

پیش بینی فعالیت ضدسرطانی مشتقات ۸-ا - نفتیریدین توسط روش الگوریتم ژنتیک-رگرسیون خطی چندگانه

شهین احمدی^۱، روح الله خانی^۲، مریم مقدس^۳

^۱ استادیار دانشگاه گروه شیمی، واحد کرمانشاه، دانشگاه آزاد اسلامی، کرمانشاه، ایران

^۲ دانشجوی ارشد دانشگاه - گروه شیمی دارویی، علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران

^۳ استادیار دانشگاه - گروه شیمی، واحد صفادشت، دانشگاه آزاد اسلامی، صفادشت، ایران

چکیده

سابقه و هدف: این مطالعه به مقایسه مدل سازی QSAR فعالیت ضد سرطانی ترکیبات ۱-۴-دی هیدرو-۴-اکسو-۱- (۲-تيازولیل)-۸-ا-نفتیریدین و مشتقات آن با روش رگرسیون خطی چندگانه مرحله ای (S-MLR) و روش الگوریتم ژنتیک-رگرسیون خطی چندگانه (GA-MLR) پرداخت.

روش بررسی: مجموعه ای از ۱۰۰ ترکیب با فعالیت ضد سرطانی مشخص از مقاله معتبر بین المللی انتخاب شد و روش میدان نیروی آلینجر MM2 برای کمینه کردن انرژی مولکولها استفاده شد. ساختار هندسی مولکولها از طریق روش کوانتوم نیمه تجربی روش آوستین با استفاده از الگوریتم پلاک-ریبایر (Polak-Ribiere) با استفاده از نرم افزار موپک بهینه سازی شدند. تعداد زیادی از توصیفگرهای تئوری برای هر مولکول با استفاده از نرم افزار دراگون محاسبه شد. به منظور انتخاب بهترین دسته از توصیفگرها برای مدل سازی QSAR از دو روش انتخاب متغیر ترکیب الگوریتم ژنتیک-رگرسیون خطی چندگانه و رگرسیون خطی چندگانه مرحله ای استفاده شد. برای مدل سازی ابتدا نمونه برداری تصادفی دسته آموزش (۸۰ درصد از داده ها) ۲۰ بار به صورت تصادفی صورت گرفته و مولکولهای باقیمانده (۲۰ درصد باقیمانده از داده ها) به عنوان دسته پیشگویی برای اعتبارسنجی خارجی استفاده شدند. در میان نمونه های تصادفی، یکی از نمونه ها با بالاترین Q^2_{CV} ، Q^2_{cal} و Q^2_{test} به عنوان بهترین دسته یادگیری و آموزش انتخاب شد. با استفاده از این دسته یادگیری هر بار مدل به دو روش S-MLR و GA-MLR ایجاد شد.

یافته ها: مدل های QSAR به دست آمده با GA-MLR مجذور ضریب همبستگی اعتبارسنجی بزرگتری نسبت به روش S-MLR داشتند.

نتیجه گیری: نتایج این مقایسه نشان می دهد که می توان با استفاده از مدل حاصل، فعالیت ترکیبات ضد سرطانی مشابه را پیشگویی کرد.

واژگان کلیدی: مدل سازی QSAR، فعالیت ضدسرطان، انتخاب متغیر، GA-MLR، Stepwise-MLR

مقدمه

سرطان، بیماری هولناک ایدز، بیماری های مشترک انسان و دام و مقاوم شدن ویروس ها در برابر آنتی بیوتیک ها همه از جمله مواردی هستند که ذهن دانشمندان را در جهت یافتن داروهای مؤثر و کارآمد برای مقابله با این بیماری ها معطوف خود کرده اند (۱).

روندی که در گذشته منجر به کشف داروهای جدید می شد به روش آزمون و خطا صورت می گرفت که روش وقت گیر و هزینه بر است. مشکل دیگری که در این راه دانشمندان را آزار می دهد عدم اطلاع آن ها از فعالیت دارویی ترکیبات، قبل از انجام سنتز و بررسی تجربی آن ها است و به همین دلیل یکی از

یکی از مشکلاتی که جامعه بشری همیشه با آن روبه رو است، مقابله با انواع بیماری هایی است که سلامت انسان ها را به مخاطره انداخته و همواره یکی از مهم ترین دغدغه های محققان یافتن داروهای مؤثر، برای رفع این معضل و کاهش عوارض این بیماری ها بوده است. بروز انواع بیماری ها از قبیل

آدرس نویسنده مسئول: کرمانشاه، دانشگاه آزاد کرمانشاه، گروه شیمی تجزیه، شهین احمدی

(email: ahmadi.chemometrics@gmail.com)

تاریخ دریافت مقاله: ۹۶/۹/۲۷

تاریخ پذیرش مقاله: ۹۶/۱۱/۲۹

توصیفگرها روش الگوریتم ژنتیک است که بر اساس تکامل سیستم‌های بیولوژیکی است. روش الگوریتم ژنتیک برای پیدا کردن مینیمم‌های جهانی (Global minima) برای مسأله‌های چند بعدی مانند انتخاب توصیفگر در QSAR/QSPR زمانی که رویه‌ی پاسخ چندین بهینه موضعی (Local optima) دارد بسیار سودمند است (۸). لوکاسیوس (Lucasius) و همکارانش ثابت کردند که روش الگوریتم ژنتیک به طور کلی عملکرد بهتری نسبت به رگرسیون مرحله‌ای و تبرید تدریجی دارد (۹).

از جدیدترین تحقیقات در زمینه مدلسازی فعالیت ضد سرطانی، مطالعات مدلسازی QSAR ارائه شده توسط شایانفر و همکارانش در سال ۲۰۱۳، برای مهارکننده فarnesyltransferase (Farnesyltransferase) به عنوان دسته‌ای از داروهای ضد سرطان به کمک روشهای مختلف اشاره کرد (۱۰). در سال ۲۰۱۱، بوهاری و همکارانش مدل‌های QSAR، مربوط به ۲۶۶ ترکیب برای ۲۹ سلول سرطانی مختلف با تعداد توصیفگرهای متفاوت بسط دادند (۱۱). در سال ۲۰۱۰، برتوسا و همکارانش تجزیه و تحلیل آماری مدل‌های QSAR، مربوط به فعالیت ضد تومور ۵۹ آمید و کینولین از سری تیوفن را مطالعه کردند (۱۲). هدف از انجام این پروژه پیش بینی فعالیت ضد سرطانی مشتقات او ۴-دی هیدرو-۴-اکسو-۱- (۲-تيازول)-۸و۱- نفتیریدین با استفاده از الگوریتم ژنتیک-رگرسیون خطی چندگانه و رگرسیون خطی چندگانه مرحله‌ای و در نهایت مقایسه عملکرد این دو روش در انتخاب توصیفگرهای مولکولی بود.

مواد و روشها

دسته داده‌ها

در این مطالعه یک سری صدماتی از مشتقات ۳ و ۷ استخلاف شده از او ۴-دی هیدرو-۴-اکسو-۱- (۲-تيازول)-۸و۱- نفتیریدین ها مورد تحقیق قرار گرفتند. این ترکیبات قبلاً توسط تومیتا و همکارانش سنتز و آزمایش شدند (۱۳-۱۵). جدول ۱ ساختارهای شیمیایی ترکیبات مورد مطالعه را نشان می‌دهد. از IC_{50} ترکیبات در مقابل رده سلولی P380 که مربوط به لوسمی موش است در مطالعه استفاده شد. در نهایت، از مقادیر $\log(1/IC_{50})$ ترکیبات در جدول ۱، به عنوان متغیر وابسته در این تحقیق استفاده شد.

مهم‌ترین اهداف شیمیدان‌ها و محققان دارویی پیش‌بینی فعالیت ترکیبات، قبل از سنتز و یا انجام آزمایش بر روی آن‌ها است. چرا که انجام بسیاری از آزمایشات مستلزم صرف زمان و هزینه‌های زیادی است. از این‌رو نیاز به استفاده از روش‌های تئوری و محاسباتی که بدون انجام آزمایش بتواند ویژگی و یا فعالیت ترکیبات را پیشگویی کند ضروری به نظر می‌رسد. ظهور علم کمومتری توانسته راه‌حلی برای رفع این مشکلات باشد (۴-۲).

روابط کمی ساختار - فعالیت (QSAR) برای چندین دهه در توسعه روابط بین خواص فیزیکی ترکیبات شیمیایی و فعالیت بیولوژیکی آنها برای به دست آوردن یک مدل آماری قابل اعتماد برای پیش بینی فعالیت ترکیبات جدید استفاده شده است. یکی از مهم‌ترین کاربردهای QSAR استفاده از این روش در پیش بینی فعالیت داروها و کمک به طراحی و سنتز ترکیبات دارویی جدید با احتمال تاثیر بالاتر برای درمان بسیاری از بیماری‌های صعب‌العلاج است. توسعه داروهای ضدسرطان از چهار دهه پیش، با کشف فعالیت ضدسرطانی داروها و استفاده موفقیت آمیزشان در درمان سلول‌های سرطانی گوناگون شروع شده است. از آن پس ترکیبات بی‌شماری سنتز شده و به عنوان نامزدهای بالقوه برای داروهای ضد سرطان استفاده شده‌اند، اما تنها تعداد انگشت شماری از آنها به عنوان داروهای بالینی موثر عمل می‌کنند (۵).

