

مطالعه‌ی رابطه‌ی ساختمان- فعالیت کمی دوبعدی و داکینگ مولکولی مشتقات ۴- آنیلینو کینازولین به عنوان مهارکننده‌ی آنزیم تیروزین کیناز رسپتور فاکتور رشد اپیدرمی

مهتاب قاسمی دوگاهه^۱ (Pharm.D)، حشمت‌اله ابراهیمی نجف‌آبادی^۱ (Ph.D)، فاطمه یوسف‌بیک^۲ (Ph.D)، سعید قاسمی^{۱*} (Ph.D)

۱- گروه شیمی دارویی، دانشکده داروسازی، دانشگاه علوم پزشکی گیلان، رشت، ایران

۲- گروه فارماکونگوزی، دانشکده داروسازی، دانشگاه علوم پزشکی گیلان، رشت، ایران

تاریخ دریافت: ۱۴۰۰/۲/۲۹ تاریخ پذیرش: ۱۴۰۰/۶/۲۸

ghasemi_saeed@yahoo.com

* نویسنده مسئول، تلفن: ۰۱۳-۳۳۴۸۶۴۷۰

چکیده

هدف: مشتقات مهارکننده‌ی آنزیم تیروزین کیناز رسپتور فاکتور رشد اپیدرمی (EGFR TKIs) نقش مهمی را در درمان سرطان ایفا می‌کنند. هدف از این مطالعه بررسی رابطه ساختمان-فعالیت دوبعدی (2D-QSAR) و داکینگ مشتقات ۴-آنیلینو کینازولین به عنوان EGFR TKIs و دستیابی به یک مدل مناسب جهت پیش‌بینی فعالیت این ترکیبات می‌باشد. مواد و روش‌ها: ۱۲۲ ترکیب حاوی ساختار ۴-آنیلینو کینازولین جهت بسط مدل‌های رابطه‌ی ساختمان-فعالیت کمی دوبعدی (2D-QSAR) با استفاده از روش رگرسیون خطی چندگانه (MLR)، و روش‌های غیر خطی شبکه‌های عصبی مصنوعی (ANN) و ماشین‌های بردار پشتیبانی (SVM) استفاده شد. مدل‌ها با استفاده از روش‌های اعتبارسنجی داخلی و خارجی معتبرسازی شدند. داکینگ مولکولی بر روی ترکیبات ۳۲، ۷۵ و ۹۸ با بیش‌ترین pIC_{50} با استفاده از نرم‌افزار AutoDock 4.2 انجام گرفت. یافته‌ها: توصیف‌گرهای انتخاب شده نشان دادند که الکترون‌گاتیویته اتمی، قابلیت ایجاد پیوند هیدروژنی، میزان لیپوفیلیسیته و شکل و حجم مولکولی از عوامل مؤثر بر فعالیت این ترکیبات هستند. معیارهای آماری مربوط به تفسیر و اعتبارسنجی مدل‌ها در بازه مناسبی قرار گرفتند. مدل به دست آمده با روش ANN با کم‌ترین میانگین خطای مطلق ۰/۳۶۵ برای سری آموزش و آزمون بهترین مدل انتخابی بود. در مطالعات داکینگ مولکولی ترکیب ۷۵ به عنوان قوی‌ترین ترکیب در مهار TK، انرژی اتصال ۸/۳۳- کیلوکالری بر مول را نشان داد. نتیجه‌گیری: نتایج نشان داد که همه مدل‌ها به خوبی می‌توانند فعالیت ترکیبات را پیش‌بینی کنند. هم‌چنین ترکیبات با قدرت مناسب به جایگاه فعال رسپتور متصل شدند. تمامی فرآیندها از اعتبار کافی برخوردار بودند.

واژه‌های کلیدی: رابطه ساختمان-فعالیت کمی، رسپتورهای ErbB، تحلیل رگرسیون، ماشین بردار پشتیبان، شبکه‌های عصبی

مقدمه

سیگنالینگ غیر طبیعی پروتئین کینازها به ویژه پروتئین کینازهای گیرنده‌ای که در بسیاری از مسیرهای سیگنالینگ دخیل هستند می‌تواند منجر به پرولیفراسیون، آنژیوژنز و پیشرفت تومور گردد [۲،۱]. فعال‌سازی مسیرهای مختلف انتقال سیگنال که در نتیجه بیان بیش از حد و جهش رسپتور فاکتور رشد اپیدرمی (EGFR)، به عنوان یکی از اعضای خانواده رسپتورهای تیروزین کینازی خانواده‌ی ErbB، ایجاد می‌شود می‌تواند منجر به ایجاد تومور گردد [۴،۳،۱]. بنابراین مهار EGFR می‌تواند به عنوان یک هدف مناسب برای طراحی مشتقات جدید ضد سرطان مطرح گردد [۵]. مولکول‌های

کوچک سنتزی با اتصال به محل اتصال ATP در بخش تیروزین کینازی گیرنده مانع از فسفریلاسیون آن می‌شوند. در بین ترکیبات سنتز شده، ترکیبات با هسته‌ی ۴-آنیلینو کینازولین شامل جفیتینیب، ارلوتینیب و لاپاتینیب (شکل ۱) به عنوان مهم‌ترین و انتخابی‌ترین مهارکننده‌های تیروزین کیناز EGFR به شمار می‌روند و از ترکیبات قوی با خاصیت ضد سرطانی هستند [۸-۶]. فرایند توسعه دارو یک پروسه زمان‌بر و پرهزینه است. روش‌های مختلفی جهت کاهش این مولفه‌ها وجود دارد که در این بین می‌توان به روش‌های درون رایانه‌ای (in-silico) جهت پیش‌بینی فعالیت و خصوصیات فیزیکوشیمیایی مشتقات جدید سنتز شده و طراحی مولکول‌های جدید اشاره کرد [۹].

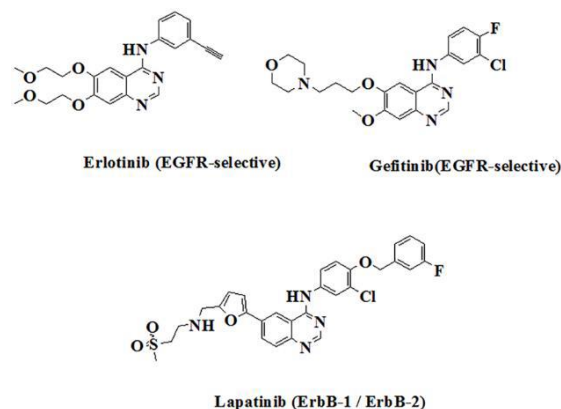
فراهم می‌کند [۱۶-۱۴]. داکینگ مولکولی جهت شناسایی برهمکنش میان لیگاند و رسپتور در جایگاه اتصال مورد نظر می‌رود تا در نهایت بهترین کاندید ماسیون اتصال بین لیگاند و رسپتور با کم‌ترین انرژی به‌دست آید. در این مطالعه ۱۲۲ ترکیب با ساختار ۴-آنیلینوکینازولین جهت ساخت مدل‌های 2D-QSAR انتخاب شدند. جهت گسترش مدل‌ها از روش‌های خطی MLR و غیرخطی ANN و SVM استفاده شد. جهت مطالعات داکینگ مولکولی ترکیبات ۳۲، ۷۵ و ۸۹ به عنوان قوی‌ترین ترکیبات انتخاب شدند.

مواد و روش‌ها

انتخاب و دسته‌بندی داده‌ها. مقادیر IC_{50} (منفی لگاریتم غلظت مورد نیاز جهت مهار ۵۰٪ فعالیت آنزیم) ۱۲۲ ترکیب از مشتقات ۴-آنیلینوکینازولین به عنوان مهارکننده‌های تیروزین کیناز فاکتور رشد اپیدرمی (EGFR TKIs) از مقالات انتخاب شدند [۱۷-۲۰]. این مجموعه داده از هفت گروه مختلف TKIs تشکیل شده بود که همگی دارای ساختار ۴-آنیلینوکینازولین بودند. جدول ۱ ساختار این ترکیبات و مقادیر IC_{50} آن‌ها را نشان می‌دهد.

توصیف‌گرهای مولکولی. جهت محاسبه توصیف‌گرها ابتدا ساختار ترکیبات با استفاده از نرم‌افزار HyperChem رسم و انرژی آن‌ها نیز با استفاده از همین نرم‌افزار بهینه گردید. پیش بهینه‌سازی ساختارها با استفاده از روش Molecular Mechanics force field (MM^+) و بهینه کردن ساختار سبب‌تری ترکیبات با استفاده از محاسبات semiempirical AM1 با استفاده از الگوریتم Polak-Ribière (conjugate gradient) انجام گرفت [۲۱]. ساختارهای بهینه شده در نرم‌افزار Dragon 3.0 بارگزاری شده و توصیف‌کننده‌های مولکولی این ترکیبات محاسبه شد [۲۲، ۲۳].

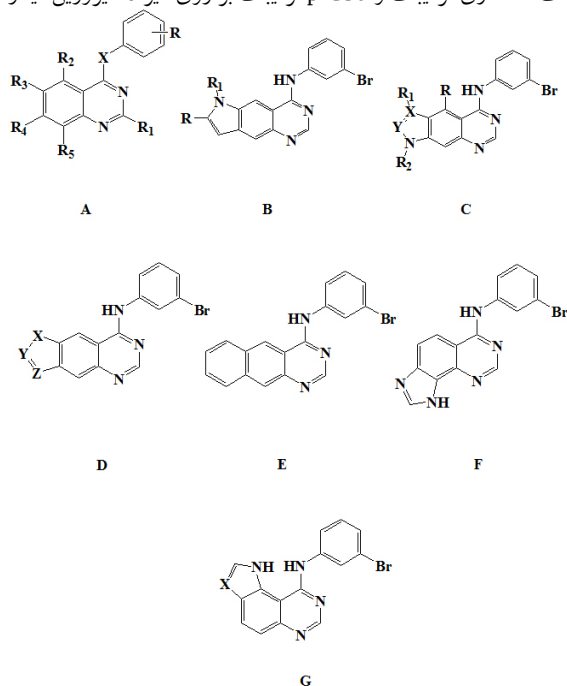
انتخاب سری آموزش و تست. با استفاده از الگوریتم کنار-استون و نرم‌افزار Matlab، ترکیبات به دو دسته‌ی سری آموزش (training set) و سری آزمون (test set) تقسیم شدند. ۹۸ ترکیب به‌عنوان training set و ۲۴ ترکیب به‌عنوان test set انتخاب شدند [۲۴].



