

## مطالعه رابطه ساختار - خاصیت برای پیش بینی $\log P$ مشتقات سیکلوپروپان کربوکسیلیک اسید (پیرتروئیدها) با استفاده از مدل رگرسیون خطی چندگانه

مصطفی صادقی<sup>۱</sup>، عصمت محمدی نسب<sup>۲\*</sup>، طاهره مومنی اصفهانی<sup>۳</sup>

۱- دانشجوی دکتری شیمی فیزیک، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

۲- استادیار شیمی فیزیک، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

۳- استادیار شیمی تجزیه، گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

### چکیده

در این مطالعه، قدرت پیش بینی ضریب تقسیم آب-اکتانل ( $\log P$ ) برای 34 نوع از مشتقات پیرتروئیدی با استفاده از رابطه کمی ساختار-خاصیت مورد مطالعه قرار گرفت. مقدار  $\log P$  پیرتروئیدهای مورد مطالعه با کمک الگوریتم ژنتیک بر اساس روش رگرسیون خطی چندگانه (GA-MLR) مدل سازی شد و معلوم گردید که سه توصیفگر موثر GATS4P، ZM1V و PW3 همبستگی معقولی با  $\log P$  دارند و منجر به ایجاد مدلی با ضریب رگرسیون بالا و خطای کم شدند. ارزیابی توانایی پیش بینی  $\log P$  با مدل (GA-MLR) توسط پارامترهای آماری:  $R^2_{adj} = 0.848$ ،  $R^2 = 0.862$ ،  $F=62.296$  و  $MSE = 0.503$  برای مجموعه آزمایشی انجام شد. همچنین مقدار  $Q^2_{LOO} = 0.861$  در روش اعتبارسنجی تقاطعی و نیز مقادیر  $R^2$  برابر با 0.880 و 0.929 به ترتیب برای سری های آموزش و آزمایش در روش اعتبارسنجی خارجی، همبستگی بسیار خوبی را بین مقادیر تجربی و مقادیر پیش بینی نشان داد. مشخص گردید که مدل MLR در پیش بینی  $\log P$  حشره کش های پیرتروئیدی قابل اعتماد بوده و با در نظر داشتن خطای بسیار کم از دقت کافی برخوردار است.

واژه های کلیدی: "ضریب توزیع آب-اکتانل"; "روش رگرسیون خطی چندگانه"; "مطالعه کمی رابطه ساختار-خاصیت"; "الگوریتم ژنتیک"; "پیرتروئیدها".

\* نویسنده رابط، پست الکترونیکی:

Es.Mohammadinasab@iau.ac.ir

تاریخ دریافت مقاله: ۱۴۰۱/۰۹/۱۶ - تاریخ پذیرش مقاله: ۱۴۰۱/۱۰/۱۰



## مقدمه

از دهه ۱۹۲۰، مطالعات نشان داده است که پیرترین<sup>۱</sup> می تواند به عنوان ماده خام برای تولید آفت کش های مصنوعی استفاده شود [Aloizou, et al., 2020; Zhu, et al., 2020; Ravula & Yenugu, 2021]. آلتترین یک استر سیکلو پروپان کربوکسیلیک، اولین آفت کش پیرتروئیدی<sup>۲</sup> است که در سال ۱۹۴۹ توسعه یافته است [Aznar-Alemanly & Eljarrat, 2020]. حشره کش های پیرتروئید بیش از ۳۰ سال است که در فرمولاسیون های کشاورزی و خانگی مورد استفاده قرار می گیرند [Shafer & Crofton, 2005] و ۲۵٪ درصد بازار حشره کش های جهان را به خود اختصاص داده اند. [Casida & Gary, 1998] سموم پیرتروئید<sup>۳</sup>، آنالوگ های مصنوعی پیرترین های طبیعی هستند که از نظر شیمیایی با ثبات تر از آنها می باشند و کاربردهای زیادی در صنایع کشاورزی، دامپزشکی و محصولات خانگی مانند شامپوهای ضد شپش و پشه کش ها دارند [Schleier III, et al., 2011; Zuluaga & Peterson, 2016]. موثرترین و ایمن ترین حشره کش های طبیعی، پیرترین ها هستند که از گل های تاناستیم سینراریفولیم<sup>۴</sup> استخراج شده اند. پیرتروئیدها با توجه به اثربخشی و ایمنی نسبی که برای انسان ها و پستانداران دارند جایگزین محصولات قدیمی تر و خطرناک تر مانند ارگانوکلرین ها، ارگانوفسفات ها و کاربامات ها شده اند [Katsuda, 2012; Cycoń & Piotrowska-Seget, 2016; Khambay & P. J. Jewess, 2005]. این ترکیبات به آسانی تجزیه می شوند و دوام کوتاهی دارند. اکثریت آنها سمیت کمی برای پستانداران دارند و برای کودکان خطر بیشتری نسبت به بزرگسالان دارند و انواع آنها می توانند بر سیستم تولید مثل تأثیر بگذارند [Chrutek, et al., 2018; Marettova, et al., 2017]. پیرتروئیدها نسبت به بسیاری از حشره کش های دیگر میل ترکیبی بالایی برای اتصال به ذرات خاک و رسوب دارند، ضریب تقسیم آب اکتانول بالایی دارند و تجزیه نمی شوند و باعث آلودگی فوری محیط و خاک می شوند و ارگانسیم های غیر هدف را تحت تأثیر قرار می دهند. [Palmquist, et al., 2012] در حال حاضر پیرتروئیدهای سنتزی نسبت به پیرتروم گیاهی فعالیت بیشتری در از بین بردن حشرات دارند [Breckenridge, et al., 2009]. تا بحال روشهای محاسباتی زیادی در خصوص پیش بینی خواص فیزیکی - شیمیایی حشره کش ها گزارش شده است. به مجموعه ابزارها و روش هایی که به این منظور مورد استفاده قرار می گیرند روش های پارامتری می گویند. در روش های پارامتری مانند رابطه کمی بین ساختار - فعالیت<sup>۵</sup> (QSAR) و یا رابطه کمی بین ساختار - خاصیت<sup>۶</sup> (QSPR) سعی می شود بین یک سری توصیف کننده های مولکولی با فعالیت یا خاصیت مورد نظر ارتباط منطقی برقرار گردد [Toropov, et al., 2015; Cappelli, et al., 2015]. مطالعات QSPR/ QSAR شامل مراحل اصلی زیر می باشد:

۱. جمع آوری داده های معتبر مربوط به فعالیت یا خاصیت ترکیبات ۲. محاسبه توصیف گرهای مولکولی ۳. گزینش توصیف کننده های موثر ۴. برقراری ارتباط بین توصیف گرها و داده ها با ایجاد مدل مناسب ۵. ارزیابی و اعتبارسنجی مدل نهایی جهت پیش بینی. منظور از توصیف کننده های مولکولی، مقادیر عددی هستند که جنبه های مختلف ساختاری مولکول را به طور کمی نشان می دهند. شیمی کوانتوم از طریق محاسبات هارتری-فاک<sup>۷</sup> ابزاری مناسب برای استخراج توصیفگرهای مولکولی ترکیبات شیمیایی به شمار می رود. دقت پیش بینی مدل های QSAR/QSPR با نوع توصیفگرها مرتبط می باشد. این رابطه می تواند برای پیش بینی پاسخ بیولوژیکی یا شیمیایی ساختارهای ناشناخته مورد استفاده قرار گیرد. با توجه به اینکه مطالعه ویژگی های ترکیبات شیمیایی یک

