



## کاربرد رگرسیون خطی چند متغیره و شبکه های عصبی مصنوعی جهت پیش بینی فعالیت ضد میکروبی برخی مشتقات آنیلیدها به روش ارتباط کمی ساختار - فعالیت (QSAR)

مهدی نکویی\*<sup>۱</sup>، سید حامد موسوی<sup>۲</sup>، پرستو پورعلی<sup>۲</sup>

<sup>۱</sup>گروه شیمی، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

<sup>۲</sup>گروه زیست شناسی، واحد شاهرود، دانشگاه آزاد اسلامی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۴۰۰/۰۶/۲۵، تاریخ دریافت نسخه اصلاح شده: ۱۴۰۰/۰۹/۲۴، تاریخ پذیرش قطعی: ۱۴۰۰/۱۰/۰۵

### چکیده

مطالعه ارتباط کمی ساختار-فعالیت (QSAR) جهت پیش بینی فعالیت ضد میکروبی برخی مشتقات آنیلیدها با استفاده از روشهای رگرسیون خطی چند متغیره (MLR) و شبکه های عصبی مصنوعی (ANN) انجام شد. در ابتدا ساختار ترکیبات، 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 ضد میکروبی ترکیبات مورد نظر داشتند استفاده گردید. با این روش ۵ توصیف کننده انتخاب و در ابتدا مدل خطی MLR ساخته شد. سپس برای به دست آوردن نتایج بهتر از شبکه عصبی مصنوعی استفاده گردید. مقادیر ضریب تعیین ( $R^2$ ) و ریشه میانگین مربعات خطا (RMSE) برای سری تست به ترتیب برابر ۰/۰۷۷ و ۰/۰۷۳ برای مدل خطی MLR و ۰/۶۱۳ و ۰/۰۲۱ برای مدل غیرخطی ANN بدست آمد. داده های آماری، برتری روش ANN را نسبت به روش MLR نشان می دهد.

واژه های کلیدی: ارتباط کمی ساختار-فعالیت، مشتقات آنیلیدها، رگرسیون خطی چند گانه، شبکه عصبی مصنوعی

### ۱. مقدمه

آنیلیدها یا فنیل آمیدها دسته ای از ترکیبات شیمیایی هستند که مشتق آسیل از ترکیب آروماتیک آنیلین محسوب می شوند. از واکنش آنیلین با آسیل کلریدها یا کربوکسیلیک آنیدریدها، آنیلیدها تولید می شوند. به عنوان مثال، واکنش آنیلین با استیل

\*عهده دار مکاتبات: مهدی نکویی

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

پست الکترونیک: E-mail:m\_nekoei1356@yahoo.com

تلفن: ۰۲۳۲۳۹۴۲۸۹

کلرید منجر به تولید استانیلید می‌شود. آنیلین و کربوکسیلیک اسیدها در دمای بالا واکنش نشان داده و آنیلید می‌دهند. یکی از کاربردهای مهم آنیلیدها در تهیه علف کش ها و ترکیبات ضد میکروبی است. [۵-۱]. MIC غلظتی از یک آنتی‌بیوتیک است که می‌تواند رشد باکتری را در شرایط آزمایشگاهی مهار کند. روش‌های مختلفی برای اندازه‌گیری این پارامتر وجود دارد اما از آنجاییکه این بررسی‌ها وقت گیر و هزینه بر می‌باشند استفاده از روش‌هایی برای تخمین این پارامترها ضروری به نظر می‌رسد. برای پیش‌بینی فعالیت ضد میکروبی ترکیبات دارویی روش ارتباط کمی ساختار - فعالیت (QSAR) روش مطمئن و مناسبی می‌باشد. QSAR ارتباط ریاضی بین ساختار و فعالیت دسته‌ای از ترکیبات دارویی را توصیف می‌کند. روشهای مختلفی از جمله رگرسیون خطی چندگانه (MLR)، کمترین مربعات جزئی (PLS)، شبکه‌های عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) در مدل‌سازی های QSAR مورد استفاده قرار گرفته است [۱۵-۶].

هدف از انجام این تحقیق پیش‌بینی فعالیت ضد میکروبی برخی مشتقات آنیلیدها با استفاده از روش انتخاب مرحله‌ای<sup>۱</sup> به عنوان روش انتخاب متغیر و روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) به عنوان روش‌های مدل‌سازی می‌باشد.

## ۲. روش‌های محاسباتی

### ۲-۱. انتخاب سری داده‌ها

در این کار تعداد ۴۴ ترکیب از مشتقات آنیلید به عنوان ترکیبات ضد میکروبی توسط روش‌های QSAR مورد بررسی قرار گرفت [۱۶]. در این مقاله قدرت بازدارندگی این ترکیبات به صورت MIC گزارش شده است. طبق تعریف MIC غلظتی از یک آنتی‌بیوتیک است که می‌تواند رشد باکتری را در شرایط آزمایشگاهی مهار کند. این مقادیر به مقیاس لگاریتمی تبدیل شده -  $\log(\text{MIC}) = \text{pMIC}$  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 ۳۵ مولکول (۸۰٪) و سری پیش‌بینی شامل ۹ مولکول (۲۰٪) می‌باشد. مقادیر pMIC به عنوان متغیر وابسته و توصیف کننده‌ها به عنوان متغیر مستقل انتخاب شد. سری آموزش جهت ایجاد یک مدل مناسب و سری پیش‌بینی جهت ارزیابی مدل مورد استفاده قرار گرفت. در شبکه عصبی مصنوعی علاوه بر سری آموزش و تست از سری ارزیابی نیز استفاده گردید.