شکل ۱. ساختار مهارکننده‌های آنزیم تیروزین کیناز فاکتور رشد اپیدرمی

از میان این روش‌ها، مطالعات رابطه ساختمان-فعالیت کمی (QSAR) شامل روش‌های دوبعدی و سه‌بعدی مجموعه‌ای از مطالعات می‌باشد که به بررسی یک رابطه کمی بین خصوصیات ساختاری و فعالیت بیولوژیک می‌پردازد [۱۱، ۱۰]. از جنبه عملی، توسعه مدل‌های 2D-QSAR آسان‌تر و سریع‌تر از مدل‌های سه‌بعدی است. مدل‌های 2D-QSAR به انطباق ساختارهای سه‌بعدی مولکول‌ها که در 3D-QSAR ضروری است، و جزء موانع روش‌های سه‌بعدی است نیازی ندارد [۱۲]. روش‌های مختلف کمومتریک خطی و غیر خطی شامل رگرسیون خطی چندگانه (MLR)، شبکه عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) می‌توانند برای استخراج مدل‌های QSAR استفاده شوند. ANN پردازش اطلاعات توسط مغز انسان را تقلید می‌کند. ANN دارای ساختاری چند لایه (لایه ورودی، لایه پنهان و لایه خروجی) است. توصیف‌گرهای انتخاب شده، نوروهای لایه ورودی و مقادیر IC_{50} ترکیبات، نوروهای خروجی هستند. نوروهای لایه پنهان، نوروهای لایه‌های ورودی و خروجی را بهم متصل می‌کنند و بین این دو یک رابطه غیر خطی ایجاد می‌کند [۲۶]. SVM یک روش آموزش آماری است که توسط Vapnik معرفی شد [۱۳]. این روش، یک روش غیر خطی در مطالعات QSAR است. این روش یک ابر صفحه (Hperplane) در یک فضای چندبعدی ایجاد می‌کند که با به کارگیری یک تابع کرنل غیر خطی برای طبقه‌بندی یا بررسی هم‌بستگی، حداقل خطا را

جدول ۱: مشخصات ساختاری ترکیبات و pIC₅₀ ترکیبات بر روی گیرنده تیروزین کیناز فاکتور رشد اپیدرمی



شماره ترکیب	نوع	X	Y	Z	R ₁	R ₂	R ₃	R ₄	R ₅	R	pIC ₅₀
۱	A	NH	-	-	H	H	OCH ₂ CH ₂ OMe	OCH ₂ CH ₂ OMe	H	CN	۷/۸
۲	A	NH	-	-	H	H	H	H	H	H	۴۶/۶
۳	A	NH	-	-	H	H	H	H	H	3'-Cl	۲۵/۷
۴	A	NH	-	-	H	H	H	H	H	3'-I	۶۴/۷
۵	A	NH	-	-	H	H	H	H	H	3'-CF ₃	۲۴/۶
۶	A	NH	-	-	H	H	OMe	H	H	H	۲۶/۷
۷	A	NH	-	-	H	H	OMe	H	H	3'-Br	۵۲/۷
۸	A	NH	-	-	H	H	NH ₂	H	H	H	۱۱/۶
۹	A	NH	-	-	H	H	NH ₂	H	H	3'-CF ₃	۲۴/۶
۱۰	A	NH	-	-	H	H	NO ₂	H	H	H	۷/۸
۱۱	A	NH	-	-	H	H	H	OMe	H	H	۶/۹
۱۲	A	NH	-	-	H	H	H	NH ₂	H	3'-F	۴۸/۸
۱۳	A	NH	-	-	H	H	H	NH ₂	H	3'-CF ₃	۲۱/۵
۱۴	A	NH	-	-	H	H	H	NO ₂	H	3'-F	۰۹/۶
۱۵	A	NH	-	-	H	H	H	NO ₂	H	3'-I	۲۷/۶
۱۶	A	NH	-	-	H	H	OMe	OMe	H	H	۵۴/۷
۱۷	A	NH	-	-	H	H	OMe	OMe	H	3'-F	۴۲/۸
۱۸	A	NH	-	-	H	H	OMe	OMe	H	3'-I	۵۱/۹
۱۹	A	NH	-	-	H	H	OMe	OMe	H	3'-CF ₃	۰۵/۹
۲۰	A	NH	-	-	H	H	NMe ₂	H	H	3'-Br	۶۲/۹
۲۱	A	NH	-	-	H	H	NHCO ₂ Me	H	H	3'-Br	۰۸/۷
۲۲	A	NH	-	-	H	H	H	NHAc	H	3'-Br	۹۲/۷

۲۳	A	NH	-	-	H	H	H	NHMe	H	3'-Br	۳۳/۸
۲۴	A	NH	-	-	H	H	H	NHEt	H	3'-Br	۴/۷
۲۵	A	NH	-	-	H	H	H	NMe ₂	H	3'-Br	۱۵/۸
۲۶	A	NH	-	-	H	H	NH ₂	OMe	H	3'-Br	۹۲/۷
۲۷	A	NH	-	-	H	H	NO ₂	NH ₂	H	3'-Br	۹۶/۷
۲۸	A	NH	-	-	H	H	NO ₂	NHAc	H	3'-Br	۱۶/۹
۲۹	A	NH	-	-	H	H	NO ₂	OMe	H	3'-Br	۴۲/۸
۳۰	A	NH	-	-	H	H	NO ₂	Cl	H	3'-Br	۱۹/۸
۳۱	A	NH	-	-	H	H	OH	OH	H	3'-Br	۲۸/۷
۳۲	A	NH	-	-	H	H	OEt	OEt	H	3'-Br	۵۵/۷
۳۳	A	NH	-	-	H	H	OPr	OPr	H	3'-Br	۸۲/۷
۳۴	A	NH	-	-	H	OMe	OMe	H	H	3'-Br	۶/۷
۳۵	A	NH	-	-	NH ₂	H	OMe	OMe	H	3'-Br	۷۷/۹
۳۶	A	NMe	-	-	H	H	OMe	OMe	H	3'-Br	۲/۱۱ ۲
۳۷	A	NH	-	-	H	OMe	OMe	OMe	H	3'-Br	۷۷/۹
۳۸	A	NH	-	-	H	H	OMe	OMe	H	2'-Br	۸۶/۵
۳۹	A	NH	-	-	H	H	OMe	OMe	H	3',5'-diBr	۳۳/۶
۴۰	A	NH(CH ₂) ₂	-	-	H	H	H	H	H	H	۸۲/۶
۴۱	A	NHCH ₂	-	-	H	H	H	H	H	H	۱۷/۹
۴۲	A	NHCH ₂	-	-	H	H	H	H	H	4'-OMe	۸۹/۶
۴۳	A	NMe	-	-	H	H	H	H	H	H	۹۵/۶
۴۴	A	NH	-	-	H	H	H	H	H	3'-OMe	۳۹/۵
۴۵	A	NH	-	-	H	H	H	H	H	3'-Me	۵
۴۶	A	NHCH ₂	-	-	H	NO ₂	H	H	H	H	۴
۴۷	A	NHCH ₂	-	-	H	OMe	H	H	H	H	۰۷/۶
۴۸	A	NHCH ₂	-	-	H	H	OMe	H	H	H	۰۴/۶
۴۹	A	NHCH ₂	-	-	H	H	OMe	OH	H	H	۱/۵
۵۰	A	NHCH ₂	-	-	H	H	OH	OMe	H	H	۲۴/۷
۵۱	A	NH	-	-	H	NO ₂	H	H	H	3'-Br	۲۳/۶
۵۲	A	NH	-	-	H	OMe	H	H	H	3'-Br	۲۵/۷
۵۳	A	NH	-	-	H	H	H	H	NH ₂	3'-Br	۴۵/۶
۵۴	A	NH	-	-	H	H	H	H	OMe	3'-Br	۸۶/۶
۵۵	B	-	-	-	H	-	-	-	-	Me	۱۱
۵۶	B	-	-	-	Me	-	-	-	-	H	۸۸/۸
۵۷	B	-	-	-	(CH ₂) ₂ NMe ₂	-	-	-	-	H	۶/۱۰

۵۸	C	N	CH	-	-	Me	-	-	-	H	۶۶/۷
۵۹	C	N	CH	-	-	(CH ₂) ₂ NMe ₂	-	-	-	H	۹۲/۷
۶۰	C	CH	N	-	H	CH ₂ CH(OH)CH ₂ O H	-	-	-	H	۴/۷
۶۱	C	CH	N	-	H	(CH ₂) ₂ NMe ₂	-	-	-	H	۴۳/۸
۶۲	C	CH	N	-	H	(CH ₂) ₂ Nmorpholide	-	-	-	H	۲۸/۷
۶۳	C	CH	N	-	H	CH ₂ COOH	-	-	-	H	۱/۹
۶۴	C	CH	CH	-	H	Me	-	-	-	H	۸/۸
۶۵	C	CH	CH	-	H	CH ₂ CH(OH)CH ₂ O H	-	-	-	H	۴۳/۸
۶۶	C	CH	CH	-	H	(CH ₂) ₂ Nmorpholide	-	-	-	H	۰۶/۸
۶۷	C	CH	CH	-	H	(CH ₂) ₃ Nmorpholide	-	-	-	H	۲۹/۸
۶۸	C	CH	CH	-	CH ₂ CH(OH)CH ₂ OH	H	-	-	-	H	۴۶/۸
۶۹	C	CH	CH	-	(CH ₂) ₂ Nmorpholide	H	-	-	-	H	۳۲/۸
۷۰	C	CH	CH	-	CH ₂ N(Me)CH ₂ COOH	H	-	-	-	H	۴۷/۸
۷۱	C	C	N		H	H	-	-	-	H	۱۴/۹
۷۲	D	NH	N	N	-	-	-	-	-	-	۳۶/۹
۷۳	D	S	CH	N	-	-	-	-	-	-	۳۹/۸
۷۴	D	NH	N	C H	-	-	-	-	-	-	۳۶/۷
۷۵	E	-	-	-	-	-	-	-	-	-	۴۷/۹
۷۶	F	-	-	-	-	-	-	-	-	-	۵/۱۱ ۲
۷۷	A	NH	-	-	H	H	NO ₂	H	-	3'- Br	۵۷/۶
۷۸	A	NH	-	-	H	H	NHMe	H	-	3'- Br	۰۵/۶
۷۹	A	NH	-	-	H	H	NH ₂	NH ₂	-	3'- Br	۴/۸
۸۰	A	NH	-	-	H	H	NH ₂	NMe ₂	-	3'- Br	۹۲/۹
۸۱	A	NH	-	-	H	H	NO ₂	NHMe	-	3'- Br	۸/۶
۸۲	A	NH	-	-	H	H	NO ₂	NMe ₂	-	3'- Br	۱۷/۷
۸۳	A	NH	-	-	H	H	-OCH ₂ O-		-	3'- Br	7.82
۸۴	A	NH	-	-	H	H	OBu	OBu	-	3'- Br	۸۲/۷
۸۵	A	NH	-	-	H	H	OMe	OMe	-	4'- Br	۹۸/۶
۸۶	A	NH	-	-	H	H	OMe	OMe	-	3',4'- diBr	۱/۱۰ ۴
۸۷	A	NHCH ₂	-	-	H	H	H	H	-	4'-Cl	۱۵/۵
۸۸	A	O	-	-	H	H	H	H	-	3'- Br	۵
۸۹	A	NHCH ₂	-	-	H	H	NO ₂	H	-	H	۸۵/۵
۹۰	A	NHCH ₂	-	-	H	H	NH ₂	H	-	H	۳۶/۶
۹۱	A	NHCH ₂	-	-	H	H	OMe	OMe	-	H	۴۳/۹
۹۲	A	NH	-	-	H	NH ₂	H	H	-	3'- Br	۳۹/۷
۹۳	C	C	N	-	H	Me	-	-	-	-	۶۸/۷
۹۴	C	C	CH	-	H	(CH ₂) ₃ NMe ₂	-	-	-	-	۱۲/۸
۹۵	C	C	CH	-	CH ₂ N(Me)(CH ₂) ₂ NMe 2	H	-	-	-	-	۶۹/۶
۹۶	C	N	CHC H	-	-	-	-	-	-	-	۳۶/۹