<sup>1</sup> pyrethrin

<sup>2</sup> Pyrethroidi

<sup>3</sup> Pyrethroid

<sup>4</sup> Tanacetum Cinerariifolium

<sup>5</sup> Quantitative Structure Activity Relationship

<sup>6</sup> Quantitative Structure Property Relationship

<sup>7</sup> Hartree - Fock

فرآیند پرهزینه و وقت گیر است، لذا پیشگویی این ویژگی ها با مدل های محاسباتی قبل از تولیدشان برای کاهش هزینه ها امری ضروری به نظر می رسد. برخی از روش های پیشرفته QSAR/QSPR به طور گسترده برای طراحی ترکیبات جدید مورد استفاده قرار گرفتند [Amiri, et al., 2020; Mudasir, et al., 2016] یکی از این تکنیک ها روش رگرسیون خطی چندگانه<sup>۱</sup> (MLR) است [Souyei, et al., 2019; Zapadka, et al., 2019]. مطالعات خطی و غیرخطی QSAR/QSPR زیادی تاکنون برای مدل سازی و پیش بینی فعالیت/خاصیت ترکیبات شیمیایی انجام یافته اند. برمودز-سالدانا<sup>۲</sup> و کرونین<sup>۳</sup> و به مطالعه سمیت ماهی قزل آلی رنگین کمان<sup>۴</sup> ۷۵ نوع آفت کش با استفاده از تکنیک های MLR و PLS<sup>۵</sup> پرداختند [Bermudez-Saldana & Cronin, 2006]. واکوندیو<sup>۶</sup> و همکاران یک رابطه کیفی بین ساختار مولکولی و توانایی هیدرولیز متابولیک کاربامات های دارویی برقرار کردند [Vacondio, et al., 2010]. رابطه کمی ساختار-پایداری برای دسته ای از آریل استرهای سیکلوهگزیل کاربامیک اسید که به عنوان مهارکننده های هیدرولاز آمید اسید چرب عمل می کنند، بر اساس پایداری آزمایشگاهی آنها در محیط های شیمیایی و بیولوژیکی مورد مطالعه قرار گرفت [Vacondio, et al., 2009]. تخمین ضریب تقسیم اکتانول/آب (logP) ترکیبات شیمیایی در بسیاری از علوم مانند: پزشکی، داروسازی، غذاها، عطرها، صنایع شیمیایی و حفاظت از محیط زیست مطرح می شود. logP به عنوان نسبت غلظت ماده در فازهای آلی و آبی در شرایط تعادل تعریف می شود که می تواند به روش های مختلف به صورت تجربی یا به کمک نرم افزارهای مختلف [Bahmani, et al., 2017; Valko, 2016; Lopez, et al., 2021] با روش های محاسباتی گوناگون [Lanevskij & Didziapetris, 2019] اندازه گیری شود. روش های آزمایشی برای تعیین logP مانند شیک فلاسک، کروماتوگرافی لایه نازک فاز معکوس و کروماتوگرافی مایع با کارایی بالا<sup>۷</sup> (HPLC) همیشه قابل دسترسی نیستند. از این رو، استفاده از روش های محاسباتی برای یافتن مقادیر logP ترکیبات جدید امری ضروری به نظر می رسد [Wang, et al., 2017; Bouakkadia, et al., 2021]. گل محمدی روش های MLR، PLS و ANN<sup>۸</sup> را برای پیش بینی logP ترکیبات آلی مورد مطالعه قرار داد و نتیجه گرفت که مدل ANN می تواند در طراحی کاربامات های جدید و مشتقات آنها کارایی لازم را داشته باشد [Golmohammadi, 2009]. با استفاده از پارامترهای فیزیکی - شیمیایی و ساختاری ترکیبات آروماتیکی، مشخص شد که رابطه خطی بین پتانسیل جریان دم گل های داوودی با قدرت حشره کشی پیرتراتها وجود دارد [Nishimura, et al., 1991]. رابطه ساختار-فعالیت برای حشره کش های بنزیل-ترانس پروپان-کربوکسیلات با ترکیبی از تکنیک های آماری چندمتغیره و مدل سازی مولکولی توسط فورد و همکارانش مورد مطالعه قرار گرفت. روش های مکانیک مولکولی و کوانتومی برای محاسبه مجموعه ای از هفتاد توصیفگر برای نشان دادن خصوصیات مولکولی ترکیبات مورد استفاده قرار گرفت [Ford, et al., 1989]. اثر دیپلاریزه شدن یک سری پیرتروئیدها بر غشاهای محوری سوسک های آمریکایی با تکنیک ثبت داخل سلولی اندازه گیری شد. رابطه کمی بین فعالیت دیپلاریزه شدن با فعالیت انسداد عصبی (در مسیر عصب مرکزی) و نیز فعالیت کشتن حشرات به روش حداقل مربعات مورد تجزیه و تحلیل قرار گرفت و مشخص شد که هرچه فعالیت دیپلاریزاسیون بیشتر و آبگریزی بیشتر باشد،

<sup>1</sup> Multiple linear regression

<sup>2</sup> Bermudez-Saldana

<sup>3</sup> Cronin

<sup>4</sup> oncorhynchus mykiss Walbaum

<sup>5</sup> Partial least squares regression

<sup>6</sup> Vacondio

<sup>7</sup> High Performance Liquid Chromatography

<sup>8</sup> Artificial Neural Network

فعالیت انسداد عصبی نیز بالاتر است [Matsuda, et al., 1991]. امیری و همکاران وی با کمک تکنیک کمی رابطه ساختار-خاصیت بر روی ۴۳ ترکیب حشره کش با استفاده از روش های MLR و ماشین های بردار پشتیبان (SVM) توانستند به پیش بینی ضریب تقسیم آب-اکتانال ۴۳ حشره کش ارگانوفسفره پردازند [Amiri, et al., 2019]. هدف این پژوهش کاربرد رابطه کمی ساختار-خاصیت QSPR با کمک مدل رگرسیون خطی چندگانه MLR توسط نرم افزار SPSS برای پیش بینی logP مشتقات پیرتروئید با استفاده از محاسبات مکانیک کوانتمی و بکارگیری الگوریتم ژنتیکی برای انتخاب بهترین توصیفگرها میباشد.

## روش کار

ابتدا ساختار ۳۴ نوع حشره کش پیرتروئیدی از دسته استرهای سیکلوپروپان کربوکسیلیک اسید با استفاده از سایت سیگما آلدریج <https://www.sigmaldrich.com> استخراج گردید و مقدار logP آنها از یک پایگاه حاوی داده های تجربی استخراج شد <http://chem.sis.nlm.nih.gov/chemidplus>. نام و مقادیر تجربی logP این ترکیبات در جدول ۱ فهرست شده اند.

