با کمک مدل GA-MLR برای مجموعه‌های آموزشی و پیش‌بینی بر حسب مقادیر تجربی



شکل ۲. نمودار تغییرات خطا برای مقادیر اندیس کوتاس محاسبه شده توسط مدل GA-MLR برای مجموعه‌های آموزشی و پیش‌بینی

۳-۳. ارزیابی اعتبار مدل‌های انتخاب‌شده

یکی از مواردی که می‌توان با استناد به آن از معتبر بودن مدل انتخابی اطمینان پیدا کرد، به‌دست آوردن ضریب تعیین (R^2) می‌باشد که هرچه این مقدار به یک نزدیک‌تر باشد مدل ما معتبرتر خواهد بود. مقدار ضریب تعیین، برابر 0.9416 حاکی از اعتبار مدل می‌باشد. مورد دیگری که می‌توان برای اعتبار مدل به آن استناد کرد رسم مقادیر باقی‌مانده‌ی حاصل از اختلاف مقادیر پیش‌بینی

شده‌ی اندیس‌های بازداري بر حسب مقادير تجربي آن مي‌باشد که شکل ۲ آن را به‌خوبي نشان مي‌دهد. در نمودار فوق مي‌توان پراکندگي نسبتاً يکسان نقاط را حول مقادير صفر، ناشي از عدم وجود خطاي معين در روش و اعتبار مدل دانست. درصد خطاي نسبي براي سري پيش بيني (REP) نيز جهت بررسي اعتبار مدل مورد استفاده قرار مي‌گيرد. هر چه مقدار REP کوچکتر باشد مدل، مدل مناسب‌تری خواهد بود. مقدار اين آماره برابر ۵/۵۷٪ مي‌باشد که مقدار نسبتاً کوچک آن اعتبار مدل را نشان مي‌دهد.

يکي ديگر از روش‌ها اعتبارسنجي تقاطعي است که يکي از رايج‌ترين تکنیک‌ها جهت اعتبارسنجي مدل محسوب مي‌شود. در اين روش هر بار يک (LOO) يا تعداد معينی (LGO) از مولکولها کنار گذاشته مي‌شوند و مدل با مولکول‌های باقی مانده ساخته مي‌شود، سپس کميت مدل‌سازي براي مولکول‌های کنار گذاشته شده، توسط مدل حاصله پيش بيني مي‌شود. اين عمل به صورت چرخه‌اي براي تمام مولکول‌ها تکرار مي‌شود به طوري که همه مولکولها يک بار در سري پيش بيني قرار مي‌گيرند. پارامتر آماری Q^2 براي ارزيابي مدل حاصل از ارزيابي تقاطعي بکار مي‌رود. مقدار Q^2_{LOO} برابر ۰/۹۴۱۲ و مقدار Q^2_{LGO} برابر ۰/۹۳۲ نشان دهنده اعتبار مدل مي‌باشد.

و در نهايت براي اينکه نشان دهيم مدل ارايه شده يک مدل شانس‌ي نيست از روش Y-randomization استفاده گرديد. در اين روش تمام انديس‌های کوتاس، بطور تصادفي بهم ريخته و سپس مدل‌سازي و پيش بيني انجام ميشود. مسلماً بايد ضريب همبستگي بسيار پايين بدست آيد که اگر اينطور شد نشان دهنده اين است که مدل اوليه شانس‌ي نبوده و داراي اعتبار مي‌باشد. نتايج ۱۰ بار تکرار در جدول (۷-۱۸) آورده شده است. همانطور که مشاهده مي‌شود در همه تکرارها ضريب همبستگي بسيار پايين مي‌باشد که نشان دهنده اعتبار مدل ارايه شده و شانس‌ي نبودن مدل مي‌باشد.

جدول ۵. مقادير R^2 و Q^2 براي ۱۰ بار Y-randomization

Iteration	R^2	Q^2
1	0.1568	0.0031
2	0.1217	0.0013
3	0.2186	0.0122
4	0.0729	0.0096
5	0.1354	0.0197
6	0.0097	0.2751
7	0.1877	0.0026
8	0.3235	0.1193
9	0.0814	0.0506
10	0.2099	0.0059

۳-۴. پارامترهای آماری

جدول ۶ پارامترهای آماری مختلف را برای مدل GA-MLR نشان می دهد. به عنوان مثال مقادیر R^2 برای مدل برای سری آموزش و پیش بینی به ترتیب عبارتند از: ۰/۹۴۱۶ و ۰/۸۹۴ و همچنین درصد خطای نسبی پیشگویی برای اندیس کوتاس محاسبه شده برای مجموعه پیش بینی با کمک این مدل برابر ۵/۵۷٪ می باشند. همانگونه که مشاهده می شود مدل GA-MLR توانسته است فقط با ۳ توصیف کننده مدل مناسبی ارائه نماید که می توان از این مدل، برای محاسبه اندیس کوتاس سایر ترکیبات مجهول استفاده نمود.