۹۷	G	N	-	-	-	-	-	-	-	-	۷۷/۸
۹۸	D	NH	CH	C H	-	-	-	-	-	-	۰/۴۴
سری آزمون											
۹۹	A	NH	-	-	H	H	OMe	OMe	H	3'-Br	۰/۱۴۸
۱۰۰	A	NH	-	-	H	H	H	H	H	3'-F	۰/۷۱۸
۱۰۱	A	NH	-	-	H	H	H	H	H	3'-Br	۰/۵۱۸
۱۰۲	A	NH	-	-	H	H	NH ₂	H	H	3'-Br	۰/۱۰۶
۱۰۳	A	NH	-	-	H	H	H	OMe	H	3'-Br	۰/۳۸۶
۱۰۴	A	NH	-	-	H	H	H	NH ₂	H	H	۰/۶۴۳
۱۰۵	A	NH	-	-	H	H	H	NH ₂	H	3'-Cl	۰/۱۷۱
۱۰۶	A	NH	-	-	H	H	H	NH ₂	H	3'-Br	۰/۱۷۴
۱۰۷	A	NH	-	-	H	H	H	NH ₂	H	3'-I	۰/۵۱۳
۱۰۸	A	NH	-	-	H	H	H	NO ₂	H	H	۰/۲۱۸
۱۰۹	A	NH	-	-	H	H	H	NO ₂	H	3'-Cl	۰/۰۴۵
۱۱۰	A	NH	-	-	H	H	H	NO ₂	H	3'-Br	۰/۱۵۴
۱۱۱	A	NH	-	-	H	H	OMe	OMe	H	3'-Cl	۰/۷۲۰
۱۱۲	A	NH	-	-	H	H	H	OH	H	3'-Br	۰/۰۲۵
۱۱۳	A	NH	-	-	H	H	NH ₂	NHMe	H	3'-Br	۰/۷۱۵
۱۱۴	A	NH	-	-	H	H	NH ₂	Cl	H	3'-Br	۰/۵۸۷
۱۱۵	A	NHCH ₂	-	-	H	H	H	NO ₂	H	H	۰/۰۷۶
۱۱۶	A	NHCH ₂	-	-	H	H	H	OMe	H	H	۰/۰۰۱
۱۱۷	C	CH	CH	-	H	CH ₂ COOH	-	-	-	H	۰/۵۹۰
۱۱۸	C	CH	CH	-	CH ₂ N(Me)CH ₂ COOMe	H	-	-	-	H	۰/۵۴۷
۱۱۹	C	C	CH	-	H	(CH ₂) ₂ NMe ₂	-	-	-	H	۰/۶۲۴
۱۲۰	C	N	CH	-	-	(CH ₂) ₂ NMe ₂	-	-	-	H	۰/۶۸۹
۱۲۱	C	C	CH	-	H	H	-	-	-	H	۰/۰۵۳
۱۲۲	G	CH	-	-	-	-	-	-	-	-	۰/۲۷۲

و GA برای کاهش تعداد توصیف‌گرها استفاده شد [۲۸، ۲۷]. سپس بیش‌ترین هم‌بستگی با pIC₅₀ و کم‌ترین هم‌بستگی درونی به وسیله ضریب هم‌بستگی پیرسون جهت انتخاب توصیف‌گرها استفاده شد و در انتها با استفاده از نرم‌افزار SPSS و روش رگرسیون گام‌به‌گام، تعداد توصیف‌گرهای نهایی به ۱۳ عدد رسید.

ساخت مدل

مدل MLR جهت مدل‌سازی به روش MLR، از نرم‌افزار SPSS استفاده شد. پس از به‌دست‌آوردن مقادیر pIC₅₀ پیش‌بینی‌شده، ارزیابی مدل با استفاده از آماره‌های ضریب هم‌بستگی (R)، ضریب هم‌بستگی تعدیل‌شده (Radj)، خطای استاندارد برآورد شده (SEE)، و ارزش احتمال (p-value) هر توصیف‌گر، و آماره فیشر یا نسبت واریانس (F) که توسط

انتخاب توصیف‌گرها. ۱۲۳۶ توصیف‌گر توسط نرم‌افزار Dragon 3 محاسبه شد که در اولین مرحله‌ی حذف توصیف‌گرها، توصیف‌گرهایی که هم‌بستگی آن‌ها بیش از ۰/۹ بود و مقادیرشان نیز تا ۵۰٪ تکراری بود، حذف شدند و تعداد توصیف‌گرها به ۴۱۶ عدد کاهش یافت [۲۵]. سپس با استفاده از روش الگوریتم ژنتیکی - رگرسیون حداقل مربعات جزئی (GA-PLS) و توسط نرم‌افزار متلب و برنامه نوشته شده توسط Leardi تعداد توصیف‌گرها به ۵۳ عدد کاهش یافت. اندازه جمعیت در الگوریتم ژنتیکی ۱۰۰ و نهایتاً ۱۰٪ از داده‌ها با امتیاز بالا انتخاب شدند [۲۶]. GA روند تکامل طبیعی را شبیه‌سازی می‌کند و نشان داده شده است که یک روش قابل قبول برای کاهش تعداد توصیف‌گرها می‌باشد [۲۵]. هم‌چنین ترکیبی از PLS، به عنوان ابزاری ارزشمند برای کاهش داده‌ها،

$$AAE = \frac{\sum |Calculated pIC_{50} - Experimental pIC_{50}|}{N}$$

داکینگ مولکولی:

ترکیبات ۷۵،۳۲ و ۹۸ که بیش‌ترین پوتنسی را در مطالعات تجربی از خود نشان دادند جهت انجام فرایند داکینگ انتخاب شدند. برای انجام فرایند داکینگ، ساختار کریستالی آنزیم تیروزین کیناز فاکتور رشد اپیدرمی به همراه ارلوتینیب به عنوان لیگاند داخلی (PDB: 1M17) از پایگاه اطلاعاتی Protein Data Bank دریافت گردید (<http://www.rcsb.org>). از AutoDock 4.2 و AutoDock Tools 1.5.4 (ADT) برای مطالعه برهمکنش‌ها استفاده شد. قبل از داکینگ، ارلوتینیب و مولکول‌های آب از 1M17 حذف شدند، سپس هیدروژن‌ها و بارهای کلمن اضافه شدند و هیدروژن‌های غیر قطبی ادغام شدند. ساختار ترکیبات ترسیم شد و بهینه‌سازی ساختار توسط مکانیک مولکولی (MM⁺) و سپس روش‌های نیمه‌تجربی AM1 توسط نرم‌افزار HyperChem 8.0 انجام شد. مکعب حاوی نقاط جست‌وجو (spacing) با ابعاد ۰/۳۷۵×۰/۳۷۵×۰/۳۷۵ استفاده شد. پس از تنظیم گرید، اعداد مربوط به مراکز و اندازه کارتریجی جعبه گرید (Grid Box) ۹۰×۹۰×۹۰ انتخاب شد. الگوریتم جست‌وجوی ژنتیکی لامارکین با ۱۰۰ بار اجرای GA استفاده شد. لیگاند (erlotinib) برای اعتبارسنجی روش اتصال به روش فوق‌الذکر استفاده شد [۳۳].

نتایج

در این مطالعه از ۱۳ توصیف‌گر دوبعدی و سه‌بعدی که متعلق به ۶ دسته‌ی مختلف از توصیف‌گرها هستند، استفاده شد. جدول ۲ جزئیات توصیف‌گرهای انتخاب شده با استفاده از GA-PLS و رگرسیون گام به گام را نشان می‌دهد. توصیف‌گرهای انتخاب شده برای توسعه مدل‌های QSAR با استفاده از روش‌های ANN، MLR و SVM استفاده شدند. یک مدل خطی MLR به عنوان ساده‌ترین مدل ارائه شد. جدول ۴ ضرایب، SEE و p-value توصیف‌گرهای انتخاب شده با این روش را نشان می‌دهد. اطلاعات آماری لازم جهت اعتبارسنجی تمامی مدل‌های QSAR برای مدل‌های بسط یافته در این مطالعه در جدول ۵ آورده شده است. پارامترهای آماری مدل ANN توسعه‌یافته برای مجموعه داده‌ها که به مجموعه‌های آموزش، اعتبارسنجی و آزمون تقسیم شده بود، در جدول ۵ نشان داده شده است. مقادیر بهینه شده C، ε و γ به برای مدل SVM به

Dearden و همکاران توصیه شده است انجام شد [۱۰]. از روش the leave one-out (LOO) جهت تأیید اعتبار مدل ساخته شده استفاده شد.

مدل ANN. در این مطالعه، از نرم‌افزار MATLAB 7.8 برای آموزش شبکه استفاده شد. این نرم‌افزار از الگوریتم‌های پس انتشار Levenberg-Marquardt برای آموزش ANN استفاده می‌کند. برای آموزش یک شبکه معتبر برای جلوگیری از بیش‌برازش (Overfitting)، داده‌های سری آموزش به طور تصادفی در سه سری آموزش (۷۰٪)، اعتبارسنجی (۱۵٪) و آزمون (۱۵٪) با استفاده از نرم‌افزار طبقه‌بندی شدند. نهایتاً، یک شبکه سه لایه با ۴ نورون در لایه پنهان طراحی شد [۲۹].

مدل SVM. فاکتور ظرفیت و اسیلون با استفاده از روش اعتبارسنجی متقابل leave-many-out بهینه شدند. از نرم‌افزار STATISTICA 7 جهت انجام فرایند SVM استفاده شد. طی فرایند بهینه‌سازی پارامترهای C، ε و γ به ترتیب ۱۰، ۰/۱ و ۰/۰۸۷ به دست آمد [۳۰، ۱۴، ۱۰].

اعتبارسنجی خارجی مدل‌های پیشنهادی و مقایسه مدل‌ها به منظور بررسی اعتبار مدل‌های پیشنهادی و جهت مقایسه قابلیت پیش‌بینی مدل‌ها، از یک مجموعه داده خارجی (سری آزمون) متشکل از ۲۴ داده استفاده شد. در این مطالعه، از مجموعه‌ای از آماره‌ها ارائه شده در مقالات برای بررسی اعتبار مدل‌های بسط یافته استفاده شد [۳۱]. این آماره‌ها عبارتند از:

$$1. \text{ ضریب تعیین بیش‌تر از } 0/6 (R^2 > 0/6)$$

R^2 نشان‌دهنده ضریب تعیین (ضریب هم‌بستگی) بین pIC_{50} پیش‌بینی شده و pIC_{50} مشاهده شده است.

$$2. (R^2 - R_0^2)/R^2 < 1, (R^2 - R_0'^2)/R^2 < 1$$

R_0^2 نشان‌دهنده ضریب هم‌بستگی بین pIC_{50} پیش‌بینی شده و pIC_{50} مشاهده شده زمانی که خط رگرسیون از مبدا عبور کند و $R_0'^2$ نشان‌دهنده ضریب هم‌بستگی بین pIC_{50} مشاهده شده و pIC_{50} پیش‌بینی شده است زمانی که خط رگرسیون از مبدا عبور کند.

$$3. 0/85 < K' < 1/15 \text{ یا } 0/85 < K < 1/15$$

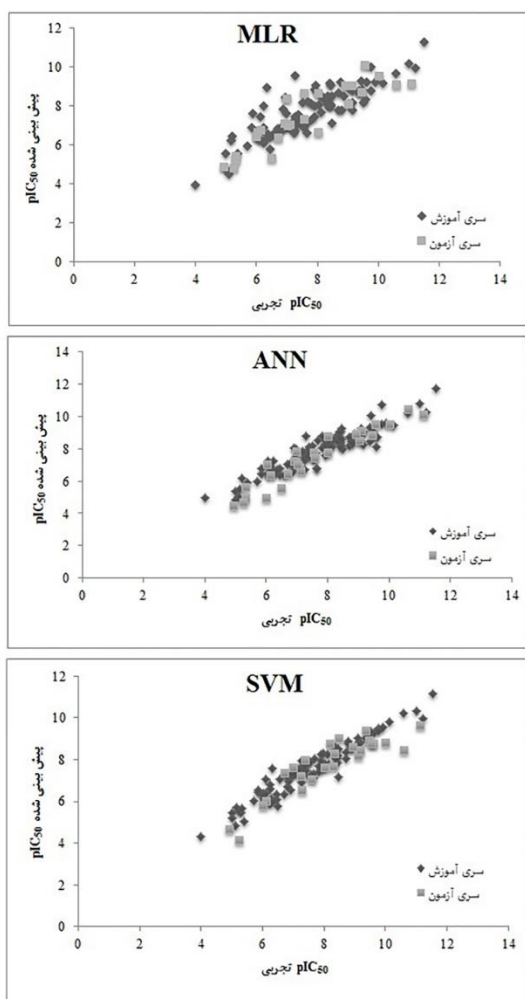
K و K' شیب‌های خطوط رگرسیون عبور کرده از مبدا به ترتیب برای داده‌های تجربی و پیش‌بینی شده هستند [۳۱].

