



مدل سازی و مطالعه ارتباط کمی ساختار-فعالیت برخی از مشتقات کربوکسیلات به عنوان داروهای ضد سرطان با استفاده از رگرسیون خطی چند متغیره و شبکه های عصبی مصنوعی

پرستو فتاحی صدر^۱، محمود ابراهیمی^۱، مهدی نکوئی*^۲، بهزاد چهکندی^۱

^۱گروه شیمی، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

^۲گروه شیمی، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۴۰۱/۰۴/۱۸، تاریخ دریافت نسخه اصلاح شده: ۱۴۰۱/۰۷/۲۲، تاریخ پذیرش قطعی: ۱۴۰۱/۰۸/۱۴

چکیده

شیمی درمانی، یکی از مؤثرترین درمان ها برای سرطانها به شمار می رود. اما بسیاری از سرطانها در طی درمان با داروهای شیمی درمانی نسبت به اثرات درمانی داروی مصرفی، مقاوم می شوند که به آن مقاومت به داروهای چندگانه یا Multi Drug Resistance می گویند. در حال حاضر از برخی داروهای جدید از جمله مشتقات کربوکسیلات جهت کاهش مقاومت دارویی استفاده شده است. در تحقیق حاضر، مطالعه ارتباط کمی ساختار-فعالیت (QSAR) جهت پیش بینی فعالیت دارویی برخی از مشتقات کربوکسیلات با استفاده از روش رگرسیون خطی چند متغیره (MLR) و شبکه های عصبی مصنوعی (ANN) انجام شد. در ابتدا ساختار ترکیبات دارویی، رسم و گروه مناسبی از توصیف کننده ها محاسبه شدند. سپس از روش انتخاب مرحله ای برای بدست آوردن بهترین توصیف کننده ها که بیشترین ارتباط را با فعالیت دارویی ترکیبات مورد نظر داشتند استفاده گردید. در ابتدا مدل خطی رگرسیون خطی چندگانه (MLR) ساخته شد. سپس برای به دست آوردن نتایج بهتر از ANN استفاده گردید. داده های آماری، برتری روش ANN را نسبت به روش MLR نشان می دهد.

واژه های کلیدی: ارتباط کمی ساختار-فعالیت، مقاومت دارویی، مشتقات کربوکسیلات، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی.

۱. مقدمه

شیمی درمانی، یکی از مؤثرترین درمان ها برای سرطانها به شمار می رود اما بسیاری از سرطانها در طی درمان با داروهای شیمی درمانی نسبت به اثرات درمانی داروی مصرفی، مقاوم می شوند که به آن مقاومت به داروهای چندگانه یا Multi Drug Resistance

*عهده دار مکاتبات: مهدی نکوئی

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

پست الکترونیک: E-mail: m_nekoei1356@yahoo.com

تلفن: ۰۲۳۳۲۳۹۴۲۸۹

می‌گویند. سه دهه تحقیق درباره مقاومت به دارو، راه‌های مختلفی را برای برطرف کردن این مانع بزرگ روشن کرده است. طبق مطالعات گزارش شده مقاومت به دارو، در مقابل هر نوع داروی مؤثر و جدیدی نیز به وجود می‌آید [۵-۱]. بنابراین، توانایی پیش‌بینی و غلبه کردن بر مقاومت به دارو، احتمالاً در بهبود شیمی درمانی بسیار مؤثر خواهد بود. در حدود ۵۰٪ بیماران سرطانی که به جراحی یا رادیوتراپی جواب نداده و نیز بیمارانی که دچار بدخیمی‌های بافت مایع مانند خون می‌باشند، با شیمی درمانی، ایمونوتراپی و دیگر روش‌های تغییردهنده پاسخ بیولوژیکی، تحت درمان قرار می‌گیرند. طی سالها تحقیق، هنوز مشخص نشده است که چرا درمان بعضی از بیماران سرطانی با داروهای شیمی درمانی با موفقیت همراه بوده و چرا بعضی نیز موقتی به درمان پاسخ می‌دهند و یا اینکه بعضی از افراد، اصلاً به طور کامل درمان نمی‌شوند؟ به نظر می‌رسد که تغییرات ژنتیکی تومور و نیز متفاوت بودن ساختار ژنتیکی بیماران، تغییرات اپی‌ژنتیکی و عوامل محیطی تومور، همگی در داستان پیچیده مقاومت به داروی سرطانی نقش دارند. عوامل دخیل در مقاومت سلول به داروهای ضدسرطانی به دو دسته کلی تقسیم می‌شوند:

دسته اول عواملی که دسترسی سلول‌های توموری به داروهای ضد سرطان را تحت تأثیر قرار می‌دهند. در این صورت، تداخل در رسیدن دارو به سلول، می‌تواند در نتیجه جذب ضعیف داروهای مصرف شده، افزایش متابولیسم داروها و یا افزایش دفع آنها و در نهایت کاهش سطح دارو در خون و نیز کاهش انتشار آن از خون به توده توموری گردد.

دسته دوم عواملی که به دلیل تغییرات ژنتیکی، در خود سلول‌های سرطانی حساسیت به دارو را تحت تأثیر قرار می‌دهند. هم‌چنین عوامل محیطی مانند ماتریکس خارج سلولی یا موقعیت قرارگیری تومور نیز، در مقاومت به دارو نقش دارند.