### ۲-۲. رسم و بهینه سازی 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 بهینه سازی تا زمانی ادامه

<sup>1</sup> Stepwise

یافت که جذر میانگین مربعات گرادیان انرژی به  $0/001$  کیلوکالری بر مول برسد. با استفاده از این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم ها، انرژی تشکیل مولکول و... را بدست آورد.

### ۲-۳. محاسبه توصیف کننده ها

در اولین قدم باید مجموعه آزمایشی و ساختارهای متناظر جمع آوری شده و به صورت قابل پردازش برای کامپیوتر درآید. در مورد داده های آزمایشی که مقادیر عددی می باشند مشکلی نداریم، مساله مهم بیان ساختارهای شیمیایی به شکل قابل پذیرش برای کامپیوتر می باشد. برای محاسبه توصیف گرهای نظری ابتدا ساختارهای مولکولی به کمک نرم افزار Hyper Chem 0.7 رسم می شوند. این ساختارها به نرم افزار Dragon ویرایش  $2/1$  وارد شده و توصیف گرها به تعداد ۱۴۸۱ مورد به وسیله این نرم افزار محاسبه می شوند. این نرم افزار برای محاسبه هیجده دسته از توصیف گرهای مولکولی مورد استفاده قرار می گیرد. اولین ویرایش این نرم افزار در سال ۱۹۹۷ مورد استفاده قرار گرفته است. این نرم افزار هیچگونه بهینه سازی در ساختار را انجام نمی دهد و می تواند ۱۴۸۱ توصیف گر را برای مولکول های تا ۱۵۰ اتم محاسبه نماید. همانطور که گفته شد، خروجی نرم افزار HyperChem برای هر ترکیب به برنامه Dragon منتقل و تعداد ۱۴۸۱ توصیف گر مولکولی برای هر ترکیب محاسبه شد.

### ۲-۴. کاهش تعداد توصیف کننده ها

یکی از مشکلاتی که در هنگام ایجاد مدل های QSAR با آن مواجه می شویم، تعداد زیاد متغیرهای مستقل می باشد. امروزه روش های QSAR با این حقیقت که تعداد ترکیبات شیمیایی اغلب در مقایسه با تعداد متغیرهای مستقل کمتر است، محدود شده اند. در اغلب موارد تعداد توصیف گرها از تعداد مولکول ها بیشتر است. در این صورت استفاده از روش های حداقل مربعات باعث ایجاد مشکلاتی نظیر انتخاب شانسی و همبستگی تصادفی می گردد. با توجه به این که بعضی از متغیرهای مستقل ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می دهند، لذا به روش زیر بعضی از متغیرها حذف می شوند. توصیف گرهای محاسبه شده برای ترکیبات، مورد بررسی قرار می گیرند تا آنهایی را که مقادیر ثابت و یا تقریباً ثابت دارند (بیش از  $90\%$  داده های ثابت دارند)، حذف شوند. در این مرحله تعداد ۳۲۷ توصیف گر کنار گذاشته شدند. بدین ترتیب ۱۱۵۴ توصیف گر باقی می ماند. با توجه به اینکه در برخی از موارد بعضی از متغیرهای مستقل با یکدیگر همبستگی بالایی دارند و وجود تنها یکی از این متغیرها در مدل کردن کافی بوده و نیازی به حضور بقیه متغیرها نمی باشد. لذا داده ها از این نظر مورد بررسی قرار گرفتند. به منظور حذف متغیرهایی که همبستگی خطی بزرگتر از  $0/9$  با یکدیگر دارند، ماتریس  $1154 \times 1154$  بین همه متغیرهای مستقل و نیز متغیر وابسته تشکیل می شود. متغیر مستقلی که همبستگی بیشتری از  $0/9$  با متغیرهای دیگر دارند، بایستی حذف شوند. اما بایستی دقت داشت که هنگام حذف متغیر اضافی، متغیری کنار گذاشته می شود که همبستگی کمتری با متغیر وابسته داشته باشد. بدین ترتیب تعداد ۹۴۱ توصیف گر اضافی کنار گذاشته می شوند و تعداد ۲۱۳ توصیف گر باقی می ماند.