معیار دیگری نیز توسط Roy و Roy برای ارزیابی قدرت پیش‌بینی مدل‌های QSAR پیشنهاد شد [۳۲].

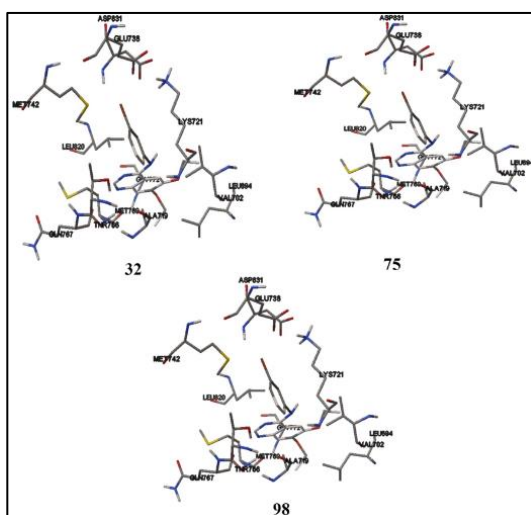
$$4. R_m^2 = R^2(1 - \sqrt{R^2 - R_0^2})$$

که در آن $R_m^2 > 0/5$ قدرت پیش‌بینی خارجی بالای مدل‌های QSAR را نشان می‌دهد.

دقت مدل‌های پیشنهادی با استفاده از میانگین خطای مطلق (AAE) مقایسه شد که به صورت زیر تعریف می‌شود:

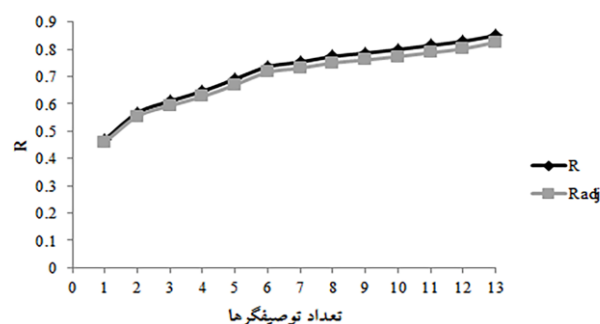


شکل ۳. مقادیر تجربی در برابر مقادیر پیش بینی شده pIC_{50} با استفاده از مدل های ANN، MLR و SVM.



شکل ۴. برهمکنش ترکیبات ۳۲، ۷۵ و ترکیب ۹۸ با رسپتور تیروزین کیناز فاکتور رشد اپیدرمی

ترتیب ۱۰، ۱/۰ و ۰/۰۸۷ بودند و خصوصیات آماری مدل پیشنهادی SVM برای مجموعه آموزش در جدول ۵ ذکر شده است. مقادیر پیش‌بینی شده و خطای مطلق (AE) با استفاده از مدل‌های MLR، ANN و SVM در جدول ۶ ذکر شده است. مقادیر پیش‌بینی شده برای سری آزمون جهت انجام اعتبارسنجی برای مدل‌های مختلف در جدول ۶ آورده شده است. شکل ۳ مقادیر تجربی در مقابل مقادیر پیش‌بینی شده برای سری آموزش (۹۸ داده) و سری آزمون (۲۴ داده) را نشان می‌دهد. مقادیر AAE سری آموزش و آزمون در جدول ۷ خلاصه شده است. برای تأیید بیش‌تر این مشاهده، برخی از معیارهای آماری برای ارزیابی توانایی پیش‌بینی و مقاومت مدل برای سری آزمون محاسبه شد (جدول ۸) و نتایج نشان داد که مدل‌های پیشنهادی ساخته شده با استفاده از ANN، MLR و SVM برای پیش‌بینی خارجی بسیار مقاوم و معتبر بودند. شکل ۴ نحوه برهمکنش ترکیبات ۳۲، ۷۵ و ۹۸ با جایگاه فعال گیرنده EGFR-TK (PDB ID: 1M17) را نشان می‌دهد. جدول ۹ انرژی برهمکنش‌های هیدروژنی و فواصل پیوند هیدروژنی تشکیل شده بین آمینواسید Met769 و N1 کینازولین هر سه ترکیب را نشان می‌دهد.



شکل ۲. اثرات تعداد توصیف کنندگرا (مطابق جدول ۱) بر R و Radj برای ارزیابی تأثیر توصیف کنندگراهای انتخاب شده

جدول ۲. توصیفگرهای انتخاب شده با استفاده از GA-PLS و رگرسیون گام به گام

ردیف	نماد	تعریف	دسته
۱	Mor15e	signal 15 / weighted by Sanderson electronegativity	3D-MorSE descriptors
۲	MATS3e	Moran autocorrelation of lag 3 weighted by Sanderson electronegativity	2D autocorrelations
۳	MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
۴	RDF090m	Radial Distribution Function - 090 / weighted by mass	RDF descriptors
۵	BELe1	lowest eigenvalue n. 1 of burden matrix/ weighted by Sanderson electronegativity	BCUT descriptors
۶	RDF085p	Radial Distribution Function - 085 / weighted by polarizability	RDF descriptors
۷	BEHv2	Highest eigenvalue n. 2 of burden matrix/weighted by atomic van der waals volumes	BCUT descriptors
۸	Mor12u	signal 12 / unweighted	3D-MorSE descriptors
۹	Mor10p	signal 10 / weighted by polarizability	3D-MorSE descriptors
۱۰	G1s	1st component symmetry directional WHIM index / weighted by I-state	WHIM descriptors
۱۱	DELS	molecular electrotopological variation	Topological indices
۱۲	G2s	2nd component symmetry directional WHIM index / weighted by I-state	WHIM descriptors
۱۳	RDF040p	Radial Distribution Function - 040 / weighted by polarizability	RDF descriptors

جدول ۳. ماتریس همبستگی توصیف گرهای انتخاب شده

	Mor15e	MATS3e	MATS1e	RDF090m	BELe1	RDF085p	BEHv2	Mor12u	Mor10p	G1s	DELS	G2s	RDF040p
Mor15e	۱/۰۰												
MATS3e	۰/۱۸	۱/۰۰											
MATS1e	-۰/۴۰	۰/۰۰	۱/۰۰										
RDF090m	۰/۲۷	-۰/۱۶	-۰/۱۷	۱/۰۰									
BELe1	۰/۰۷	-۰/۱۸	۰/۱۴	۰/۲۵	۱/۰۰								
RDF085p	۰/۲۸	۰/۰۴	۰/۰۱	۰/۳۲	۰/۱۰	۱/۰۰							
BEHv2	۰/۲۵	۰/۱۲	-۰/۰۱	۰/۲۲	۰/۱۱	۰/۰۴	۱/۰۰						
Mor12u	۰/۱۸	-۰/۱۶	۰/۰۶	۰/۰۲	۰/۰۰	-۰/۰۲	-۰/۲۲	۱/۰۰					
Mor10p	۰/۲۶	-۰/۱۱	-۰/۰۳	۰/۲۲	۰/۲۴	۰/۲۴	۰/۱۸	۰/۰۶	۱/۰۰				
G1s	-۰/۵۴	-۰/۰۸	-۰/۰۸	-۰/۴۲	-۰/۲۹	-۰/۵۲	-۰/۳۴	-۰/۰۱	-۰/۵۰	۱/۰۰			
DELS	۰/۱۴	۰/۱۷	۰/۲۸	۰/۱۴	۰/۱۵	۰/۱۲	۰/۱۰	۰/۱۶	۰/۱۵	-۰/۴۴	۱/۰۰		
G2s	-۰/۵۷	-۰/۱۰	-۰/۰۸	-۰/۴۶	-۰/۲۵	-۰/۵۱	-۰/۳۹	۰/۰۴	-۰/۴۵	۰/۸۱	-۰/۴۴	۱/۰۰	
RDF040p	۰/۴۱	۰/۲۶	-۰/۰۵	۰/۲۷	۰/۰۰	۰/۱۲	۰/۴۵	-۰/۲۲	۰/۱۳	-۰/۵۱	۰/۰۰	-۰/۵۳	۱/۰۰

جدول ۴. ضرایب و داده های آماری توصیفگرهای انتخاب شده در مدل MLR

توصیف گرها	ضرایب	SEE	p-value
Constant	-۲۱۷/۹۰۹	۳۷/۵۶۸	۰/۰۰۰
Mor15e	۱/۴۹۹	۰/۳۴۳	۰/۰۰۰
MATS3e	-۶/۵۶۸	۱/۲۲۸	۰/۰۰۰
MATS1e	-۴/۶۰۵	۱/۱۶۹	۰/۰۰۰
RDF090m	-۰/۱۵۴	۰/۰۳۷	۰/۰۰۰
BELe1	۵۱/۳۱۸	۹/۶۰۶	۰/۰۰۰
RDF085p	۰/۳۱۵	۰/۰۴۰	۰/۰۰۰
BEHv2	۲۴/۰۵۶	۸/۴۱۹	۰/۰۰۵
Mor12u	۰/۷۸۰	۰/۳۰۹	۰/۰۱۴
Mor10p	۲/۲۴۶	۰/۴۸۶	۰/۰۰۰
G1s	۸۲/۸۷۳	۲۴/۳۵۸	۰/۰۰۱
DELS	۰/۰۵۹	۰/۰۱۶	۰/۰۰۰

G2s	۸۸/۷۵۶	۲۷/۰۷۸	۰/۰۰۲
RDF040p	۰/۲۸۰	۰/۰۶۵	۰/۰۰۰

جدول ۵. معیارهای آماری سری آموزش برای مدل های بسط یافته

MLR					
	N	R	R _{adj}	SEE	F
سری آموزش	۹۸	۰/۸۵۱	۰/۸۲۵	۰/۸۳۳	۹۸/۱۶
ANN					
سری آموزش	۶۸	۰/۹۰۷	۰/۹۰۵	۰/۵۸۸	۹۴/۳۰۵
سری معتبر سازی	۱۵	۰/۹۴۶	۰/۹۴۱	۰/۶۴۹	۸۳/۱۰۹
سری تست	۱۵	۰/۸۶۱	۰/۸۴۹	۰/۷۰۰	۲۳/۳۷
کل	۹۸	۰/۹۲۸	۰/۹۲۶	۰/۵۵۴	۹۱/۵۹۲
SVM					
سری آموزش	۹۸	۰/۹۶۲	۰/۹۶۱	۰/۴۱۳	۷/۱۱۸۴

جدول ۶. مقادیر تجربی، پیش بینی شده و خطای مطلق (AE) ۹۸ ترکیب سری آموزش و ۲۴ ترکیب سری آزمون