برخی از داروها بر پایه مشتقات کربوکسیلات برای کاهش اثر مقاومت دارویی و در نهایت درمان بهتر و سریعتر سرطان طراحی، سنتز و مورد ارزیابی واقع شده‌اند. از آنجاییکه این بررسی‌ها بسیار وقت‌گیر و هزینه‌بر می‌باشد، استفاده از روشهایی برای تخمین فعالیت داروها قبل از سنتز آنها ضروری به نظر می‌رسد. برای پیش‌بینی فعالیت و یا میزان بازدارندگی ترکیبات دارویی و همچنین طراحی داروهای جدید استفاده از روشهای کمومتریکس از جمله ارتباط کمی ساختار - فعالیت (QSAR) روش مطمئن و مناسبی بوده که امروزه بطور وسیع از این روش‌ها استفاده می‌گردد [۱۰-۶]. یک مدل QSAR ارتباط ریاضی بین ساختار و فعالیت دسته‌ای از ترکیبات دارویی را توصیف می‌کند. بنابراین استفاده از اطلاعات مدل QSAR قبل از سنتز و تولید ترکیبات دارویی توسط کمپانی‌های سازنده‌ی این داروها بسیار حائز اهمیت می‌باشد. روشهای مختلفی از جمله رگرسیون خطی چندگانه (MLR)، کمترین مربعات جزئی (PLS)، شبکه‌های عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) در مدل‌سازی‌های QSAR مورد استفاده قرار گرفته است [۱۱-۱۵].

هدف از انجام این تحقیق پیش‌بینی فعالیت دارویی مشتقات کربوکسیلات با استفاده از روش‌های رگرسیون خطی چندگانه (MLR) و شبکه‌های عصبی مصنوعی (ANN) می‌باشد.

۲. روش‌های محاسباتی

۲-۱. انتخاب سری داده‌ها

در این کار تعداد ۴۱ ترکیب از مشتقات کربوکسیلات (CXL017) به عنوان ترکیبات ضد سرطان توسط روشهای کموتریکس مورد بررسی قرار گرفت [۱۶]. در این مقاله قدرت بازدارندگی این ترکیبات به صورت IC_{50} گزارش شده است. IC_{50} عبارتست از مینیمم غلظتی از ترکیب دارویی که باعث ۵۰٪ اثر بازدارندگی می‌شود. این مقادیر به مقیاس لگاریتمی تبدیل شده $-\log(IC_{50})=pIC_{50}$ و مورد استفاده قرار گرفته است. در این کار این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری پیش‌بینی تقسیم شده است، سری آموزش شامل ۲۹ مولکول (۷۰٪ داده‌ها) و سری پیش‌بینی شامل ۱۲ مولکول (۳۰٪ داده‌ها) می‌باشد. مقادیر pIC_{50} به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیر مستقل انتخاب شد. سری آموزش جهت ایجاد یک مدل مناسب و سری پیش‌بینی جهت ارزیابی مدل مورد استفاده قرار گرفت.

۲-۲. رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، ساختار مولکولی هر ترکیب ابتدا در نرم افزار HyperChem07 ترسیم شد. سپس با احتساب اتم‌های هیدروژن، ساختار سه بعدی ترکیبات با استفاده از روش‌های نیمه تجربی کوانتومی AM1 بهینه گردید و این بهینه‌سازی تا زمانی ادامه یافت که جذر میانگین مربعات گرادیان انرژی به ۰/۰۰۱ کیلوکالری بر مول برسد. با استفاده از این نرم افزار می‌توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم‌ها، انرژی تشکیل مولکول و... را بدست آورد.

۲-۳. محاسبه توصیف‌کننده‌ها

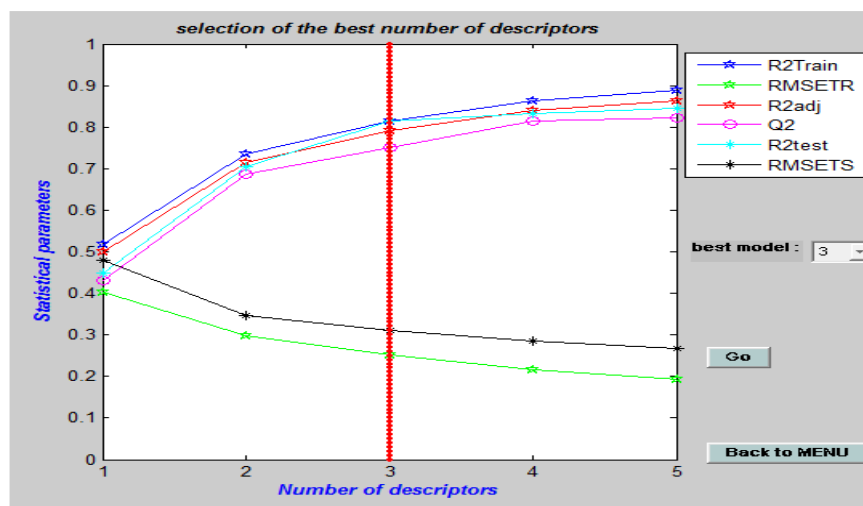
در اولین قدم باید مجموعه آزمایشی و ساختارهای متناظر جمع‌آوری شده و به صورت قابل پردازش برای کامپیوتر درآید. در مورد داده‌های آزمایشی که مقادیر عددی می‌باشند مشکلی نداریم، مساله مهم بیان ساختارهای شیمیایی به شکل قابل پذیرش برای کامپیوتر می‌باشد. برای محاسبه توصیف‌گرهای نظری ابتدا ساختارهای مولکولی به کمک نرم افزار Hyper Chem 0.7 رسم می‌شوند. این ساختارها به نرم افزار Dragon ویرایش ۵/۱ وارد شده و توصیف‌گرها به تعداد ۵۳۰۰ مورد به وسیله این نرم افزار محاسبه می‌شوند.