در حال حاضر، هزاران توصیفگر مولکولی برای مطالعات QSAR/QSPR وجود دارد. وارد کردن همه توصیفگرها در مدل‌های QSAR/QSPR می‌تواند منجر به بیش برآزش (Overfitting)، پیچیدگی مدل و کاهش تفسیرپذیری آن شود و بنابراین عملکرد مدل و قابلیت پیشگویی آن را کاهش می‌یابد. روش‌های انتخاب متغیر می‌توانند با حذف تعدادی از توصیفگرها و انتخاب توصیفگرهای مرتبط با ویژگی یا فعالیت بر تعدادی از این ضعفها غلبه کنند. تاکنون چندین روش انتخاب متغیر، از قبیل الگوریتم ژنتیک، رگرسیون مرحله‌ای و شبیه سازی تبرید تدریجی (Simulated annealing) به طور گسترده استفاده شده‌اند. روش‌های بهینه سازی هوش ازدحامی (Swarm intelligence optimizations)، از قبیل بهینه سازی ازدحامی جزیی (Partial swarm optimization)، از جمله روشهای انتخاب متغیر هستند که معمولاً بر اساس رفتار زیستی حیوان و حشره به منظور پیدا کردن کوتاه ترین مسیر بین یک منبع غذا و لانه هایشان شبیه سازی شده‌اند، اخیراً در روشهای QSAR/QSPR به کار می‌روند (۶، ۷). یکی از بهترین روش‌های جستجو برای پیدا کردن با اهمیت‌ترین

جدول ۱. ساختارهای شیمیایی ترکیبات ۱ و ۴-دی هیدرو-۴-اکسو-۱-(۲-تiazول)-۸و۱-نفتریدین و مشتقات آن

ساختار عمومی ۱					ساختار عمومی ۴				
شماره	X	Y	Z		شماره	R1		R1	شماره
۲۰	N	N	CH		۱۱				۱
۲۱	CF	N	N		۱۲				۲
۲۲	N	CH	CH		۱۳				۳
۲۳	N	CH	CH		۱۴				۴
۲۴	CF	CF	CH		۱۵				۵
					۱۶				۶
					۱۷				۷
					۱۸				۸
					۱۹				۹
									۱۰
شماره	R2		R2	شماره					
۲۵				۳۰					
۲۶				۳۱					
۲۷				۳۲					
۲۸				۳۳					
۲۹				۳۴					

مولکولی میدان نیروی آلاینجر MM2 و روش کوانتوم نیمه تجربی روش آوستین (AM1) با استفاده از الگوریتم پلاک-ریبار (Polak-Ribier) تا گردایان ریشه مربع میانگین ۰/۰۱ با استفاده از نرم افزار موپک (MOPAC Version 6.00) بهینه سازی شدند. مولکول‌های بهینه شده به نرم افزار دراگون (Dragon Version 5.4) انتقال داده شد و توصیفگرهای

بهینه سازی مولکولی و محاسبه توصیفگرها

تمامی محاسبات توسط لپ تاپ ۷ هسته‌ای با ویندوز ۷ به عنوان سیستم عامل انجام شد. ابتدا با استفاده از نرم افزار کم در (ChemDraw Ultra Version 8.0) ساختار مولکول‌ها رسم شد و ساختار هندسی مولکول‌ها از طریق روش مکانیک

ادامه جدول ۱. ساختارهای شیمیایی ترکیبات او ۴-دی هیدرو-۴-اکسو-۱-(۲-تiazول)-۸و۱-نفتیریدین و مشتقات آن

شماره	R2	شماره	R2	شماره	R2	شماره	R2	شماره	R2
۳۵		۳۹		۴۳		۴۷		۵۱	
۳۶		۴۰		۴۴		۴۸		۵۲	
۳۷		۴۱		۴۵		۴۹		۵۳	
۳۸		۴۲		۴۶		۵۰		۵۴	
شماره	U	V	شماره	U	V				
۵۵	H	H	۵۹	Cl	H				
۵۶	H	NO ₂	۶۰	H	H				
۵۷	H	NH ₂	۶۱	H	CF ₃				
۵۸	H	OH	۶۲	H	NH ₂				

شدند. این توصیفگرهای مولکولی به حدود ۴۶۸ توصیفگر با حذف متغیرهایی که همبستگی کمی با فعالیت دارند، متغیرهای ثابت و توصیفگرهای همبسته کاهش پیدا کردند.

شکافت داده‌ها (Data splitting)

در قدم بعدی سری داده‌ها به دو دسته آموزشی و آزمون به صورت تصادفی توسط نرم افزار اکسل (Microsoft Excel

مولکولی محاسبه شدند. تمامی توصیفگرهای مولکولی به دست آمده از اطلاعات ساختارهای مولکولی به دست آمدند؛ بنابراین نیازی به داده های آزمایشگاهی برای محاسبه نبود. حدود ۱۲۱۲ توصیفگر مولکولی از قبیل GETAWAY, RDF, WHIM, توصیفگرهای شمارشی گروه عاملی و توصیفگرهای 3D-MorSE برای مشخص کردن ساختارهای مولکولی مشتق

ادامه جدول ۱. ساختارهای شیمیایی ترکیبات ۱ و ۴-دی هیدرو-۴-اکسو-۱-(۲-تiazول)-۸۱-نفتیریدین و مشتقات آن

شماره		R3		شماره		R3		شماره	
شماره		R3		شماره		R3		شماره	
۶۳		۷۰		۷۱		۷۲		۷۳	
۶۴		۷۴		۷۵		۷۶		۷۷	
۶۵		۷۸		۷۹		۸۰		۸۱	
۶۶		۸۲		۸۳		۸۴		۸۵	
۶۷		۸۶		۸۷		۸۸		۸۹	
۶۸		۹۰						شماره	
۶۹		۹۱						شماره	R4
۹۱	COOEt	۹۲	H	۹۳	CH2Ph	۹۴	CHO	۹۵	CONH2
۹۲	H	۹۳	COOEt	۹۶	OH	۹۷	CHO	۹۸	CH2OH
۹۹	۹۹	۱۰۰	COCH3	۱۰۰	CH2CH=CHCOOCH3				

سری آزمون در هیچ یک از مراحل مدل سازی شرکت نمی کند و در واقع همان طور که از نامش پیداست به منظور ارزیابی قدرت پیش بینی و اعتبار مدل های به دست آمده مورد استفاده قرار می گیرد. شایان ذکر است که هر قدر مجموعه داده ها بزرگ تر و متنوع تر باشد مدل حاصل از آن معتبر بوده و قدرت پیش بینی آن بالا خواهد بود. در روش انتخاب تصادفی،

(2013) تقسیم می شوند. سری آموزشی که حاوی بیشترین تعداد مولکول ها است باید نماینده مناسبی از کل مولکول ها باشد، چون این دسته از مولکول ها برای مدل سازی استفاده می شوند. مولکول های سری آزمون طوری انتخاب می شوند که نماینده مولکول های موجود در سری آموزش باشند و معمولاً ۲۰٪ از کل داده ها را به عنوان سری آزمون انتخاب می کنند.