شماره ترکیبات	تجربی pIC ₅₀	MLR	خطای مطلق	ANN	خطای مطلق	SVM	خطای مطلق
		پیش بینی pIC ₅₀ شده		پیش بینی pIC ₅₀ شده		پیش بینی pIC ₅₀ شده	
سری آموزش							
۱	۸/۷	۸/۶۵	-/۰۴۵	۸/۷۱	-/۰۰۹	۸/۳۵	-/۳۴۹
۲	۶/۴۶	۵/۷۸	-/۶۸۴	۶/۳۵	-/۱۰۹	۶/۳۵	-/۱۰۸
۳	۷/۲۵	۶/۷۵	-/۴۹۹	۷/۲۹	-/۰۳۵	۷/۱۰	-/۵۴۵
۴	۷/۶۴	۷/۰۷	-/۵۷۲	۶/۷۹	-/۸۵۲	۷/۴۵	-/۳۵۰
۵	۶/۲۴	۶/۱۲	-/۱۲۴	۶/۵۴	-/۳۰۴	۵/۸۹	-/۳۵۱
۶	۷/۲۶	۷/۰۰	-/۲۵۸	۷/۴۶	-/۱۹۵	۷/۶۱	-/۳۴۸
۷	۷/۵۲	۸/۱۹	-/۶۶۵	۸/۱۵	-/۶۳۳	۷/۸۷	-/۳۵۱
۸	۶/۱۱	۷/۴۶	۱/۳۴۹	۶/۸۹	-/۷۸۲	۷/۰۸	-/۹۷۳
۹	۶/۲۴	۸/۰۰	۱/۷۵۶	۶/۴۲	-/۱۷۶	۶/۵۹	-/۳۵۲
۱۰	۸/۷	۷/۸۱	-/۸۸۸	۸/۲۰	-/۵۰۳	۵/۴۴	-/۱۴۳
۱۱	۹/۶	۸/۳۴	۱/۲۶۰	۸/۱۴	۱/۴۵۷	۶/۵۷	-/۳۵۲
۱۲	۸/۴۸	۷/۱۲	۱/۳۵۶	۷/۹۷	-/۵۱۲	۸/۰۵	-/۶۵۴
۱۳	۵/۲۱	۶/۴۶	۱/۲۵۳	۶/۱۷	-/۹۶۰	۷/۱۸	۱/۲۹۷
۱۴	۶/۰۹	۶/۳۱	-/۲۱۶	۷/۰۴	-/۹۴۶	۵/۵۶	-/۳۵۰
۱۵	۶/۲۷	۶/۲۴	-/۰۲۷	۶/۶۲	-/۳۵۰	۶/۱۳	-/۱۴۵
۱۶	۷/۵۴	۶/۸۹	-/۶۵۳	۷/۳۲	-/۲۱۹	۷/۱۹	-/۳۵۰
۱۷	۸/۴۲	۸/۱۵	-/۲۷۴	۹/۰۷	-/۶۴۷	۸/۲۰	-/۲۱۸
۱۸	۹/۵۱	۸/۱۵	۱/۳۵۷	۸/۷۶	-/۷۴۹	۸/۷۴	-/۳۰۵
۱۹	۹/۰۵	۸/۱۴	-/۹۱۰	۸/۴۹	-/۵۶۴	۹/۲۷	-/۳۴۷
۲۰	۹/۶۲	۹/۲۵	-/۳۷۰	۸/۷۴	-/۸۸۵	۷/۴۸	-/۴۰۳
۲۱	۷/۰۸	۶/۸۹	-/۱۹۰	۶/۸۱	-/۲۶۹	۷/۵۷	-/۳۵۲
۲۲	۷/۹۲	۷/۴۴	-/۴۸۲	۷/۵۹	-/۳۳۵	۷/۴۷	-/۰۷۰
۲۳	۸/۳۳	۸/۵۳	-/۲۰۵	۸/۸۶	-/۵۳۲	۷/۸۰	-/۳۵۱
۲۴	۷/۴	۷/۱۸	-/۲۱۹	۸/۰۹	-/۶۹۰	۸/۲۷	-/۳۵۱
۲۵	۸/۱۵	۸/۴۶	-/۳۰۵	۸/۱۳	-/۰۲۰	۸/۳۱	-/۳۵۱
۲۶	۷/۹۲	۸/۸۵	-/۹۲۷	۸/۴۹	-/۵۷۰	۷/۸۲	-/۶۰۰
۲۷	۷/۹۶	۹/۰۸	۱/۱۲۳	۸/۴۳	-/۴۶۷	۷/۶۳	-/۳۵۲
۲۸	۹/۱۶	۸/۷۸	-/۳۸۰	۹/۰۳	-/۱۳۵	۷/۲۰	-/۳۵۰
۲۹	۸/۴۲	۸/۴۷	-/۰۵۱	۸/۲۶	-/۱۶۲	۷/۴۱	-/۴۱۱

۳۰	۸/۱۹	۸/۴۵	۰/۲۵۷	۸/۵۲	۰/۳۳۰	۷/۲۵	۰/۳۵۱
۳۱	۷/۲۸	۷/۴۴	۰/۱۶۴	۷/۵۴	۰/۲۶۴	۹/۵۱	۰/۲۶۱
۳۲	۷/۵۵	۷/۱۲	۰/۴۳۳	۷/۵۴	۰/۰۱۱	۹/۹۵	۱/۲۷۵
۳۳	۷/۸۲	۸/۱۴	۰/۳۲۳	۸/۷۸	۰/۹۶۰	۹/۴۲	۰/۳۵۰
۳۴	۷/۶	۷/۲۱	۰/۳۹۲	۷/۷۴	۰/۱۳۶	۶/۴۲	۰/۵۷۳
۳۵	۹/۷۷	۸/۷۷	۱/۰۰۵	۹/۵۳	۰/۲۴۵	۷/۵۹	۱/۲۶۳
۳۶	۱۱/۲۲	۹/۹۸	۱/۲۴۱	۱۰/۲۵	۰/۹۷۳	۷/۱۷	۰/۳۵۱
۳۷	۹/۷۷	۱۰/۰۳	۰/۲۵۶	۱۰/۷۲	۰/۹۵۲	۸/۸۲	۰/۳۴۹
۳۸	۵/۸۶	۷/۶۳	۱/۷۷۲	۶/۷۵	۰/۸۹۱	۷/۲۸	۰/۳۸۸
۳۹	۶/۳۳	۸/۹۵	۲/۶۱۶	۶/۶۵	۰/۳۱۷	۷/۳۰	۰/۳۵۰
۴۰	۶/۸۲	۶/۹۲	۰/۰۹۷	۷/۱۴	۰/۳۱۶	۵/۰۴	۰/۳۴۹
۴۱	۹/۱۷	۷/۸۱	۱/۳۶۰	۸/۲۲	۰/۹۵۱	۵/۷۶	۰/۷۳۴
۴۲	۶/۸۹	۷/۸۵	۰/۹۵۷	۸/۰۳	۱/۱۳۷	۵/۴۷	۰/۴۶۶
۴۳	۶/۹۵	۸/۴۵	۱/۵۰۱	۷/۴۵	۰/۵۰۲	۴/۳۵	۰/۳۴۹
۴۴	۵/۳۹	۵/۵۷	۰/۱۷۶	۵/۸۹	۰/۴۹۷	۶/۰۴	۰/۰۲۸
۴۵	۵	۵/۵۵	۰/۵۵۳	۵/۳۹	۰/۳۹۲	۶/۱۳	۰/۰۸۷
۴۶	۴	۳/۹۵	۰/۰۴۹	۴/۹۶	۰/۹۶۱	۴/۸۲	۰/۲۷۶
۴۷	۶/۰۷	۶/۳۷	۰/۳۰۰	۶/۴۱	۰/۳۳۷	۵/۶۶	۰/۳۵۲
۴۸	۶/۰۴	۶/۷۳	۰/۶۹۴	۶/۴۵	۰/۴۰۸	۶/۳۵	۰/۳۵۱
۴۹	۵/۱	۴/۵۴	۰/۵۶۵	۵/۰۱	۰/۰۹۵	۶/۸۰	۰/۵۷۲
۵۰	۷/۲۴	۶/۶۳	۰/۶۱۱	۶/۷۲	۰/۵۱۵	۶/۹۰	۰/۳۴۹
۵۱	۶/۲۳	۶/۹۱	۰/۶۸۵	۷/۲۴	۱/۰۰۵	۶/۱۰	۰/۳۵۲
۵۲	۷/۲۵	۷/۱۰	۰/۱۴۷	۷/۶۶	۰/۴۰۶	۶/۹۷	۰/۱۱۱
۵۳	۶/۴۵	۶/۳۹	۰/۰۵۶	۶/۸۰	۰/۳۵۲	۷/۳۳	۰/۳۵۰
۵۴	۶/۸۶	۶/۸۶	۰/۰۰۴	۶/۶۰	۰/۲۵۹	۶/۳۶	۰/۳۵۲
۵۵	۱۱	۱۰/۱۷	۰/۸۳۲	۱۰/۷۷	۰/۲۳۴	۹/۳۳	۰/۲۰۹
۵۶	۸/۸۸	۸/۵۰	۰/۳۸۲	۸/۹۲	۰/۰۴۳	۱۰/۳۵	۰/۶۴۹
۵۷	۱۰/۶	۹/۶۶	۰/۹۳۸	۱۰/۱۹	۰/۴۱۳	۸/۵۳	۰/۳۵۲
۵۸	۷/۶۶	۶/۶۰	۱/۰۵۸	۶/۸۰	۰/۸۵۸	۱۰/۲۵	۰/۳۵۲
۵۹	۷/۹۲	۸/۰۰	۰/۰۸۱	۸/۲۶	۰/۳۳۷	۷/۳۱	۰/۳۴۹
۶۰	۷/۴	۷/۳۲	۰/۰۷۹	۸/۰۰	۰/۶۰۲	۷/۹۴	۰/۰۲۰
۶۱	۸/۴۳	۸/۰۳	۰/۳۹۵	۸/۳۷	۰/۰۵۷	۷/۷۵	۰/۳۵۲
۶۲	۷/۲۸	۹/۵۸	۲/۳۰۰	۸/۷۷	۱/۴۹۵	۸/۴۷	۰/۰۴۱
۶۳	۹/۱	۸/۹۰	۰/۲۰۵	۹/۲۷	۰/۱۷۱	۷/۹۶	۰/۶۸۵
۶۴	۸/۸	۷/۸۱	۰/۹۸۸	۸/۲۸	۰/۵۱۸	۹/۰۵	۰/۰۵۳
۶۵	۸/۴۳	۸/۰۹	۰/۳۴۰	۸/۱۴	۰/۲۹۱	۸/۷۹	۰/۰۰۹
۶۶	۸/۰۶	۸/۰۸	۰/۰۱۷	۸/۳۶	۰/۳۰۲	۸/۱۵	۰/۲۸۴
۶۷	۸/۲۹	۸/۳۹	۰/۰۹۷	۸/۶۷	۰/۳۸۳	۷/۷۸	۰/۲۷۸
۶۸	۸/۴۶	۹/۰۹	۰/۶۲۷	۹/۲۳	۰/۷۷۰	۸/۵۱	۰/۰۴۷
۶۹	۸/۳۲	۸/۱۵	۰/۱۷۳	۸/۵۰	۰/۱۷۶	۸/۶۷	۰/۳۵۰
۷۰	۸/۴۷	۸/۵۳	۰/۰۶۱	۸/۵۵	۰/۰۷۹	۸/۷۷	۰/۳۷۱
۷۱	۹/۱۴	۸/۲۱	۰/۹۳۴	۸/۸۹	۰/۲۵۲	۹/۰۱	۰/۳۴۹
۷۲	۹/۳۶	۸/۸۰	۰/۵۵۷	۹/۳۱	۰/۰۵۴	۸/۰۴	۰/۳۵۳
۷۳	۸/۳۹	۸/۰۰	۰/۳۸۶	۸/۷۴	۰/۳۴۶	۷/۷۱	۰/۳۵۱
۷۴	۷/۳۶	۷/۲۹	۰/۰۶۸	۷/۵۸	۰/۲۱۸	۹/۱۲	۰/۳۴۹
۷۵	۹/۴۷	۸/۸۳	۰/۶۳۷	۹/۰۴	۰/۴۳۱	۱۱/۱۷	۰/۳۵۱
۷۶	۱۱/۵۲	۱۱/۲۸	۰/۲۳۸	۱۱/۷۰	۰/۱۸۲	۷/۰۸	۰/۵۱۱