۲-۴. انتخاب مناسب‌ترین توصیف‌کننده‌ها

جهت انتخاب مناسب‌ترین توصیف‌کننده‌ها از روش رگرسیون مرحله‌ای^۱ استفاده شد. در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل شدند در این حالت، ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته

¹ Stepwise

دارد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معناداری خود را از دست بدهند، قبل از ورود متغیر جدید از مدل خارج می‌شود. به این ترتیب داده‌های pIC_{50} به عنوان متغیر وابسته و توصیفگرها به عنوان متغیر مستقل در نظر گرفته شده و تکنیک رگرسیون مرحله‌ای انجام شد. همانطور که می‌دانیم روش رگرسیون مرحله‌ای تعداد زیادی مدل ارائه می‌کند. که مدل اول شامل یک توصیفگر، مدل دوم شامل دو توصیفگر و ... می‌باشد. با افزایش تعداد توصیفگرها بالطبع مقدار R^2 افزایش و $RMSE$ (خطای جذر میانگین مربعات) کاهش می‌یابد. اما بدلیل پیچیدگی مدل، نمی‌توانیم تعداد زیادی توصیفگر را جهت مدلسازی انتخاب کنیم. بدین منظور و جهت انتخاب تعداد توصیفگرهای مناسب، نمودار پارامترهای مختلف آماری از جمله R^2_{train} , $RMSE_{train}$, R^2_{test} , $RMSE_{test}$ بر حسب تعداد توصیفگرها رسم گردید که در شکل ۱ نشان داده شده است. بر طبق این نمودار تعداد ۳ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت دارویی ترکیبات دارند، انتخاب شدند. این ۳ توصیفگر به همراه طبقه آنها در جدول ۱ ارائه شده است.



شکل ۱. نمودار پارامترهای آماری از جمله (R^2_{train} , R^2_{test} , $RMSE_{train}$, $RMSE_{test}$) بر حسب تعداد توصیفگرها

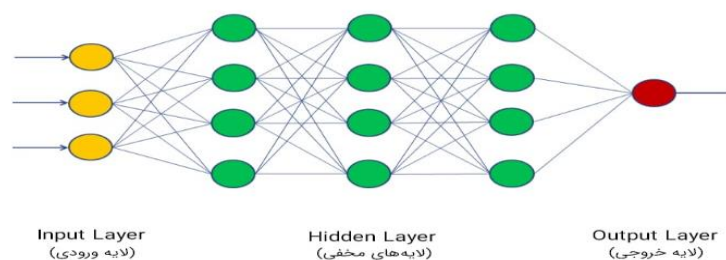
جدول ۱. توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله

نشانه توصیف کننده	مفهوم توصیف کننده	نوع توصیف کننده
R3u	R autocorrelation of lag 3/unweighted.	GETAWAY descriptors
R7e+	R maximal autocorrelation of lag 7/ weighted by atomic Sanderson electronegativity.	GETAWAY descriptors
nHDon	Number of donor atoms for H-bonds (with N and O).	Functional groups

¹Root-mean-square-error

۲-۵. شبکه های عصبی مصنوعی

شبکه های عصبی مصنوعی، سیستم ها و روش های محاسباتی نوین برای یادگیری، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت پیش بینی پاسخ های خروجی از سامانه های پیچیده هستند. ایده اصلی این گونه شبکه ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده ها و اطلاعات به منظور یادگیری و ایجاد دانش میباشد. شبکه عصبی مصنوعی روشی است که دانش ارتباط بین چند مجموعه داده را از طریق آموزش فراگرفته و برای استفاده در موارد مشابه ذخیره می کند. یک شبکه عصبی مصنوعی، از سه لایه ورودی، خروجی و پنهان تشکیل می شود. هر لایه شامل گروهی از سلول های عصبی (نورون) است که عموماً با کلیه نورون های لایه های دیگر در ارتباط هستند، مگر این که کاربر ارتباط بین نورون ها را محدود کند؛ ولی نورون های هر لایه با سایر نورون های همان لایه، ارتباطی ندارند. با استفاده از دانش برنامه نویسی رایانه می توان ساختار داده ای طراحی کرد که همانند یک نورون عمل نماید. سپس با ایجاد شبکه ای از این نورون های مصنوعی به هم پیوسته، ایجاد یک الگوریتم آموزشی برای شبکه و اعمال این الگوریتم به شبکه آن را آموزش داد. نورون کوچک ترین واحد پردازشگر اطلاعات است که اساس عملکرد شبکه های عصبی را تشکیل می دهد. یک شبکه عصبی مجموعه ای از نورون ها است که با قرار گرفتن در لایه های مختلف، معماری خاصی را بر مبنای ارتباطات بین نورون ها در لایه های مختلف تشکیل می دهند. نورون می تواند یک تابع ریاضی غیرخطی باشد، در نتیجه یک شبکه عصبی که از اجتماع این نورون ها تشکیل می شود، نیز می تواند یک سامانه کاملاً پیچیده و غیرخطی باشد. در شبکه عصبی هر نورون به طور مستقل عمل می کند و رفتار کلی شبکه، برآیند رفتار نورون های متعدد است. به عبارت دیگر، نورون ها در یک روند همکاری، یکدیگر را تصحیح می کنند [۲۰-۱۷]. شکل ۲ نمایی از یک شبکه عصبی مصنوعی را نشان می دهد



شکل ۲. نمایی از یک شبکه عصبی مصنوعی

۳. نتایج و بحث

۳-۱. مدل سازی به روش رگرسیون خطی چندگانه (MLR)

