

# مطالعه ارتباط کمی ساختار - فعالیت برای پیش‌بینی فعالیت مهارکنندگی PIM مشتق‌های تری آزولوپیریدین با استفاده از الگوریتم ژنتیک - برازش خطی چندگانه

اسلام پوربشیر\*<sup>+</sup>، ژیلا مهاجری اول

گروه شیمی، دانشگاه پیام نور، تهران، ایران

مهدی نکوئی

گروه شیمی، دانشگاه آزاد اسلامی، واحد شاهرود، شاهرود، ایران

سمیه حمیدوند

گروه شیمی، دانشگاه پیام نور، تهران، ایران

**چکیده:** مطالعه ارتباط کمی ساختار - فعالیت (QSAR) برای مدل‌سازی و پیش‌بینی فعالیت مهارکنندگی PIM، ۳۹ ترکیب از مشتق‌های تری آزولوپیریدین به کار گرفته شد. نخست پس از رسم و بهینه‌سازی ساختار ترکیب‌ها و محاسبه توصیف‌کننده‌ها، مجموعه داده‌ها به صورت تصادفی به دو مجموعه آموزش و آزمون تقسیم شد. در انتخاب مهم‌ترین توصیف‌کننده‌ها روش گام به گام (SW) و الگوریتم ژنتیک (GA) استفاده شد. سپس برای مدل‌سازی و پیش‌بینی فعالیت‌های ترکیب‌ها از روش برازش خطی چندگانه (MLR) استفاده شد. مقایسه مدل‌های SW-MLR و GA-MLR نشان داد که GA-MLR نسبت به SW-MLR نتیجه‌های بهتری به دست می‌دهد. سپس برای بررسی اعتبار مدل‌ها از روش‌های اعتبارسنجی تقاطعی (LOO) و (LGO)  $Y$ -تصادفی و آزمون خارجی استفاده شد. این مطالعه نشان داد که مدل GA-MLR با پنج توصیف‌کننده می‌تواند برای پیش‌بینی فعالیت ترکیب‌های دارویی همانند به کار رود و همچنین از تحلیل توصیف‌کننده‌های مدل می‌توان در طراحی ترکیب‌های دارویی نوین استفاده کرد.

**واژه‌های کلیدی:** ارتباط کمی ساختار - فعالیت؛ الگوریتم ژنتیک؛ برازش خطی چندگانه؛ مهارکننده‌های PIM؛ مشتق‌های تری آزولوپیریدین.

**KEYWORDS:** QSAR; Genetic algorithms (GA); Multiple linear regressions (MLR); PIM inhibitors; Triazolopyridine derivatives.

## مقدمه

آنزیم PIM متعلق به یک خانواده از پروتئین کیناز سرین /  
ترئونین است که در طی تکامل در موجودات چند سلولی حفظ شده است. کینازهای PIM متشکل از سه ژن بسیار همولوگ PIM1، PIM2 و PIM3 هستند [۱، ۲]. این آنزیم به طور مستقیم در تنظیم پیشرفت

\*عهدہ دار مکاتبات

+E-mail: pourbasheer@ut.ac.ir

از ترکیب دارویی که باعث ۵۰٪ اثر مهارکنندگی می‌شود. این مقادارهای به مقیاس لگاریتمی ( $IC_{50}$ ) تبدیل و مورد استفاده قرار گرفت. کل مجموعه داده شامل ۳۹ ترکیب برحسب مقادارهای فعالیت تجربی بطور صعودی در اکسل مرتب شد و به صورت تصادفی به دو مجموعه‌ی آموزش شامل ۳۱ ترکیب (۸۰٪) و مجموعه‌ی کل مجموعه داده (۲۰٪) تقسیم شد. ترکیب‌های کل مجموعه داده و مجموعه‌ی آزمون شامل ۸ ترکیب (۲۰٪) ترکیب‌های کل مجموعه داده) تقسیم شد. مجموعه‌ی آموزش برای ساخت مدل برازش و مجموعه‌ی آزمون برای ارزیابی قدرت پیش‌بینی مدل به‌دست آمده به کار گرفته شد. ترکیب‌های مجموعه‌ی آموزش و آزمون در جدول ۱ نشان داده شده است.

### محاسبه توصیف کننده‌ها

برای محاسبه توصیف کننده‌ها، نخست ساختار ترکیب‌ها به کمک نرم افزار Hyperchem رسم شدند. سپس ساختارهای مولکولی رسم شده، به وسیله الگوریتم AM1 بهینه شدند. با استفاده از این نرم افزار می‌توان اطلاعات فراوانی مانند زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم‌ها، انرژی تشکیل مولکول و... را به‌دست آورد. ساختارهای بهینه شده به نرم افزار دراگون منتقل و توصیف کننده‌ها به تعداد ۱۴۹۷ مورد به کمک این نرم افزار محاسبه شدند. این نرم افزار برای محاسبه هیجده دسته از توصیف کننده‌های مولکولی مورد استفاده قرار می‌گیرد.

### کاهش تعداد توصیف کننده‌های تئوری

یکی از مشکل‌هایی که در هنگام ایجاد مدل‌های QSAR با آن روبه‌رو می‌شویم، تعداد زیاد متغیرهای مستقل می‌باشد. در بیش‌تر موارد تعداد توصیف کننده‌ها بسیار بیش‌تر از تعداد مولکول‌ها است. در این صورت استفاده از روش‌های کم‌ترین مربع‌ها باعث ایجاد مشکل‌هایی نظیر انتخاب شانسی و همبستگی تصادفی می‌شود. با توجه به این که برخی از متغیرهای مستقل، ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می‌دهند، در نتیجه به روش زیر بعضی از متغیرها حذف شدند.

