



مطالعه ارتباط کمی ساختار- بازداري (QSRR) مقادير اندیس کواتس ترکیبات تشکیل دهنده گیاه *Nepeta macrosiphon Boiss*

اعظم وفايي*^۱

اگره شیمی، واحد گچساران، دانشگاه آزاد اسلامی، گچساران، ایران

تاریخ ثبت اولیه: ۱۴۰۰/۰۶/۰۴، تاریخ دریافت نسخه اصلاح شده: ۱۴۰۰/۰۸/۲۰، تاریخ پذیرش قطعی: ۱۴۰۰/۰۹/۱۱

چکیده

مطالعات ارتباط کمی ساختار- بازداري (QSRR) بر روی شاخص بازداري (اندیس کواتس) ترکیبات تشکیل دهنده گیاه *Nepeta macrosiphon* انجام گرفت. از روش برازش مرحله ای برای انتخاب توصیف کننده های مناسب استفاده شد. توصیف کننده های انتخاب شده از این روش برای مدل سازی و پیش بینی شاخص بازداري این ترکیبات توسط روش رگرسیون خطی چندگانه (MLR) استفاده گردید. به منظور بررسی اعتبار مدل از روش های مختلفی مانند به کارگیری سری تست و رد مرحله ای تک تک و گروهی داده ها استفاده گردید. نتایج به دست آمده بیانگر توانایی خوب روش SW-MLR برای پیش بینی شاخص بازداري می باشد. ضرایب تعیین سری آموزش و تست به ترتیب ۰/۹۹۱ و ۰/۹۸۲ بود.

واژه های کلیدی: ارتباط کمی ساختار- بازداري، اندیس کواتس، رگرسیون خطی چند گانه، *Nepeta macrosiphon Boiss*

۱. مقدمه

امروزه به طور وسیع از روش های آماری برای بررسی چگونگی تغییرات خواص یا فعالیت، به صورت تابعی از توصیف کننده هایی که نشان دهنده ساختار شیمیایی مولکول ها هستند به کار گرفته می شود. به این روش ها مطالعات کمی ساختار خاصیت یا فعالیت می گویند [۱-۵]. نتایج این مطالعات علاوه بر شفاف سازی نحوه ارتباط بین خواص مولکولها و ویژگیهای ساختاری آنها، به پژوهشگران در پیش بینی رفتار مولکولهای جدید بر اساس رفتار مولکولهای مشابه کمک میکند. به مجموعه ابزارها و روشهایی که

*عهده دار مکاتبات: اعظم وفايي

نشانی: گروه شیمی، واحد گچساران، دانشگاه آزاد اسلامی، گچساران، ایران

پست الکترونیک: E-mail: a.vafaei11@yahoo.com

تلفن: ۰۲۳۳۲۳۹۴۲۸۹

به این منظور مورد استفاده قرار میگیرند، روشهای پارامتری میگویند. در روشهای پارامتری سعی میشود بین یک سری توصیف کننده های مولکولی با فعالیت یا خاصیت مورد نظر ارتباط منطقی برقرار شود. توصیف کننده های مولکولی که به این منظور استفاده میشوند، مقادیر عددی می باشند که جنبه های مختلف ساختاری مولکول را بطور کمی نشان میدهند. استفاده از چنین روش هایی برای پیش بینی خصوصیات ترکیبات مختلف، مانع از اتلاف وقت و هزینه خواهد شد. ایجاد ارتباط بین ویژگی یک ترکیب و ساختار آن با استفاده از چند توصیف کننده مناسب امکان پذیر است. به طوری که با محاسبه چند توصیف کننده به راحتی می توان خصوصیت یا فعالیت بیولوژیکی یک ترکیب مجهول را پیش بینی کرد [۱۰-۶].

شاخص بازداري معياري کمي است که ميزان بازداري نسبي هر یک از اجزای نمونه را نسبت به آلکانهای نرمال (هیدروکربنها) بر روی یک فاز ساکن و در یک دمای مشخص نشان میدهد. شاخص بازداري، متغیرهای مختلف سیستم کروماتوگرافی را نرمالیزه میکند، به گونه ای که امکان مقایسه نتایج سیستمهای مختلف فراهم می شود. شاخص بازداري n -آلکان ها را می توان به صورت تعداد اتم های کربن $\times 100$ (به عنوان مثال، هگزان = ۶۰۰) تعریف نمود. شاخص بازداري (RI) ماده مورد تجزیه با استفاده از معادله ۱ محاسبه می گردد که در این معادله t_{Ra} ، t_{Rn} و t_{Rn+1} به ترتیب عبارتند از: زمان های بازداري تنظیم شده ماده مورد تجزیه، آلکان های حاوی n و $n+1$ کربن که بلافاصله قبل و بعد از ماده مورد تجزیه از ستون خارج می شوند [۱۳-۱۱].