۷۷	۶/۵۷	۶/۵۶	۰/۰۰۷	۶/۳۶	۰/۲۱۳	۶/۴۰	۰/۳۵۱
۷۸	۶/۰۵	۶/۶۲	۰/۵۶۵	۷/۲۶	۱/۲۰۷	۸/۰۵	۰/۳۵۰
۷۹	۸/۴	۹/۱۵	۰/۷۵۴	۹/۰۳	۰/۶۳۴	۹/۵۷	۰/۳۵۱
۸۰	۹/۹۲	۹/۲۵	۰/۶۷۵	۹/۵۹	۰/۳۲۸	۶/۶۸	۰/۱۱۹
۸۱	۶/۸	۶/۷۲	۰/۰۸۰	۷/۱۲	۰/۳۱۵	۷/۵۲	۰/۳۵۱
۸۲	۷/۱۷	۷/۱۳	۰/۰۴۲	۷/۸۷	۰/۷۰۲	۶/۰۵	۰/۳۵۱
۸۳	۵/۷	۵/۹۵	۰/۲۴۸	۶/۰۰	۰/۲۹۸	۷/۶۹	۰/۱۳۰
۸۴	۷/۸۲	۷/۲۷	۰/۵۴۶	۸/۰۸	۰/۲۵۶	۷/۳۳	۰/۳۵۰
۸۵	۶/۹۸	۷/۶۵	۰/۶۶۷	۷/۹۶	۰/۹۸۵	۸/۶۷	۰/۳۵۱
۸۶	۱۰/۱۴	۹/۱۶	۰/۹۷۶	۹/۴۴	۰/۷۰۳	۹/۸۱	۰/۳۳۲
۸۷	۵/۱۵	۶/۲۴	۱/۰۹۳	۵/۴۷	۰/۳۲۳	۵/۷۳	۰/۵۸۳
۸۸	۵	۴/۷۷	۰/۲۲۸	۴/۸۳	۰/۱۷۱	۶/۳۷	۰/۲۴۹
۸۹	۵/۸۵	۶/۸۸	۱/۰۳۵	۶/۴۷	۰/۶۱۶	۵/۲۱	۰/۲۱۵
۹۰	۶/۳۶	۶/۵۹	۰/۲۲۸	۶/۶۹	۰/۳۲۹	۶/۵۶	۰/۷۱۴
۹۱	۹/۴۳	۹/۲۹	۰/۱۴۳	۱۰/۰۵	۰/۶۲۴	۷/۶۵	۰/۳۵۱
۹۲	۷/۳۹	۷/۵۶	۰/۱۶۶	۷/۷۶	۰/۳۷۳	۶/۰۱	۰/۳۵۲
۹۳	۷/۶۸	۸/۲۵	۰/۵۶۸	۸/۵۳	۰/۸۵۲	۹/۰۸	۰/۳۵۳
۹۴	۸/۱۲	۷/۶۷	۰/۴۵۴	۸/۱۶	۰/۰۴۴	۸/۰۳	۰/۳۵۱
۹۵	۶/۶۹	۶/۸۵	۰/۱۶۲	۷/۰۵	۰/۳۶۵	۸/۲۴	۰/۱۱۷
۹۶	۹/۳۶	۸/۶۵	۰/۷۰۶	۸/۴۲	۰/۹۳۸	۸/۸۷	۰/۰۹۷
۹۷	۸/۷۷	۹/۲۲	۰/۴۵۲	۸/۴۶	۰/۳۱۱	۷/۸۹	۰/۳۴۹
۹۸	۷/۵۴	۸/۶۹	۱/۱۵۱	۷/۷۹	۰/۲۵۴	۱۰/۶۶	۰/۴۴۰
سری آزمون							
۹۹	۱۱/۱	۹/۲۰	۱/۹۰۳	۱۰/۱۴	۰/۹۵۷	۱۰/۴۵	۰/۱۴۸
۱۰۰	۱۰/۶	۹/۱۲	۱/۴۸۳	۱۰/۴۷	۰/۱۳۰	۶/۵۳	۰/۷۱۸
۱۰۱	۷/۵۷	۷/۳۳	۰/۲۴۰	۷/۴۳	۰/۱۳۹	۷/۰۵	۰/۵۱۸
۱۰۲	۷/۱	۷/۰۹	۰/۰۱۰	۶/۶۵	۰/۴۴۶	۹/۲۲	۰/۱۰۶
۱۰۳	۹/۱۱	۹/۰۶	۰/۰۵۰	۹/۱۴	۰/۰۲۶	۷/۶۱	۰/۳۸۶
۱۰۴	۵/۳	۵/۲۱	۰/۰۹۰	۴/۹۸	۰/۳۲۲	۷/۶۴	۰/۶۴۳
۱۰۵	۶/۹۲	۷/۱۴	۰/۲۲۰	۷/۲۲	۰/۲۹۶	۹/۷۷	۰/۱۷۱
۱۰۶	۸	۸/۶۶	۰/۶۵۷	۸/۷۵	۰/۷۴۹	۹/۸۳	۰/۱۷۴
۱۰۷	۷	۷/۰۰	۰/۰۰۰	۷/۰۵	۰/۰۵۴	۸/۹۵	۰/۵۱۳
۱۰۸	۱۰	۹/۵۷	۰/۴۳۳	۹/۵۱	۰/۴۸۸	۴/۷۰	۰/۲۱۸
۱۰۹	۹/۴۶	۸/۷۵	۰/۷۰۵	۸/۸۳	۰/۶۳۴	۶/۰۴	۰/۰۴۵
۱۱۰	۴/۹۲	۴/۹۱	۰/۰۱۴	۴/۴۸	۰/۴۳۵	۵/۸۵	۰/۱۵۴
۱۱۱	۶	۶/۴۴	۰/۴۴۱	۴/۹۴	۱/۰۶۴	۸/۷۸	۰/۷۳۰
۱۱۲	۶/۴۹	۵/۳۶	۱/۱۳۴	۵/۵۹	۰/۸۹۹	۸/۳۱	۰/۰۲۵
۱۱۳	۵/۳۱	۵/۴۶	۰/۱۵۰	۵/۶۲	۰/۳۱۵	۸/۴۵	۰/۷۱۵
۱۱۴	۶/۷	۶/۴۲	۰/۲۷۶	۶/۴۰	۰/۲۹۶	۸/۷۸	۰/۵۸۷
۱۱۵	۵/۲۳	۴/۸۵	۰/۳۸۲	۴/۷۲	۰/۵۰۶	۵/۱۵	۰/۰۷۶
۱۱۶	۶/۹۸	۸/۴۱	۱/۴۳۴	۷/۸۴	۰/۸۵۷	۷/۲۴	۰/۰۰۱
۱۱۷	۶/۰۱	۶/۶۷	۰/۶۶۴	۷/۰۱	۱/۰۰۳	۷/۷۰	۰/۵۹۰
۱۱۸	۹/۵۴	۱۰/۱۳	۰/۵۸۷	۹/۵۱	۰/۰۳۱	۹/۰۲	۰/۵۴۷
۱۱۹	۹/۰۲	۸/۱۲	۰/۸۹۷	۸/۵۴	۰/۴۷۹	۸/۰۱	۰/۶۲۴
۱۲۰	۶/۱۲	۶/۸۰	۰/۶۷۶	۶/۳۹	۰/۲۷۵	۷/۳۸	۰/۶۸۹
۱۲۱	۸	۶/۶۶	۱/۳۴۰	۷/۸۰	۰/۱۹۸	۹/۴۱	۰/۰۵۳
۱۲۲	۸/۹۱	۹/۰۵	۰/۱۳۷	۸/۹۴	۰/۰۳۴	۸/۶۴	۰/۲۷۲

جدول ۷. AAE سری آموزش و آزمون مدل‌های پیشنهادی با روش‌های مختلف کمومتریک

میانگین خطای مطلق سری آزمون (تعداد=۲۴)	میانگین خطای مطلق سری آموزش (تعداد=۹۸)	
۰/۵۸۰±۰/۵۳۷	۰/۵۸۱±۰/۵۱	MLR
۰/۴۴۳±۰/۳۲۹	۰/۴۷۳±۰/۳۳۱	ANN
۰/۳۶۵±۰/۲۵۸	۰/۳۶۵±۰/۲۳۰	SVM

جدول ۸. معیارهای آماری برای اعتبارسنجی خارجی سه مدل توسعه یافته

معیارهای آماری	MLR	ANN	SVM
$R^2 > 0/6$	۰/۸۱۲	۰/۹۱۱	۰/۹۲۴
$(R^2 - R_c^2) / R^2 < 0/1$	۰/۰۵	۰/۰۰۱	۰/۰۰۳
$0/85 < K < 1/15$	۰/۹۶۸	۰/۹۷۹	۰/۹۸۶
$R_m^2 > 0/5$	۰/۶۴۶	۰/۸۸۱	۰/۹۲۰

جدول ۹: انرژی برهمکنش و فاصله پیوند هیدروژنی بین N1 کینازولین و Met769 برای ترکیبات ۳۲، ۷۵ و ۹۸

ترکیبات	انرژی برهمکنش (کیلوکالری بر مول)	فاصله باند هیدروژنی (آنگستروم)
۳۲	-۷/۸۴	۲/۱۰
۷۵	-۸/۳۳	۱/۸۵
۹۸	-۷/۷۹	۱/۸۹

بحث و نتیجه‌گیری

نتایج نشان داد که ترکیبی از توصیف‌گرهای دوبعدی و سه‌بعدی برای پیش‌بینی pic_{50} ساختارهای مورد مطالعه انتخاب شدند. دو توصیف‌گر انتخاب شده شامل MATS3e و MATS1e توصیف‌گرهای هم‌بستگی دوبعدی بودند. از این توصیف‌گرها در مطالعات QSAR جهت محاسبه توزیع فضایی خواص مولکولی استفاده می‌شود. کابالرو و همکاران از توصیف‌گرهای هم‌بستگی دوبعدی برای ساخت مدل‌های QSAR از مشتقات پیریدوپیریمیدین به عنوان مهارکننده‌های تیروزین کیناز استفاده کردند [۳۴]. علاوه بر این، توصیف‌گرهای دوبعدی دیگری مانند شاخص‌های توپولوژیکی و BCUT انتخاب شدند. دسته‌ی توصیف‌گرهای BCUT نیز زیرمجموعه‌ای از توصیف‌گرهای توپولوژیکی هستند که اطلاعات مربوط به اتصالات اتمی را با سایر ویژگی‌های اتمی مثل بار اتمی، قطبیت‌پذیری اتمی و پتانسیل ایجاد پیوند هیدروژنی ترکیب می‌کنند. توصیف‌گرهای توپولوژیکی با توجه به نحوه‌ی آرایش اتمی ترکیبات تعریف می‌شوند و نشان‌دهنده‌ی اتصال اتم‌ها در مولکول به واسطه‌ی پیوندهای شیمیایی هستند. این توصیف‌گر دربردارنده اطلاعات مربوط به اندازه‌ی مولکولی، شکل مولکولی، انشعابات، هترواتم‌ها و انواع پیوندهای می‌باشد [۳۶، ۳۵]. هم‌چنین، ۴ توصیف‌گر سه‌بعدی شامل 3D-MoRSE ، RDF و WHIM در میان توصیف‌گرها انتخاب شده وجود داشتند. دسته‌ی توصیف‌گرهای 3D-MoRSE، توزیع سه‌بعدی ویژگی‌های مختلف مولکولی را در یک ترکیب بیان می‌کنند [۳۷]. اساس این توصیف‌گرها، دریافت اطلاعات از مختصات