۱- توصیف کننده‌هایی که مقادارهای ثابت و یا به تقریب ثابت داشتند (بیش از ۹۰٪ داده‌های ثابت داشتند)، حذف شدند. در این مرحله تعداد ۴۱۵ توصیف کننده حذف و بدین ترتیب ۱۰۸۲ توصیف کننده باقی ماند.

چرخه سلولی و آپوپتوز نقش دارد [۳، ۴] و در سرطان‌های بسیاری از جمله سرطان پروستات، لنفوم بورکیت و سرطان دهان، و همچنین به عنوان لنفوم خون‌ساز بسیار نقش دارد [۵، ۶]. پلی مورفیسیم تک نوکلئوتیدی در ژن PIM1 با افزایش خطر ابتلا به سرطان ریه در بیماران کره‌ای همراه بوده است. مشتق‌های ۴-۶-دی (هترو) آریل-۲-ایمینو، ۲-دی هیدرو پیریدین-۳-کربو نیتریل به عنوان مهارکننده PIM1 کیناز امروزه به‌عنوان داروهای ضد سرطان مطرح می‌باشند. بنابراین، مهار دارویی این آنزیم می‌تواند برنامه‌های کاربردی در درمان بیماری‌هایی مانند سرطان، در اختلال‌های التهابی و بیماری‌های ایسکمیک داشته باشد [۷، ۸].

از آنجا که آزمایش‌های تعیین فعالیت مهارکنندگی ترکیب‌های تازه طراحی شده مشکل، گران و وقت گیر هست بنابراین یک روش محاسباتی برای آنالیز و پیش‌بینی ویژگی‌های ترکیب‌های گوناگون، پیش از سنتز ترکیب، مهم و لازم است. QSAR به عنوان یک روش محاسباتی با ایجاد رابطه دقیق بین داده‌های تجربی و نظری یک سری ترکیب‌ها، برای پیش‌بینی فعالیت‌های دارویی ترکیب‌های جدید استفاده می‌شود [۹-۱۳].

برای آغاز روش QSAR لازم است پارامترهای نظری به نام توصیف کننده‌ها که با مقدار جبری ساختار یا اشکال هر یک از مولکول‌های انتخاب شده توصیف شده محاسبه شود [۸، ۱۱]. بین تعداد زیادی از توصیف کننده‌های محاسبه شده برای هر مولکول تنها تعدادی از توصیف کننده‌ها نقش مهمی در فعالیت زیستی دارند. بنابراین لازم است از ابزار انتخاب متغیر برای انتخاب توصیف کننده‌ها مؤثر در ایجاد روش استفاده شود.

در این کار روش‌های انتخاب متغیر گام به گام (SW) و الگوریتم ژنتیک (GA) همراه با برازش خطی چندگانه (MLR) از فناوری‌های به کار رفته در ایجاد مدل QSAR بودند. هدف اصلی کار حاضر توسعه یک مدل QSAR خطی جدید با استفاده از فناوری برازش خطی الگوریتم ژنتیک چندگانه برای پیش‌بینی فعالیت‌های مهارکنندگی PIM1 از مشتق‌های تری آزلوپیریدین است.

### محاسبه‌ها

#### انتخاب مجموعه داده‌ها

مجموعه داده‌ها شامل فعالیت دارویی ۳۹ ترکیب از مشتق‌های تری آزلوپیریدین توسط روش‌های کمومتریکس مورد بررسی قرار گرفت [۱۴]. قدرت مهارکنندگی این ترکیب‌ها به صورت  $IC_{50}$  گزارش شده است.  $IC_{50}$  عبارت است از کم‌ترین غلظتی

جدول ۱- ساختارهای شیمیایی و مقادارهای تجربی و پیش‌بینی شده  $pIC_{50}$  با استفاده از مدل GA-MLR.

شماره	ساختار عمومی	R <sup>1</sup>	R <sup>2</sup>	تجربی	پیش‌بینی
۱				۶,۵۵	۶,۹۸
۲				۵,۴۱	۶,۰۱
۳				۶,۰۷	۶,۱۹
۴				۵,۵۹	۵,۵۶
۵				۶,۵۰	۶,۴۹
۶				۶,۳۶	۶,۲۰
۷				۶,۴۰	۶,۳۰
۸				۷,۱۶	۷,۲۲
۹				۶,۷۹	۶,۱۸
۱۰ <sup>a</sup>				۶,۴۹	۶,۸۷
۱۱ <sup>a</sup>				۵,۴۹	۵,۴۵

ادامه جدول ۱- ساختارهای شیمیایی و مقدارهای تجربی و پیش‌بینی شده  $pIC_{50}$  با استفاده از مدل GA-MLR.

شماره	ساختار عمومی	R <sup>1</sup>	R <sup>2</sup>	تجربی	پیش‌بینی
۱۲			-	۶٫۸۱	۷٫۱۹
۱۳			-	۸٫۲۲	۷٫۴۳
۱۴ <sup>a</sup>			-	۷٫۴۶	۷٫۱۴
۱۵			-	۷٫۱۱	۷٫۱۴
۱۶ <sup>a</sup>			-	۶٫۶۵	۷٫۲۴
۱۷			-	۶٫۵۸	۷٫۰۳
۱۸			-	۶٫۶۲	۶٫۴۳
۱۹			-	۶٫۳۶	۶٫۶۲
۲۰ <sup>a</sup>			-	۶٫۰۷	۶٫۰۷
۲۱			-	۷٫۴۲	۷٫۲۵
۲۲		-	۷٫۰۶	۶٫۷۵	

ادامه جدول ۱- ساختارهای شیمیایی و مقادیرهای تجربی و پیش‌بینی شده  $pIC_{50}$  با استفاده از مدل GA-MLR.