$$RI = n \times 100 + 100 \frac{\log t'_{Ra} - \log t'_{Rn}}{\log t'_{Rn+1} - \log t'_{Rn}} \quad \text{معادله (۱)}$$

شاخص بازداري زمانی مفید است که پیش بینی ترتیب خروج نسبی اجزاء نمونه از ستون، مورد نظر باشد. اگر از شاخص بازداري با چنین منظوری استفاده شود، باید در نظر داشت که شاخص بازداري در یک شرایط معین و برای یک ستون خاص قابل کاربرد است. از شاخص بازداري، برای مقایسه بازداري و گزینش پذیری ستون های مشابه (ستون هایی که از لحاظ جنس و ابعاد لوله موئین، نوع و ضخامت فاز ساکن و سایر موارد کاملاً مشابه اند) نیز میتوان استفاده کرد [۱۵-۱۴].

هدف از این تحقیق، مطالعه و بررسی ارتباط کمی ساختار ترکیبات با شاخص بازداري آنها در یکسری از ترکیبات استحصال شده از گیاه *Nepeta macrosiphon* با استفاده از روش انتخاب مرحله ای به عنوان روش انتخاب متغیر و روش رگرسیون خطی چندگانه (MLR) می باشد.

۲. بخش محاسباتی

۲-۱. انتخاب سری داده ها

سری داده ها که برای مطالعه ارتباط کمی ساختار-بازداري انتخاب گردید، مربوط به اندیس کوآتس ۵۴ ترکیب از روغن های اسانسی به دست آمده از گیاه *Nepeta macrosiphon* می باشد که در آزمایشگاه تخصصی شیمی تجزیه دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران استخراج شده است. در این کار این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری تست

تقسیم شدند، سری آموزش شامل ۴۳ مولکول و سری تست شامل ۱۱ مولکول می‌باشند. مقادیر شاخص بازداری (اندیس کوتاس) به عنوان متغیر وابسته و توصیف کننده‌ها به عنوان متغیر مستقل انتخاب شدند. سری آموزش جهت ایجاد یک مدل مناسب و سری تست جهت ارزیابی مدل مورد استفاده قرار گرفت.

۲-۲. محاسبه توصیف گر‌ها

در اولین قدم باید مجموعه آزمایشی (اندیس کوتاس ترکیبات) و ساختارهای متناظر جمع آوری شده و به صورت قابل پردازش برای کامپیوتر درآید. در مورد داده‌های آزمایشی که مقادیر عددی می‌باشند مشکلی نداریم، مساله مهم بیان ساختارهای شیمیایی به شکل قابل پذیرش برای کامپیوتر می‌باشد. برای محاسبه توصیف گرهای نظری ابتدا ساختارهای مولکولی به کمک نرم افزار هایپرکم (Hyperchem) رسم می‌شوند. شکل‌ها و پیکربندی‌های مولکولی نقش بسیار مهمی در پیش بینی و توصیف خاصیت یا فعالیت‌های بیولوژیکی بازی می‌کنند. بنابراین ساختارهای مولکولی به وسیله الگوریتم AM1 بهینه می‌شوند. با استفاده از این نرم افزار می‌توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم‌ها، انرژی تشکیل مولکول و... را بدست آورد. سپس توصیف گر‌ها به تعداد ۱۴۸۱ مورد به وسیله نرم افزار دراگون (Dragon) نسخه ۲/۱ محاسبه شدند. این نرم افزار برای محاسبه هیجده دسته از توصیف گرهای مولکولی مورد استفاده قرار می‌گیرد (جدول ۱).

۲-۳. کاهش تعداد توصیف کننده‌های شیمیایی

یکی از مشکلاتی که در هنگام ایجاد مدل‌های QSAR/QSPR با آن مواجه می‌شویم، تعداد زیاد متغیرهای مستقل می‌باشد در اکثر مواقع تعداد توصیف گر‌ها از تعداد مولکول‌ها بیشتر است. در این صورت استفاده از روش‌های حداقل مربعات باعث ایجاد مشکلاتی نظیر انتخاب شانسی و همبستگی تصادفی می‌گردد. با توجه به این که بعضی از متغیرهای مستقل ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می‌دهند، لذا به روش زیر بعضی از متغیرها حذف می‌شوند. توصیف گرهای محاسبه شده برای ترکیبات مورد بررسی قرار می‌گیرند تا آنهایی را که مقادیر ثابت و یا تقریباً ثابت دارند (بیش از ۹۰٪ داده‌های ثابت دارند) حذف می‌شوند. در این مرحله تعداد ۲۷۱ توصیف گر کنار گذاشته شدند. بدین ترتیب ۱۱۷۸ توصیف گر باقی می‌مانند. با توجه به اینکه در برخی از موارد بعضی از متغیرهای مستقل با یکدیگر همبستگی کاملی دارند، و وجود تنها یکی از این متغیرها در مدل کردن کافی بوده و نیازی به حضور بقیه متغیرها نمی‌باشد. لذا داده‌ها از این نظر مورد بررسی قرار گرفتند. به منظور حذف متغیرهایی که همبستگی خطی بزرگتر از ۰/۹ با یکدیگر دارند، ماتریس 1179×1179 بین همه متغیرهای مستقل و نیز متغیر وابسته تشکیل می‌شود. متغیر مستقلی که همبستگی بیشتری از ۰/۹ با متغیرهای دیگر دارند، بایستی حذف شوند. اما بایستی دقت داشت که هنگام حذف متغیر اضافی، متغیری کنار گذاشته می‌شود که همبستگی کمتری با متغیر وابسته داشته باشد. بدین ترتیب تعداد ۸۸۱ توصیف گر اضافی کنار گذاشته می‌شوند و تعداد ۲۸۳ توصیف گر باقی می‌مانند.