سه‌بعدی اتمی توسط تکنیک پراش الکترونی است. در نهایت این توصیف‌گرها با جمع کردن ضرایب اتمی مشاهده‌شده توسط تابع پراکندگی زاویه‌ای محاسبه می‌شوند [۳۸]. انواع ضرایب اتمی مورد استفاده در این محاسبات شامل جرم اتمی، حجم و اندروالسی اتمی، الکترون‌گاتیویته‌ی اتمی و قطبیت‌پذیری اتمی می‌باشند [۳۹]. دسته‌ی توصیف‌گرهای WHIM (Weighted Holistic Invariant Molecular)، اطلاعات شیمیایی مختلفی را که به اندازه، شکل، تقارن، توزیع و تراکم اتم‌ها در مولکول مربوط هستند، بیان می‌کند. این شاخص‌ها از مختصات سه‌بعدی مولکول‌هایی که معمولاً از نظر کنفورماسیون فضایی دارای حداقل انرژی هستند، با استفاده از ضرایب مولکولی مختلف محاسبه می‌شوند. انواع ضرایب مولکولی مختلف شامل وزن‌های واحد، جرم اتمی، حجم و اندروالسی اتمی، الکترون‌گاتیویته‌ی اتمی، قطبیت‌پذیری اتمی و حالت توپولوژیکی اتمی می‌باشند. در نتیجه با به‌کاربردن ضرایب مولکولی مختلف، شاخص‌های مختلف نیز طراحی می‌شوند [۴۰-۴۲]. توصیف‌گرهای RDF (Radial distribution function) تمام الزامات توصیف‌گرهای سه‌بعدی را تأمین می‌کند. این توصیف‌گرها مستقل از تعداد اتم‌ها و اندازه یک مولکول هستند و از نظر آرایش سه‌بعدی اتم‌ها، منحصر به فرد هستند و مقدار آن در برابر چرخش کل مولکول ثابت است. علاوه بر این، توصیف‌گرهای RDF می‌توانند به انواع خاص اتم یا دامنه فاصله محدود شوند تا اطلاعات خاص را در یک فضای ساختار سه‌بعدی خاص نشان دهند. به عنوان مثال این توصیف‌گرها برای توصیف ممانعت فضایی یا خصوصیات ساختمان-فعالیت یک مولکول بسیار کارا هستند

این ساختارها و EGFR یک پیوند هیدروژنی بود که بین آمینواسید Met769 و N1 کینازولین هر سه ترکیب تشکیل شد. این برهمکنش برای اتصال مشتقات کینازولین در محل اتصال ATP در ساختار EGFR ضروری هستند. مشخص شد که هر سه ترکیب به بهترین شکل در محل اتصال ATP در ساختار EGFR قرار گرفتند که به ترتیب انرژی اتصال برای ترکیبات ۳۲، ۷۵ و ۹۸، $-7/84$ و $-8/33$ و $-7/79$ کیلوکالری بر مول بود (جدول ۹).

در این مطالعه از روش‌های مختلف کمومتریک برای ساخت مدل‌هایی جهت پیش‌بینی فعالیت مهارکننده‌های آنزیم تیروزین کیناز رستور فاکتور رشد اپیدرمی حاوی گروه ۴-آنیلینوکینازولین استفاده شد. برای نمایش ساختارهای TKIs از مجموعه بزرگی از توصیف‌گرها استفاده شد. نتایج نشان داد که الکترون‌گاتیویته‌ی اتمی، قابلیت ایجاد پیوند هیدروژنی، میزان لیوفیلیسیته و شکل و حجم مولکولی از عوامل مؤثر بر فعالیت این ترکیبات هستند و مدل‌های پیشنهادی 2D-QSAR که با استفاده از روش‌های خطی (MLR) و غیرخطی (ANN و SVM) ساخته شده‌اند می‌توانند برای پیش‌بینی دقیق فعالیت‌های TKIs استفاده شوند. مدل‌های غیر خطی در این تحقیق از مدل خطی نتایج بهتری نشان دادند. مدل‌های پیشنهادی می‌توانند در طراحی دارو برای ارزیابی فعالیت TKIs جدید حاوی کینازولین استفاده شوند. ترکیبات ۳۲، ۷۵ و ۹۸ از طریق تشکیل اتصال هیدروژنی N1 کینازولین و برهم کنش آب‌گریز حلقه آنیلین به درستی در جایگاه فعال آنزیم تیروزین کیناز قرار گرفتند.

تشکر و قدردانی

از معاونت تحقیقات و فن‌آوری دانشگاه علوم پزشکی گیلان جهت حمایت از این طرح کمال تشکر را داریم. این تحقیق با کد اخلاق IR.GUMS.REC.1397.191 در تاریخ ۹۷/۵/۲۷ در معاونت تحقیقات و فن‌آوری دانشگاه علوم پزشکی گیلان به ثبت رسیده است.

مشارکت و نقش نویسندگان

سعید قاسمی: ایده و طراحی مطالعه و آنالیز و تفسیر نتایج، حشمت اله ابراهیمی نجف آبادی و فاطمه یوسف بیک: آنالیز و تفسیر نتایج، مهتاب قاسمی دوگاهه: جمع‌آوری داده‌ها و نگارش نسخه اولیه مقاله. همه نویسندگان نتایج را بررسی نموده و نسخه نهایی مقاله را تایید نمودند.

[۴۳]. بر اساس توصیف‌گرهای انتخاب شده، الکترون‌گاتیویته‌ی اتمی، قابلیت ایجاد پیوند هیدروژنی، میزان لیوفیلیسیته و شکل و حجم مولکولی از عوامل مؤثر بر فعالیت این ترکیبات هستند. ماتریس هم‌بستگی بین توصیف‌گرها نشان داد که هیچ رابطه متقابل ($R < 0.6$) بین توصیف‌گرهای انتخاب شده وجود نداشت (جدول ۳) و توصیف‌گرهای انتخاب شده به طور خطی مستقل هستند و از این رو می‌توانند در توسعه مدل‌های QSAR در کنار هم استفاده شوند. نتایج نشان داد که ضریب هم‌بستگی برای تمامی مدل‌ها قابل قبول بود و تفاوت معنی‌داری بین R و R_{adj} وجود نداشت. اعتبارسنجی داخلی با استفاده از روش LOO ($q_{loo} = 0.727$) نشان می‌دهد که مدل‌های MLR پیشنهادی پیش‌بینی مناسبی دارد. شکل ۲ تأثیر تعداد توصیف‌گرها بر روی R و R_{adj} را برای مدل توسعه‌یافته نشان می‌دهد. افزایش در R_{adj} پس از اضافه شدن هر توصیف‌گر، تأثیر همه توصیف‌گرهای انتخاب شده را تأیید می‌کرد. در مرحله بعدی، از داده‌ها برای تهیه یک مدل ANN استفاده شد که در آن تعداد نورون‌های پنهان بهینه، چهار نورون بود. با توجه به نتایج هیچ تفاوت قابل توجهی بین خصوصیات آماری مجموعه‌های آموزش، اعتبارسنجی و آزمون وجود نداشت. علاوه بر این، بر اساس نتایج اعتبارسنجی خارجی مدل آموزش‌دیده معتبر بود و برازش بیش از حد رخ نداده بود. مدل SVM با استفاده از توصیف‌گرهای انتخاب شده توسعه یافت. بهینه‌سازی پارامترهای SVM (C) و γ با اعتبارسنجی متقاطع ۱۰ باره (10-fold cross-validation) با استفاده از نرم‌افزار STATISTICA 7 انجام شد. یک مدل معتبر با انتخاب پارامترهایی که کم‌ترین خطا را دارند ایجاد شد. برای تأیید این‌که این مدل‌ها برای کاربرد در ترکیبات دیگری غیر از مجموعه آموزش قابل استفاده‌اند و برای مقایسه سه مدل پیشنهادی، از ۲۴ مولکول مجموعه آزمون که در ساخت مدل استفاده نشده بودند، جهت انجام اعتبارسنجی خارجی استفاده شد. نتایج اعتبارسنجی خارجی و بررسی آماره‌های K ، R_m^2 ، $(R^2 - R_0^2)/R^2$ و R^2 در هر سه مدل نشان داد که هر سه مدل بسط‌یافته از اعتبار کافی برخوردار بودند. بررسی دقیق مقادیر AAE نشان داد که مدل‌های توسعه‌یافته از قابلیت پیش‌بینی خوبی برخوردار بودند. مطالعات داکینگ برای بررسی نحوه و قدرت اتصال قوی‌ترین مشتقات شامل ترکیبات ۳۲، ۷۵ و ۹۸ با جایگاه فعال گیرنده EGFR-TK (PDB ID: 1M17) انجام شد و اعتبارسنجی با استفاده از کافورماسیون فعال لیگاند‌های کوکریستالی انجام شد. حلقه‌های آنیلین هر ۳ ترکیب برهمکنش آب‌گریز با آمینو اسیدهای Leu764، Val702 و Lys721 در پاکت آب‌گریز، انجام دادند. علاوه بر این، برهمکنش حیاتی بین

منابع

Chem 2010; 45: 3970-3980.

<https://doi.org/10.1016/j.ejmech.2010.05.052>

PMid:20561721

[16] Darnag R, Mazouz EM, Schmitzer A, Villemain D, Jarid A, Cherqaoui D. Support vector machines: Development of qsar models for predicting anti-hiv-1 activity of tibo derivatives. Eur J Med Chem 2010; 45: 1590-1597. <https://doi.org/10.1016/j.ejmech.2010.01.002>

PMid:20110136

[17] Bridges AJ, Zhou H, Cody DR, Rewcastle GW, McMichael A, Showalter HD, et al. Tyrosine kinase inhibitors. 8. An unusually steep structure-activity relationship for analogues of 4-(3-bromoanilino)-6,7-dimethoxyquinazoline (pd 153035), a potent inhibitor of the epidermal growth factor receptor. Eur J Med Chem 1996; 39: 267-276. <https://doi.org/10.1021/jm9503613>

PMid:8568816

[18] Palmer BD, Trumpp-Kallmeyer S, Fry DW, Nelson JM, Showalter HD, Denny WA. Tyrosine kinase inhibitors. 11. Soluble analogues of pyrrolo- and pyrazoloquinazolines as epidermal growth factor receptor inhibitors: Synthesis, biological evaluation, and modeling of the mode of binding. Eur J Med Chem 1997; 40: 1519-1529. <https://doi.org/10.1021/jm960789h>

PMid:9154973

[19] Rewcastle GW, Denny WA, Bridges AJ, Zhou H, Cody DR, McMichael A, et al. Tyrosine kinase inhibitors. 5. Synthesis and structure-activity relationships for 4-[(phenylmethyl)amino]- and 4-(phenylamino)quinazolines as potent adenosine 5'-triphosphate binding site inhibitors of the tyrosine kinase domain of the epidermal growth factor receptor. Eur J Med Chem 1995; 38: 3482-3487. <https://doi.org/10.1021/jm00018a008>

PMid:7658435

[20] Rewcastle GW, Palmer BD, Bridges AJ, Showalter HD, Sun L, Nelson J, et al. Tyrosine kinase inhibitors. 9. Synthesis and evaluation of fused tricyclic quinazoline analogues as atp site inhibitors of the tyrosine kinase activity of the epidermal growth factor receptor. Eur J Med Chem 1996; 39: 918-928. <https://doi.org/10.1021/jm950692f>

PMid:8632415

[22] Mauri A, Consonni V, Pavan M, Todeschini R. Dragon software: An easy approach to molecular descriptor calculations. Match 2006; 56: 237-248.

[21] Teppen BJ. Hyperchem, release 2: Molecular modeling for the personal computer. J Chem Inform Comput Sci 1992; 32: 757-759. <https://doi.org/10.1021/ci00010a025>

[23] Ghanbarzadeh S, Ghasemi S, Shayanfar A, Ebrahimi-Najafabadi H. 2d-qsar study of some 2, 5-diaminobenzophenone farnesyltransferase inhibitors by different chemometric methods. EXCLI J 2015; 14: 484.