شماره	ساختار عمومی	R <sup>1</sup>	R <sup>2</sup>	تجربی	پیش‌بینی
۲۳			-	۶,۴۷	۶,۰۸
۲۴			-	۸,۰۰	۸,۰۵
۲۵ <sup>a</sup>			-	۷,۹۶	۷,۸۷
۲۶			-	۸,۴۰	۸,۰۳
۲۷			-	۷,۱۷	۷,۵۱
۲۸			-	۶,۹۰	۶,۸۰
۲۹ <sup>a</sup>			-	۶,۹۷	۷,۱۱
۳۰			-	۶,۷۸	۶,۹۴
۳۱			-	۵,۶۲	۶,۴۱
۳۲			-	۷,۷۷	۷,۸۳
۳۳				۸,۳۰	۷,۴۶
۳۴				۷,۸۵	۷,۶۱

ادامه جدول ۱- ساختارهای شیمیایی و مقدرهای تجربی و پیش‌بینی شده با  $pIC_{50}$  استفاده از مدل GA-MLR.

شماره	ساختار عمومی	R <sup>1</sup>	R <sup>2</sup>	تجربی	پیش‌بینی
۳۵				۸,۰۵	۷,۹۹
۳۶				۷,۸۵	۷,۸۱
۳۷				۶,۸۸	۷,۱۲
۳۸				۶,۵۰	۵,۹۲
۳۹				۶,۳۹	۶,۵۶

جستجو در میان توصیف کننده‌ها به منظور یافتن مؤثرترین توصیف کننده‌ها برای انجام مدلسازی است. الگوریتم ژنتیک، یک روش جستجوی هوشمند و تصادفی است که با به کارگیری عملگرهای ژنتیک از یک فرایند تکامل تدریجی پیروی می‌کند. از الگوریتم ژنتیک در شیمی به‌طور گسترده‌ای در زمینه مدل‌سازی QSAR و کمومتریکس استفاده شده است [۲۱ - ۱۶].

### نتیجه‌ها و بحث

در ابتدا با استفاده از نرم افزار SPSS و به روش گام به گام مناسب‌ترین توصیف کننده‌ها انتخاب شد. با این روش ۵ توصیف کننده انتخاب و مدل SW-MLR به‌صورت زیر به‌دست آمد.

$$pIC_{50} = -9.5885(\pm 4.314) + (1) \\ 10.127(\pm 1.951)BELm - 3.875(\pm 1.221)GATS4p - \\ 1.640(\pm 0.484)Mor23m - 10.9326(\pm 42.013)R6p - \\ 3.998(\pm 1.666)MATSYv$$

۲- توصیف کننده‌هایی که همبستگی بالای ۰/۹ با یکدیگر داشتند مورد بررسی قرار گرفتند و بین آن‌ها، توصیف کننده‌ای که همبستگی کمتری با متغیر مستقل داشت حذف شد. بدین ترتیب تعداد ۶۵۷ توصیف کننده، کنار گذاشته شد و سرانجام تعداد ۴۲۵ توصیف کننده باقی ماند.

### انتخاب توصیف کننده‌های مؤثر

مهم‌ترین بخش در ایجاد یک مدل کارآمد، انتخاب توصیف کننده‌های مناسب است. پس از محاسبه توصیف کننده‌های گوناگون، تعدادی از آن‌ها به عنوان توصیف کننده‌های مناسب برای ساخت مدل انتخاب می‌شوند. این مرحله شامل یافتن توصیف کننده‌های دارای اطلاعات مفید است به طوری که قدرت پیش‌بینی مدل در سطح قابل پذیرش باشد [۱۵، ۱۴، ۱۱].

در این کار از روش کلاسیک گام به گام و همچنین از روش به نسبت نوین الگوریتم ژنتیک برای انتخاب مناسب‌ترین توصیف کننده‌ها استفاده شد. کاربرد الگوریتم ژنتیک در مدل‌سازی QSAR

سپس از معادله به دست آمده برای پیش‌بینی فعالیت‌های مجموعه آزمون استفاده شد. مقدارهای واقعی و پیش‌بینی شده فعالیت‌ها برای کلیه ترکیب‌های مجموعه آموزش و آزمون در جدول ۱، آورده شده است. همچنین مقدارهای فعالیت‌های پیش‌بینی شده و تجربی برای ترکیب‌ها براساس مدل MLR در دو مجموعه آموزش و آزمون در شکل ۱ آورده شده است. همان‌گونه که در شکل دیده می‌شود نزدیکی نتیجه‌ها به یکدیگر، اعتبار مدل را نشان می‌دهد. برای ارزیابی اهمیت و میزان تأثیر توصیف‌کننده‌هایی که در مدل وارد شدند، اثر متوسط هر توصیف‌کننده به صورت زیر محاسبه شد.

$$ME_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j \sum_i \beta_j d_{ij}} \quad (3)$$

در این معادله  $ME_j$  اثر متوسط توصیف‌کننده  $j$ ،  $\beta_j$  ضریب توصیف‌کننده در معادله MLR،  $d_{ij}$  مقدار توصیف‌کننده  $j$ ام برای مولکول  $i$ ام،  $m$  تعداد توصیف‌کننده‌های وارد شده در مدل و  $n$  تعداد مولکول‌هاست. مقدارهای  $ME_j$  برای هر توصیف‌کننده در جدول ۲ نشان داده شده است.