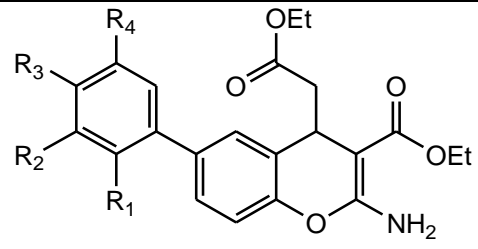
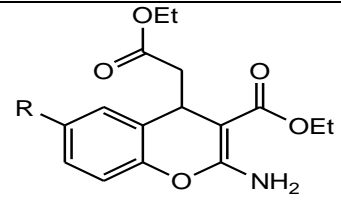
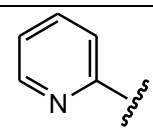
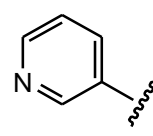
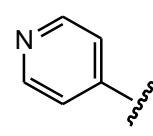
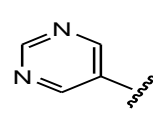
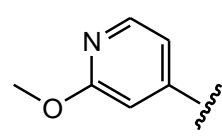
پس از انتخاب مناسب ترین توصیفگرها توسط روش مرحله ای، مرحله بعدی، ایجاد مدل میان توصیفگرهای انتخاب شده و pic_{50} می باشد. بین توصیف کننده ها و فعالیت دارویی ترکیبات برای سری آموزش با استفاده از روش MLR رابطه زیر به عنوان مدل خطی بدست آمد

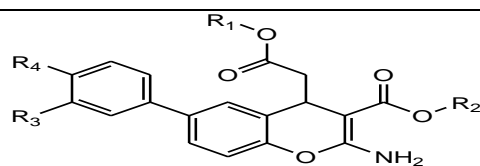
$$pIC_{50} = 11.476 - 3.204 (R3u) - 120.672 (R7e^+) + 0.190 (nHDon)$$

سپس از معادله بدست آمده برای پیش‌بینی فعالیت دارویی ترکیبات سری تست استفاده گردید. مقادیر تجربی و پیش‌بینی شده برای pIC_{50} برای کلیه ترکیبات مجموعه آموزش و تست در جدول (2) آورده شده است. شکل (3) نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی را نشان می‌دهد. شکل (4) نمودار مقادیر باقیمانده‌ها (اختلاف مقادیر پیش‌بینی شده و مقادیر تجربی) را بر حسب مقادیر تجربی نشان می‌دهد. در این شکل پراکندگی نقاط در حول خط صفر نشان می‌دهد که خطای سیستماتیک در مدل وجود ندارد.

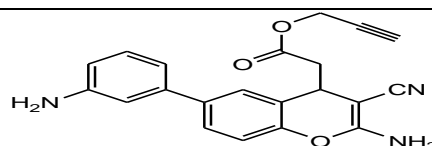
جدول 2. مقادیر تجربی و محاسبه شده pIC_{50} ترکیبات مختلف برای مجموعه‌های آموزشی و پیش‌بینی در مدل‌های SW-MLR، SW-ANN

NO	R ₁	R ₂	R ₃	R ₄	Exp (pIC_{50})	SW-MLR	SW-ANN
1	Me	Me	-	-	5.62	5.67	5.62
2	Et	Me	-	-	4.97	5.09	5.02
3 ^a	Et	Et	-	-	4.97	4.85	4.89
4 ^a	n-Pr	n-Pr	-	-	5.12	4.89	4.78
5	n-Bu	n-Bu	-	-	4.49	4.24	4.49
6	allyl	allyl	-	-	5.33	5.50	5.33
7	C ₃ H ₃	C ₃ H ₃	-	-	5.82	6.11	5.82
8 ^a	C ₃ H ₃	Et	-	-	5.33	5.33	5.32
9 ^a	cyclopropylmethyl	cyclopropylmethyl	-	-	5.17	4.99	4.85
10	-	-	-	-	4.71	4.58	4.71

							
11	F	H	H	H	4.14	4.39	4.32
12	H	F	H	H	4.12	3.96	4.12
13 ^a	H	H	F	H	4.14	4.34	4.35
14 ^a	F	H	F	H	4.12	3.84	4.15
15 ^a	H	F	H	F	4.44	4.48	4.41
							
16		-	-	-	4.11	4.44	4.46
17 ^a		-	-	-	4.67	4.46	4.48
18		-	-	-	4.99	4.90	4.99
19		-	-	-	4.42	4.24	4.42
20		-	-	-	4.99	5.12	4.99