جدول ۱. انواع توصیف گرها به همراه تعداد آنها که توسط نرم افزار Dragon محاسبه می گردد

شماره	نام گروه توصیف کننده	بعد توصیف کننده	تعداد توصیف کننده ها
۱	توصیف کننده های ساختاری	۰	۴۷
۲	توصیف کننده های توپولوژیکی	۲	۲۶۶
۳	توصیف کننده های شمارنده گام مولکولی	۲	۲۱
۴	توصیف کننده های BCUT	۲	۶۴
۵	اندیس های بار توپولوژیکی	۲	۲۱
۶	توصیف کننده های خود ارتباطی دو بعدی	۲	۹۶
۷	توصیف کننده های مربوط به بار	۳	۱۴
۸	اندیس های آروماتیکی	۳	۴
۹	پروفایل های مولکولی Randic	۳	۴۱
۱۰	توصیف کننده های هندسی	۳	۷۰
۱۱	توصیف کننده های RDF	۳	۱۵۰
۱۲	توصیف کننده های 3D-MORSE	۳	۱۶۰
۱۳	توصیف کننده های WHIM	۳	۹۹
۱۴	توصیف کننده های GETAWAY	۳	۱۹۳
۱۵	گروه های عاملی	۱	۱۱۵
۱۶	اجزای متمرکز اتمی	۱	۱۱۶
۱۷	توصیف کننده های تجربی	۱	۳
۱۸	ویژگی ها	۱	۳
مجموع توصیف کننده ها			۱۴۸۱

۲-۴. انتخاب توصیف کننده های موثر

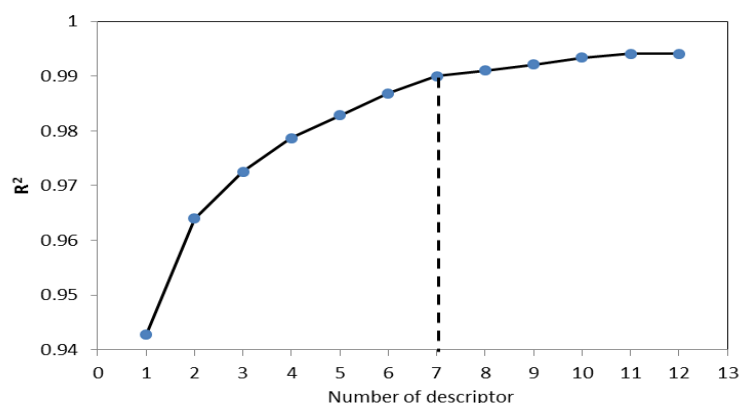
مهم ترین بخش در ایجاد یک مدل کار آمد، انتخاب توصیف کننده های مناسب است. پس از محاسبه توصیف کننده های مختلف، تعدادی از آنها به عنوان توصیف کننده های مناسب برای ساخت مدل انتخاب می شوند. این مرحله شامل یافتن توصیف کننده های حاوی اطلاعات مفید است به طوری که قدرت پیش بینی مدل در سطح قابل قبولی باشد. در این قسمت توصیف کننده

ها به عنوان متغیرهای مستقل و خاصیت مورد نظر به عنوان متغیر وابسته در نظر گرفته می شود. مدل‌های خطی با استفاده از روش MLR ایجاد می شوند.

در این کار از روش افزایش مرحله ای برای انتخاب توصیف کننده های مناسب استفاده شد. با آزمایش همه توصیف کننده ها، فرآیند انتخاب تا زمانی ادامه می یابد که مدلی با ضریب همبستگی بالا (معمولاً در حدود ۰/۷ تا ۰/۹) به دست آید. اگر ضریب همبستگی به مقدار بالای ۰/۹۵ برسد، می توان گفت که مدل خطی، مدل مناسبی برای توصیف سیستم مورد بررسی است و با استفاده از این مدل می توان پیش بینی را به نحو مطلوب انجام داد.