### ارزیابی مدل‌ها با استفاده از پارامترهای آماری

سه پارامتر آماری مهم، برای مقایسه و ارزیابی توانایی پیش‌بینی مدل ساخته شده به روش GA-MLR و SW-MLR در جدول ۴ نمایش داده شده است. با توجه به جدول ۴، می‌توان دید که نتیجه‌های روش الگوریتم ژنتیک هم برای مجموعه آموزش و هم برای مجموعه آزمون، بهتر از روش گام به گام می‌باشد. بنابراین در مرحله‌های بعد مدل GA-MLR بحث و بررسی خواهد شد.

### ارزیابی مدل‌ها با استفاده از آزمون Y-تصادفی

این تکنیک ارزیابی مدل، با هدف بررسی هرگونه ارتباط تصادفی بین داده‌ها انجام شد. در این آزمون، متغیر وابسته به‌طور تصادفی بهم ریخته شد. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقدارهای تصادفی از متغیر وابسته توسعه یافت. اگر در مدل اصلی هیچ گونه ارتباط تصادفی وجود نداشته باشد، تفاوت چشمگیری بین مقدار ضریب تعیین مدل اصلی و مدل QSAR که با پاسخ تصادفی توسعه یافته، وجود خواهد داشت. نتیجه‌های به دست آمده از چندین بار اجرای

$$N_{train}=31, R^2_{train}=0.777, R^2_{test}=0.228, R^2_{adj}=0.726, \\ F_{train}=169.07, F_{test}=0.195, RMSE_{train}=0.362, RMSE_{test}=0.794, \\ Q^2_{LOO}=0.655$$

همان‌گونه که از نتیجه‌های به دست آمده از مدل دیده می‌شود پارامترهای آماری از جمله ضریب برازش ( $R^2$ ) برای مجموعه آموزش به نسبت مناسب بوده اما برای مجموعه آزمون اصلاً مناسب نیست به همین خاطر و برای حصول نتیجه‌های بهتر از ژنتیک الگوریتم برای انتخاب مناسب‌ترین توصیف‌کننده‌ها استفاده شد.

با استفاده از الگوریتم ژنتیک از بین ۴۲۵ توصیف‌کننده باقی مانده تعداد ۵ توصیف‌کننده به عنوان مناسب‌ترین آن‌ها انتخاب شد. لیست توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک به همراه توصیف مختصری از آن‌ها در جدول ۲ آورده شده است.

### ارزیابی توصیف‌کننده‌های انتخاب شده

به منظور ارزیابی توصیف‌کننده‌های انتخاب شده مبنی بر مستقل بودن از همدیگر در جدول ۳ ضریب‌های همبستگی توصیف‌کننده‌های انتخاب شده نسبت به یکدیگر آورده شده است. نتیجه‌ها نشان می‌دهد که بین توصیف‌کننده‌های انتخاب شده به جزء دو توصیف‌کننده BELV7 و H2e همبستگی به نسبت پایینی وجود داشته و توصیف‌کننده‌ها به تقریب مستقل از هم هستند.

### ایجاد مدل با استفاده از MLR

پس از انتخاب مناسب‌ترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک، مرحله بعدی ایجاد مدل، میان توصیف‌کننده‌های انتخاب شده و فعالیت‌های ترکیب‌ها می‌باشد [۲۲-۲۴]. از نرم افزار SPSS برای این منظور استفاده شد. بین توصیف‌کننده‌ها و فعالیت‌های مهارکنندگی مشتق‌های تری آزلوپیریدین و مجموعه آموزش با استفاده از روش MLR رابطه زیر به دست آمد:

$$pIC_{50} = +4.336(\pm 1.30) + \\ 6.557(\pm 1.357)BELV7 - 3.277(\pm 1.304)GATS7V + \\ 0.799(\pm 0.19)Mor0.2m - 0.803(\pm 0.320)H2e - \\ 0.43(\pm 0.08)PSA \\ N_{train}=31, R^2_{train}=0.790, R^2_{test}=0.792, R^2_{adj}=0.748, \\ F_{train}=18.845, F_{test}=0.275, RMSE_{train}=0.3467, \\ RMSE_{test}=0.4096, Q^2_{LOO}=0.705$$

جدول ۲- توصیف کننده‌های انتخاب شده توسط الگوریتم ژنتیک به همراه توصیف مختصری از آن‌ها.