### جدول ۱- نام و مقادیر تجربی logP مشتقات پیرتروئید

Table 1. The name and logP values of pyrethroid pesticides

No	Compound	logP	No	Compound	logP
1	Cyclopropanecarboxylic acid	0.63	18	Bioresmethrin	4.79
2	Cyclopropanecarboxylic acid, 1(phenylsulfonyl)-, methyl ester	1.43	19	Prallethrin	4.94
3	Levomilnacipran	1.71	20	Tralomethrin(Scout)	5
4	Cyclopropanecarboxylic acid, 1-((4methylphenyl)sulfonyl)-, methyl ester	1.77	21	Transfluthrin	5.46
5	Cyclopropanecarboxylic acid, 1-((4chlorophenyl)sulfonyl)-, methyl ester	2.03	22	Fenpropathrin	5.7
6	Cyclopropanecarboxylic acid, 1-((4methylphenyl)sulfonyl)-, ethyl ester	2.23	23	Cyclopropanecarboxylic acid, 2,2dimethyl-3-(2-methyl-1-propenyl)-, (3ethoxyphenyl)methyl ester, (1R,3R)-	5.87
7	Cyclopropanecarboxylic acid, 1-((4bromophenyl)sulfonyl)-, methyl ester	2.32	24	Pyrethrin I	5.9
8	Cyclopropanecarboxylic acid, 1-((4methylphenyl)sulfonyl)-, 1methylethyl ester	2.52	25	Cyfluthrin	5.95
9	Cyclopropanecarboxylic acid, 1-((4chlorophenyl)sulfonyl)-, 1methylethyl ester	2.64	26	Bifenthrin	6
10	Cyclopropanecarboxylic acid, 1-((4chlorophenyl)sulfonyl)-, 1,1dimethylethyl ester	2.68	27	Phenothrin	6.01
11	Imiprothrin	2.9	28	beta-Cypermethrin	6.05
12	Cyclanilide	3.25	29	Resmethrin	6.14
13	Methyl 3-(2,2-dichlorovinyl)-2,2dimethylcyclopropanecarboxylate	3.63	30	Deltamethrin	6.2
14	Cyclopropanecarboxylic acid, 2,2dimethyl-3-(2-methyl-1-propenyl)-, (3-(methylsulfonyl)phenyl)methyl ester, (1R-trans)-	3.83	31	Cyphenothrin	6.29
15	Pyrethrin II	4.3	32	Permethrin	6.5
16	Tetramethrin	4.73	33	Cypermethrin	6.6
17	Allethrin	4.78	34	Cyhalothrin	6.8

<sup>1</sup> Support Vector Machine

با استفاده از نرم افزار گوس ویو ۵،۰<sup>۱</sup> و گوسین ۲۰۹، ساختار شیمیایی همه ترکیبات مورد مطالعه ترسیم و بهینه شدند [Frisch, et al., 2009]. بهینه سازی ترکیبات با دستور محاسبات کوانتمی B<sub>3</sub>LYP<sup>۲</sup> [Tsuneda, 2014] و سری مجموعه پایه-6\*311G [Tsuneda & Hirao, 2014]. سپس ساختار بهینه شده تمامی ترکیبات به نرم افزار دراگون ۵،۴<sup>۳</sup> [Todeschini, et al., 2005; Mauri, et al., 2006]. انتقال داده شد تا انواع توصیفگرهای مولکولی مانند شاخص های ساختاری، مقادیر ویژه بار، مجاورت لبه، هندسی، دو و سه بعدی، خودهمبستگی، اطلاعاتی و همچنین خواص مولکولی با استفاده از آن ها استخراج گردد. توصیفگرهای استخراج شده، به نرم افزار متلب ۲۰۱۷<sup>۴</sup> [Todeschini & Consonni, 2008; Todeschini & Consonni, 2010] انتقال داده شدند. سپس با کمک الگوریتم ژنتیک، توصیفگرهای نامناسب که با هم همخطی داشتند و نیز توصیف کننده های با مقادیر مشابه و یکسان حذف گردیدند. الگوریتم ژنتیک ابزاری مناسب برای بهینه سازی تعداد توصیفگرها و به حداقل رسانیدن تعداد آنها برای استخراج مناسبترین آنهاست. این الگوریتم، از فرآیند تکامل بیولوژیکی و قوانین فیزیک الهام گرفته شده و از مفهوم بقای بهترین ها برای انتخاب توصیفگرهای مناسب در مدل سازی QSPR استفاده می کند [Lin, et al., 2018; Niazi & Leardi, 2012; Sarkhosh, et al., Leardi, 2003; Niazi & Leardi, 2012]. بعد از استخراج مناسبترین توصیفگرها، مدل سازی QSPR با کمک نرم افزار آماری اس پی اس ورژن ۲۲<sup>۵</sup> [Randic & Basak, 2000; Hocking, 2013] با استفاده از تکنیک ژنتیک الگوریتم-رگرسیون خطی چندگانه بروش برگشتی<sup>۶</sup> GA-BWMLR انجام گرفت.

### روش رگرسیون خطی چندگانه

روش رگرسیون خطی چندگانه یکی از تکنیک های پیچیده آماری است که به پیش بینی مقدار یک متغیر وابسته بر اساس مشارکت دو یا چند متغیر مستقل می پردازد. در این پژوهش، مقادیر logP مشتقات پیرتروئید بعنوان متغیر وابسته و توصیفگرهای مولکولی نقش متغیرهای مستقل را دارا می باشند. بعد از انتقال ساختار بهینه شده 34 نوع ترکیب مورد مطالعه به نرم افزار دراگون، تعداد ۶۰۶ توصیفگر مولکولی بدست آمد. با حذف توصیفگرهای نامناسب به کمک برنامه متلب و روش ژنتیک الگوریتم<sup>۷</sup>، ۲۲ توصیفگر استخراج گردیدند که برای مدل سازی QSPR توسط نرم افزار اس پی اس اس، به روش رگرسیون خطی چندگانه برگشتی بکار گرفته شدند. چهارده مدل با استفاده از روش گام به گام به دست آمد. تعداد توصیفگرها از مدل اول تا آخرین مدل بدست آمده، از ۲۲ به ۹ توصیفگر کاهش یافت. پارامترهای آماری مربوط به آنالیز ترکیبات مورد مطالعه برای همه مدل های بدست آمده با کمک نرم افزار اس پی اس اس، شامل: ضریب تعیین<sup>۸</sup> R<sup>2</sup> [Seber & Lee, 2003; Shanmuganathan, 2016]، ضریب

<sup>1</sup> Gauss View (5.0)

<sup>2</sup> Gaussian (09)

<sup>3</sup> B<sub>3</sub>LYP

<sup>4</sup> Dragon(4,5)

<sup>5</sup> MATLAB 2010a

<sup>6</sup> Statistical package for social science(22)

<sup>7</sup> Genetic Algorithm - multiple linear regression analysis returned

<sup>8</sup> Genetic Algorithm

<sup>9</sup>Coefficient of determination

همبستگی<sup>۱</sup> R، ضریب تعیین تعدیل یافته<sup>۲</sup> R<sup>2</sup>adj، آماره فیشر<sup>۳</sup> F، سطح معناداری<sup>۴</sup> Sig [Chai & Draxler, 2014]، خطای مربعات میانگین<sup>۵</sup> (MSE) و جذر خطای میانگین مربعات<sup>۶</sup> (RMSE) در جدول (۲) آورده شده اند.