۳. نتایج و بحث

برای ساختن مدلی که بیانگر ارتباط ساختاری ترکیبات مورد بررسی با اندیس کوآتس آن‌ها باشد، از روش رگرسیون خطی چندگانه (MLR) استفاده شد. همانطور که در بالا توضیح داده شد، طبیعی است که تعداد زیاد توصیف کننده‌ها باعث پیچیدگی محاسبات شده و همچنین احتمال وجود فاکتورهایی دارای برهم کنش با هم را افزایش می دهد. لذا تعدادی از این توصیف کننده‌ها که دارای ۹۰٪ مقادیر یکسان بودند حذف شدند و همچنین از بین توصیف کننده‌هایی که دارای همبستگی بیش از ۰/۹ بودند توصیف کننده‌ای که دارای همبستگی کمتری با متغیر وابسته بود، از سری داده‌ها حذف شد. سپس مقادیر توصیف کننده‌های محاسبه شده توسط نرم افزار Dragon پس از کاهش تعداد متغیرها بعنوان متغیرهای مستقل و مقادیر اندیس کوآتس مولکول‌های مورد نظر بعنوان متغیرهای وابسته بعنوان ورودی به نرم افزار SPSS وارد شد. در نهایت با استفاده از منوی آنالیز، گزینه‌ی رگرسیون خطی و روش مرحله ای (Stepwise) انتخاب و نهایتاً چندین مدل مختلف به طور جداگانه به دست آمد، که با توجه به خصوصیات آماری آن‌ها از جمله ضریب رگرسیون (R)، آماره F و خطای استاندارد و پس از رسم مقادیر R و R² و SE بر حسب تعداد توصیف کننده‌ها بهترین مدل که دارای بیشترین مقدار R و F و کمترین مقدار خطای استاندارد و شامل توصیف کننده‌های تا حد امکان قابل توجه باشد، به عنوان مدل نهایی برای ارتباط اندیس کوآتس مولکول‌ها با ساختار آن‌ها انتخاب شد. با این روش هفتم با تعداد ۷ توصیف کننده به عنوان مناسب ترین آنها انتخاب شد و توسط روش MLR مدل سازی و مورد ارزیابی قرار گرفت.



شکل ۱. تأثیر تعداد توصیف کننده ها بر مقدار R²

شکل ۱ تأثیر تعداد توصیف کننده ها را بر مقدار R^2 نشان می دهد. همانطور که ملاحظه می شود، تغییرات R^2 بعد از ۷ توصیف کننده خیلی کم می باشد. بنابراین ۷ توصیف کننده جهت مدل سازی انتخاب گردید.

فهرست توصیف کننده های انتخاب شده توسط نرم افزار SPSS به همراه توصیف مختصری از آنها در جدول (۲) آورده شده است.

جدول ۲. توصیف کننده های انتخاب شده

توصیف کننده	نوع توصیف کننده	نماد	ضریب
Modified Randic chi-1 index	Topological	XMOD	29.00
Moran autocorrelation-lag1/weighted by atomic Sanderson electronegativities	2D autocorrelations	MATS1e	-953.68
Mean topological charge index of order1	Galvez topol.charge indices	JGI1	-454.51
The symbol C-002 corresponds to: CH ₂ R ₂	Atom-centred fragments	C-002	23.30
H autocorrelation of lag 7/weighted by atomic masses	GETAWAY	H7m	-1402.27
Distance/detour ring index of order 6	Topological	D/Dr06	0.58
2 nd component accessibility directional WHIM index/ weighted by atomic masses	WHIM	E2m	348.37
Constant			231.43

۳-۱. ارزیابی توصیف کننده های انتخاب شده

به منظور ارزیابی توصیف کننده های انتخاب شده مبنی بر مستقل بودن از همدیگر در جدول (۳) ضرایب همبستگی توصیف کننده های انتخاب شده نسبت به یکدیگر آورده شده است. همانطور که از جدول مشاهده می شود بیشترین ضریب همبستگی بین توصیف کننده XMOD و توصیف کننده MATS1e با مقدار ضریب همبستگی ۰/۶۵۴ 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 نشان دهنده وابستگی توصیف کننده ها نمی باشد.