توصیف کننده	شرح مختصر	اثرات میانگین (MF)
BELv7	پایین‌ترین مقادارهای ویژه ماتریس بوردن/ با حجم اتمی و اندوالس وزن می‌شود.	۳/۰۶۵
GATS1v	توصیف کننده همبستگی گیری با فاصله توپولوژیک ۱ می‌باشد/ توسط حجم اتمی و اندوالس وزن شده است.	-۱/۳۴۳
Mor02m	3D مورس - سیگنال ۲/ با وزن اتمی وزن می‌شود.	۰/۹۷۴
H2e	توصیف کننده همبستگی H با فاصله توپولوژیک ۲/ توسط الکترون‌گاتیوتیه اتمی ساندسون وزن شده است.	-۱/۲۱۱
PSA	ناحیه‌های سطح قطبی	۰/۴۸۵

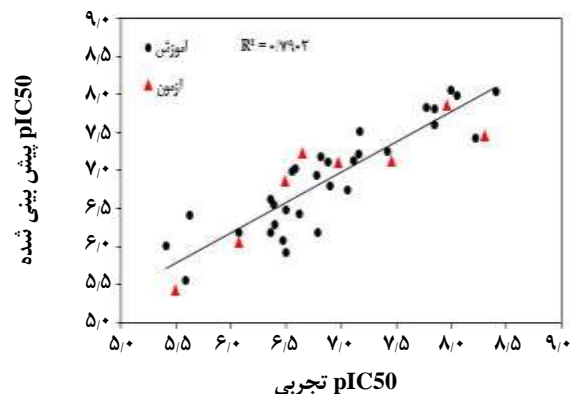
جدول ۳- ماتریس ضرایب همبستگی توصیف کننده‌های انتخاب شده.

	BELv7	GATS1v	Mor02m	H2e	PSA
BELv7	۱	۰	۰	۰	۰
GATS1v	۰/۶۶۱	۱	۰	۰	۰
Mor02m	۰/۵۰۷	۰/۲۰۷	۱	۰	۰
H2e	۰/۸۳۹	۰/۵۹۱	۰/۶۱۲	۱	۰
PSA	۰/۰۹۰	۰/۱۵۰	۰/۳۷۶	۰/۲۲۲	۱

بنابراین توسط طبیعت ترکیب‌های شیمیایی در مجموعه آموزش و توسط توصیف کننده‌های مولکولی هر مدل نمایش داده می‌شود. یک روش ساده و در عین حال قابل کاربرد در مشکل توزیع نرمال داده‌ها در روش‌های مدلسازی اندازه گیری فاصله مشتق‌های ترکیب‌های تا Leverage است. مقدارهای Leverage فاصله ترکیب‌ها را از مرکز فضای متغیرها مشخص می‌کند. با روش Leverage می‌توان فهمید که آیا ترکیب شیمیایی نوین در محدوده ساختاری ترکیب‌های مدل شده می‌باشد یا این‌که از محدوده ساختاری ترکیب‌های مدل بیرون شده است.

به منظور نمایش دادن محدوده کاربرد پذیری مدل GA-MLR، نمودار باقیمانده‌های استاندارد شده با روش ارزیابی متقاطع، بر حسب Leverage در شکل ۲، رسم شده است (نمودار ویلیامز). از روی نمودار ویلیامز می‌توان دید که تنها مولکول‌های شماره ۳، ۷ و ۱۰ مقدار Leverage بالاتر از مقدار بحرانی دارد. مقدار Leverage بحرانی در این مدل ۰/۵۸ می‌باشد. بنابراین می‌توان نتیجه گرفت که این ترکیب‌ها در ساخت مدل تأثیر بیشینه‌ای داشته است.

هرچند که این ترکیب‌ها را می‌توان به عنوان یک ترکیب پرت یا بیرون از رده دانست ولی خوشبختانه برای این ترکیب



شکل ۱- مقدارهای فعالیت‌های پیش‌بینی شده و تجربی برای هر دو مجموعه‌ی آموزش و آزمون براساس مدل GA-MLR.

آزمون Y- تصادفی در جدول ۵، نشان داده شده است. مقدارهای کوچک ضریب تعیین ( $R^2$ ) بیانگر عدم ارتباط شانس در مدل توسعه یافته توسط برازش خطی چندگانه می‌باشد.

### ب کاربرد پذیری

محدوده کاربرد پذیری یک محدوده نظری از فضای شیمیایی است که توسط توصیف کننده‌های مدل و پاسخ مدل تعریف می‌شود و



جدول ۴ - پارامترهای آماری گوناگون برای مدل‌های انتخاب شده.

	R <sup>2</sup>		F		RMSE	
	آموزش	آزمون	آموزش	آزمون	آموزش	آزمون
GA-MLR	۰٫۷۹۰	۰٫۷۹۲	۱۸٫۸۴۵	۱٫۳۷۵	۰٫۳۴۷۶	۰٫۴۰۹۶
SW-MLR	۰٫۷۷۲	۰٫۲۲۸	۱۶٫۹۰۷	۰٫۱۹۵	۰٫۳۶۲	۰٫۷۹۴

جدول ۵ - نتیجه‌های به دست آمده از ۱۰ بار آزمون Y- تصادفی.

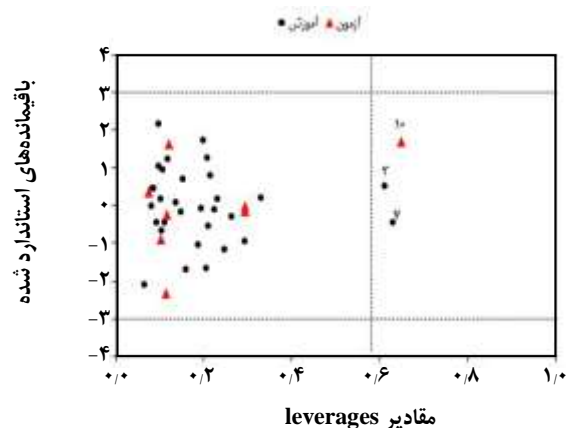
تکرار	R <sup>2</sup> <sub>training</sub>	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
		GA-MLR	۰٫۰۳۲	۰٫۰۰۶	۰٫۰۰۲	۰٫۰۶۰	۰٫۳۰۴	۰٫۰۰۴	۰٫۰۰۰	۰٫۰۹۹	۰٫۰۵۰
SW-MLR	۰٫۱۴۵	۰٫۲۴۴	۰٫۱۷۷	۰٫۰۹۹	۰٫۳۰۶	۰٫۱۸۷	۰٫۱۹۶	۰٫۱۰۹	۰٫۱۱۲	۰٫۰۷۶	