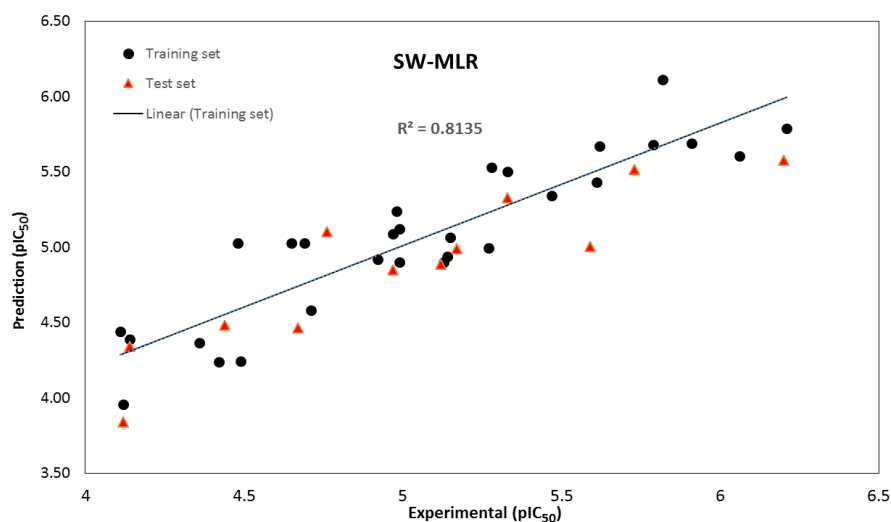


21	Et	Et	NH ₂	H	4.69	5.03	4.95
22	C ₃ H ₃	Et	NH ₂	H	5.47	5.34	5.47
23	C ₃ H ₃	C ₃ H ₃	NH ₂	H	5.91	5.69	5.89
24	Et	Et	H	NH ₂	4.65	5.03	4.95
25	C ₃ H ₃	Et	H	NH ₂	4.98	5.24	5.29
26	C ₃ H ₃	C ₃ H ₃	H	NH ₂	6.21	5.79	5.96
27	Et	Et	NHMe	H	5.13	4.90	5.09
28	C ₃ H ₃	Et	NHMe	H	5.15	5.06	5.15
29	C ₃ H ₃	C ₃ H ₃	NHMe	H	6.06	5.61	5.70
30	C ₃ H ₃	C ₃ H ₃	NMe ₂	H	4.92	4.92	4.92
31 ^a	C ₃ H ₃	Et	NHEt	H	4.76	5.10	5.17
32 ^a	C ₃ H ₃	C ₃ H ₃	NHEt	H	5.73	5.51	5.52
33 ^a	C ₃ H ₃	C ₃ H ₃	NHCH ₂ CH ₂ OH	H	6.20	5.58	5.56
34	C ₃ H ₃	C ₃ H ₃	NHCOCH ₂ Cl	H	5.28	5.53	5.59
35	Et	Et	CONH ₂	H	5.27	4.99	5.08
36	C ₃ H ₃	C ₃ H ₃	tetrazole	H	5.79	5.68	5.66
37	C ₃ H ₃	C ₃ H ₃	H	NHMe	5.61	5.43	5.61
38 ^a	C ₃ H ₃	C ₃ H ₃	H	NMe ₂	5.59	5.01	5.01
39	C ₃ H ₃	Et	H	NHEt	4.48	5.03	5.06
40	C ₃ H ₃	C ₃ H ₃	H	NHEt	5.14	4.94	5.14

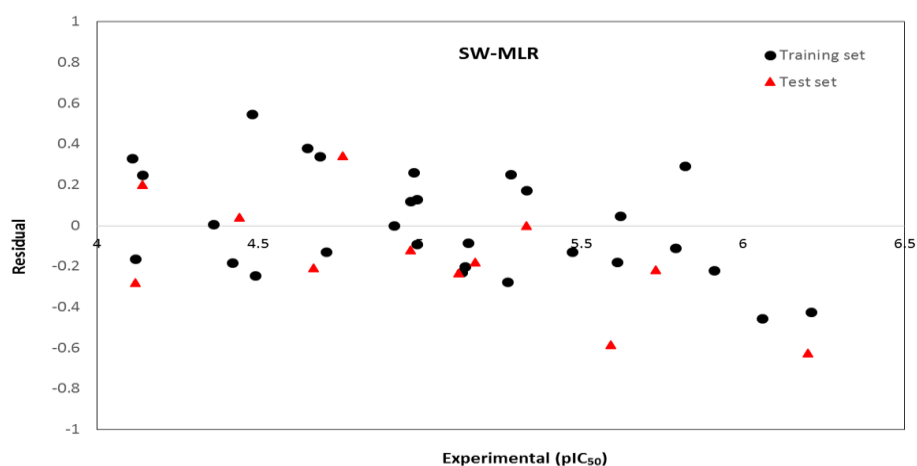


41	-	-	-	-	4.36	4.37	4.36
----	---	---	---	---	------	------	------

^aسری تست و بقیه سری آموزش



شکل ۳. نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری آموزش و تست به روش SW-MLR



شکل ۴. نمودار تغییرات باقیمانده‌ها بر حسب مقادیر تجربی برای مقادیر pIC_{50} بر اساس مدل SW-MLR در دو مجموعه آموزشی و تست

۲-۳. مدل سازی و پیش‌بینی توسط شبکه‌های عصبی مصنوعی

در این روش توصیف‌کننده‌های انتخاب‌شده وارد شبکه عصبی مصنوعی می‌شوند. پردازش داده‌ها در محیط ویندوز ۱۰ و با استفاده از نرم‌افزار MATLAB انجام شد. یک شبکه سه‌لایه با تابع انتقال سیگموئیدی برای نرون‌ها طراحی شده است. مقادیر اولیه وزن‌ها بطور تصادفی از بازه [0, 1] بوده و قبل از عمل آموزش مقادیر ورودی و خروجی در فاصله [0.1, 0.9] نرمال شده است. بهینه‌سازی و بهنگام‌کردن وزن‌ها و بایاس‌ها بوسیله الگوریتم BP^۱ انجام شده است. مجموعه داده‌ها به سه گروه تقسیم شده است: مجموعه آموزش، مجموعه ارزیابی و مجموعه تست. مجموعه آموزش (۵۰٪ داده‌ها) جهت آموزش دادن شبکه عصبی مصنوعی،