جدول ۲. نتایج اعتبار سنجی مدل ترکیبات تصادفی انتخاب شده برای مجموعه آزمون

آزمایه	دسته آزمون تصادفی	S-MLR			GA-MLR			
		Q ² CV	Q ² Cal	Q ² Test	Q ² CV	Q ² Cal	Q ² Test	
۱		۳, ۱۲, ۱۳, ۲۰, ۲۳, ۲۶, ۳۲, ۳۳, ۳۹, ۴۶, ۴۹, ۵۶, ۶۳, ۶۹, ۷۰, ۷۶, ۷۹, ۸۳, ۸۶, ۸۸	۰/۷۰۲۵	۰/۷۶۱۹	۰/۳۲۱۲	۰/۷۶۸۹	۰/۸۰۸۴	۰/۲۳۴۷
۲		۷, ۸, ۱۵, ۲۶, ۳۱, ۳۸, ۳۹, ۴۱, ۴۴, ۴۵, ۵۱, ۵۴, ۵۹, ۶۱, ۶۳, ۶۴, ۶۵, ۶۷, ۷۸, ۸۹	۰/۷۱۶۵	۰/۷۷۱۶	-۰/۳۰۲۹	۰/۷۰۷۹	۰/۷۶۱۲	۰/۳۹۶۵
۳		۵, ۶, ۱۲, ۱۶, ۲۰, ۲۳, ۳۲, ۵۰, ۵۶, ۶۱, ۶۴, ۷۲, ۷۳, ۸۲, ۸۵, ۸۷, ۹۰, ۹۱, ۹۴, ۹۵	۰/۶۶۷۷	۰/۷۱۷۸	۰/۶۶۴۲	۰/۶۸۳۱	۰/۷۳۴۲	۰/۶۲۷۸
۴		۲, ۱۹, ۲۷, ۳۲, ۳۳, ۴۱, ۴۷, ۴۸, ۵۳, ۵۵, ۶۵, ۶۶, ۶۷, ۷۰, ۷۸, ۸۷, ۸۹, ۹۰, ۹۲, ۹۷	۰/۶۶۰۳	۰/۷۲۶۰	۰/۵۹۷۱	۰/۶۹۲۰	۰/۷۴۸۷	۰/۵۹۴۲
۵		۲, ۷, ۱۵, ۳۰, ۳۳, ۳۸, ۴۴, ۴۶, ۵۳, ۵۶, ۶۳, ۶۶, ۶۹, ۷۵, ۷۶, ۸۰, ۸۲, ۸۴, ۸۹, ۹۲	۰/۶۲۴۱	۰/۶۹۵۲	۰/۲۶۲۵	۰/۶۸۱۷	۰/۷۳۲۵	۰/۵۵۶۵
۶		۳, ۴, ۱۳, ۱۴, ۲۰, ۲۱, ۲۶, ۲۷, ۳۷, ۴۴, ۴۵, ۴۹, ۵۲, ۵۳, ۵۴, ۶۶, ۷۶, ۸۸, ۸۹, ۹۹	۰/۶۴۷	۰/۷۱۱۶	۰/۲۸۲۲	۰/۷۱۳۷	۰/۷۷۲۸	-۰/۲۵۴
۷		۱۶, ۲۱, ۲۲, ۳۳, ۴۱, ۴۲, ۴۴, ۴۶, ۵۶, ۵۷, ۵۹, ۶۵, ۶۶, ۷۲, ۷۷, ۸۹, ۹۱, ۹۵, ۹۷, ۱۰۰	۰/۶۷۵۱	۰/۷۳۱۸	-۰/۰۰۴۵	۰/۷۳۰۳	۰/۷۸۳۱	۰/۱۹۰۹
۸		۲, ۶, ۱۴, ۱۶, ۱۷, ۲۷, ۳۱, ۳۲, ۳۵, ۳۶, ۴۲, ۴۸, ۵۲, ۵۵, ۵۹, ۷۵, ۷۹, ۸۲, ۸۳, ۹۵	۰/۷۱۱۳	۰/۷۶۲۸	-۰/۷۶۱۰	۰/۷۰۴۲	۰/۷۷۱۶	-۱/۹۳۴۱
۹		۲, ۱۱, ۲۱, ۲۵, ۲۶, ۳۸, ۴۱, ۴۴, ۵۱, ۵۳, ۵۷, ۶۲, ۶۷, ۸۱, ۸۴, ۸۷, ۸۸, ۹۲, ۹۴, ۹۵	۰/۶۹۲۵	۰/۷۴۵۹	۰/۳۵۸۹	۰/۷۱۳۸	۰/۷۶۶۳	۰/۳۵۸۸
۱۰		۱, ۲, ۶, ۹, ۱۰, ۱۹, ۲۳, ۲۵, ۳۰, ۳۹, ۴۱, ۴۵, ۵۲, ۵۴, ۵۶, ۶۴, ۶۶, ۷۹, ۹۲, ۹۸	۰/۶۶۴۳	۰/۷۳۷۸	۰/۴۷۱۸	۰/۷۲۴۸	۰/۷۸۰۲	۰/۴۲۷۳
۱۱		۲, ۱۲, ۱۴, ۱۶, ۱۷, ۱۹, ۲۰, ۲۹, ۳۴, ۳۸, ۳۹, ۴۰, ۵۸, ۶۰, ۶۷, ۷۶, ۸۵, ۹۴, ۹۶, ۹۹	۰/۷۰۷۹	۰/۷۶۱۰	-۰/۴۷۶۲	۰/۷۱۰۵	۰/۷۶۲۳	-۰/۱۹۷۱
۱۲		۹, ۱۱, ۱۷, ۲۱, ۲۴, ۲۶, ۳۰, ۳۸, ۳۹, ۴۰, ۵۰, ۵۲, ۵۳, ۵۹, ۶۱, ۶۷, ۶۸, ۸۱, ۹۷, ۹۸	۰/۵۸۰۲	۰/۶۶۵۰	۰/۵۲۶۰	۰/۶۵۱۴	۰/۷۰۹۲	۰/۴۷۹۰
۱۳		۳, ۷, ۱۳, ۱۶, ۲۸, ۳۰, ۳۴, ۳۹, ۴۴, ۴۵, ۵۰, ۵۳, ۶۲, ۶۷, ۷۸, ۸۰, ۸۵, ۸۸, ۹۱, ۹۲	۰/۶۱۷۹	۰/۶۹۰۷	۰/۳۲۸۳	۰/۶۸۹۵	۰/۷۴۸۹	۰/۵۱۳۳
۱۴		۶, ۷, ۸, ۱۱, ۱۲, ۱۵, ۱۶, ۲۲, ۲۳, ۴۱, ۴۸, ۴۹, ۶۰, ۶۳, ۶۷, ۷۰, ۷۶, ۸۵, ۹۳, ۹۹	۰/۶۶۶۱	۰/۷۲۹۷	۰/۴۶۱۶	۰/۷۱۴۴	۰/۷۶۴۱	۰/۳۲۹۶
۱۵		۲, ۴, ۲۰, ۲۷, ۳۳, ۳۴, ۳۵, ۳۷, ۴۰, ۴۷, ۵۲, ۵۳, ۶۶, ۶۹, ۷۱, ۷۷, ۸۳, ۸۷, ۸۹, ۹۷	۰/۶۱۰۸	۰/۶۸۸۹	۰/۴۱۱۰	۰/۶۶۵۰	۰/۷۱۸۳	۰/۶۷۳۴
۱۶		۳, ۱۰, ۱۸, ۲۰, ۲۱, ۳۱, ۴۰, ۴۲, ۴۹, ۵۵, ۶۷, ۷۲, ۷۳, ۷۶, ۷۹, ۸۰, ۸۵, ۹۰, ۹۷, ۱۰۰	۰/۶۷۵۶	۰/۷۴۳۱	۰/۵۳۲۹	۰/۶۹۵۶	۰/۷۵۵۶	۰/۳۰۴۸
۱۷		۳, ۱۰, ۳۱, ۳۴, ۳۵, ۳۷, ۴۴, ۴۵, ۴۹, ۵۵, ۶۱, ۶۲, ۶۴, ۷۲, ۷۶, ۸۴, ۸۶, ۹۰, ۹۳, ۹۶	۰/۶۹۳۶	۰/۷۶۲۱	-۰/۳۷۶۹	۰/۷۴۹۶	۰/۷۹۴۰	۰/۴۶۵۵
۱۸		۱, ۱۰, ۲۰, ۲۳, ۲۵, ۴۱, ۵۶, ۵۸, ۶۰, ۶۶, ۶۷, ۶۸, ۷۰, ۷۱, ۸۴, ۸۷, ۸۹, ۹۶, ۹۷, ۹۹	۰/۷۴۴۴	۰/۷۹۱۲	۰/۳۱۳۷	۰/۷۵۳۹	۰/۸۰۱۳	۰/۴۱۰۹
۱۹		۲, ۶, ۱۸, ۱۹, ۳۸, ۴۳, ۴۴, ۵۰, ۵۱, ۵۵, ۶۲, ۷۰, ۷۱, ۷۲, ۷۴, ۷۸, ۸۶, ۸۹, ۹۳, ۹۶	۰/۶۶۳۳	۰/۷۳۰۹	-۰/۱۰۴۲	۰/۶۹۲۹	۰/۷۳۷۸	۰/۵۸۲۱
۲۰		۱, ۳, ۵, ۶, ۱۰, ۱۱, ۱۲, ۲۱, ۳۰, ۳۱, ۳۲, ۳۳, ۳۷, ۴۰, ۴۱, ۴۶, ۴۷, ۹۳, ۹۸, ۱۰۰	۰/۷۶۰۷	۰/۸۱۲۹	-۰/۰۴۴۰	۰/۷۶۴۸	۰/۸۱۷۶	-۰/۱۲۱۳

مولکولها را برحسب فعالیتشان بصورت صعودی و یا نزولی مرتب می کنیم و سپس از بین آنها نسبت های ذکر شده در بالا را طوری انتخاب می کنیم که از لحاظ آماری دارای توزیع نرمال باشد. در این مطالعه ۲۰ بار انتخاب تصادفی انجام شد و دسته آموزش و آزمون به ترتیب شامل ۸۰ و ۲۰ مولکول است. جدول ۲ تمامی ترکیبات انتخاب شده برای دسته آزمون و نتایج مدلسازی آنها را نشان می دهد.

انتخاب متغیر توسط الگوریتم ژنتیک

روش ترکیبی الگوریتم ژنتیک- رگرسیون خطی چندگانه (GA-MLR) یکی از روش هایی است که در این مطالعه به عنوان روش انتخاب متغیر استفاده شده است. الگوریتم ژنتیک یک جستجوی قدرتمند براساس تکامل سیستم های بیولوژیکی است. در مطالعات QSAR/QSPR، دستیابی به مدلی با تعداد کمی از توصیفگرهای ساختاری دارای اهمیت است، زیرا این مسأله منجر به ایجاد مدلی ساده و قابل پیشگویی خواهد شد. در حقیقت روش الگوریتم ژنتیک یکی از

رایج ترین روش های انتخاب متغیر است که به میزان گسترده ای برای مشکلات بهینه سازی پیچیده در زمینه های گوناگون از جمله روش های QSAR/QSPR استفاده می شود (۲۴-۱۶). در اینجا از جعبه ابزار الگوریتم ژنتیک موجود در نرم افزار MATLAB 2013a (gatool) برای انتخاب توصیفگرهای مناسب استفاده شده است. همان طور که در بخش گذشته اشاره کردیم متغیرها از ۱۲۱۲ متغیر به حدود ۴۶۸ متغیر با حذف متغیرهایی که همبستگی کمی با فعالیت دارند، متغیرهای ثابت و توصیفگرهای همبسته کاهش پیدا کردند. روش الگوریتم ژنتیک به کار گرفته در این مطالعه از یک نمایش دوتایی به عنوان روش رمزگذاری استفاده می کند؛ حضور و یا عدم حضور یک توصیفگر در یک کروموزوم به وسیله ۱ یا ۰ رمزگذاری می شود. در الگوریتم ژنتیک بهینه سازی از طریق تغییر و انتخاب و به واسطه ارزیابی تابع شایستگی [J] انجام می شود. تابع شایستگی مورد استفاده در این مطالعه متوسط ریشه میانگین مربع خطای واسنجی و

جدول ۳. نتایج مدل S-MLR برای مشتقات ۱ و ۴-دی هیدرو-۴-اکسو-۱-(۲-تيازول)-۱-ا و ۸-نفتیریدین برای بهترین دسته تصادفی

متغیر	نوع توصیفگر	تعریف	b	S_b	b_s	VIF
عرض از مبدا	—	—	13/046	2/567	—	—
RDF020p	توصیفگرهای RDF	تابع توزیع شعاعی -۲۰ / توزین شده توسط الکتروننگاتیویته ساندرسون	0/514	0/096	0/408	1/315
C-002	قطعات اتم محور	CH2R2	0/564	0/161	0/268	1/334
Mor16m	توصیفگرهای 3D-MoRSE	سیگنال ۱۶ / توزین شده توسط جرم	-1/418	0/416	-0/246	1/185
R4m	توصیفگرهای GETAWAY	همبستگی R لایه ۴ / توزین شده توسط جرم	-4/582	0/983	-0/326	1/116
Mor21m	توصیفگرهای 3D-MoRSE	سیگنال ۲۱ / توزین شده توسط جرم	-1/797	0/495	-0/274	1/300
G2e	توصیفگرهای WHIM	دومین جز تقارن جهتی ضریب WHIM / توزین شده توسط الکتروننگاتیویته ی ساندرسون	65/981	14/679	0/355	1/422
nRNR2	شمارش های گروه عاملی	تعداد آمین های نوع سوم (آلیفاتیک)	۰/۶۸۰	0/235	۰/۲۳۰	1/437
Mor22m	توصیفگرهای 3D-MoRSE	سیگنال ۲۲ / توزین شده توسط جرم	1/040	0/406	0/182	1/149

b ^۱ ضریب رگرسیونی غیراستاندارد توصیفگرها؛ S_b ^۲ خطای استاندارد رگرسیونی؛ b_s ^۳ ضریب رگرسیونی استاندارد شده؛ VIF ^۴ فاکتور تورم واریانس برای سنجش همبستگی خطی چندگانه بین توصیف گرها در مدل مورد نظر