شماره ۷ از ماتریس بار/ توزین شده به‌وسیله حجم اتمی و اندروالس) هست که از دسته توصیف‌کننده‌های دو بعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده (ماتریس بوردن) می‌باشند. ماتریس بوردن یک گراف تهی از هیدروژن را ارائه می‌دهد که در آن B<sub>ii</sub>، یعنی عنصرهای قطری در ارتباط با خاصیت اتمی مانند الکترونگاتیویته، حجم و اندروالس، جرم اتمی و عناصر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند. BELv7 ظاهر شده در مدل بالاترین مقدار عددی و علامت مثبت دارد. بنابراین این توصیف‌کننده نقش مهمی در افزایش فعالیت دارد، زیرا مقدار PIC<sub>50</sub> به طور مستقیم با این توصیف‌کننده مرتبط است. در نتیجه با افزایش حجم و اندروالس اتمی مقدار این توصیف‌کننده و همچنین مقدار PIC<sub>50</sub> افزایش می‌یابد.

دومین توصیف‌کننده GATS<sub>1v</sub> می‌باشد که یکی از توصیف‌کننده‌های ضریب همبستگی دو بعدی (2D)-autocorrelation است. توصیف‌کننده‌های دو بعدی از گراف مولکولی و با محاسبه وزن‌های اتم‌های انتهایی کل مسیرها با طول مسیر مورد نظر، به دست می‌آیند. نوعی از این توصیف‌کننده‌ها گروه Gearyautocorrelation است که ضریب گری نام دارد و بدین صورت محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} (w_i - w_j)^2}{\frac{1}{A-1} \sum_{i=1}^A (w_i - \bar{w}_j)^2} \quad (4)$$

که w یک ویژگی اتم،  $\bar{w}$  میانگین مقدار آن ویژگی روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و  $\delta_{ij}$  نیز



شکل ۲- نمودار ویلیامز برای مدل GA-MLR.

مقدار فعالیت پیش‌بینی شده خیلی خوب است و بنابراین این ترکیب را می‌توانیم در مدل نگه داریم. همان‌طوری که از شکل ۲ دیده می‌شود هیچ کدام از ترکیب‌های باقیمانده استاندارد بالاتر از سه برابر انحراف استاندارد باقیمانده‌ها ندارند.

### بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با اثر مهارکنندگی

تفسیر هر توصیف‌کننده ظاهر شده در مدل مطابق با معانی شیمیایی آن‌ها، علامت و ضریب‌ها می‌تواند اطلاعات شیمیایی مفید در مورد مکانیسم فعالیت مهارکنندگی ترکیب‌ها ارائه دهد. به همین منظور با استفاده از اطلاعات داده شده در جدول ۲، توصیف‌کننده‌های موجود در مدل GA-MLR تفسیر شدند. اولین توصیف‌کننده در مدل به‌دست آمده BELv7 (کم‌ترین مقدار ویژه

اتم‌ها در مولکول مرتبط است. هر چه الکترونگاتیویته، اندازه اتم‌ها و فاصله بین اتم و مرکز مولکول بیش تر باشد مقدار توصیف کننده بیش تر است. علامت منفی این توصیف کننده نشان می‌دهد که مقدار  $pIC_{50}$  با  $H2e$  رابطه معکوس دارد. بنابراین افزایش الکترونگاتیویته ساندروسون اتمی، منجر به افزایش این توصیف کننده و سرانجام کاهش مقدار  $pIC_{50}$  می‌شود. توصیف کننده نهایی در مدل سطح فضایی قطبی (PSA) هست که توسط اتم‌های قطبی یک مولکول تشکیل شده است. این توصیف کننده است که با انتقال مولکولی از طریق غشاء مرتبط است و ویژگی‌های انتقالی داروها از طریق این توصیف کننده تخمین زده می‌شود [۲۵]. توصیف کننده PSA علامت منفی دارد. این نشان دهنده آن است که این توصیف کننده با مقدار  $pIC_{50}$  رابطه عکس دارد. بنابراین، افزایش این توصیف کننده منجر به کاهش مقدار  $pIC_{50}$  می‌شود. در نتیجه حجم واندروالس اتمی، جرم اتمی، الکترونگاتیوی ساندروسون اتمی و سطح‌های قطبی نقش مهمی در مهار PIM از این نوع ترکیب‌ها بازی می‌کند.

### نتیجه‌گیری

در این مطالعه، مدل QSAR برای پیش‌بینی فعالیت مهارکنندگی PIM، ۳۹ ترکیب از مشتق‌های تری آزلوپیریدین به کار گرفته شد. دو مدل خطی چند متغیره با ۵ توصیف کننده انتخاب شده به وسیله روش‌های گام به گام (SW) و الگوریتم ژنتیک (GA) ایجاد شد. مقایسه پارامترهای آماری به دست آمده برای مدل‌های SW-MLR و GA-MLR نشان داد که مدل GA-MLR برای هر دو مجموعه آموزش و آزمون نتیجه‌های خوبی ارائه می‌دهد. سرانجام با استفاده از تکنیک‌های ارزیابی مشخص شد که مدل GA-MLR ایجاد شده با بالاترین کیفیت آماری و همچنین کم‌ترین خطای پیش‌بینی می‌تواند فعالیت مهارکنندگی PIM این نوع ترکیب‌ها را پیش‌بینی کند.