۱. Back-propagation

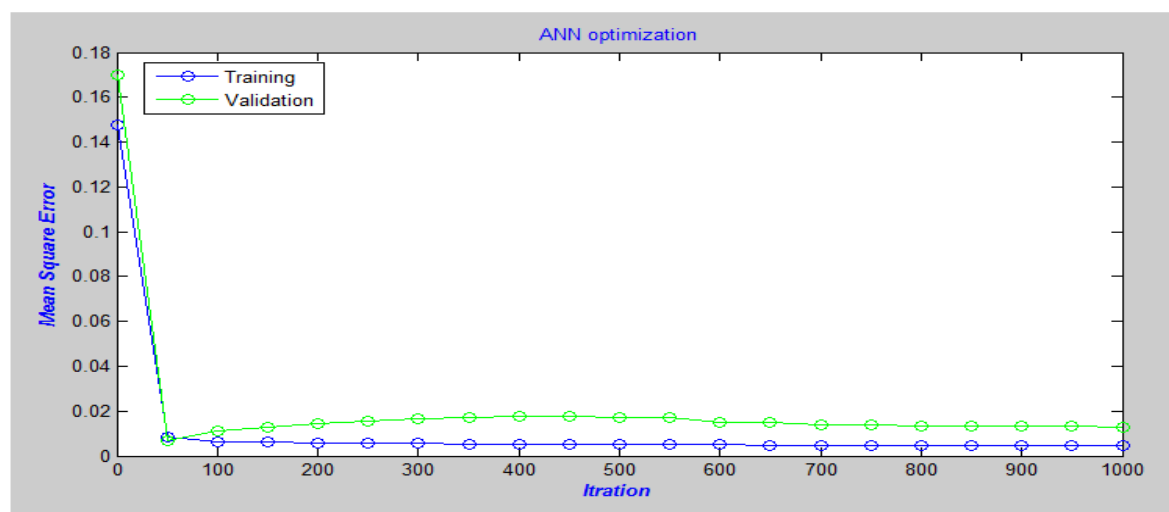
مجموعه ارزیابی (۲۰٪ داده ها) برای ارزیابی مدل در طی آموزش دادن شبکه و ایجاد مدل مناسب و مجموعه پیش‌بینی (۳۰٪ داده ها) برای تست مدل ایجاد شده به کار رفت.

تعداد نرون‌ها در لایه ورودی با تعداد توصیف‌کننده‌های وارد شده به شبکه‌های عصبی مصنوعی برابر است. به ازای هر تعداد توصیف‌کننده وارد شده به شبکه عصبی، تعداد نرون‌ها در لایه مخفی بهینه می‌شود. بدین ترتیب که به ازای هر مدل ANN، تعداد نرون‌ها در لایه مخفی از ۱ تا ۱۰ تغییر داده شده و مقادیر RMSE برای مجموعه‌های آموزشی و پیش‌بینی محاسبه گردید. از رسم مقادیر RMSE برحسب تعداد نرون‌ها در لایه مخفی، تعداد نرون‌های لایه مخفی بهینه شد. سپس بقیه پارامترها از جمله وزن‌ها و بایاس‌ها، سرعت یادگیری و مومنتوم نیز بهینه گردید. جدول ۳ مشخصات شبکه عصبی مصنوعی بهینه شده را نشان می‌دهد.

جدول ۳. ساختار و مشخصات ANN تولید شده

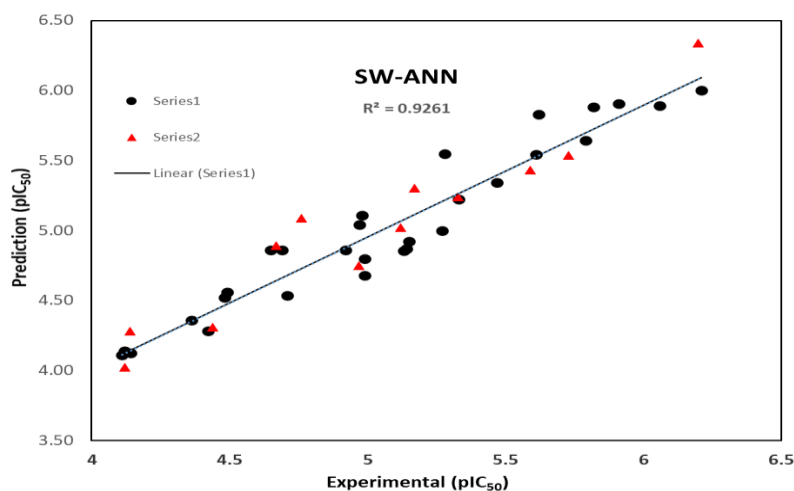
No. of nodes in the input layer	۳
No. of nodes in the hidden layer	۴
No. of nodes in the output layer	۱
Learning rate	۰/۵
Momentum	۰/۳
Number of epochs	۱۰۰۰
Transfer function	Sigmoid

برای جلوگیری از Overfitting در طول آموزش مقادیر RMSE بعد از هر ۵۰ بار چرخه آموزشی، محاسبه و ثبت گردید. شکل ۵ نمودار میزان خطا بر حسب تعداد دورها برای این داده‌ها را نشان می‌دهد. همانطور که ملاحظه می‌شود میزان خطا برای سری آموزش همواره در حال کاهش است اما برای سری ارزیابی در ۵۰ چرخه آموزش، کمترین خطا مشاهده می‌شود و بعد از آن افزایش می‌یابد. بنابراین این مقدار به عنوان مقدار بهینه تعداد چرخه‌های آموزش انتخاب شد.