جدول ۲. ضرایب آماری مدل های بدست آمده logP با استفاده از نرم افزار SPSS

Table2. The statistical coefficients of obtained models by using SPSS software

Model	Independent variables	R	R <sup>2</sup>	R <sup>2</sup> adj	RMSE	MSE	F	Sig
1	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor08v, Mor13u, Mor25u, BEHv2, GATS4p, HATS3v, HATS4v, Mor31u, HATS4u, Vindex, PW3, H1v, S3K	.986 <sup>a</sup>	0.971	0.914	0.534	0.285	16.924	0.000
2	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor13u, Mor25u, BEHv2, GATS4p, HATS3v, HATS4v, Mor31u, HATS4u, Vindex, PW3, H1v, S3K	.986 <sup>b</sup>	0.971	0.921	0.510	0.261	19.323	0.000
3	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor13u, Mor25u, BEHv2, GATS4p, HATS3v, Mor31u, HATS4u, Vindex, PW3, H1v, S3K	.986 <sup>c</sup>	0.971	0.927	0.492	0.242	21.933	0.000
4	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor13u, Mor25u, BEHv2, GATS4p, HATS3v, HATS4u, Vindex, PW3, H1v, S3K	.985 <sup>d</sup>	0.971	0.932	0.474	0.225	24.821	0.000
5	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor13u, Mor25u, BEHv2, GATS4p, HATS3v, HATS4u, PW3, H1v, S3K	.985 <sup>e</sup>	0.971	0.936	0.460	0.212	27.839	0.000
6	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor25u, BEHv2, GATS4p, HATS3v, HATS4u, PW3, H1v, S3K	.985 <sup>f</sup>	0.970	0.938	0.451	0.204	30.524	0.000
7	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, G1m, ZM1V, Mor25u, BEHv2, GATS4p, HATS4u, PW3, H1v, S3K	.984 <sup>g</sup>	0.969	0.940	0.446	0.199	33.190	0.000
8	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, ZM1V, Mor25u, BEHv2, GATS4p, HATS4u, PW3, H1v, S3K	.983 <sup>h</sup>	0.967	0.939	0.448	0.201	34.992	0.000
9	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v, S3K	.981 <sup>i</sup>	0.963	0.935	0.463	0.215	34.907	0.000
10	BLTF96, X2A, R3m, IVDE, MEcc, HATS1e, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v	.981 <sup>j</sup>	0.962	0.938	0.455	0.207	39.107	0.000
11	BLTF96, X2A, R3m, IVDE, MEcc, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v	.979 <sup>k</sup>	0.959	0.936	0.460	0.212	41.239	0.000
12	BLTF96, X2A, R3m, IVDE, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v	.978 <sup>l</sup>	0.956	0.933	0.470	0.221	42.966	0.000
13	BLTF96, X2A, R3m, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v	.976 <sup>m</sup>	0.953	0.932	0.474	0.225	46.330	0.000
14	BLTF96, R3m, H0v, ZM1V, Mor25u, BEHv2, GATS4p, PW3, H1v	.976 <sup>n</sup>	0.952	0.934	0.468	0.219	52.718	0.000

<sup>1</sup> Correlation coefficient

<sup>2</sup> R<sup>2</sup> adjusted

<sup>3</sup> Fisher

<sup>4</sup> Significance level

<sup>5</sup> Mean Squared Error

<sup>6</sup> Root Mean Squared Error

در معادله ۱، تعداد نه توصیف کننده مولکولی و ضرایب تاثیر آنها و پارامترهای آماری مدل چهاردهم به صورت زیر بدست آمده است:

$$\text{LogP} = -2.990 + 0.006 \text{ ZM1V} - 12.726 \text{ PW3} + 1.032 \text{ GATS4P} + 1.740 \text{ BEHV2} + 1.17 \text{ Mor25u} + 4.476 \text{ H0v} - 5.464 \text{ H1v} - 0.570 \text{ R3m} - 0.578 \text{ BLTF96} \quad (\text{معادله ۱})$$

$$R = 0.976; R^2 = 0.952; R^2_{\text{adj}} = 0.934; \text{RMSE} = 0.468; F = 52.718; \text{DW} = 1.634; \text{MSE} = 0.219; \text{Sig} = 0.00$$

سطح معناداری  $(\text{sig}) < 0.05$  به عنوان معیاری از قدرت مدل بدست آمده نشان داد که رابطه بین  $\log P$  و توصیفگرها از نظر آماری معنادار است. مقدار قابل قبول برای تغییرات آماره دوربین واتسون (DW) بین صفر تا ۴ است. مقادیر صفر، دو و چهار به ترتیب نشان دهنده همبستگی مثبت، غیر خود همبستگی و همبستگی منفی میان توصیفگرهاست [Kim, 2022; Mohammaei, & Mohammadinasab, 2018]. مطابق جدول ۲ مقدار  $\text{DW} = 1.634$  و آماره فیشر  $F = 52.718$  در مدل چهاردهم، نشان دهنده عدم وجود خود همبستگی بین توصیفگرهاست. مقادیر ضریب تعیین رگرسیون ( $R^2$ ) برازندگی مدل رگرسیون را نشان میدهد و نسبت واریانس را در متغیر وابسته تعیین میکند. مقدار  $R^2$  نشان میدهد که مدل رگرسیون تا چه اندازه بر داده های تجربی منطبق است. بنابراین، مقدار  $R^2$  نزدیک به یک به طور کلی نماد خوبی برای یک مدل پیش بینی است. مقدار  $R^2 = 0.952$  نشان داد که توصیفگرهای اضافه شده به مدل به درستی انتخاب گردیده اند. همچنین، جذرمیانگین مربعات خطا (RMSE) یا خطای پیش بینی، پراکندگی داده ها را از خط رگرسیون نشان داد و برای تأیید نتایج تجربی استفاده گردید.

به منظور مطالعه همبستگی بین توصیفگرها، رابطه خطی بین متغیرهای مستقل از طریق بررسی مقادیر ضریب تورم واریانس (VIF) [Kelley & Lai, 2011; Khan, et al., 2021] و ضرایب همبستگی پیرسون (PCC) بین نه توصیفگر بدست آمده در معادله ۱ [Goodarzi, et al., 2013; Chirico & Gramatica, 2011] مورد بررسی قرار گرفت. نتایج این بررسی بین توصیفگرها در جداول ۳ و ۴ نشان داده شده است.

جدول ۳ مقادیر ضرایب تولرنس و واریانس توصیفگرها

Table 3. The values of variance inflation factor and tolerance of descriptors

		Model 1	Model 2	Model ۳	Model ۴	Model ۵	Model ۶	Model ۷
Descriptor	Tolerance	VIF	VIF	VIF	VIF	VIF	VIF	VIF
ZM1V	۰.۲۶۰۰	۲.۲۹۰۰	۰.۲۰۲۶	2.299	1.799	1.741	1.405	1.190
PW3	۰.۲۰۹۹	۰.۲۰۰۰	۰.۲۰۱۰	2.913	1.586	1.537	1.102	1.088
GATS4p	۰.۲۰۰۶	۰.۲۲۰۶	۰.۲۲۹۶	3.676	2.016	1.593	1.101	1.100
BEHV2	۰.۲۶۱۰	۱.۲۶۶۲	۰.۲۰۲۱	4.002	3.093	3.093	-	-
Mor25u	۰.۲۹۰۰	۶.۹۲۱۷۰	۷.۲۲۹۰	7.495	-	-	-	-
H0v	۰.۲۶۶۰	۲.۲۷۲۹	۰.۲۰۰۰	2.399	2.044	1.343	1.251	-
H1v	۰.۲۹۲۹	۶.۰۲۰۲۶	-	-	-	-	-	-
R3m	۰.۲۰۷۱	۰.۲۱۲۰	۰.۲۱۲۰	2.317	2.316	-	-	-
BLTF96	۰.۲۶۰۹	۷.۲۶۰۲	۱.۲۶۷۰	-	-	-	-	-

جدول ۴ مقادیر ضرایب همبستگی پیرسون بین توصیفگرها

Table 4. The values of pearson correlation coefficients between descriptors

Descriptor	BLTF96	PW3	R3m	H0v	GATS4p	BEHv2	ZM1V	Mor25u	H1v
BLTF96	1.000								
PW3	-0.248	1.000							
R3m	0.314	-0.221	1.000						
H0v	0.312	0.045	-0.309	1.000					
GATS4p	0.368	-0.430	0.427	0.007	1.000				
BEHv2	0.401	-0.491	0.048	0.469	0.049	1.000			
ZM1V	0.734	-0.324	0.112	0.321	0.423	0.259	1.000		
Mor25u	0.175	0.617	-0.016	0.295	-0.562	-0.031	-0.103	1.000	
H1v	-0.367	-0.146	0.047	-0.852	0.080	-0.547	-0.328	-0.545	1.000