جدول ۳. ماتریس ضرایب همبستگی توصیف کننده های انتخاب شده

	MATS1e	JGI1	C-002	H7m	D/Dr06	E2m	XMOD
MATS1e	1						
JGI1	0.115	1					
C-002	-0.365	-0.419	1				
H7m	0.559	0.189	0.021	1			
D/Dr06	0.492	-0.071	-0.187	0.191	1		
E2m	0.048	0.062	-0.137	-0.184	0.129	1	
XMOD	0.654	0.058	-0.283	0.333	0.231	-0.112	1

۳-۲. ایجاد مدل با استفاده از SW-MLR

پس از انتخاب مناسب ترین توصیف کننده ها توسط روش مرحله ای، مرحله بعدی ایجاد مدل، میان توصیف کننده های انتخاب شده و اندیس کواتس ترکیبات می باشد. از نرم افزار SPSS برای این منظور استفاده گردید. بین توصیف کننده ها و اندیس کواتس ترکیبات سری آموزش با استفاده از روش MLR رابطه زیر بدست آمد:

$$RI=231.432+348.373E2m-1402.278H7m-454.513 JGI1-953.683 MATS1e+0.589 D/Dr06+29.004 XMOD$$

$$R^2_{calibration}=0.991, R^2_{prediction}=0.982, Q^2_{LOO}=0.988, Q^2_{LGO}=0.986, REP(\%)=2.07, F_{train}=570.62, F_{test}=24.85, RMSEP=29.015$$

سپس از معادله بدست آمده برای پیش بینی اندیس کوتاس سری تست استفاده گردید. مقادیر واقعی و پیش بینی شده اندیس کوتاس و همچنین خطاهای پیش بینی برای کلیه ترکیبات مجموعه آموزش و پیش بینی در جدول ۴ آورده شده است. شکل ۲ نمودار مقادیر اندیس کوتاس محاسبه شده برای سری آموزش و تست بر حسب مقادیر تجربی را نشان می دهد. در این شکل، میزان نزدیکی داده ها به خط راست قدرت پیشگویی مدل را نشان می دهد. شکل ۳ مقادیر باقیمانده خطاها (اختلاف مقدار پیش بینی شده و مقدار واقعی اندیس کوتاس) را نسبت به مقادیر تجربی نشان می دهد. میزان پراکنندگی خطاها در اطراف محور نشان دهنده این است که خطای سیستماتیک در مدل وجود ندارد.

جدول ۴. مقادیر تجربی و محاسبه شده اندیس کوتاس برای ترکیبات مختلف برای مجموعه های آموزشی و پیش بینی در مدل SW-MLR همراه با

مقادیر خطای نسبی

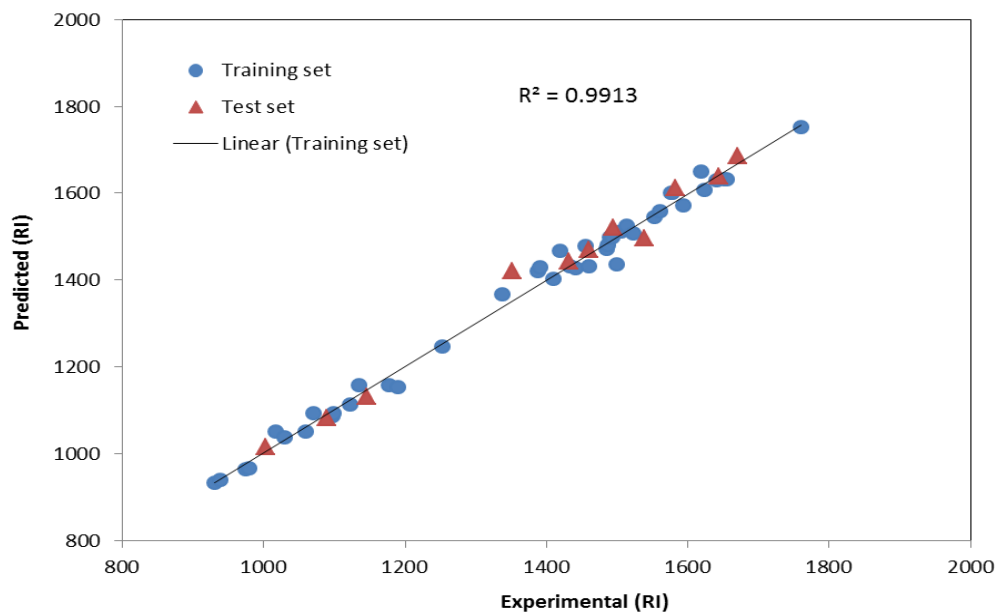
No.	Compound	RI (Exp) ^a	RI (SW-MLR) ^b	E (%) ^c
Training set				
1	Alpha thujene	930	934.17	0.45
2	Alpha pinene	939	939.98	0.10
3	Sabinene	975	963.40	-1.19
4	Beta pinene	979	967.47	-1.18
5	Alpha terpinene	1017	1051.76	3.42
6	Beta phellandrene	1030	1037.13	0.69
7	Gamma terpinene	1060	1050.37	-0.91
8	Cis -sabinene hydrate	1070	1093.84	2.23
9	Linalool	1097	1085.95	-1.01
10	Trans- sabinene hydrate	1098	1093.84	-0.38
11	Menthe-2-en-1-ol	1122	1114.56	-0.66
12	Terpinene -1	1134	1158.58	2.17
13	Terpinene-4-ol	1177	1159.01	-1.53
14	Alpha terpineol	1189	1153.34	-3.00
15	Geraniol	1253	1246.10	-0.55
16	Delta elemene	1338	1366.67	2.14
17	Beta bourbonene	1388	1420.65	2.35
18	Beta elemene	1391	1429.25	2.75
19	Alpha gurjunene	1410	1402.35	-0.54
20	Beta caryophyllene	1419	1466.50	3.35
21	Beta gurjunene	1434	1431.43	-0.18
22	aromadendrene	1441	1428.23	-0.89
23	Alpha humulene	1455	1478.09	1.59
24	Alloaromadendrene	1460	1432.41	-1.89
25	Germacron-D	1485	1471.25	-0.93