جدول ۴. نتایج مدل GA-MLR برای مشتقات ۱ و ۴-دی هیدرو-۴-اکسو-۱-(۲-تيازولیل)-۱-ا و ۸-نفتیریدین برای بهترین دسته تصادفی

متغیر	نوع توصیفگر	تعریف	b	S_b	b_s	VIF
عرض از مبدا	—	—	25/437	8/325	—	—
RDF055e	توصیفگرهای RDF	تابع توزیع شعاعی -۵۵ / توزین شده توسط الکتروننگاتیویته ی ساندرسون	-0/111	0/016	-0/555	1/562
nRNR2	شمارش های گروه عاملی	تعداد آمین های نوع سوم (آلیفاتیک)	0/556	0/2۰۰	0/187	1/144
Mor03p	توصیفگرهای 3D-MoRSE	سیگنال ۳ / توزین شده توسط قطبش پذیری	-0/777	0/144	-0/4۰۰	1/389
nCp	شمارش های گروه عاملی	تعداد C(sp3) نوع اول انتهایی	0/439	0/094	0/383	1/713
nRSR	شمارش های گروه عاملی	تعداد سولفیدها	0/84۰	0/243	0/240	1/223
Mor07e	توصیفگرهای 3D-MoRSE	سیگنال ۷ / توزین شده توسط الکتروننگاتیویته ساندرسون	0/289	0/096	0/212	1/247
ISH	توصیفگرهای GETAWAY	مقدار اطلاعات استاندارد شده روی اهرم برابری	21/324	8/322	0/164	1/035
Mor16p	توصیفگرهای 3D-MoRSE	سیگنال ۱۶ / توزین شده توسط قطبش پذیری	-1/654	0/479	-0/251	1/337

b ^۱ ضریب رگرسیونی غیراستاندارد توصیفگرها؛ S_b ^۲ خطای استاندارد رگرسیونی؛ b_s ^۳ ضریب رگرسیونی استاندارد شده؛ VIF ^۴ فاکتور تورم واریانس برای سنجش همبستگی خطی چندگانه بین توصیف گرها در مدل مورد نظر

پارامتر در یک معادله واحد، معادله QSAR را محاسبه می کند (۲۶، ۲۵). به منظور انجام محاسبات در روش S-MLR، نرم افزار IBM SPSS مورد استفاده قرار گرفت. مجموعه داده ها به مجموعه های کاهش یافته ترکیبات ضد سرطان به مجموعه آموزش و آزمون، در مرحله نخست، باید مجموعه آزمون از کل داده ها حذف گردیده و سپس محاسبات با استفاده از روش رگرسیونی خطی گام به گام انجام شوند. پس از بررسی، مدل های دارای چهار، پنج، شش و ... بهترین مدل خطی چندمتغیره دارای ۸ توصیفگر بود.

در نهایت باید اعتبار مدل از لحاظ آماری مورد ارزیابی قرار گیرد و مدل مناسب انتخاب شود. پارامتر آماری Q2 برای ارزیابی مدل حاصل از اعتبار سنجی همگداری به کار می رود و

اعتبارسنجی متقاطع، $(RMSEC+RMSEC)/2$ ، است. اندازه جمعیت در این بررسی ۷۰۰۰ بود و سایر گزینه ها از جمله هم گذری، جهش و مهاجرت، مطابق با پیش فرض بخش مربوط به الگوریتم ژنتیک در نرم افزار MATLAB 2012 است. پس از ارزیابی مدل های متفاوت مشاهده می شود که بهترین مدل خطی چندمتغیره ای دارای هشت توصیفگر است. انتخاب متغیر توسط رگرسیون خطی چندگانه گام به گام

گام رگرسیون خطی چندگانه (MLR) روش گسترش یافته رگرسیون کلاسیک است که بیش از یک بعد را در محاسبات در نظر می گیرد. روش MLR به واسطه انجام محاسبات رگرسیونی چند متغیره استاندارد و از طریق استفاده از چندین

جدول ۵. ماتریس ضریب همبستگی برای توصیف کننده ها و ثابت $\text{Log}(1/IC_{50})$ ترکیبات او ۴-دی هیدرو-۴-اکسو-۱-(۲-تيازول)-او۸- نفتیریدین و مشتقات آن با اثر ضد سرطان در مدل S-MLR

Mor22m	nRNR2	G2e	Mor21m	R4m	Mor16m	C-002	RDF020p	$\text{Log}(1/IC_{50})$	
								۱	$\text{Log}(1/IC_{50})$
							۱	۰/۴۹۲	RDF020p
						۱	-۰/۱۲۸	۰/۳۰۷	C-002
					۱	-۰/۰۶۵	-۰/۳۴۷	-۰/۴۲۴	Mor16m
				۱	-۰/۰۰۵	-۰/۲۲۶	۰/۰۵۲	-۰/۲۸۵	R4m
			۱	-۰/۱۱۸	۰/۳۱۱	۰/۰۵۹	-۰/۱۶۳	-۰/۳۰۶	Mor21m
		۱	۰/۳۱۶	۰/۱۹۶	-۰/۱۴۴	-۰/۲۲۴	-۰/۱۸۱	-۰/۰۶۸	G2e
	۱	-۰/۴۳۲	-۰/۲۱۳	-۰/۰۸۷	-۰/۰۸۶	۰/۳۸۲	۰/۱۴۷	۰/۳۶۳	nRNR2
۱	۰/۰۹۴	-۰/۰۲۳	-۰/۰۳۶	۰/۰۰۰	-۰/۰۲۴	۰/۰۵۵	۰/۳۳۲	۰/۳۶۱	Mor22m

جدول ۶. ماتریس ضریب همبستگی برای توصیف کننده ها و ثابت $\text{Log}(1/IC_{50})$ ترکیبات او ۴-دی هیدرو-۴-اکسو-۱-(۲-تيازول)-او۸- نفتیریدین و مشتقات آن با اثر ضد سرطان در مدل GA-MLR

Mor16p	ISH	Mor07e	nRSR	nCp	Mor03p	nRNR2	RDF055e	$\text{Log}(1/IC_{50})$	
								۱	$\text{Log}(1/IC_{50})$
							۱	-۰/۰۷۸	RDF055e
						۱	۰/۳۱۴	۰/۳۶۳	nRNR2
					۱	-۰/۲۲۳	-۰/۳۲۳	-۰/۴۱۹	Mor03p
				۱	-۰/۰۳۴	۰/۲۵۱	۰/۴۸۵	۰/۳۶۶	nCp
			۱	۰/۱۰۸	-۰/۳۳۴	۰/۰۹۸	۰/۰۷۸	۰/۴۳۳	nRSR
		۱	-۰/۰۶۲	۰/۳۱۹	-۰/۲۱۴	۰/۱۷۲	۰/۳۴۶	۰/۲۸۱	Mor07e
	۱	-۰/۰۳۴	۰/۰۱۷	-۰/۰۶۹	۰/۰۲۵	-۰/۰۴۷	-۰/۰۹۲	۰/۱۲۷	ISH
۱	۰/۱۵۸	-۰/۱۶۷	-۰/۲۱۱	-۰/۴۳۵	۰/۰۹	-۰/۲۲۸	-۰/۱۷۷	-۰/۴۵۸	Mor16p

نیست. مدل QSAR توسط دسته واسنجی ترکیبات شیمیایی به تنهایی ایجاد شده سپس برای اعتبار سنجی برونی دسته ترکیبات شیمیایی و تحقیق اعتبار سنجی بیشتر قدرت پیشگویی مدل، به کار گرفته می شود. سرانجام مدل با استفاده از دسته اعتبارسنجی برونی (دسته آزمون) اعتبارسنجی می شود و نتایج اعتبار سنجی مدل در جدول ۲ نشان داده شده است.

فرمول برای محاسبه Q_{ext}^2 به صورت زیر است:

(۲)

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{t_{test}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{t_{test}} (y_i - \bar{y}_{test})^2}$$

که \hat{y}_i و y_i به ترتیب مقادیر پیشگویی و اندازه گیری فعالیت ترکیب هستند، و \bar{y}_{test} مقدار میانگین فعالیت برای دسته آزمون است؛ جمع تمام ترکیبات دسته ی آزمون را پوشش می دهد.

مقدار Q^2 آزمون خوبی برای داده هایی است که به طور یکنواخت توزیع شده اند. در این بخش با نگاهی گذرا بر پارامترهای آماری خواهیم پرداخت. اعتبار سنجی همگذری تک به تک (Leave One Out - Cross Validation) یکی از اعتبار سنجی های داخلی مدل QSAR است. توانایی پیشگویی مدل QSAR با استفاده از روش LOO-CV تعیین می شود. واریانس اعتبار سنجی همگذری Q_{CV}^2 توسط معادله زیر محاسبه می شود.