بر اساس ضرایب همبستگی پیرسون و واریانس که در جداول ۳ و ۴ نشان داده شده است، همبستگی قوی بین توصیفگرهای بدست آمده در آخرین مدل وجود دارد. لذا اقدام به حذف مرحله به مرحله توصیفگرها گردید. ابتدا توصیفگری حذف شد که در ضمن اینکه بالاترین ضریب همبستگی پیرسون را با توصیفگرهای دیگر دارد، ضریب واریانس بالاتری هم دارا باشد. با بررسی های به عمل آمده مشخص گردید که توصیفگرهای حذف شده به ترتیب عبارتند از: BLTF96, H1V, H0V, Mor25u, BEHV2, R3m. بنابر این معادله نهایی بدست آمده همراه با ضرایب رگرسیون مدل نهایی به شرح زیر بدست آمد:

$$\text{LogP} = 5.197 + 2.193 \text{ GATS4P} + 0.012 \text{ ZM1V} - 20.371 \text{ PW3} \quad (\text{معادله ۲})$$

$$N_{\text{total}}=34; R=0.928; R^2=0.862; R^2_{\text{adj}}=0.848; F=62.296; DW=2.006; \text{MSE}=0.503; \text{Sig}=0.00$$

ضرایب رگرسیون بدست آمده با حذف 6 توصیفگر نامناسب نشان داد معادله ۲ علی رغم در برداشتن تعداد توصیفگر کمتر، دارای توانایی پیش بینی مناسب می باشد. این معادله نشان داد که تغییرات مقدار توصیفگر GATS4P و ZM1V تاثیر مستقیم بر روی logP مشتقات پیرتروئید گذاشته و با افزایش مقدار این توصیفگرها مقدار logP هم زیاد می شود. در حالی که علامت منفی توصیفگر PW3 در معادله مذکور نشان دهنده اثر معکوس این توصیفگر بر روی مقدار logP ترکیبات مورد مطالعه می باشد. توصیفگرهای PW3 و ZM1V از دسته توصیفگرهای ساختاری و توصیفگر GATS4P جزء توصیفگرهای خودهمبستگی 2D می باشد. با توجه به نتایج درج شده در جدول ۴ معلوم می گردد که توصیفگرهای ساختاری<sup>۱</sup> و خودهمبستگی 2D نسبت به دیگر توصیفگرهای مولکولی از اهمیت بیشتری برای مدلسازی logP مشتقات پیرتروئید برخوردار می باشند. تعریف و نوع سه توصیف گر نهایی در جدول ۵ شرح داده شده است.

<sup>1</sup> Topological



## جدول ۵ شرح و نوع توصیف گرهای نهایی

Table 5. The name, block and explanation of final descriptors

توصیفگر	شرح	نوع توصیفگر
PW3	path/walk 3 - Randic shape index	Topological indices
GATS4P	Geary autocorrelation of lag 4 weighted by polarizability	2D autocorrelations
ZM1V	first Zagreb index by valence vertex degrees	Topological indices

## اعتبارسنجی

پارامترهای کیفی مورد استفاده برای ارزیابی اعتبار مدل باید برای توانایی پیش بینی مدل بدست آمده را داشته باشد. پارامترهای اعتبارسنجی در رگرسیون خطی چندگانه مشخص می کنند که چگونه متغیرهای مستقل می توانند متغیر وابسته را توضیح دهند و برای پیش بینی مشاهدات جدید استفاده می شوند. توانایی خوب مدل با روش های اعتبارسنجی داخلی و خارجی نشان داده شد. در این مطالعه، در اعتبارسنجی خارجی، گروه داده ها شامل 34 مولکول به یک گروه آموزشی شامل 25 ترکیب (۷۵ درصد) و یک گروه آزمایشی شامل 9 ترکیب (۲۵ درصد) تقسیم شدند.

قابلیت پیش بینی مجموعه آزمایشی برای اندازه گیری اعتبار مدل ساخته شده مورد آزمایش قرار گرفت. در اعتبارسنجی داخلی توانایی پیش بینی مدل غالباً برحسب ضریب  $Q^2$  محاسبه می شود که با روش هایی مانند ترک یک خروجی (LOO)<sup>۱</sup> و یا ترک تعداد خروجی بیشتر (LMO)<sup>۲</sup> محاسبه می شود [Roy, et al., 2015; Saghiaie, et al., 2013]. اعتبارسنجی متقاطع (LOO) شناخته شده ترین و مورد استفاده ترین روش برای تعیین اعتبار داخلی مدل بکار رفته است. این روش توانایی مدل را برای پیش بینی مواد شیمیایی در مجموعه داده ها یک به یک ارزیابی می کند و آنها را به طور مکرر در مجموعه آزمایش قرار می دهد. فرمول پارامتر محاسبه شده  $Q_{LOO}^2$  که به طور گسترده توسط اکثر مدل سازان QSPR پذیرفته شده و اعمال می شود به صورت معادله (۳-۹) می باشد.

$$Q_{LOO}^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_{pred} - y_{exp})^2}{\sum_{i=1}^n (y_{exp} - y_{ave})^2} =$$

فرمول های محاسبه پارامترهای آماری:  $R^2$ , RMSE در زیر آورده شده است. در این معادلات منظور از  $y_{obs}$ ,  $y_{pred}$  and  $y_{ave}$  به ترتیب مقدار  $\log P$  میانگین، پیش بینی و مشاهده می باشد و  $n$  اشاره به تعداد داده ها دارد.

<sup>1</sup> Leave-One-Out (LOO) method<sup>2</sup> Leave-Many-Out (LMO) technique

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{obs} - y_{pred})^2}$$

مقادیر محاسبه شده مجذور ضریب رگرسیون ( $R^2$ )، میانگین مربعات خطا (MSE)،  $R^2_{adj}$ ، ریشه میانگین مربعات خطا (RMSE)، برای مجموعه داده های آموزشی و آزمایشی و همچنین مقدار  $Q^2_{LOO}$  (برای کل داده ها) در مدل نهایی پیشنهادی در جدول ۶ گزارش شده است.

جدول ۶ مقادیر ضرایب اعتبارسنجی

**Table 6. The values of validation coefficients**

	Total MLR	Train MLR	Test MLR
N	34	25	9
$R^1$	0.862	0.880	0.931
$R^2_{adj}$	0.848	0.862	0.890
MSE	0.503	0.482	0.340
RMSE	0.709	0.694	0.583
$Q^2$ (date set)	0.861	-	-

مقدار ضریب رگرسیون  $R=0.964$  در سری آزمایش، میزان ارتباط خطی بین مقادیر logP و توصیفگرها را بیان می کند. مقدار بدست آمده برای ضریب تعیین در سری آزمایش نشان می دهد که ۹۳٫۱٪ تغییرات logP به وسیله توصیف گرهای وابسته به آن تبیین می شوند.  $R^2_{adj}$  مقدار ضریب رگرسیون تعیین را با توجه به متغیرهای مستقل اضافه شده به خط رگرسیون و با توجه به عرض از مبدأهای جدید، تعدیل و اصلاح می کند. مقدار محاسبه شده برای RMSE بیانگر میزان تفاوت ناچیز میان مقدار پیش بینی شده logP توسط مدل با مقدار تجربی آن می باشد. مقادیر بالای پارامترهای  $Q^2$ ،  $R^2$  نشانه رضایت بخشی مدل رگرسیون خطی چندگانه برای پیش بینی مقادیر logP مشتقات پیرتروئید و نیز انتخاب صحیح توصیفگرهای مدل انتخابی می باشد. همچنین مقادیر بسیار کم خطای RMSE مطابقت بسیار خوب مقادیر تجربی با مقادیر پیش بینی را نشان می دهد.