که به تابع کروئکر معروف است، در حالتی که  $d=d_{ij}$  باشد مقدار این تابع یک است و در غیر این صورت صفر است و  $\Delta$  نیز مجموعه‌ی  $\delta$  هاست. مقدار این توصیف کننده‌ها که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است. از میان این توصیف کننده‌ها  $GATS1v$  وارد مدل برتر شد که نقش خاصیت فیزیکی و الکترونگاتیویته‌ی ساندروسون را بیان می‌کند. این توصیف کننده دارای اثر متوسط منفی بوده، بدین معنی که با افزایش مقدار این توصیف کننده،  $pIC_{50}$  کاهش می‌یابد.

سومین توصیف کننده  $Mor02m$  می‌باشد که جزء توصیف کننده‌های 3D-MORSE می‌باشد. توصیف کننده‌های 3D-MORSE (نمایش سه بعدی ساختار مولکول بر اساس تفرق الکترون) از طریق معادله تبدیلی که در پراش الکترون استفاده می‌شود، محاسبه می‌گردند:

$$I(S) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (5)$$

$I$  شدت پراکنده شده،  $A_i$  و  $A_j$  خاصیت (جرم) اتمی  $i$  و  $j$ ،  $s$  زاویه پراکندگی،  $r_{ij}$  فاصله بین اتم‌های  $i$  و  $j$  و  $N$  نیز تعداد کل اتم‌ها را نشان می‌دهد. این روش باعث می‌شود که ساختار سه بعدی مولکول به یک کد ثابت تبدیل شود. چون این توصیف کننده‌ها آرایش سه بعدی اتم‌ها را بدون اینکه به اندازه مولکول ارتباط داشته باشند، بیان می‌کنند بنابراین برای تعداد زیادی مولکول با تفاوت‌های ساختاری زیاد قابل کاربرد است.

$Mor02m$  علامت مثبت دارد که نشان دهنده ارتباط مستقیم این توصیف کننده با مقدار  $pIC_{50}$  است. بنابراین با افزایش جرم اتمی، مقدار این توصیف کننده و سرانجام مقدار  $pIC_{50}$  افزایش می‌یابد. توصیف کننده بعدی  $H2e$  است ( $H$  همبستگی ناحیه  $2/2$  توزین شده به وسیله الکترونگاتیویته ساندروسون اتمی) که متعلق به توصیف کننده GETAWAY می‌باشد. این توصیف کننده، یکی از توصیف کننده‌هایی است که با الکترونگاتیویته، اندازه و موقعیت

تاریخ دریافت: ۱۳۹۶/۱/۳۰ ؛ تاریخ پذیرش: ۱۳۹۶/۷/۲۴

### مراجع

- [1] Narlik-Grassow M., Blanco-Aparicio C., Cecilia Y., Peregrina S., Serelde B.G., Muñoz-Galvan S., Cañamero M., Carnero A., [Essential Role of PIM Kinases in Sarcoma Growth and Bone Invasion](#), *Carcinogenesis*, **33**: 1479-1486 (2012).

- [2] Holder S., Zemskova M., Zhang C., Tabrizid M., Bremer R., Neidigh J.W., Lilly M.B., [Characterization of a Potent and Selective Small-Molecule Inhibitor of the PIM1 Kinase](#), *Molecular Cancer Therapeutics*, **6**: 163-172 (2007).
- [3] Amaravadi R., Thompson C.B., [The Survival Kinases Akt and Pim as Potential Pharmacological Targets](#), *The Journal of clinical investigation*, **115**: 2618-2624 (2005).
- [4] Davis R.A., Simpson M.M., Nugent R.B., Carroll A.R., Avery V.M., Rali T., Chen H., Qurallo B., Quinn R.J., [PIM2 Inhibitors from the Papua New Guinean Plant \*Cupaniopsis Macropetala\*](#), *Journal of Natural Products*, **71**: 451-452 (2007).
- [5] Baytel D., Shalom S., Madgar I., Weissenberg R., Don J., [The Human PIM-2 Proto-Oncogene and Its Testicular Expression](#), *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, **1442**: 274-285 (1998).
- [6] Allen J.D., Verhoeven E., Domen J., van der Valk M., Berns A., [PIM-2 Transgene Induces Lymphoid Tumors, Exhibiting Potent Synergy with c-myc](#), *Oncogene*, **15**: 1133-1141 (1997).
- [7] Pourbasheer E., Riahi S., Ganjali M.R., Norouzi P., [Quantitative Structure-Activity Relationship \(QSAR\) Study of Interleukin-1 Receptor Associated Kinase 4 \(IRAK-4\) Inhibitor Activity by the Genetic Algorithm and Multiple Linear Regression \(GA-MLR\) Method](#), *Journal of Enzyme Inhibition and Medicinal Chemistry*, **25**: 844-853 (2010).
- [8] Sahebamee H., Yaghmaei P., Abdolmaleki P., Foroumadi A.R., [Quantitative Structure - Activity Relationships Study of Carbonic Anhydrase Inhibitors Using Logistic Regression Model](#), *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, **32**: 19-29 (2013).
- [9] Gajare S.P., Mahajan S.S., [QSAR Studies of Phthalimide Derivatives for Their Potent Anxiolytic Activity](#), *International Journal of Current Research and Review*, **4**: 151-156 (2012).
- [10] Ivan D., Crisan L., Funar-Timofei S., Mracec M., [A Quantitative Structure-Activity Relationships Study for the Anti-HIV-Lactivities of 1-\[\(2-hydroxyethoxy\) methyl\]-6--\(phenylthio\) Thymine Derivatives Using the Multiple Linear Regression and Partial Least Squares Methodologies](#), *Journal of Serbian Chemical Society*, **78**: 495-506 (2013).
- [11] Barahona-Urbina C., Nunez-Gonzalez S., Gomez-Jeria J.S., [Model-Based Quantum-Chemical Study of the Uptake of Some Polychlorinated Pollutant Compounds by Zucchini Subspecies](#), *Journal of the Chilean Chemical Society*, **57**: 1497-1503 (2012).
- [12] Pourbasheer E., Aalizadeh R., Ganjali M. R., [Analysis of B-RafV600E Inhibitors Using 2D and 3D-QSAR, Molecular Docking and Pharmacophore Studies](#), *Molecular Diversity*, **19**: 915-930 (2015).
- [13] Pourbasheer E., Aalizadeh R., Shiri H., Banaei A., Ganjali M., [2D and 3D-QSAR Analysis of Pyrazole-Thiazolinone Derivatives as EGFR Kinase Inhibitors by CoMFA and CoMSIA](#), *Current Computer-Aided Drug Design*, **11**: 279-290 (2015).
- [14] Pastor J., Oyarzabal J., Saluste G., Alvarez R.M., Rivero V., Ramos F., Cendón E., Blanco-Aparicio C., Ajenjo N., Cebriá A., [Hit to Lead Evaluation of 1, 2, 3-triazolo \[4, 5-b\] Pyridines as PIM Kinase Inhibitors](#), *Bioorganic & Medicinal Chemistry Letters*, **22**: 1591-1599 (2012).