جدول ۶. مقایسه پارامترهای آماری مختلف برای مدل GA-MLR

Model	$R^2_{\text{calibration}}$	$R^2_{\text{prediction}}$	REP(%)	F	RMSEP
GA-MLR	0.9416	0.894	5.57	646.57	66.473

۴. نتیجه گیری

در این تحقیق، الگوریتم ژنتیک به عنوان یک روش موفق در انتخاب مناسب ترین توصیف کننده ها به کار گرفته شد. سپس برای مدل سازی از روش رگرسیون خطی چندگانه استفاده گردید. نتایج نشان داد که پیش بینی اندیس بازداري ترکیبات با این روش با ضریب همبستگی بالا و خطای کم قابل انجام است.

۶. مراجع

- [1] Nazzaro, F., Fratianni, F., De Martino, L., Coppola, R., & De Feo, V. (2013). Effect of essential oils on pathogenic bacteria. *Pharmaceuticals*, 6(12), 1451-1474.
- [2] Romeo, F. V., De Luca, S., Piscopo, A., & Poiana, M. (2008). Antimicrobial effect of some essential oils. *Journal of Essential Oil Research*, 20(4), 373-379.
- [3] Zhang, N., & Yao, L. (2019). Anxiolytic effect of essential oils and their constituents: a review. *Journal of agricultural and food chemistry*, 67(50), 13790-13808.
- [4] Heghes, S. C., Vostinaru, O., Rus, L. M., Mogosan, C., Iuga, C. A., & Filip, L. (2019). Antispasmodic effect of essential oils and their constituents: A review. *Molecules*, 24(9), 1675.
- [5] Cinteza, L. O., Scmoroscenco, C., Voicu, S. N., Nistor, C. L., Nitu, S. G., Trica, B., ... & Petcu, C. (2018). Chitosan-stabilized Ag nanoparticles with superior biocompatibility and their synergistic antibacterial effect in mixtures with essential oils. *Nanomaterials*, 8(10), 826.
- [6] Gonzalez, F. R., & Nardillo, A. M. (1999). Retention index in temperature-programmed gas chromatography. *Journal of Chromatography A*, 842(1-2), 29-49.
- [7] Stein, S. E., Babushok, V. I., Brown, R. L., & Linstrom, P. J. (2007). Estimation of Kovats retention indices using group contributions. *Journal of chemical information and modeling*, 47(3), 975-980.
- [8] Tan, P., Wei, X., Zhang, H., Niu, M., & Zhang, D. (2019). A practical method for rapid screening and quantitative analysis of 130 pesticide residues in herbal medicines based on the Kovats retention