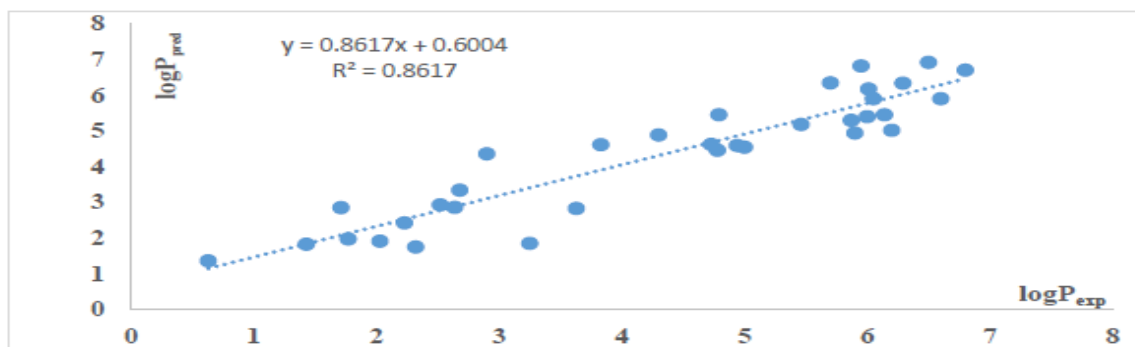
جدول ۷ مقادیر تجربی، پیش بینی شده و باقیمانده logP

**Table 7. The values of experimental, predicted and residual logP**

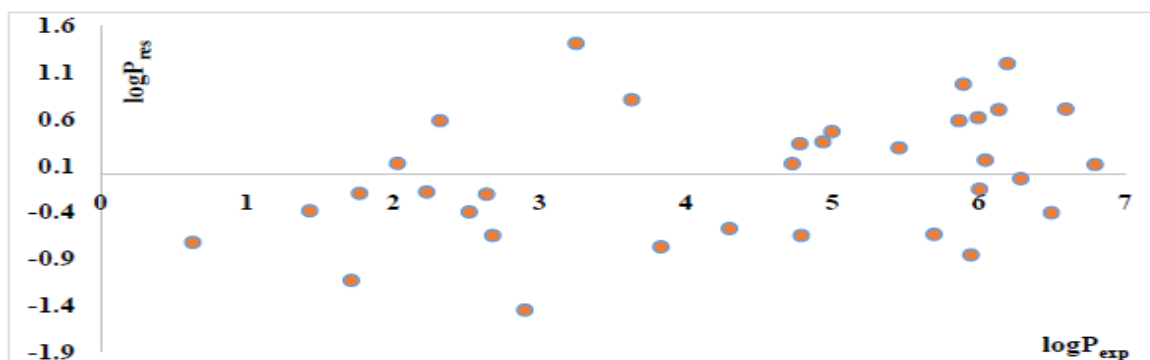
No	logPexp	logPpred	logPres
1.	0.63	1.35689	-0.72689
2.	1.43	1.81889	-0.38889
3.	1.713	2.84732	-1.13432
4.	1.77	1.96995	-0.19995
5.	2.03	1.91053	0.11947

6.	2.23	2.4156	-0.1856
7.	2.32	1.74199	0.57801
8.	2.52	2.92078	-0.40078
9.	2.64	2.85039	-0.21039
10.	2.68	3.33364	-0.65364
11.	2.9	4.35472	-1.45472
12.	3.25	1.84256	1.40744
13.	3.63	2.82471	0.80529
14.	3.83	4.60674	-0.77674
15.	4.3	4.88026	-0.58026
16.	4.73	4.61315	0.11685
17.	4.78	4.44819	0.33181
18.	4.79	5.442	-0.652
19.	4.94	4.58735	0.35265
20.	5	4.53902	0.46098
21.	5.46	5.17396	0.28604
22.	5.7	6.34068	-0.64068
23.	5.87	5.29243	0.57757
24.	5.9	4.92927	0.97073
25.	5.95	6.81063	-0.86063
26.	6	5.38761	0.61239
27.	6.01	6.1665	-0.1565
28.	6.05	5.89548	0.15452
29.	6.14	5.442	0.698
30.	6.2	5.01002	1.18998
31.	6.29	6.33236	-0.04236
32.	6.5	6.90894	-0.40894
33.	6.6	5.89548	0.70452
34.	6.8	6.69294	0.10706

میزان پراکندگی نقاط برای دسته های آموزش و آزمون در نمودار مشاهده شده برحسب پیش بینی شده حاکی از همبستگی خوب بین توصیف گرهای نهایی با  $\log P$  و در نتیجه کارآیی روش به کاررفته در پیش بینی  $\log P$  مشتقات پیرتروئید می باشد. همچنین نمودار تغییرات مقادیر مشاهده شده برحسب باقیمانده نشان می دهد که خطاها در اطراف محور X توزیع یکسانی دارند و این نمایانگر این است که الگوی مذکور برای پیش بینی  $\log P$  در ترکیبات مورد مطالعه از دقت بالا و بسیار خوبی برخوردار می باشد. شکل های ۱ و ۲ که به روند تغییرات مقادیر تجربی برحسب پیش بینی و نیز نحوه تغییرات مقادیر تجربی برحسب مقادیر باقیمانده اشاره میکند حاکی از انتخاب درست مدل بکار رفته و نیز انطباق بسیار خوب مقادیر مشاهده و محاسبه می باشد.



شکل ۱- نمودار تغییرات مقادیر  $\log P$  تجربی درمقابل پیش بینی  
 Fig 1. The curve of the experimental versus the predicted  $\log P$



شکل 2 نمودار مقادیر  $\log P$  تجربی برحسب باقیمانده  
 Fig 2. The curve of the experimental versus the residual  $\log P$

## نتیجه گیری

هدف اصلی این مطالعه پیش‌بینی شاخص  $\log P$  ساختارهای مولکولی پیرتروئیدها با استفاده از مدل QSPR می‌باشد. در حال حاضر هیچ‌گونه مدلی بر اساس تکنیک QSPR برای پیش‌بینی شاخص  $\log P$  مشتقات پیرتروئیدی مورد مطالعه وجود ندارد. توصیفگرهای PW3، ZM1V و GATS4P می‌توانند ماهیت فیزیکوشیمیایی  $\log P$  پیرتروئیدها را توجیه کنند. در مدلسازی به روش رگرسیون خطی چندگانه مقدار پارامتر  $Q^2$  در اعتبارسنجی داخلی و همچنین مقادیر  $R^2$ ،  $R^2_{adj}$ ،  $RMSE$  در اعتبارسنجی خارجی برای نمونه‌های آموزشی و آزمایشی، انطباق خوب مقادیر پیش‌بینی و تجربی  $\log P$  را نشان داد. روش رگرسیون خطی چندگانه، با مقادیر اعتبارسنجی عالی، دقت و راندمان بالا و پایداری خوب، می‌تواند  $\log P$  مشتقات پیرتروئیدی را پیش‌بینی کند، که نشان‌دهنده دقت بالا و خطای پایین این مدل در پیش‌بینی  $\log P$  این ترکیبات می‌باشد. نتایج این تحقیق نشان داد که روش رگرسیون خطی از میزان توانایی بسیار خوب و قابل ملاحظه‌ای برای پیش‌بینی ارتباط بین متغیر وابسته و متغیرهای مستقل برخوردار بوده است و قابلیت تعمیم برای مدل‌سازی خواص ترکیبات مشابه را دارد.