(۱)

$$Q_{CV}^2 = 1 - \frac{\sum_{i=1}^{Cal} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{Cal} (y_i - \bar{y})^2}$$

y_i ، \hat{y}_i و \bar{y} به ترتیب مقادیر اندازه گیری، پیشگویی و میانگین (کل داده ها ی کالیبراسیون) متغیر وابسته هستند؛ عمل جمع تمام ترکیبات در دسته واسنجی را در بر می گیرد. روش LOO-CV برای تعیین استحکام و قدرت پیشگویی کافی

جدول ۷. مقادیر توصیفگرهای مولکولی، مقادیر $\log(1/IC_{50})$ آزمایشی و پیشگویی شده توسط مدل GA-MLR

دسته	$\log(1/IC_{50})_{Pred}$	$\log(1/IC_{50})_{EXP}$	Mor16p	ISH	Mor07e	nRSR	nCp	Mor03p	nNR2	RDF055e	شماره
واسنجی	-۱/۳۵۰۰	-۰/۸۱۳۹	-۰/۰۱۸	۱	۳/۵۱۸	۰	۰	-۳/۴۱۸	۲	۱۸/۴۸۰	۱
آزمون	-۰/۷۱۸۰	-۰/۷۳۵۲	-۰/۱۲۹	۱	۳/۹۵۴	۰	۰	-۳/۲۶۵	۲	۱۴/۵۰۴	۲
واسنجی	-۰/۷۱۸۰	-۱/۴۳۲۶	-۰/۱۲۹	۱	۳/۹۵۴	۰	۰	-۳/۲۶۵	۲	۱۴/۵۰۴	۳
آزمون	-۱/۸۱۰۱	-۱/۳۷۶۱	-۰/۱۸۸	۰/۹۸۲	۳/۵۵۷	۰	۰	-۲/۸۳۰	۲	۱۷/۶۸۶	۴
واسنجی	۰/۲۸۵۳	-۰/۵۰۵۹	-۰/۰۴۹	۱	۲/۹۰۵	۱	۰	-۳/۶۱۹	۲	۱۱/۵۸۸	۵
واسنجی	-۱/۱۹۳۲	-۱/۴۱۷۷	-۰/۲۳۳	۰/۹۸۴	۳/۳۰۹	۰	۰	-۳/۰۱۳	۲	۱۳/۸۱۸	۶
واسنجی	-۰/۴۴۲۵	-۰/۱۹۱۳	-۰/۲۶۷	۱	۲/۹۱۲	۰	۱	-۳/۰۹۷	۲	۱۴/۱۴۵	۷
واسنجی	-۰/۷۱۲۱	-۱/۱۶۴۸	-۰/۱۵۲	۰/۹۸۴	۳/۱۳۴	۰	۰	-۳/۴۸۱	۲	۱۱/۰۹۷	۸
واسنجی	-۱/۷۱۴۵	-۱/۴۲۹۰	-۰/۰۲۳	۱	۳/۵۰۱	۰	۱	-۲/۵۲۴	۲	۱۹/۴۹۱	۹
واسنجی	۰/۵۶۷۹	-۱/۴۰۹۶	-۰/۱۸۲	۱	۳/۰۶۱	۱	۰	-۳/۶۴۴	۲	۱۱/۶۰۵	۱۰
واسنجی	۰/۹۵۱۹	۰/۸۸۲۷	-۰/۲۵۴	۱	۲/۸۶۰	۱	۱	-۳/۴۲۴	۲	۱۱/۱۱۰	۱۱
واسنجی	۰/۷۶۹۵	۰/۱۷۵۴	-۰/۲۰۵	۱	۲/۵۸۸	۱	۱	-۳/۳۵۴	۲	۱۰/۸۲۵	۱۲
واسنجی	۰/۲۰۸۱	۰/۳۵۰۵	-۰/۰۳۴	۱	۳/۱۰۸	۱	۲	-۳/۰۳۶	۲	۱۶/۴۱۷	۱۳
واسنجی	-۰/۲۴۵۵	-۰/۳۰۹۵	-۰/۴۳۲	۰/۹۸۶	۲/۱۰۹	۱	۳	-۴/۰۴۵	۲	۳۲/۱۶۲	۱۴
واسنجی	۰/۵۸۲۷	۰/۹۳۸۵	-۰/۱۹۴	۱	۴/۲۳۷	۱	۰	-۴/۱۱۷	۲	۱۸/۰۲۳	۱۵
واسنجی	۰/۲۴۸۳	۰/۶۳۹۴	-۰/۱۷۴	۱	۲/۴۳۱	۱	۰	-۳/۴۵۶	۲	۱۱/۴۰۹	۱۶
واسنجی	۰/۳۸۳۰	۰/۷۲۷۹	-۰/۱۶۵	۱	۲/۵۶۶	۱	۰	-۳/۴۶۵	۲	۱۰/۴۷۶	۱۷
واسنجی	۰/۰۱۷۸	۰/۰۶۳۸	-۰/۱۶۷	۱	۳/۵۱۵	۱	۱	-۳/۳۵۱	۲	۱۹/۴۲۳	۱۸
واسنجی	۰/۶۳۵۸	۰/۹۳۸۵	-۰/۳۲۵	۱	۴/۷۲۹	۱	۰	-۳/۹۷۳	۲	۱۹/۷۷۰	۱۹
آزمون	۰/۴۳۸۴	۰/۸۸۲۴	-۰/۰۱۸	۱	۳/۳۴۱	۱	۰	-۳/۵۹۲	۲	۱۰/۶۹۳	۲۰
واسنجی	-۱/۰۷۳۹	-۱/۴۲۴۴	-۰/۱۱۳	۱	۳/۳۱۱	۱	۰	-۳/۴۴۷	۱	۱۹/۶۳۱	۲۱
واسنجی	۰/۱۸۴۵	-۰/۴۴۶۹	-۰/۱۳۸	۱	۲/۹۸۳	۱	۰	-۳/۵۹۰	۲	۱۳/۸۲۲	۲۲
واسنجی	-۰/۸۵۲۲	-۱/۴۲۶۷	-۰/۲۶۲	۱	۲/۵۳۸	۱	۰	-۲/۵۴۹	۲	۱۶/۵۶۴	۲۳
واسنجی	-۱/۶۴۱۷	-۲/۱۴۶۷	-۰/۳۰۸	۰/۹۸۴	۲/۲۱۷	۱	۰	-۲/۵۲۰	۲	۲۰/۲۵۰	۲۴
واسنجی	-۱/۵۳۵۲	-۱/۴۵۶۸	-۰/۰۲۴	۱	۱/۳۹۷	۱	۰	-۲/۹۷۴	۱	۱۴/۱۶۶	۲۵
واسنجی	-۰/۱۱۹۸	۱/۰۱۹۱	-۰/۱۹۳	۱	۳/۱۳۸	۱	۰	-۲/۹۰۲	۲	۱۲/۹۷۱	۲۶
آزمون	۰/۶۲۶۶	۰/۷۶۲۷	-۰/۲۵۴	۱	۳/۰۴۳	۱	۰	-۳/۹۰۹	۲	۱۳/۹۵۷	۲۷
واسنجی	۱/۰۳۶۱	۱/۲۷۴۹	-۰/۳۷۲	۱	۳/۰۲۸	۱	۰	-۴/۰۰۱	۲	۱۲/۶۳۱	۲۸
واسنجی	۰/۶۷۷۸	۱/۳۷۰۶	-۰/۳۱۶	۱	۲/۸۹۸	۱	۰	-۴/۰۰۲	۲	۱۴/۶۹۳	۲۹
واسنجی	۰/۶۲۷۱	۱/۱۲۷۳	-۰/۲۹۵	۱	۲/۹۸۶	۱	۰	-۳/۹۲۵	۲	۱۴/۵۲۷	۳۰
واسنجی	۱/۱۱۹۴	۰/۸۴۹۹	-۰/۲۰۱	۱	۳/۱۴۳	۱	۲	-۳/۰۵۹	۳	۱۵/۹۵۷	۳۱
واسنجی	۱/۱۹۵۵	۱/۱۷۵۲	-۰/۱۸۶	۱	۲/۹۷۱	۱	۱	-۴/۲۰۰	۲	۱۳/۶۲۳	۳۲
آزمون	۱/۰۸۴۴	۱/۵۴۹۸	-۰/۳۲۳	۱	۲/۸۲۶	۱	۱	-۴/۰۵۵	۲	۱۵/۲۷۳	۳۳
آزمون	۱/۲۰۸۹	۰/۸۳۴۵	-۰/۲۸۲	۱	۲/۵۹۹	۱	۱	-۳/۹۲۶	۲	۱۲/۰۴۶	۳۴

برای هر توصیف کننده در مدل از $0/9$ تجاوز نمی کند (۲۷، ۲۸).

یافته‌ها

مدلسازی S-MLR و GA-MLR

بعد از شکافت داده‌ها به دو دسته واسنجی و آزمون، به منظور انتخاب متغیرهای مهم برای توصیف فعالیت ترکیبات

به علاوه، آنالیز فاکتور تورم واریانس (VIF = Variance Inflation Factor) انجام می شود تا ببینیم آیا هم بستگی خطی چند گانه بین توصیف کننده‌ها در مدل وجود دارد. مقدار VIF از رابطه $1/1-r^2$ محاسبه می شود، که r^2 ضریب هم بستگی چند گانه یک توصیف گر بر روی بقیه توصیف کننده های مولکولی است. اگر مدل‌ها دارای توصیف کننده های مولکولی با مقدار VIF بزرگتر از ۱۰ باشند پذیرفته نخواهند شد و تضمین می کند که مربع ضریب هم بستگی چند گانه

ادامه جدول ۷. مقادیر توصیفگرهای مولکولی، مقادیر $\log(1/IC_{50})$ آزمایشی و پیشگویی شده توسط مدل GA-MLR