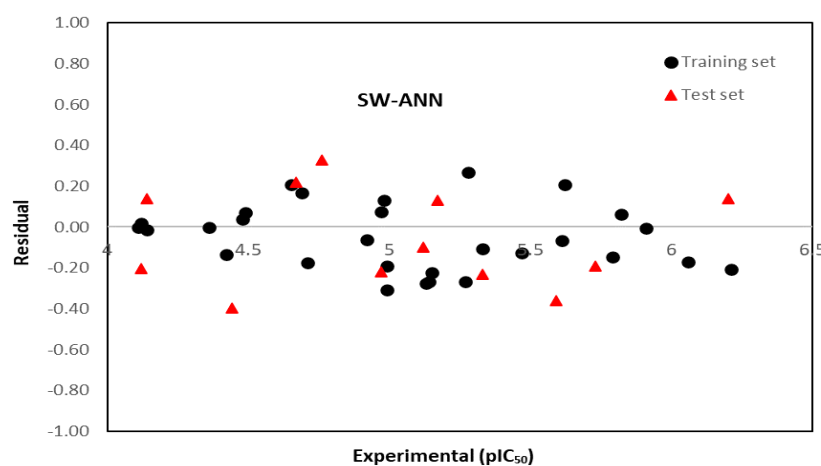


شکل ۵. مقادیر RMSE برای مجموعه‌های آموزشی و ارزیابی برحسب تعداد چرخه‌های آموزش

با استفاده از مدل ANN بهینه شده مقادیر فعالیتهای بازدارندگی (pIC_{50}) ترکیبات مورد نظر در مجموعه آموزشی، ارزیابی و پیش‌بینی (تست) مورد محاسبه قرار گرفت و در جدول (۲) نشان داده شده است. در شکل (۶) نیز مقادیر محاسبه شده pIC_{50} ترکیبات مورد نظر در مجموعه‌های مختلف برحسب مقادیر تجربی رسم شده‌اند. به‌منظور بررسی مقبولیت مدل ارائه شده برای مقادیر pIC_{50} محاسبه شده، نمودار تغییرات مقدار خطا برحسب مقادیر تجربی pIC_{50} در شکل (۷) رسم شده است. همانگونه که مشاهده می‌شود توزیع خطا کاملاً تصادفی بوده و مدل ارائه شده از نظر آماری قابل قبول می‌باشد.



شکل ۶. مقادیر pIC_{50} محاسبه شده براساس مدل SW-ANN در دو مجموعه آموزشی و تست برحسب مقادیر تجربی



شکل ۷. نمودار تغییرات باقیمانده‌ها برحسب مقادیر تجربی برای مقادیر pIC_{50} براساس مدل SW-ANN در دو مجموعه آموزشی و تست

۳-۳. ارزیابی مدل با استفاده از پارامترهای آماری

مطابق جدول ۴ چهار پارامتر آماری، جهت ارزیابی توانایی پیش‌بینی مدل‌های ساخته شده به روش‌های ANN، MLR به کار گرفته شد. همانطور که در این جدول مشاهده می‌شود تمام پارامترهای آماری برای روش غیرخطی شبکه عصبی مصنوعی بهتر از روش خطی رگرسیون خطی چندگانه است.

جدول 4. پارامترهای آماری برای مدل‌های انتخاب شده

		SW-MLR	SW-ANN
R ²	سری آموزش	0/813	0/926
	سری تست	0/742	0/919
RMSE	سری آموزش	0/250	0/167
	سری تست	0/309	0/175
REP	سری آموزش	4/954	3/305
	سری تست	6/163	3/490
AARD	سری آموزش	4/308	2/689
	سری تست	4/897	3/283

4. نتیجه گیری

در این تحقیق از دو روش رگرسیون خطی چندگانه و شبکه‌های عصبی مصنوعی جهت مدل‌سازی و پیش‌بینی فعالیت دارویی برخی مشتقات کربوکسیلات در درمان سرطان استفاده شد. توسط روش مرحله‌ای، سه توصیف‌کننده‌ایی که بیشترین ارتباط را با فعالیت دارویی داشتند شامل R3u، R7e⁺ و nHDOn انتخاب شدند. جهت مدل‌سازی از دو روش رگرسیون خطی چندمتغیره (MLR) و شبکه‌های عصبی مصنوعی (ANN) استفاده گردید. نتایج نشان از برتری روش ANN نسبت به MLR دارد بنابراین از این روش می‌توان برای پیش‌بینی فعالیت دارویی ترکیبات مشابه استفاده کرد. یعنی با محاسبه توصیف‌کننده‌های یک ترکیب دارویی جدید که با نرم افزار دراگون قابل انجام است و وارد کردن آنها به شبکه عصبی بهینه شده می‌توان مقدار فعالیت دارویی آن را محاسبه نمود.

5. مراجع

- [1] Bukowski, K., Kciuk, M., & Kontek, R. (2020). Mechanisms of multidrug resistance in cancer chemotherapy. *International journal of molecular sciences*, 21(9), 3233.
- [2] Kumar, A., & Jaitak, V. (2019). Natural products as multidrug resistance modulators in cancer. *European Journal of Medicinal Chemistry*, 176, 268-291.
- [3] Jernigan, J. A., Hatfield, K. M., Wolford, H., Nelson, R. E., Olubajo, B., Reddy, S. C., ... & Baggs, J. (2020). Multidrug-resistant bacterial infections in US hospitalized patients, 2012–2017. *New England Journal of Medicine*, 382(14), 1309-1319.
- [4] Cui, Q., Wang, J. Q., Assaraf, Y. G., Ren, L., Gupta, P., Wei, L., ... & Chen, Z. S. (2018). Modulating ROS to overcome multidrug resistance in cancer. *Drug Resistance Updates*, 41, 1-25.
- [5] for the Meta, T. C. G., Ahmad, N., Ahuja, S. D., Akkerman, O. W., Alffenaar, J. W. C., Anderson, L. F., ... & Menzies, D. (2018). Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *The Lancet*, 392(10150), 821-834.