- [15] Azar P.A., Nekoei M., Riahi S., Ganjali M.R., Zare K., [A Quantitative Structure–Retention Relationship for the Prediction of Retention Indices of the Essential Oils of \*Ammoides Atlantica\*](#), *Journal of Serbian Chemical Society*, **76**: 891-902 (2011).
- [16] Adimi M., Salimi M., Nekoei M., Pourbasheer E., Beheshti A., [A Quantitative Structure-Activity Relationship Study on Histamine Receptor Antagonists Using the Genetic Algorithm-Multi-Parameter Linear Regression Method](#), *Journal of Serbian Chemical Society*, **77**: 639-650 (2012).
- [17] Li J., Lei B., Liu H., Li S., Yao X., Liu M., Gramatica P., [QSAR Study of Malonyl-CoA Decarboxylase Inhibitors Using GA-MLR and a New Strategy of Consensus Modeling](#), *Journal of Computational Chemistry*, **29**: 2636-2647 (2008).
- [18] Zhu W., Chen G., Hu L., Luo X., Gui C., Luo C., Pua C.M., Chen K., Jiang H., [QSAR Analyses on Ginkgolides and Their Analogues Using CoMFA, CoMSIA, and HQSAR](#), *Bioorganic & Medicinal Chemistry*, **13**: 313-322 (2005).
- [19] Pourbasheer E., Aalizadeh R., Ganjali M.R., Norouzi P., [QSAR Study of IKK \$\beta\$  Inhibitors by the Genetic Algorithm: Multiple Linear Regressions](#), *Medicinal Chemistry Reserch*, **23**: 57-66 (2014).
- [20] Pourbasheer E., Aalizadeh R., Ganjali M.R., Norouzi P., Banaei A., [QSAR Study of mGlu5 Inhibitors by Genetic Algorithm-Multiple Linear Regressions](#), *Medicinal Chemistry Reserch*, **23**: 3082-3091 (2014).
- [21] Pourbasheer E., Aalizadeh R., Ganjali M.R., Norouzi P., Shadmanesh J., Methenitis C., [QSAR Study of Nav1.7 Antagonists by Multiple Linear Regression Method Based on Genetic Algorithm \(GA-MLR\)](#), *Medicinal Chemistry Reserch*, **23**: 2264-2276 (2014).
- [22] Pourbasheer E., Beheshti A., Khajehsharifi H., Ganjali M.R., Norouzi P., [QSAR Study on hERG Inhibitory Effect of Kappa Opioid Receptor Antagonists by Linear and Non-Linear Methods](#), *Medicinal Chemistry Reserch*, **22**: 4047-4058 (2013).
- [23] Firoozpour L., Sadatnezhad K., Dehghani S., Pourbasheer E., Foroumadi A., Shafiee A., Amanlou M., [An Efficient Piecewise Linear Model for Predicting Activity of Caspase-3 Inhibitors](#), *DARU Journal of Pharmaceutical Science*, **20**: 1-6 (2012).
- [24] Nekoei M., Mohammad hosseini M., Pourbasheer E., [QSAR Study of VEGFR-2 Inhibitors by Using Genetic Algorithm-Multiple Linear Regressions \(GA-MLR\) and Genetic Algorithm-Support Vector Machine \(GA-SVM\): A Comparative Approach](#), *Medicinal Chemistry Reserch*, **24**: 3037-3046 (2015).
- [25] Bartzatt R., [In Silico Optimized Mechlorethamine Based Drug Structures Targeting Brain and Spinal Cord Tumors](#), *British Journal of Pharmaceutical Research*, **3**: 1058-1069 (2013).