دسته	$\log(1/IC_{50})_{Pred}$	$\log(1/IC_{50})_{EXP}$	Mor16p	ISH	Mor07e	nRSR	nCp	Mor03p	nNR2	RDF055e	شماره
آزمون	۰/۴۹۰۷	-۰/۷۲۷۵	-۰/۲۴۲	۰/۹۸۴	۳/۲۱۹	۱	۰	-۴/۱۰۵	۲	۱۳/۷۵۹	۳۵
واسنجی	۰/۹۹۷۷	۰/۷۷۵۵	-۰/۴۷۰	۱	۳/۴۳۰	۱	۱	-۳/۴۴۴	۲	۱۵/۵۴۰	۳۶
آزمون	۰/۷۹۶۸	۱/۲۴۶۴	-۰/۳۱۳	۱	۳/۳۷۴	۱	۱	-۳/۹۶۵	۲	۱۸/۵۱۲	۳۷
واسنجی	۰/۸۱۰۶	۰/۶۸۳۶	-۰/۴۳۱	۱	۳/۲۵۶	۱	۱	-۳/۴۱۰	۲	۱۵/۹۵۳	۳۸
واسنجی	۱/۲۰۶۹	۱/۳۶۷۵	-۰/۴۶۱	۰/۹۸۶	۳/۶۰۶	۱	۲	-۴/۰۷۴	۲	۱۹/۶۵۵	۳۹
آزمون	۰/۷۵۹۳	۰/۶۸۳۶	-۰/۱۳۸	۱	۳/۰۰۹	۱	۰	-۴/۳۴۶	۲	۱۴/۰۰۴	۴۰
واسنجی	۰/۵۳۱۴	-۰/۳۷۹۴	-۰/۲۹۹	۱	۲/۳۷۵	۱	۰	-۴/۶۵۰	۲	۱۸/۹۳۳	۴۱
واسنجی	۰/۱۶۴۵	۰/۱۳۵۶	-۰/۲۲۸	۰/۹۸۶	۳/۴۲۳	۱	۰	-۴/۲۶۶	۲	۱۸/۵۳۲	۴۲
واسنجی	۱/۰۶۶۱	۱/۰۳۶۷	-۰/۳۵۹	۰/۹۸۶	۳/۶۲۴	۱	۱	-۳/۵۳۳	۳	۱۶/۷۱۷	۴۳
واسنجی	۰/۵۸۶۰	۰/۴۵۴۸	-۰/۱۸۳	۱	۳/۸۰۳	۱	۰	-۴/۵۹۲	۲	۲۰/۰۲۵	۴۴
واسنجی	۰/۱۱۳۲	۰/۵۸۶۰	-۰/۲۵۶	۰/۹۶۹	۲/۵۰۰	۱	۰	-۳/۹۸۰	۲	۱۱/۷۴۰	۴۵
واسنجی	۰/۵۱۵۴	۰/۴۳۴۰	-۰/۲۶۰	۱	۱/۵۰۶	۱	۰	-۴/۱۸۹	۲	۱۳/۰۰۷	۴۶
آزمون	-۰/۲۰۶۵	۰/۶۶۳۹	-۰/۱۴۰	۰/۹۸۴	۲/۰۰۲	۱	۰	-۳/۷۵۲	۲	۱۲/۸۸۱	۴۷
واسنجی	-۰/۷۲۲۲	-۰/۶۲۱۸	-۰/۰۵۶	۱	۲/۰۱۹	۱	۰	-۳/۸۹۲	۲	۲۰/۳۷۳	۴۸
واسنجی	-۰/۴۰۳۴	-۰/۳۱۷۴	-۰/۱۱۶	۰/۹۸۸	۳/۹۳۷	۱	۱	-۴/۴۲۸	۲	۲۸/۷۹۰	۴۹
واسنجی	-۰/۶۷۵۳	-۰/۶۳۰۲	-۰/۱۵۵	۰/۹۸۵	۱/۵۴۰	۱	۰	-۳/۸۲۲	۲	۱۶/۸۰۷	۵۰
واسنجی	-۰/۲۲۴۹	-۰/۶۴۶۸	-۰/۰۷۴	۱	۲/۶۷۱	۱	۰	-۴/۰۴۴	۲	۱۸/۹۲۳	۵۱
واسنجی	۰/۰۹۱۳	-۱/۴۴۶۹	-۰/۲۹۷	۱	۳/۲۸۳	۱	۰	-۳/۱۵۰	۲	۱۴/۷۳۲	۵۲
واسنجی	-۱/۷۳۶۸	-۱/۴۳۶۱	-۰/۱۵۵	۰/۹۸۲	۲/۴۸۶	۱	۰	-۳/۷۴۱	۱	۲۲/۶۸۱	۵۳
آزمون	-۱/۸۳۴۰	-۱/۵۱۲۵	-۰/۰۰۸	۱	۲/۳۶۸	۰	۰	-۲/۸۳۷	۱	۱۰/۶۲۱	۵۴
آزمون	۰/۷۵۰۱	۱/۵۵۲۸	-۰/۱۴۷	۱	۳/۳۳۱	۱	۰	-۳/۷۳۳	۲	۱۰/۷۶۸	۵۵
واسنجی	-۰/۹۴۹۴	-۰/۷۵۵۲	-۰/۵۰۴	۰/۹۸۵	۲/۶۴۹	۱	۰	-۳/۳۶۵	۲	۲۴/۱۶۵	۵۶
واسنجی	۰/۴۲۰۳	-۰/۴۳۳۳	-۰/۱۵۴	۰/۹۸۴	۳/۱۶۴	۱	۰	-۴/۰۰۲	۲	۱۲/۲۱۸	۵۷
واسنجی	۰/۴۲۵۰	-۱/۰۶۲۳	-۰/۱۲۸	۱	۳/۳۲۷	۱	۰	-۳/۶۸۸	۲	۱۳/۰۸۸	۵۸
واسنجی	۰/۱۳۹۰	۰/۲۳۷۱	-۰/۳۴۶	۱	۲/۱۰۲	۱	۰	-۳/۵۲۷	۲	۱۴/۵۹۷	۵۹
واسنجی	۰/۸۶۶۹	۰/۹۰۲۷	-۰/۲۶۹	۱	۳/۳۴۶	۱	۰	-۴/۲۶۱	۲	۱۵/۲۶۹	۶۰
واسنجی	۰/۷۷۰۶	۰/۶۳۷۵	-۰/۲۰۵	۱	۳/۵۵۵	۱	۰	-۴/۳۰۲	۲	۱۶/۰۱۴	۶۱
واسنجی	۰/۵۷۹۸	۱/۶۶۱۵	-۰/۱۸۴	۱	۳/۲۴۹	۱	۰	-۴/۰۳۶	۲	۱۴/۷۶۱	۶۲
واسنجی	۱/۰۰۳۳	۱/۵۷۰۲	-۰/۱۷۷	۱	۳/۳۳۷	۱	۱	-۳/۷۷۳	۲	۱۳/۱۸۴	۶۳
واسنجی	۱/۲۳۰۲	۱/۲۹۰۷	-۰/۰۲۸	۱	۳/۳۸۶	۱	۱	-۴/۲۳۷	۲	۱۲/۲۹۶	۶۴
واسنجی	۱/۲۱۵۳	۱/۳۳۰۷	-۰/۱۶۹	۱	۲/۹۳۱	۱	۲	-۴/۱۰۲	۲	۱۶/۳۵۶	۶۵
آزمون	۰/۵۵۸۸	۱/۵۸۸۴	-۰/۱۴۳	۱	۳/۲۳۴	۱	۱	-۴/۰۰۰	۲	۱۸/۰۰۳	۶۶
واسنجی	۱/۳۲۹۵	۱/۵۰۷۲	-۰/۳۶۷	۱	۱/۱۹۵	۱	۲	-۳/۸۴۳	۲	۱۸/۰۵۰	۶۷
واسنجی	۰/۶۸۸۹	۱/۲۲۸۴	-۰/۰۹۵	۱	۲/۹۳۳	۱	۰	-۴/۵۳۳	۲	۱۵/۱۰۸	۶۸

صحت بهترین مدل نخواهد داشت. همان طور که جدول ۲ نشان می‌دهد، در این تحقیق برای روش GA-MLR و مدل‌های ۸ توصیفگر دارای مجذور همبستگی اعتبار-سنجی خارجی بین ۱/۹۳۴۱- تا ۰/۶۷۳۴ و برای روش S-MLR بین ۱/۶۱۰- تا ۰/۶۶۴۲ است. بهترین ارتباط معنی‌دار موجود برای $\log(1/IC_{50})$ این دسته از ترکیبات ضدسرطان در مدل‌های به دست آمده از روش‌های S-MLR و GA-MLR برای دسته

ضدسرطانی، روش‌های S-MLR و GA-MLR بر روی دسته واسنجی اعمال شد. برای دستیابی به بهترین مدل‌های QSAR، ابتدا بهترین مدل‌های دارای یک، دو، سه، چهار تا ده متغیره ساخته شدند و از بین این مدل‌ها، مدل هشت متغیره به عنوان بهترین مدل از نظر شاخص‌های آماری برای هر دو روش برگزیده شد. باید توجه داشت که انتخاب مدل‌های با تعداد بالای توصیفگرها ضروری نیست، زیرا افزایش در تعداد توصیفگرهای مولکولی هیچ گونه اثر معناداری بر دقت و

ادامه جدول ۷. مقادیر توصیفگرهای مولکولی، مقادیر $\log(1/IC_{50})$ آزمایشی و پیشگویی شده توسط مدل GA-MLR