## Reference

- Aloizou, A.M., Siokas, V., Vogiatzi, C., Peristeri, E., Docea, A.O., Petrakis, D., Provatas, A., Folia, V., Chalkia, C., Vinceti, M., Wilks, M. 2020.** Toxicology letters, 326: 31-51.
- Amiri, R., Messadi, D. and Bouakkadia, A. 2019.** QSPR Models for the prediction of octanol/water partition coefficient of organophosphorous insecticides. Egyptian Journal of Chemistry, 62(9): 1563-1574.
- Amiri, R., Messadi, D. and Bouakkadia, A. 2020.** QSAR Study of the octanol/water partition coefficient of organophosphorus compounds: The hybrid GA/MLR and GA/ANN approaches. Journal of the Serbian Chemical Society, 85(4): 467-480.
- Aznar-Aleman, Ö. and Eljarrat, E. 2020.** Introduction to pyrethroid insecticides: chemical structures, properties, mode of action and use. Pyrethroid Insecticides, 1-16.
- Bahmani, A., Saaidpour, S. and Rostami, A. 2017.** A simple, robust and efficient computational method for n-octanol/water partition coefficients of substituted aromatic drugs. Scientific Reports, 7(1): 1-14.
- Bermudez-Saldana, J. M. and Cronin, M. T. 2006.** Quantitative structure–activity relationships for the toxicity of organophosphorus and carbamate pesticides to the Rainbow trout *Onchorhynchus mykiss*. Pest Management Science: formerly Pesticide Science, 62(9): 819-831.
- Bouakkadia, A., Lourici, L. and Messadi, D. 2017.** Modeling and prediction of octanol/water partition coefficient of pesticides using QSPR methods. Management of Environmental Quality: An International Journal, 28(4): 579-592.
- Breckenridge, C.B., Holden, L., Sturgess, N., Weiner, M., Sheets, L., Sargent, D., Soderlund, D.M., Choi, J.S., Symington, S., Clark, J.M. and Burr, S. 2009.** Evidence for a separate mechanism of toxicity for the Type I and the Type II pyrethroid insecticides. Neurotoxicology, 30 :S17-S31.
- Cappelli, C. I., Benfenati, E. and Cester, J. 2015.** Evaluation of QSAR models for predicting the partition coefficient (log P) of chemicals under the REACH regulation. Environmental Research, 143: 26-32.
- Casida, John E., and Gary B. Quistad. 1998.** Golden age of insecticide research: past, present, or future? Annual review of entomology 43(1): 1-16.
- Chai, T. and Draxler, R. R. 2014.** Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. Geoscientific model development, 7(3): 1247-1250.
- Chirico, N. and Gramatica, P. 2011.** Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. Journal of chemical information and modeling, 51(9): 2320-2335.
- Chrutek, A., Holynska-Iwan, I., Dziembowska, I., Bogusiewicz, J., Wroblewski, M., Cwynar, A. and Olszewska-Slonina, D. 2018.** Current research on the safety of pyrethroids used as insecticides. Medicina, 54(4): 61.
- Cycoń, M. and Piotrowska-Seget, Z. 2016.** Pyrethroid-degrading microorganisms and their potential for the bioremediation of contaminated soils: a review. Frontiers in microbiology, 7: 1463.
- Ford, M. G., Greenwood, R., Turner, C. H., Hudson, B. and Livingstone, D. J. 1989.** The structure/activity relationships of pyrethroid insecticides. 1. A novel approach based upon the use of multivariate QSAR and computational chemistry. Pesticide science, 27(3): 305-326.
- Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A. and Nakatsuji, H. 2009.** Revision D. 01 Gaussian. Inc., Wallingford CT.
- Golmohammadi, H. 2009.** Prediction of octanol–water partition coefficients of organic compounds by multiple linear regression, partial least squares, and artificial neural network. Journal of computational chemistry, 30(15): 2455-2465.
- Goodarzi, M., Saeys, W., Deeb, O., Pieters, S. and Vander Heyden, Y. 2013.** Particle swarm optimization and genetic algorithm as feature selection techniques for the QSAR modeling of imidazo [1, 5.a] pyrido [3, 2 .e] pyrazines, inhibitors of phosphodiesterase 10 A. Chemical biology & drug design, 82(6): 685-696.

- Hocking, R. 2013.** Methods and Applications of Linear Models: Regression and the Analysis of Variance. New York: John Wiley & Sons.  
<https://www.sigmaaldrich.com>.
- Katsuda, Y. 2012.** Progress and future of pyrethroids. Pyrethroids: From Chrysanthemum to Modern Industrial Insecticide, 1-30.
- Kelley, K. and Lai, K. 2011.** Accuracy in parameter estimation for the root mean square error of approximation: sample size planning for narrow confidence intervals. *Multivariate Behavioral Research*, 46(1): 1-32.
- Khambay, B. P. S., and P. J. Jewess. 2005.** Pyrethroids. 1-29.
- Khan, H., Hussain, S., Hussain, S. F., Gul, S., Ahmad, A. and Ullah, S. 2021.** Multivariate modeling and optimization of Cr (VI) adsorption onto carbonaceous material via response surface models assisted with multiple regression analysis and particle swarm embedded neural network. *Environmental Technology & Innovation*, 24: 101952.
- Kim, H. 2022.** A finite sample correction for the panel Durbin–Watson test. *Applied Economics*, 54(28): 3197-3205.
- Lanevskij, K. and Didziapetris, R. 2019.** Physicochemical QSAR analysis of passive permeability across Caco-2 monolayers. *Journal of Pharmaceutical Sciences*, 108(1): 78-86.
- Leardi, R. 2003.** Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks, Amsterdam.
- Lin, W., Jiang, R., Shen, Y., Xiong, Y., Hu, S., Xu, J. and Ouyang, G. 2018.** Effect of dissolved organic matter on pre-equilibrium passive sampling: A predictive QSAR modeling study. *Science of the Total Environment*, 635: 53-59.
- Lopez, K., Pinheiro, S., and Zamora, W. J. 2021.** Multiple linear regression models for predicting the n-octanol/water partition coefficients in the SAMPL7 blind challenge. *Journal of computer-aided molecular design*, 35: 923-931.
- Marettova, E., Marett, M. and Legáth, J. 2017.** Effect of pyrethroids on female genital system. *Review. Animal reproduction science*, 184: 132-138.
- Matsuda, K., Nishikawa, M., Hamada, M., Nishimura, K. and Fujita, T. 1991.** Quantitative structure-activity studies of pyrethroids: Depolarizing effect of pyrethroids on axonal membranes of American cockroaches. *Pesticide Biochemistry and Physiology*, 41(2): 170-177.
- Mauri, A., Consonni, V., Pavan, M. and Todeschini, R. 2006.** Dragon software: An easy approach to molecular descriptor calculations. *Match*, 56(2): 237-248.
- Mohammaei, F., and Mohammadinasab, E. 2018.** Coefficient Partition Prediction of Saturated Monocarboxylic Acids Using the Molecular Descriptors. *Journal of the Chilean Chemical Society*, 63(3): 4068-4071.
- Mudasir, M., Wibowo, Y.M. and Pranowo, H.D. 2016.** *Indian J. Chem*, 5: 54-68.
- Niazi, A. and Leardi, R. 2012.** Genetic algorithms in chemometrics. *Journal of Chemometrics*, 26(6): 345-351.
- Niazi, A., and Leardi, R. 2012.** Genetic algorithms in chemometrics. *Journal of Chemometrics*, 26(6): 345-351.
- Nishimura, K., Kobayashi, T. and Fujita, T. 1991.** Quantitative structure-activity studies of pyrethroids: 23. Rate of decrease in sodium tail currents induced by substituted benzyl chrysanthemates and pyrethrates in crayfish giant axons. *Pesticide Biochemistry and Physiology*, 40(2): 99-110.
- Palmquist, K., Salatas, J., Fairbrother, A. and Perveen, F. 2012.** Insecticides—Advances in integrated pest management. 1-29.
- Randic, M. and Basak, S. C. 2000.** Multiple regression analysis with optimal molecular descriptors. SAR and QSAR in environmental research, 11(1): 1-23.
- Ravula, A. R. and Yenugu, S. 2021.** Pyrethroid based pesticides—chemical and biological aspects. *Critical Reviews in Toxicology*, 51(2): 117-140.
- Roy, K., Kar, S., Das, R. N., Roy, K., Kar, S. and Das, R. N. 2015.** Statistical methods in QSAR/QSPR. A Primer on QSAR/QSPR Modeling: Fundamental Concepts, 37-59.