26	Phenylethy butanoate	1487	1481.64	-0.36
27	Beta selinene	1490	1497.76	0.52
28	Beta-cisguainene	1493	1499.35	0.43
29	bicyclogermacrene	1500	1437.28	-4.18
30	E,E-alpha farnesene	1506	1511.66	0.38
31	Gamma cadinene	1514	1524.91	0.72
32	Delta cadinene	1523	1506.41	-1.09
33	1-hydroxy-1,7-dimethyl-4-isopropyl-2,4-cyclodecadiene	1553	1544.26	-0.56
34	Germacrene B	1561	1558.39	-0.17
35	Germacrene D-4-ol	1576	1600.41	1.55
36	Spathulenol	1578	1601.40	1.48
37	Viridiflorol	1593	1571.04	-1.38
38	Cubenol(1,10-di-epi)	1619	1649.25	1.87
39	gamma Eudesmol(10-epi)	1624	1607.81	-1.00
40	Alpha cadinol(epi)	1640	1628.88	-0.68
41	Cedr-8(15)-en-10-ol	1652	1633.04	-1.15
42	Alpha cadinl	1654	1632.00	-1.33
43	Benzyl benzoate	1760	1752.18	-0.44
Test set				
1	Alpha phellandrene	1003	1014.87	1.18
2	Carvacrol	1089	1082.74	-0.58
3	Trans verbenol	1145	1130.49	-1.27
4	Alpha cubebene	1351	1421.57	5.22
5	Beta copaene	1432	1442.84	0.76
6	Gamma gurjunene	1460	1469.41	0.64
7	Muurola	1494	1521.38	1.83
	4(14),5dien(trans)			
8	Alphcadinene	1539	1497.07	-2.72
9	Caryophyllene oxide	1583	1612.94	1.89
10	Selina 3,11dien-6-alpha-ol	1644	1639.24	-0.29
11	Caryophyllene (14-hydroxy-9-epi-E)	1670	1686.06	0.96

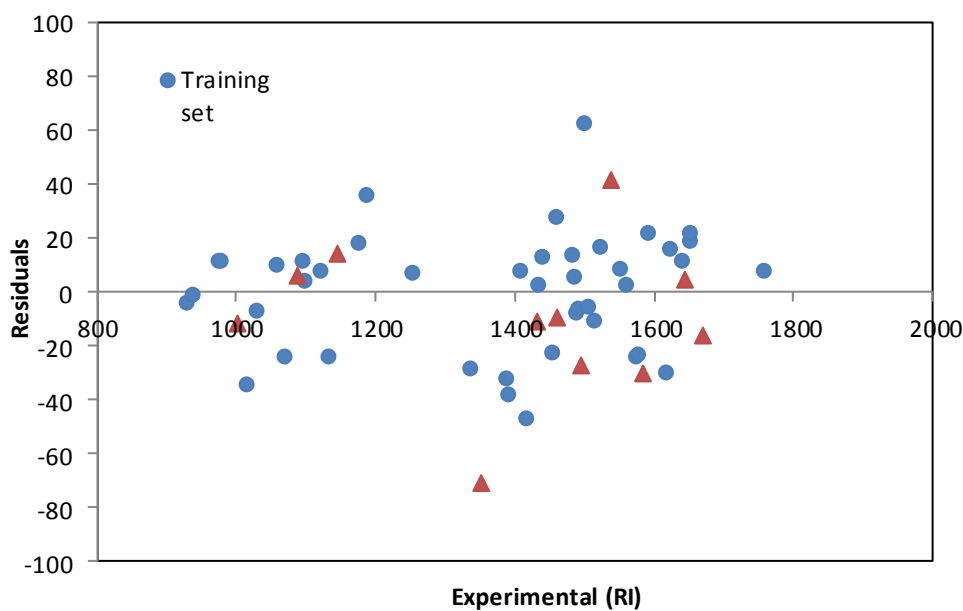
^aمقادیر تجربی اندیس کواتس

^bاندیس کواتس محاسبه شده توسط رگرسیون خطی چند گانه

^cخطای نسبی



شکل ۲. نمودار مقادیر اندیس کواتس محاسبه شده با کمک مدل SW-MLR برای مجموعه‌های آموزشی و پیش‌بینی بر حسب مقادیر تجربی