شماره	RDF055e	nNR2	Mor03p	nCp	nRSR	Mor07e	ISH	Mor16p	$\text{Log}(1/IC_{50})_{EXP}$	$\text{log}(1/IC_{50})_{PRD}$	دسته
۶۹	۱۲/۵۴۶	۲	-۴/۱۰۳	۱	۱	۲/۷۷۷	-۰/۹۸۴	-۰/۰۰۸	۰/۵۲۸۴	۰/۵۴۸۰	آزمون
۷۰	۱۱/۷۳۳	۲	-۳/۷۹۰	۱	۱	۳/۰۷۹	-۰/۹۸۴	-۰/۱۷۲	۰/۵۹۶۷	۰/۷۵۳۶	واستجی
۷۱	۱۴/۲۱۹	۲	-۴/۰۸۵	۲	۱	۲/۷۰۶	۱	-۰/۲۴۶	۱/۲۸۴۸	۱/۵۰۱۷	آزمون
۷۲	۱۶/۲۹۲	۲	-۴/۲۰۳	۰	۱	۳/۷۲۵	۱	-۰/۲۷۰	۱/۳۶۸۶	۰/۸۱۹۴	واستجی
۷۳	۱۵/۴۲۴	۲	-۳/۶۶۳	۱	۱	۳/۴۱۴	۱	-۰/۱۵۷	۱/۳۳۲۵	۰/۶۵۸۴	واستجی
۷۴	۱۳/۸۶۱	۲	-۳/۱۹۳	۲	۱	۳/۹۳۶	-۰/۹۸۵	-۰/۵۲۵	۱/۷۲۸۲	۱/۳۴۵۴	واستجی
۷۵	۱۲/۲۵۱	۲	-۳/۳۸۴	۲	۱	۳/۷۶۱	۱	-۰/۱۶۰۵	۱/۲۸۱۵	۲/۰۷۴۱	واستجی
۷۶	۱۸/۵۴۶	۲	-۳/۶۸۳	۲	۱	۳/۸۲۵	۱	-۰/۱۶۱۰	۱/۲۲۰۴	۱/۶۳۴۵	واستجی
۷۷	۱۵/۹۶۹	۲	-۳/۹۱۵	۲	۲	۲/۴۸۲	۰/۹۷	-۰/۴۲۱	۱/۳۱۹۷	۱/۶۰۰۳	آزمون
۷۸	۱۴/۹۶۳	۲	-۳/۱۵۱	۲	۱	۳/۷۵۸	۰/۹۷	-۰/۳۹۸	۱/۳۰۲۸	۰/۶۰۹۱	واستجی
۷۹	۹/۸۱۰	۲	-۳/۴۸۷	۰	۱	۳/۱۵۱	۱	-۰/۱۳۲	۰/۷۱۴۹	۰/۵۸۸۵	واستجی
۸۰	۱۲/۳۴۵	۲	-۳/۶۰۹	۰	۱	۱/۶۴۸	۱	-۰/۱۶۱	۰/۷۹۷۲	۰/۰۱۵۵	واستجی
۸۱	۱۶/۶۵۸	۲	-۳/۹۶۱	۲	۱	۱/۶۵۵	۱	-۰/۳۰۲	۰/۶۷۷۶	۰/۹۲۳۵	واستجی
۸۲	۲۳/۴۹۴	۳	-۳/۴۲۶	۱	۱	۳/۴۰۶	۱	۰/۱۴۰	-۰/۳۳۴۰	-۰/۳۵۹۱	واستجی
۸۳	۱۷/۲۷۹	۲	-۳/۸۰۶	۱	۱	۲/۱۱۰	-۰/۹۸۴	-۰/۰۹۲	-۰/۱۳۰۰	-۰/۲۶۱۹	آزمون
۸۴	۲۰/۳۸۷	۲	-۳/۹۲۲	۰	۱	۲/۹۶۷	-۰/۹۸۶	۰/۰۵۷	-۱/۳۹۳۱	-۰/۹۱۱۹	واستجی
۸۵	۱۲/۹۳۳	۲	-۳/۳۵۶	۰	۱	۱/۸۶۹	۱	-۰/۱۸۸	۰/۴۷۵۲	-۰/۱۳۷۹	واستجی
۸۶	۱۱/۶۵۰	۱	-۳/۴۸۷	۰	۱	۳/۴۴۰	-۰/۹۸۳	۰/۱۴۸	-۱/۲۵۸۱	-۰/۹۱۳۹	واستجی
۸۷	۱۵/۳۳۳	۱	-۳/۶۶۶	۲	۱	۲/۸۵۶	-۰/۹۶۹	۰/۰۷۴	-۱/۲۸۷۲	-۰/۱۶۵۰۵	آزمون
۸۸	۸/۳۶۲	۱	-۲/۱۶۸	۱	۰	۲/۷۹۴	۱	-۰/۰۵۰	-۰/۵۲۳۹	-۱/۰۷۰۵	واستجی
۸۹	۹/۴۱۹	۱	-۲/۳۵۱	۰	۱	۱/۴۹۲	۱	-۰/۱۴۷	-۰/۸۸۸۶	-۱/۲۶۱۵	آزمون
۹۰	۱۱/۱۴۵	۱	-۳/۰۰۶	۰	۱	۲/۸۴۶	۱	-۰/۱۲۵	-۱/۱۰۶۹	-۰/۵۸۹۲	واستجی
۹۱	۱۹/۵۲۵	۲	-۳/۳۵۷	۱	۱	۳/۵۸۴	۱	-۰/۳۱۱	۰/۴۷۲۲	۰/۲۶۹۳	واستجی
۹۲	۸/۹۷۰	۲	-۳/۳۰۴	۰	۱	۲/۶۸۵	-۰/۹۸۱	-۰/۳۰۳	۰/۲۷۷۳	۰/۲۸۲۵	واستجی
۹۳	۱۷/۷۸۲	۲	-۴/۴۲۳	۰	۱	۴/۵۸۸	-۰/۹۸۶	-۰/۲۳۵	۰/۷۶۸۸	۰/۷۲۵۷	واستجی
۹۴	۱۱/۹۸۲	۲	-۳/۵۴۶	۰	۱	۲/۷۰۱	۱	-۰/۳۶۶	۱/۲۲۹۱	۰/۴۸۴۸	واستجی
۹۵	۱۱/۲۷۲	۲	-۳/۳۷۷	۰	۱	۳/۲۹۹	۱	-۰/۱۶۶	۰/۱۴۳۶	۰/۴۳۹۷	واستجی
۹۶	۱۰/۱۳۷	۲	-۳/۴۵۰	۰	۱	۲/۵۲۶	۱	-۰/۳۴۹	۰/۳۱۵۵	۰/۷۰۱۷	واستجی
۹۷	۲۲/۴۷۵	۲	-۳/۷۰۹	۲	۱	۳/۸۴۵	۱	-۰/۳۷۴	۱/۰۹۵۸	۰/۸۳۴۰	آزمون
۹۸	۲۲/۸۲۴	۲	-۴/۱۱۱	۲	۱	۳/۶۲۶	-۰/۹۸۵	-۰/۴۱۴	۰/۸۴۷۱	۰/۷۹۰۶	واستجی
۹۹	۲۳/۵۰۱	۲	-۳/۴۵۹	۳	۱	۴/۰۶۸	۱	-۰/۴۵۹	۰/۸۲۸۶	۱/۱۶۹۹	واستجی
۱۰۰	۳۶/۶۸۶	۲	-۴/۴۸۳	۳	۱	۵/۴۴۸	۱	-۰/۲۸۷	۰/۶۸۲۱	۰/۶۱۶۳	واستجی

۳

$$\text{Log}\left(\frac{1}{IC_{50}}\right) = 25.437 (\pm 8.325) - 0.111 (\pm 0.016)RDF055e + 0.556 (\pm 0.200)nRNR2 - 0.777 (\pm 0.144)Mor03p + 0.439 (\pm 0.094)nCp + 0.840 (\pm 0.243)nRSR + 0.289 (\pm 0.096)Mor07e + 21.324 (\pm 8.322)ISH - 1.654 (\pm 0.479)Mor16p$$

$$Q_{CV}^2 = 0.6650; Q_{Cal}^2 = 0.7183; Q_{Test}^2 = 0.6734 \quad (۲)$$

تصادفی ۱۵ (به عنوان بهترین دسته تصادفی) به صورت زیر است.

برای روش S-MLR معادله حاصل به صورت زیر است:

$$\text{Log}\left(\frac{1}{IC_{50}}\right) = 13.046 (\pm 2.567) + 0.514 (\pm 0.096)RDF020p + 0.564 (\pm 0.161)C - 0.02 - 1.418 (\pm 0.416)Mor16m - 4.582 (\pm 0.983)R4m - 1.797 (\pm 0.495)Mor21m + 65.981 (\pm 14.679)G2e + 0.680 (\pm 0.235)nRNR2 + 1.040 (\pm 0.406)Mor22m$$

$$Q_{CV}^2 = 0.6108; Q_{Cal}^2 = 0.6889; Q_{Test}^2 = 0.4110 \quad (۳)$$

نتایج مربوط به روش‌های S-MLR و GA-MLR در جدول‌های ۳ و ۴ نشان داده شده‌اند. جدول‌های ۵ و ۶ نشان دهنده ماتریس ضرایب همبستگی خطی برای مقادیر $\log(1/IC_{50})$ و هشت متغیر موجود در مدل‌های MLR هستند. نتایج ماتریس ضریب همبستگی بیانگر این موضوع است که بین توصیفگرهای انتخاب شده همبستگی وجود نداشته و توصیف گرها مستقل از هم

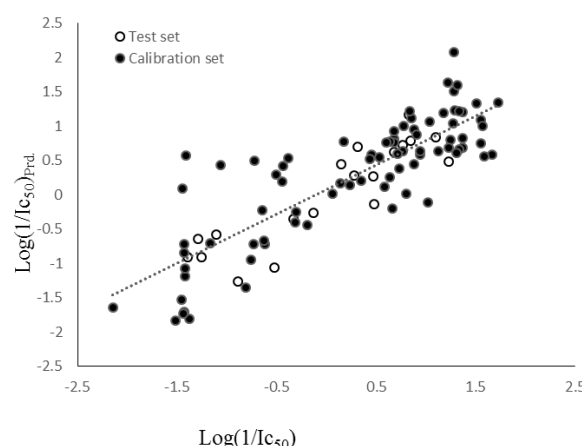
معکوس با فعالیت دارد با افزایش آن این توصیفگر، فعالیت ترکیبات کاهش می یابد. توصیفگرهای nRNR2، nCp و nRSR متعلق به توصیفگرهای شمارش گروه عاملی هستند که به ترتیب با شمارش تعداد آمین نوع سوم، تعداد کربن نوع اول انتهایی sp^3 و تعداد سولفیدها محاسبه می شوند. این گروه جزو ساده ترین توصیفگرهای دراگون هستند. علامت مثبت این ضرایب بیانگر اثر مثبت این متغیرها بر روی $\log(1/IC_{50})$ و در نتیجه با افزایش این توصیفگرها فعالیت این ترکیبات افزایش می یابد.

توصیفگرهای Mor03p، Mor16p و Mor07e متعلق به دسته توصیفگرهای سه بعدی مورس (3D-MoRSE) هستند که به ترتیب توسط قطبش و الکترونگاتیویته ساندرسون توزین شده است. توصیفگرهای سه بعدی مورس (نمایش سه بعدی مولکول از سازه های مبتنی بر پراش الکترونی) از شبیه سازی طیف های مادون قرمز با استفاده از یک تابع پراکندگی عمومی مشتق شده است (۳۰). توصیفگرهای Mor03p و Mor16p به ترتیب نمایانگر سیگنال ۳ و ۱۶ از توصیفگرهای 3D-MoRSE هستند که با استفاده از قطبش پذیری توزین شده است و ضرایب منفی این دو توصیفگر نشان دهنده اثر منفی بر روی $\log(1/IC_{50})$ و در نتیجه اثر منفی بر روی فعالیت دارو است. توصیفگر Mor07e نمایانگر سیگنال ۷ از توصیفگرهای 3D-MoRSE است که با استفاده از الکترونگاتیویته ی ساندرسون توزین شده است و ضرایب مثبت آن نشان دهنده اثر منفی بر روی فعالیت دارو است. توصیفگر ISH متعلق به توصیفگرهای GETAWAY است. توصیفگرهای GETAWAY از جمله توصیفگرهایی هستند که اخیراً ارایه شده و مربوط به ساختار شیمیایی مولکول ها هستند. این گروه از توصیفگرها به منظور جور نمودن هندسه سه بعدی مولکول طراحی شده اند و از عناصر h_{ij} از ماتریس نفوذ مولکولی (H) به دست آمده از طریق مختصات کارترین اتمی، مشتق شده- اند (۳۱،۳۲). توصیفگر ISH دارای ضریب مثبت است که نشان دهنده تاثیر مثبت این متغیر بر روی $\log(1/IC_{50})$ و در نتیجه تاثیر منفی بر روی IC_{50} است و با افزایش این توصیفگر فعالیت ترکیب افزایش می یابد.