[24] Nunes CA, Freitas MP. Aug-mia-qspr on the modeling of sweetness values of disaccharide derivatives. LWT Food Sci Technol 2013; 51: 405-408. <https://doi.org/10.1016/j.lwt.2012.11.019>

[25] Habibi-Yangjeh A. Qsar study of the 5-ht 1a receptor affinities of arylpiperazines using a genetic algorithm-artificial neural network model. Monatsh Chem 2009; 140: 523-530. <https://doi.org/10.1007/s00706-008-0084-4>

[26] Leardi R, Seasholtz MB, Pell RJ. Variable selection for multivariate calibration using a genetic algorithm: Prediction of additive concentrations in polymer films from fourier transform-infrared spectral data. Anal Chim Acta 2002; 461: 189-200. [https://doi.org/10.1016/S0003-2670\(02\)00272-6](https://doi.org/10.1016/S0003-2670(02)00272-6)

[27] Soltani S, Abolhasani H, Zarghi A, Jouyban A. Qsar analysis of diaryl cox-2 inhibitors: Comparison of feature selection and train-test data selection methods. Eur J Med Chem 2010; 45: 2753-2760. <https://doi.org/10.1016/j.ejmech.2010.02.055>

PMid:20332057

[28] Dastmalchi S, Hamzeh-Mivehroud M, Ghafourian T, Hamzeiy H. Molecular modeling of histamine h3 receptor and qsar studies on arylbenzofuran derived h3 antagonists. J Mol Graph Model 2008; 26: 834-844. <https://doi.org/10.1016/j.jmgm.2007.05.002>

PMid:17561422

[1] Garofalo A, Goossens L, Lemoine A, Ravez S, Six P, Howsam M, et al. [4-(6, 7-disubstituted quinazolin-4-ylamino) phenyl] carbamic acid esters: A novel series of dual egfr/vegfr-2 tyrosine kinase inhibitors. Med Chem Comm 2011; 2: 65-72. <https://doi.org/10.1039/C0MD00183J>

[2] Zhang HQ, Gong FH, Ye JQ, Zhang C, Yue XH, Li CG, et al. Design and discovery of 4-anilinoquinazoline-urea derivatives as dual tk inhibitors of egfr and vegfr-2. Eur J Med Chem 2017; 125: 245-254. <https://doi.org/10.1016/j.ejmech.2016.09.039>

PMid:27688180

[3] Ahmadzadeh M, Farshdari F, Nematollahi L, Maleknia S, Mohit E. Optimization of her2-based and cell-based elisa for detection of trastuzumab biosimilar. Koomesh 2020; 22: 341-350. (Persian). <https://doi.org/10.29252/koomesh.22.2.341>

[4] Osmani F, Rasekhi A, Hajizadeh E, Akbari M E. Simultaneous modeling of multiple recurrences in breast cancer patients. Koomesh 2020; 22: 359-364. (Persian). <https://doi.org/10.29252/koomesh.22.2.359>

[5] Yu H, Li Y, Ge Y, Song Z, Wang C, Huang S, et al. Novel 4-anilinoquinazoline derivatives featuring an 1-adamantyl moiety as potent egfr inhibitors with enhanced activity against nscl cell lines. Eur J Med Chem 2016; 110: 195-203. <https://doi.org/10.1016/j.ejmech.2016.01.045>

[6] Cohen RB. Epidermal growth factor receptor as a therapeutic target in colorectal cancer. Clin Colorectal Cancer 2003; 2: 246-251. <https://doi.org/10.3816/CCC.2003.n.006>

PMid:12620146

[7] Chandregowda V, Kush A, Reddy GC. Synthesis and in vitro antitumor activities of novel 4-anilinoquinazoline derivatives. Eur J Med Chem 2009; 44: 3046-3055. <https://doi.org/10.1016/j.ejmech.2008.07.023>

PMid:18771819

[8] Hennequin LF, Stokes ES, Thomas AP, Johnstone C, Plé PA, Ogilvie DJ, et al. Novel 4-anilinoquinazolines with c-7 basic side chains: Design and structure activity relationship of a series of potent, orally active, vegf receptor tyrosine kinase inhibitors. Eur J Med Chem 2002; 45: 1300-1312. <https://doi.org/10.1021/jm011022e>

PMid:11881999

[9] Hansch C, Hoekman D, Gao H. Comparative qsar: Toward a deeper understanding of chemocobiological interactions. Chem Rev 1996; 96: 1045-1076. <https://doi.org/10.1021/cr9400976>

PMid:11848780

[10] Shayanfar A, Ghasemi S, Soltani S, Asadpour-Zeynali K, J Doerksen R, Jouyban A. Quantitative structure-activity relationships of imidazole-containing farnesyltransferase inhibitors using different chemometric methods. Med Chem 2013; 9: 434-448. <https://doi.org/10.2174/1573406411309030014>

PMid:22920090

[12] Kharkar PS. Two-dimensional (2d) in silico models for absorption, distribution, metabolism, excretion and toxicity (adme/t) in drug discovery. Current topics in medicinal chemistry 2010; 10: 116-126. <https://doi.org/10.2174/156802610790232224>

PMid:19929825

[13] Vapnik V. The nature of statistical learning theory: Springer science & business media; 2013.

[11] Verma J, Khedkar VM, Coutinho EC. 3d-qsar in drug design-a review. Current topics in medicinal chemistry 2010; 10: 95-115. <https://doi.org/10.2174/156802610790232260>

PMid:19929826

[14] Shahlaei M, Fassihi A, Saghaie L. Application of pc-ann and pc-ls-svm in qsar of ccr1 antagonist compounds: A comparative study. Eur J Med Chem 2010; 45: 1572-1582. <https://doi.org/10.1016/j.ejmech.2009.12.066>

PMid:20170987

[15] Cheng Z, Zhang Y, Fu W. Qsar study of carboxylic acid derivatives as hiv-1 integrase inhibitors. Eur J Med

<https://doi.org/10.1016/j.drudis.2016.06.013>

PMid:27326911

[37] Seedher N, Bhatia S, Singh B. Quantitative correlation between theoretical molecular descriptors and drug-hsa binding affinities for various cox-2 inhibitors. *Chem Biol Drug Des* 2008; 72: 297-302.

<https://doi.org/10.1111/j.1747-0285.2008.00711.x>

PMid:18844676

[38] Gosav S, Praisler M, Dorohoi D. Ann expert system screening for illicit amphetamines using molecular descriptors. *J Mol Struct* 2007; 834: 188-194.

<https://doi.org/10.1016/j.molstruc.2006.12.059>

[40] Gramatica P, Corradi M, Consonni V. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* 2000; 41: 763-777.

[https://doi.org/10.1016/S0045-6535\(99\)00463-4](https://doi.org/10.1016/S0045-6535(99)00463-4)

[39] Devinyak O, Havrylyuk D, Lesyk R. 3d-morse descriptors explained. *J Mol Graph Model* 2014; 54: 194-203.

<https://doi.org/10.1016/j.jmglm.2014.10.006>

PMid:25459771

[41] Gramatica P. Whim descriptors of shape. *QSAR Comb Sci* 2006; 25: 327-332.

<https://doi.org/10.1002/qsar.200510159>

[42] D'Archivio AA, Maggi MA, Mazzeo P, Ruggieri F. Quantitative structure-retention relationships of pesticides in reversed-phase high-performance liquid chromatography based on whim and getaway molecular descriptors. *Anal Chim Acta* 2008; 628: 162-172.

<https://doi.org/10.1016/j.aca.2008.09.018>

PMid:18929004

[43] González MP, Terán C, Teijeira M, Helguera AM. Radial distribution function descriptors: An alternative for predicting a2 a adenosine receptors agonists. *Eur J Med Chem* 2006; 41: 56-62.

<https://doi.org/10.1016/j.ejmech.2005.08.004>

PMid:16253394

[29] Jalali-Heravi M, Asadollahi-Baboli M, Shahbazikhah P. Qsar study of heparanase inhibitors activity using artificial neural networks and levenberg-marquardt algorithm. *Eur J Med Chem* 2008; 43: 548-556.

<https://doi.org/10.1016/j.ejmech.2007.04.014>

PMid:17602800

[30] Ghandadi M, Shayanfar A, Hamzeh-Mivehroud M, Jouyban A. Quantitative structure activity relationship and docking studies of imidazole-based derivatives as p-glycoprotein inhibitors. *Med Chem Res* 2014; 23: 4700-4712.

<https://doi.org/10.1007/s00044-014-1029-6>

[31] Golbraikh A, Tropsha A. Beware of q2! *J Mol Graph Model* 2002; 20: 269-276.

[32] Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 2008; 27: 302-313.

<https://doi.org/10.1002/qsar.200710043>

[33] Moghadam FA, Evazalipour M, Kefayati H, Ghasemi S. 6, 7-disubstituted-4-anilinoquinazoline: Design, synthesis and anticancer activity as a novel series of potent anticancer agents. *Pharm Sci* 2021; 27: 209-218.

<https://doi.org/10.34172/PS.2020.72>

[34] Caballero J, Fernández M, Saavedra M, González-Nilo F D. 2d autocorrelation, comfa, and comsia modeling of protein tyrosine kinases' inhibition by substituted pyrido [2, 3-d] pyrimidine derivatives. *Bioorg Med Chem* 2008; 16: 810-821.

<https://doi.org/10.1016/j.bmc.2007.10.024>

PMid:17964795

[35] Gozalbes R, Doucet J, Derouin F. Application of topological descriptors in qsar and drug design: History and new trends. *Curr Drug Targets Infect Disord* 2002; 2: 93-102.

<https://doi.org/10.2174/1568005024605909>

PMid:12462157

[36] Khan AU. Descriptors and their selection methods in qsar analysis: Paradigm for drug design. *Drug Discov Today* 2016; 21: 1291-302.

2D-QSAR and docking studies of 4-anilinoquinazoline derivatives as epidermal growth factor receptor tyrosine kinase inhibitors

Mahtab Ghasemi Dogaheh (Pharm.D)¹, Hesmatollah Ebrahimi-Najafabadi (Ph.D)¹, Fatemeh Yousefbeyk (Ph.D)², Saeed Ghasemi (Ph.D)^{*1}

¹ – Dept. of Medicinal Chemistry, School of Pharmacy, Guilan University of Medical Sciences, Rasht, Iran

² – Dept. of Pharmacognosy, School of Pharmacy, Guilan University of Medical Sciences, Rasht, Iran

* Corresponding author. +98 13 33486470 ghasemi_saeed@yahoo.com

Received: 19 Jun 2021 ; Accepted: 19 Sep 2021

Introduction: Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor derivatives play an important role in the treatment of cancer. We aim to construct 2D-QSAR models using various chemometrics using 4-anilinoquinazoline-containing EGFR TKIs. In addition, the binding profile of these compounds was evaluated using a docking study.

Materials and Methods: In this study, 122 compounds of seven different structural categories with a 4-anilinoquinazoline scaffold were obtained from the pieces of literature. 2D-QSAR models were prepared using the linear method, including multiple linear regression (MLR), alongside non-linear methods, including artificial neural networks (ANN) and support vector machines (SVM). The validation of suggested 2D-QSAR models was performed using internal and external validation techniques. For molecular docking, three compounds, including compounds 32, 75, and 98, which had the highest pIC₅₀, were used. Molecular docking was performed using AutoDock 4.2.

Results: Selected descriptors indicated that atomic electronegativity, hydrogen bonding ability, lipophilicity, and molecular shape and volume were the factors affecting the activity of the compounds. Statistical criteria, related to interpretation and validation of the models, were in the appropriate range. The model obtained by the ANN method, with the lowest mean absolute error of 0.365 for both training and testing sets, was the best. The compound 75 as the most potent compound indicated binding energies of -8.33 kcal/mol.

Conclusion: Finally, it was found that the models could effectively predict the activity of the compounds. It was also discovered that the compounds were properly bound to the active site of the receptor. In addition, validation results showed that all processes were sufficiently valid.

Keywords: Quantitative Structure-Activity Relationship, ErbB Receptors, Regression Analysis, Support Vector Machine, Neural Networks