- [6] Cai, Z., Zafferani, M., Akande, O. M., & Hargrove, A. E. (2022). Quantitative Structure–Activity Relationship (QSAR) Study Predicts Small-Molecule Binding to RNA Structure. *Journal of medicinal chemistry*, 65(10), 7262-7277.
- [7] Xu, S., Zeng, X., Dai, S., Wang, J., Chen, Y., Song, J., ... & Zhao, Z. (2020). Turpentine derived secondary amines for sustainable crop protection: Synthesis, activity evaluation and QSAR study. *Journal of Agricultural and Food Chemistry*, 68(42), 11829-11838.
- [8] Duhan, M., Singh, R., Devi, M., Sindhu, J., Bhatia, R., Kumar, A., & Kumar, P. (2021). Synthesis, molecular docking and QSAR study of thiazole clubbed pyrazole hybrid as α -amylase inhibitor. *Journal of Biomolecular Structure and Dynamics*, 39(1), 91-107.
- [9] Soukup, O., Benkova, M., Dolezal, R., Sleha, R., Malinak, D., Salajkova, S., ... & Marek, J. (2020). The wide-spectrum antimicrobial effect of novel N-alkyl monoquaternary ammonium salts and their mixtures; the QSAR study against bacteria. *European Journal of Medicinal Chemistry*, 206, 112584.
- [10] Marquina, S., Maldonado-Santiago, M., Sánchez-Carranza, J. N., Antúnez-Mojica, M., González-Maya, L., Razo-Hernández, R. S., & Alvarez, L. (2019). Design, synthesis and QSAR study of 2'-hydroxy-4'-alkoxy chalcone derivatives that exert cytotoxic activity by the mitochondrial apoptotic pathway. *Bioorganic & Medicinal Chemistry*, 27(1), 43-54.
- [11] Chtita, S., Belhassan, A., Bakhouch, M., Taourati, A. I., Aouidate, A., Belaidi, S., ... & Lakhlifi, T. (2021). QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods. *Chemometrics and Intelligent Laboratory Systems*, 210, 104266.
- [12] Ewees, A. A., Abualigah, L., Yousri, D., Algamal, Z. Y., Al-Qaness, M. A., Ibrahim, R. A., & Abd Elaziz, M. (2022). Improved Slime Mould Algorithm based on Firefly Algorithm for feature selection: A case study on QSAR model. *Engineering with Computers*, 38(3), 2407-2421.
- [13] Mozafari, Z., Chamjangali, M. A., & Arashi, M. (2020). Combination of least absolute shrinkage and selection operator with Bayesian Regularization artificial neural network (LASSO-BR-ANN) for QSAR studies using functional group and molecular docking mixed descriptors. *Chemometrics and Intelligent Laboratory Systems*, 200, 103998.
- [14] Mozafari, Z., Arab Chamjangali, M., Arashi, M., & Goudarzi, N. (2021). Performance of smoothly clipped absolute deviation as a variable selection method in the artificial neural network-based QSAR studies. *Journal of Chemometrics*, 35(5), e3338.
- [15] Vahedi, Nafiseh, Majid Mohammadhosseini, and Mehdi Nekoei. "QSAR Study of PARP Inhibitors by GA-MLR, GA-SVM and GA-ANN Approaches." *Current Analytical Chemistry* 16, no. 8 (2020): 1088-1105.
- [16] Cong, H., Zhao, X., Castle, B. T., Pomeroy, E. J., Zhou, B., Lee, J., ... & Zhuang, C. (2018). An indole–chalcone inhibits multidrug-resistant cancer cell growth by targeting microtubules. *Molecular pharmaceutics*, 15(9), 3892-3900.
- [17] Triolascarya, K., Septiawan, R. R., & Kurniawan, I. (2022). QSAR Study of Larvicidal Phytocompounds as Anti-Aedes Aegypti by using GA-SVM Method. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(4), 632-638.
- [18] Pramana, I. K. A. P. P., Septiawan, R. R., & Kurniawan, I. (2022). QSAR Study on Diacylglycerol Acyltransferase-1 (DGAT-1) Inhibitor as Anti-diabetic using PSO-SVM Methods. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(5), 735-741.
- [19] Ai, H., Wu, X., Zhang, L., Qi, M., Zhao, Y., Zhao, Q., ... & Liu, H. (2019). QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods. *Ecotoxicology and Environmental Safety*, 179, 71-78.

Modeling and quantitative structure-activity study of some carboxylate derivatives as anticancer drugs using multivariate linear regression and artificial neural networks

Parastou Fattahi Sadr¹, Mahmoud Ebrahimi¹, Mehdi Nekoei^{*2}, Behzad Chahkandi¹

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

²Department of Chemistry, Shahrood Branch, Islamic Azad University, Shahrood, Iran

Submitted: 09 July 2022, Revised: 14 October 2022, Accepted: 05 November 2022

Abstract

Chemotherapy is one of the most effective treatments for cancers, but many cancers become resistant to the therapeutic effects of a drug during treatment with chemotherapy, which is called Multi Drug Resistance. Currently, some new drugs, including carboxylate derivatives, have been used to reduce drug resistance. In the present study, a structure-activity quantitative relationship (QSAR) study was performed to predict the drug activity of some carboxylate derivatives using multivariate linear regression (MLR) and artificial neural networks (ANN). First, the structure of drug compounds, drawing and appropriate group of descriptors were calculated. Then, the step selection method was used to obtain the best descriptors that were most related to the drug activity of the compounds. First, the linear model of multiple linear regression (MLR) was developed. ANN was then used to obtain better results. Statistical data show the superiority of ANN method over MLR method.

Keywords: *Quantitative structure-activity relationship, drug resistance, carboxylate derivatives, multiple linear regression, artificial neural network.*

*Corresponding author : Mehdi Nekoei

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

Tel: 02332394289

E-mail: m_nekoei1356@yahoo.com