شکل ۳. نمودار تغییرات خطا برای مقادیر اندیس کواتس محاسبه شده با کمک مدل SW-MLR برای مجموعه‌های آموزشی و پیش‌بینی

۳-۳. ارزیابی اعتبار مدل‌های انتخاب شده

یکی از مواردی که می‌توان با استناد به آن از معتبر بودن مدل انتخابی اطمینان پیدا کرد، به دست آوردن ضریب تعیین (R^2) می‌باشد که هرچه این مقدار به یک نزدیک‌تر باشد مدل، معتبرتر خواهد بود. برای دست‌یابی به این هدف مقادیر پیش‌بینی شده

اندیس‌های بازداري برحسب مقادير تجربی رسم شد که شکل ۲ نشان دهنده این مطلب می‌باشد. در نمودار فوق مقدار ضریب همبستگی، برابر ۰/۹۹۱ برای سری آموزش و مقدار ۰/۹۸۲ برای سری ارزیابی (تست) است که حاکی از اعتبار بالای مدل می‌باشد. مورد دیگری که می‌توان برای اعتبار مدل به آن استناد کرد رسم مقادير باقی مانده حاصل از اختلاف مقادير پیش‌بینی شده‌ی اندیس - های بازداري برحسب مقادير تجربی آن می‌باشد که شکل ۳ آن را به خوبی نشان می‌دهد. در نمودار فوق می‌توان پراکنندگی نسبتاً یکسان نقاط را حول مقادير صفر، ناشی از عدم وجود خطای معین در روش و اعتبار مدل دانست.

درصد خطای نسبی برای سری تست^۱ (REP) نیز جهت بررسی اعتبار مدل مورد استفاده قرار گرفت. هر چه مقدار REP کوچکتر باشد مدل، مدل مناسبتری خواهد بود. مقدار این آماره برابر ۲/۰۷٪ می‌باشد که مقدار نسبتاً کوچک آن اعتبار مدل را نشان می‌دهد. یکی دیگر از روش‌ها اعتبارسنجی تقاطعی است که یکی از رایج‌ترین تکنیک‌ها جهت اعتبارسنجی مدل محسوب می‌شود. در این روش هر بار یک^۲ (LOO) یا تعداد معینی^۳ (LGO) از مولکولها کنار گذاشته می‌شوند و مدل با مولکول های باقی مانده ساخته می‌شود، سپس کمیت مدلسازی برای مولکول های کنار گذاشته شده، توسط مدل حاصله پیش بینی می‌شود. این عمل به صورت چرخه ای برای تمام مولکول ها تکرار می‌شود به طوری که همه مولکولها یک بار در سری پیش بینی قرار می‌گیرند. پارامتر آماری Q^2 برای ارزیابی مدل حاصل از ارزیابی تقاطعی بکار می‌رود. مقدار Q^2_{LOO} برابر ۰/۹۸۸ و مقدار Q^2_{LGO} برابر ۰/۹۸۶ نشان دهنده اعتبار مدل می‌باشد.

۳-۴. پارامترهای آماری

جدول ۵ پارامترهای آماری مختلف را برای مدل SW-MLR نشان می‌دهد. به عنوان مثال مقادير R^2 برای سری آموزش و تست به ترتیب عبارتند از: ۰/۹۹۱ و ۰/۹۸۲ و همچنین درصد خطای نسبی پیش بینی برای اندیس کواتس محاسبه شده برای مجموعه پیش‌بینی با کمک این مدل برابر ۲/۰۷٪ می‌باشند. با توجه به پارامترهای آماری مناسب از جمله خطای کم آن، می‌توان از این مدل، برای محاسبه اندیس کواتس سایر ترکیبات مجهول استفاده نمود.