- index principle: an exemplary study using *Panacis quinquefolii radix*. *Analytical Methods*, 11(38), 4868-4874.
- [9] Melo, T. O., Marques, F. A., & Hansel, F. A. (2017). Pyrolysis-gas chromatography–mass spectrometry Kováts retention index of pyrolysis products of lignocellulosic materials. *Journal of Analytical and Applied Pyrolysis*, 126, 332-336.
- [10] Aparkin, A. M., & Pashinin, V. A. (2021). Linear Correlation between Kovats Retention Indices I and the Sum of 13 C Nuclear Magnetic Resonance Chemical Shifts in the Structural Isomers of Saturated Hydrocarbons. *Russian Journal of Physical Chemistry A*, 95(1), 101-105.
- [11] Chapman, J., Truong, V. K., Elbourne, A., Gangadoo, S., Cheeseman, S., Rajapaksha, P., ... & Cozzolino, D. (2020). Combining chemometrics and sensors: toward new applications in monitoring and environmental analysis. *Chemical Reviews*, 120(13), 6048-6069.
- [12] Brereton, R. G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., ... & Tauler, R. (2018). Chemometrics in analytical chemistry—part II: modeling, validation, and applications. *Analytical and bioanalytical chemistry*, 410(26), 6691-6704.
- [13] Brereton, R. G., Jansen, J., Lopes, J., Marini, F., Pomerantsev, A., Rodionova, O., ... & Tauler, R. (2017). Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. *Analytical and Bioanalytical Chemistry*, 409(25), 5891-5899.
- [14] Zhang, T., Tang, H., & Li, H. (2018). Chemometrics in laser-induced breakdown spectroscopy. *Journal of Chemometrics*, 32(11), 2983.
- [15] Bos, T. S., Knol, W. C., Molenaar, S. R., Niezen, L. E., Schoenmakers, P. J., Somsen, G. W., & Pirok, B. W. (2020). Recent applications of chemometrics in one-and two-dimensional chromatography. *Journal of separation science*, 43(9-10), 1678-1727.
- [16] Villaverde, J. J., Sevilla-Morán, B., López-Goti, C., Alonso-Prados, J. L., & Sandín-España, P. (2018). Considerations of nano-QSAR/QSPR models for nanopesticide risk assessment within the European legislative framework. *Science of the Total Environment*, 634, 1530-1539.
- [17] Li, M., Yu, H., Wang, Y., Li, J., Ma, G., & Wei, X. (2020). QSPR models for predicting the adsorption capacity for microplastics of polyethylene, polypropylene and polystyrene. *Scientific reports*, 10(1), 1-11.
- [18] Sepehri, B. (2020). A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view. *Journal of Molecular Liquids*, 297, 112013.
- [19] Sepehri, B. (2020). A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view. *Journal of Molecular Liquids*, 297, 112013.
- [20] Borhani, T. N., García-Muñoz, S., Luciani, C. V., Galindo, A., & Adjiman, C. S. (2019). Hybrid QSPR models for the prediction of the free energy of solvation of organic solute/solvent pairs. *Physical Chemistry Chemical Physics*, 21(25), 13706-13720.
- [21] Vrzal, T., Malečková, M., & Olšovská, J. (2021). DeepReI: Deep learning-based gas chromatographic retention index predictor. *Analytica Chimica Acta*, 1147, 64-71.
- [22] Aćimović, M., Ivanović, S., Simić, K., Pezo, L., Zeremski, T., Ovuka, J., & Sikora, V. (2021). Chemical characterization of marrubium vulgare volatiles from Serbia. *Plants*, 10(3), 600.
- [23] Laxmi, D., & Priyadarshy, S. (2002). HyperChem 6.03. *Biotech Software & Internet Report: The Computer Software Journal for Scientists*, 3(1), 5-9.

- [24] Marleau, G., Hébert, A., & Roy, R. (2011). A user guide for DRAGON Version 4. *Institute of Genius Nuclear, Department of Genius Mechanical, School Polytechnic of Montreal*.
- [25] Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006). Dragon software: An easy approach to molecular descriptor calculations. *Match*, 56(2), 237-248.
- [26] Wang, Y., Chen, J., Tang, W., Xia, D., Liang, Y., & Li, X. (2019). Modeling adsorption of organic pollutants onto single-walled carbon nanotubes with theoretical molecular descriptors using MLR and SVM algorithms. *Chemosphere*, 214, 79-84.
- [27] Réti, T., Sharafđini, R., Dregelyi-Kiss, A., & Haghbin, H. (2018). Graph irregularity indices used as molecular descriptors in QSPR studies. *MATCH Commun. Math. Comput. Chem*, 79, 509-524.

Prediction of Retention Index of *Ammoides Atlantica* Compounds Using Quantitative Structure-Retention Relationship Study (QSRR)

Saeed Nekoei^{1*}, S. Abbas Taheri²

¹Faculty of Chemistry, Shahrood University of Technology, Shahrood, Iran

²Department of Chemistry, Faculty of Science, Shahrood Branch, Islamic Azad University, Shahrood, Iran

Submitted: 04 September 2020, Revised: 15 November 2020, Accepted: 24 November 2020

Abstract

Essential oils and related compounds are widely used in traditional medicine, food industry, perfumery and pharmaceutical industries. In this study, a quantitative structure-retention relationship study was performed on the inhibition index of 32 compounds of essential oil of *Ammoides Atlantica*. Genetic algorithm was used to select the descriptors. Multiple linear regression (MLR) was used as a linear method to model the retention index of these compounds. In order to check the validity of these models, in addition to using the test series, other methods such as LOO and LGO and Y-randomization were used. Correlation coefficients for the test series were obtained by MLR model equal to 0.894 and relative error percentage equal to 5.57. Statistical data show that the GA-MLR method has provided acceptable predictions.

Keywords: *Quantitative structure-retention relationship, genetic algorithm, multiple linear regression, Ammoides Atlantica*

*Corresponding author : Saeed Nekoei

Address: Faculty of Chemistry, Shahrood University of Technology, Shahrood, Iran

Tel: 02332394289

E-mail: S_nekoei68@yahoo.com