در این تحقیق مدلسازی QSAR برای رابطه ساختار- فعالیت ترکیبات او۸-دی هیدرو-۴-کسو-۱- (۲-تيازول)-او۸- نفتیریدین و مشتقات آن به دست آمد. داده ها به روش تصادفی به دو دسته واسنجی و آزمون شکافته شدند. تاثیر انتخاب دسته های تصادفی متفاوت به روش تصادفی بر روی میزان پیشگویی مدل مورد بررسی قرار گرفت. انتخاب بهترین توصیف گرها از میان توصیفگرهای موجود توسط روش های S-MLR و GA-MLR انجام شد. آماره اعتبارسنجی خارجی گزارش شده برای هر مدل،

هستند و نتایج به دست آمده از مدلسازی دال بر وابستگی توصیف کننده ها نیست. همان طور که در جدول ۴ و ۶ مشاهده می شود ضریب همبستگی هر جفت توصیف کننده در روش GA-MLR کمتر از ۰/۴۸۵ است و برای مدل S-MLR ضریب همبستگی هر جفت توصیفگرها کمتر از ۰/۴۳۲ است.

ضریب رگرسیونی استاندارد حاصل از مدلسازی GA-MLR اهمیت نسبی توصیفگرها را نشان می دهد. مهمترین توصیفگر RDF055e است و سایر توصیفگرها به ترتیب عبارتند از: Mor03p>nCp>Mor16p>nRSR>Mor07e>nRNR2>ISH مقادیر توصیفگرهای مولکولی، مقادیر $\log(1/IC_{50})$ آزمایشی و پیشگویی شده توسط مدل GA-MLR برای دسته واسنجی و دسته آزمون در جدول ۷ می توان مشاهده کرد؛ به علاوه نمودار ۱، نمودار مقادیر $\log(1/IC_{50})$ پیشگویی شده در برابر آزمایشی مربوط به مدل GA-MLR برای دسته واسنجی و آزمون ترکیبات ضدسرطان را نشان می دهد.



نمودار ۱. نمودار مقادیر $\log(1/IC_{50})$ پیشگویی شده در برابر آزمایشی مربوط به مدل GA-MLR برای دسته واسنجی و آزمون ترکیبات ضدسرطان

بحث

با تفسیر و ارزیابی توصیفگرهای بهترین مدل GA-MLR این امکان برای بدست آوردن بینش شیمیایی مفید برای طراحی دارو وجود دارد. به همین دلیل، تفسیر قابل قبول از نتایج QSAR در زیر ارائه شده است.

توصیفگر RDF055e متعلق به دسته توصیفگرهای RDF، توصیف گر تابع توزیع شعاعی است که توسط الکترونگاتیویته ساندرسون وزن شده است (۲۹). ضریب منفی این توصیفگر در معادله QSAR نشان دهنده تاثیر منفی بر روی $\log(1/IC_{50})$ و در نتیجه تاثیر مثبت بر روی IC_{50} دارد و با توجه به آنکه IC_{50} رابطه

است. مطابق نتایج حاصل از هر دو روش می توان نتیجه گرفت که روش GA-MLR نسبت به S-MLR روش قدرتمندتری است و نتایج به دست آمده می تواند در طراحی و سنتز داروهای ضد سرطان بسیار سودمند باشد.

مبنایی برای مقایسات نهایی است. نتایج این مقایسه نشان می دهد که مدل های QSAR به دست آمده با روش GA-MLR دارای مجذور ضریب همبستگی اعتبارسنجی بزرگتری نسبت به روش S-MLR می باشند. نتایج نشان می دهند که مقادیر Q^2_{test} برای روش های S-MLR و GA-MLR به ترتیب ۰/۴۱۱۰ و ۰/۶۷۳۴

REFERENCES

1. Funatsu K, Miyao T, Arakawa M, Systematic generation of chemical structures for rational drug design based on QSAR models. *Curr Comput Aided Drug Des* 2011; 7: 1-9.
2. Ahmadi S, Habibpour E, Application of GA-MLR for QSAR modeling of the arylthioindole class of tubulin polymerization inhibitors as anticancer agents. *Anticancer Agents Med Chem* 2017; 17: 552-65.
3. Ahmadi S, Ganji S, Genetic algorithm and self-organizing maps for QSPR study of some N-aryl derivatives as butyrylcholinesterase inhibitors. *Curr Drug Discov Technol* 2017; 13: 232-53.
4. Ahmadi S, Khazaei MR, Abdolmaleki A, Quantitative structure-property relationship study on the intercalation of anticancer drugs with ct-DNA. *Med Chem Res* 2014; 23:1148-61.
5. Spiegel K, Magistrato A, Modeling anticancer drug-DNA interactions via mixed QM/MM molecular dynamics simulations. *Org Biomol Chem* 2006; 4: 2507-17.
6. Goodarzi M, Dejaegher B, Vander Heyden Y, Feature selection methods in QSAR studies. *J AOAC Int* 2012; 95: 636-651.
7. Gonzalez MP, Teran C, Saiz-Urra L, Teijeira M, Variable selection methods in QSAR: an overview. *Curr Topics Med Chem* 2008; 8: 1606-27.
8. Leardi R, Genetic algorithms in chemometrics and chemistry: a review. *J Chemometr* 2001; 15: 559-69.
9. Lucasius CB, Beckers MLM, Kateman G, Genetic algorithms in wavelength selection: a comparative study. *Anal Chim Acta* 1994; 286: 135-53.
10. Shayanfar A, Ghasemi S, Soltani S, Asadpour-Zeynali K, Doerksen RJ, Jouyban A, Quantitative structure-activity relationships of imidazole-containing farnesyltransferase inhibitors using different chemometric methods. *Med Chem* 2013; 9: 434-48.
11. Bohari MH, Srivastava HK, Sastry GN, Analogue-based approaches in anti-cancer compound modelling: the relevance of QSAR models. *Org Med Chem Lett* 2011; 1: 1-12.
12. Bertoša B, Aleksić M, Karminiski-Zamola G, Tomić S, QSAR analysis of antitumor active amides and quinolones from thiophene series. *Int J Pharm* 2010; 394: 106-14.
13. Tomita K, Tsuzuki Y, Shibamori K, Tashima M, Kajikawa F, Sato Y, et al. Synthesis and structure-activity relationships of novel 7-substituted 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridine-3-carboxylic acids as antitumor agents. Part1. *J Med Chem* 2002; 45: 5564-75.
14. Tsuzuki Y, Tomita K, Shibamori K, Sato Y, Kashimoto S, Chiba K, Synthesis and structure-activity relationships of novel 7-substituted 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridine-3-carboxylic acids as antitumor agents. Part 2. *J Med Chem* 2004; 47: 2097-109.
15. Tsuzuki Y, Tomita K, Sato Y, Shibamori K, Kashimoto S, Chiba K, Synthesis and structure-activity relationships of 3- substituted 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines as novel antitumor agents. *Bioorg Med Chem Lett* 2004; 14: 3189-93.
16. Ahmadi S, Babaee E, khazaei MR, Application of self organizing maps and GA-MLR for the estimation of stability constant of 18-crown-6 ether derivatives with sodium cation. *J Incl Phenom Macrocycl Chem* 2014; 79: 141-49.
17. Ahmadi S, Deligeorgiev TG, Vasilev A, Kubista M, The dimerization study of some cationic monomethine cyanine dyes by chemometrics method. *Russ J Phys Chem A* 2012; 86: 1974-81.
18. Ahmadi S, Application of GA-MLR method in QSPR modeling of stability constants of diverse 15-Crown-5 complexes with sodium cation. *J Incl Phenom Macrocycl Chem* 2012; 74: 57-66.
19. Ahmadi S, A QSPR study of association constants of macrocycles toward sodium cation. *Macroheterocycles* 2012; 5: 23-31.
20. Ghasemi JB, Ahmadi S, Brown SD, A quantitative structure-retention relationship study for prediction of chromatographic relative retention time of chlorinated monoterpenes. *Environ Chem Lett* 2011; 9: 87-96

21. Ghasemi JB, Ahmadi S, Ayati M, QSPR modeling of stability constants of the Li-hemispherands complexes using MLR: a theoretical host-guest study. *Macroheterocycles* 2010; 3: 234-42.
22. Ghasemi JB, Ahmadi S, Combination of genetic algorithm and partial least squares for cloud point prediction of nonionic surfactants from molecular structures. *Ann Chim Rome* 2007; 97: 69-83.
23. Rogers D, Hopfinger AJ, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships *J Chem Inf Comput Sci* 1994; 34: 854-66.
24. Cho SJ, Hermsmeier MA, Genetic algorithm guided selection: variable selection and subset selection. *J Chem Inf Comput Sci* 2002; 42: 927-36.
25. Myers RH, *Classical and modern regression with applications*. Boston: PWS-KENT Publishing company; 1990.
26. Hintze J, *NCSS – Number Cruncher Statistical System*. UT: NCSS: Kaysville; 2001.
27. Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT, Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg Med Chem Lett* 2004; 14: 3283-90.
28. Shapiro S, Guggenheim B, Inhibition of oral bacteria by phenolic compounds part1 QSAR analysis using molecular connectivity *Quant Struct Act Relat* 1998; 17: 327-37.
29. Hemmer MC, Steinhauer V, Gasteiger J, Driving the 3D structure of organic molecules from their infrared spectra. *Vib Spectrosc* 1999; 19: 151-64.
30. Gasteiger J, Sadowski J, Schuur J, Selzer P, Steinhauer L, Steinhauer V, Chemical information in 3D space. *J Chem Inf Comput Sci* 1996; 36: 1030-37.
31. Consonni V, Todeschini R, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Theory of the novel 3D molecular descriptors. *J Chem Inf Comput Sci* 2002; 42: 682-92.
32. Consonni V, Todeschini R, Pavan M, Gramatica P, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J Chem Inf Comput Sci* 2002; 42: 693-705.