جدول ۵. پارامترهای آماری مختلف برای مدل SW-MLR

Model	$R^2_{training}$	$R^2_{testing}$	REP(%)	Q^2_{LOO}	Q^2_{LGO}
SW-MLR	۰/۹۹۱	۰/۹۸۲	۲/۰۷	۰/۹۸۸	۰/۹۸۶

¹- Relative error prediction(REP)

²- Leave-one-out (LOO)

³- Leave- group-out (LGO)

۴. نتیجه گیری

در این تحقیق، مطالعه ارتباط کمی ساختار-بازداری روی ترکیبات گیاه *Nepeta macrosiphon Boiss* انجام شد. در ابتدا توصیف کننده‌های ساختاری ایجاد و مناسب‌ترین آن‌ها که بیشترین ارتباط را با اندیس کواتس ترکیبات مورد نظر داشتند به کمک رگرسیون مرحله‌ای انتخاب گردید و به روش رگرسیون خطی چندگانه (MLR) مدل مناسب ایجاد شد. مدل ایجاد شده با این روش قابلیت پیش‌بینی بسیار خوبی ارائه نمود. مقدار بالای R^2 همچنین مقادیر پایین خطای پیش‌بینی نشان دهنده کارآیی مطلوب مدل SR-MLR در پیش‌بینی اندیس کواتس می‌باشد. همچنین در این کار توصیف کننده‌های MATS1e, JGI1, C-002, H7m, D/dr06, E2m, XMOD به عنوان توصیف کننده‌های موثر در پیش‌بینی اندیس کواتس ترکیبات آشکار گردید. با دانستن مقادیر این توصیف کننده‌ها برای هر ترکیب مجهول دیگر و با استفاده از مدل بدست آمده می‌توان مقادیر اندیس کواتس آن ترکیب را بدست آورد. بنابراین از این مدل می‌توان برای پیش‌بینی اندیس کواتس ترکیبات مشابه در ستون‌ها و شرایط استفاده نمود.

۵. مراجع

- [1] Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1), 69-77.
- [2] Dearden, J. C., Cronin, M. T., & Kaiser, K. L. (2009). How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 20(3-4), 241-266.
- [3] 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.
- [4] Yousefinejad, S., & Hemmateenejad, B. (2015). Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*, 149, 177-204.
- [5] Borhani, T. N. G., Saniedanesh, M., Bagheri, M., & Lim, J. S. (2016). QSPR prediction of the hydroxyl radical rate constant of water contaminants. *Water research*, 98, 344-353.
- [6] Jin, X., Peldszus, S., & Huck, P. M. (2015). Predicting the reaction rate constants of micropollutants with hydroxyl radicals in water using QSPR modeling. *Chemosphere*, 138, 1-9.
- [7] 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.
- [8] 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.
- [9] Zhu, T., Gu, L., Chen, M., & Sun, F. (2021). Exploring QSPR models for predicting PUF-air partition coefficients of organic compounds with linear and nonlinear approaches. *Chemosphere*, 266, 128962.

- [10] Yan, F., Shi, Y., Wang, Y., Jia, Q., Wang, Q., & Xia, S. (2020). QSPR models for the properties of ionic liquids at variable temperatures based on norm descriptors. *Chemical Engineering Science*, 217, 115540.
- [11] Babushok, V. I. (2015). Chromatographic retention indices in identification of chemical compounds. *TrAC Trends in Analytical Chemistry*, 69, 98-104.
- [12] Barron, L. P., & McEneff, G. L. (2016). Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods. *Talanta*, 147, 261-270.
- [13] Ma, C., Ren, Y., Yang, J., Ren, Z., Yang, H., & Liu, S. (2018). Improved peptide retention time prediction in liquid chromatography through deep learning. *Analytical chemistry*, 90(18), 10881-10888.
- [14] Le Maux, S., Nongonierma, A. B., & FitzGerald, R. J. (2015). Improved short peptide identification using HILIC-MS/MS: Retention time prediction model based on the impact of amino acid position in the peptide sequence. *Food chemistry*, 173, 847-854.
- [15] Gussakovsky, D., Neustaeter, H., Spicer, V., & Krokhin, O. V. (2017). Sequence-specific model for peptide retention time prediction in strong cation exchange chromatography. *Analytical chemistry*, 89(21), 11795-11802.

Quantitative structure-retention relationship (QSRR) study of Covats Index of *Boiss Nepeta macrosiphon*

Azam Vafaei*¹

¹Department of Chemistry, Gascaran Branch, Islamic Azad University, Gascaran, Iran

Submitted: 26 August 2021, Revised: 11 November 2021, Accepted: 02 December 2021

Abstract

Quantitative structure-Retention relationship (QSRR) studies were performed on the Covats index of the constituents of *Boiss Nepeta macrosiphon*. The stepwise method was used to select the appropriate descriptors. Selected descriptors of this method were used to model and predict the retention index of these compounds by multiple linear regression (MLR). In order to evaluate the validity of the model, various methods such as using a test series and rejecting each data step by step were used. The results show the good ability of SR-MLR method to predict retention index. The coefficients for determining the training and test series were 0.991 and 0.982, respectively.

Keywords: *Quantitative structure-inhibition relationship, Qovats index, multiple linear regression, Boiss Nepeta macrosiphon*

*Corresponding author : Azam Vafaei

Address: Department of Chemistry, Gascaran Branch, Islamic Azad University, Gascaran, Iran

Tel: 09171404204

E-mail: a.vafaei@yahoo.com