- Saghaie, L., Sakhi, H., Sabzyan, H., Shahlaei, M. and Shamshirian, D. 2013.** Stepwise MLR and PCR QSAR study of the pharmaceutical activities of antimalarial 3-hydroxypyridinone agents using B3LYP/6-311++ G\*\* descriptors. *Medicinal Chemistry Research*, 22: 1679-1688.
- Sarkhosh, M., Khorshidi, N., Niazi, A. and Leardi, R. 2014.** Application of genetic algorithms for pixel selection in multivariate image analysis for a QSAR study of trypanocidal activity for quinone compounds and design new quinone compounds. *Chemometrics and Intelligent Laboratory Systems*, 139: 168-174.
- Schleier III, J. J. and Peterson, R. K. 2011.** Pyrethrins and pyrethroid insecticides. *Green trends in insect control*, 11: 94-131.
- Seber, G. A. and Lee, A. J. 2003.** Linear regression analysis (Vol. 330). John Wiley & Sons.
- Shafer, T.J., Meyer, D.A. and Crofton, K.M. 2005.** Developmental neurotoxicity of pyrethroid insecticides: critical review and future research needs. *Environmental health perspectives*, 113(2): 123-136.
- Shanmuganathan, A. 2016.** Artificial neural network modelling: In An introduction. Springer, Cham, 1-14.
- Souyei, B., Seyd, A. H., Zaiz, F. and Rebiai, A. 2019.** Application of inverse QSAR/QSPR analysis for pesticides structures generation. *Acta Chimica Slovenica*, 66(2): 315-325.
- Todeschini, R. and Consonni, V. 2008.** Handbook of molecular descriptors. John Wiley & Sons.
- Todeschini, R. and Consonni, V. 2010.** Molecular descriptors. *Recent Advances in QSAR Studies*, 29-102.
- Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. 2005.** DRAGON Software – version 5.4 TALETE srl.
- Toropov, A. A., Toropova, A. P., Cappelli, C. I. and Benfenati, E. 2015.** CORAL: Model for octanol/water partition coefficient. *Fluid Phase Equilibria*, 397: 44-49.
- Tsuneda, T. 2014.** Density functional theory in quantum chemistry, 978-984.
- Tsuneda, T. and Hirao, K. 2014.** Self-interaction corrections in density functional theory. *The Journal of chemical physics*, 140(18): 18A513.
- U.S. National Library of Medicine: <http://chem.sis.nlm.nih.gov/chemidplus/>.**
- Vacondio, F., Silva, C., Lodola, A., Fioni, A., Rivara, S., Duranti, A., Tontini, A., Sanchini, S., Clapper, J.R., Piomelli, D. and Mor, M. 2009.** Structure–Property Relationships of a Class of Carbamate-Based Fatty Acid Amide Hydrolase (FAAH) Inhibitors: Chemical and Biological Stability. *ChemMedChem: Chemistry Enabling Drug Discovery*, 4(9): 1495-1504.
- Vacondio, F., Silva, C., Mor, M. and Testa, B. 2010.** Qualitative structure-metabolism relationships in the hydrolysis of carbamates. *Drug metabolism reviews*, 42(4): 551-589.
- Valko, K. L. 2016.** Lipophilicity and biomimetic properties measured by HPLC to support drug discovery. *Journal of Pharmaceutical and Biomedical Analysis*, 130: 35-54.
- Wang, S., Cheng, M., Zhou, L., Dai, Y., Dang, Y. and Ji, X. 2021.** QSPR modelling for intrinsic viscosity in polymer–solvent combinations based on density functional theory. *SAR and QSAR in Environmental Research*, 32(5): 379-393.
- Zapadka, M., Kaczmarek, M., Kupcewicz, B., Dekowski, P., Walkowiak, A., Kokotkiewicz, A., Luczkiewicz, M. and Bucinski, A. 2019.** An application of QSRR approach and multiple linear regression method for lipophilicity assessment of flavonoids. *Journal of Pharmaceutical and Biomedical Analysis*, 164: 681-689.
- Zhu, Q.; Yang, Y.; Zhong, Y.; Lao, Z.; O'Neill, P.; Hong, D.; Zhang, K.; Zhao, S. 2020.** *Chemosphere*, 254: 126779.
- Zuluaga, M., Robledo, S., Osorio-Zuluaga, G. A., Yathe, L., Gonzalez, D. and Taborda, G. 2016.** Metabolomics and pesticides: systematic literature review using graph theory for analysis of references. *Nova*, 14(25): 121-138.

## Study of quantitative structure–property relationship for predicting the logP of Cyclopropane carboxylic acid (pyrethroid) derivatives using multiple linear regression method

*M. Sadeghi<sup>1</sup>, E. Mohammadinasab<sup>2\*</sup>, T. Momeni Isfahani<sup>3</sup>*

1. PHD student, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
2. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran
3. Assistant Professor, Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran

### Abstract

In this research, predicting the logP of 43 types different pyrethroid derivatives was studied using quantitative structure-property relationship. The logP of studied pyrethroids was modeled using the genetic algorithm based on linear regression method (GA-MLR). It was found that the three effective descriptors GATS4P, PW3 and ZM1V have a reasonable correlation with logP, and led to the creation of a model with the highest regression coefficient and the lower error. The evaluation of GA-MLR model predictive ability for test set was performed by statistical parameters such as  $R^2 = 0.862$ ,  $R^2_{adj} = 0.848$ ,  $F = 62.296$  and  $MSE = 0.503$ . Also, the value of  $Q^2_{LOO} = 0.861$  using the cross-validation method, and the values of  $R^2 = 0.880$  and  $0.929$  for the training and test sets respectively, in the external validation method showed a very good correlation between experimental and prediction values. It was specified that the MLR model was reliable for predicting the logP of pyrethroid insecticides, and had sufficient accuracy with the lowest error.

**Keywords:** "logP"; "Multiple linear regression method"; "Quantitative structure-property relationship"; "Genetic Algorithm" ; "Pyrethroids".

---

\* Corresponding Author, E-mail: [Es.Mohammadinasab@iau.ac.ir](mailto:Es.Mohammadinasab@iau.ac.ir)

Received: 7 Dec. 2022 – Accepted: 31 Dec. 2022