## ۲-۵. انتخاب مناسب‌ترین توصیف‌کننده‌ها

یکی از مهمترین مراحل در مطالعات QSAR انتخاب مناسب‌ترین توصیف‌کننده‌ها است که بتوانند فعالیت ضد میکروبی را به ساختار مولکول ارتباط دهند. بدین منظور، رگرسیون مرحله به مرحله توسط نرم افزار SPSS جهت انتخاب توصیفگرهای مناسب برای مدل‌سازی مورد استفاده قرار گرفت. 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 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 pMIC 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  $R^2$  نیز افزایش می‌یابد. اما بدلیل پیچیدگی مدل، نمی‌توانیم تعداد زیادی توصیفگر را جهت مدل‌سازی انتخاب کنیم. بدین منظور و جهت انتخاب تعداد توصیفگرهای مناسب، نمودار پارامترهای آماری از جمله ( $R^2_{train}$ ) و ... برحسب تعداد توصیفگرها رسم گردید و بر طبق آن، تعداد ۵ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت ضد میکروبی دارند، انتخاب شدند. این ۵ توصیفگر به همراه مشخصات آنها در جدول (۱) ارائه شده است.

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

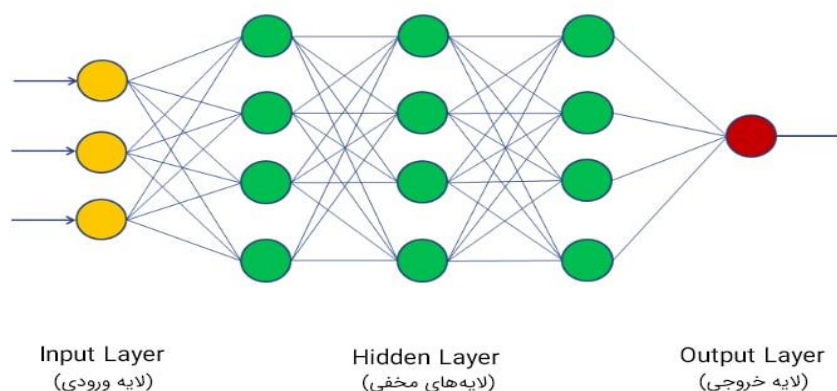
نشانه توصیف کننده	معنی توصیف کننده	نوع توصیف کننده	اثر متوسط توصیف کننده
GATS4v	Geary autocorrelation-lag4/weighted by atomic van der Waals volumes	2D autocorrelations.	۱/۸۴۵
RDF085m	Radial Distribution Function -8.5 / weighted by atomic masses.	RDF descriptors.	-۰/۰۸۵
RDF060e	Radial Distribution Function -8.0 / Weighted by atomic Sanderson electronegativities.	RDF descriptors.	-۰/۳۸۴
Mor09m	3D-MoRSE-signal 09/ weighted by atomic masses.	3D-MoRSE descriptors.	-۰/۰۹۸
Mor16e	3D-MoRSE-signal 16/ weighted by atomic Sanderson electronegativities.	3D-MoRSE descriptors.	-۰/۲۷۷

## ۲-۶. شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی مجموعه‌ای از نورون‌ها هستند که از الگوریتم‌های منحصر به فردی پیروی می‌کنند. این مجموعه که از مغز انسان الگوبرداری و الهام گرفته شده است، با هدف شناسایی الگوها طراحی می‌شوند و مورد استفاده قرار می‌گیرند. به طور کلی می‌توان گفت که شبکه عصبی شامل الگوریتم‌هایی است برای یادگیری ماشین، که منجر به طبقه‌بندی کردن داده‌های ورودی و ارائه

خروجی مطلوب می‌گردد. به همین دلیل است که می‌توان شبکه‌های عصبی را به عنوان جزئی از فرایند یادگیری ماشین در نظر گرفت.

شبکه‌های عصبی مصنوعی، سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری، نمایش دانش و در انتها اعمال دانش به دست آمده در جهت پیش‌بینی پاسخ‌های خروجی از سامانه‌های پیچیده هستند. ایده اصلی این گونه شبکه‌ها تا حدودی الهام گرفته از شیوه کارکرد سیستم عصبی زیستی برای پردازش داده‌ها 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 می‌کند. یک شبکه عصبی مصنوعی، از سه لایه ورودی، خروجی و پنهان تشکیل می‌شود. هر لایه شامل گروهی از سلول‌های عصبی (نورون) است که عموماً با کلیه نورون‌های لایه‌های دیگر در ارتباط هستند، مگر این که کاربر ارتباط بین نورون‌ها را محدود کند؛ ولی نورون‌های هر لایه با سایر نورون‌های همان لایه، ارتباطی ندارند. با استفاده از دانش برنامه‌نویسی رایانه می‌توان ساختار داده‌ای طراحی کرد که همانند یک نورون عمل نماید. سپس با ایجاد شبکه‌ای از این نورون‌های مصنوعی به هم پیوسته، ایجاد یک الگوریتم آموزشی برای شبکه و با اعمال این الگوریتم به شبکه، آن را آموزش داد. شکل ۱ نمایی از یک شبکه عصبی مصنوعی را نشان می‌دهد [۲۰-۱۷].



شکل ۱. نمایی از یک شبکه عصبی مصنوعی

یکی از پایه‌ای‌ترین مدل‌های عصبی موجود، مدل پرسپترون چند لایه<sup>۱</sup> است که عملکرد انتقالی مغز انسان را شبیه‌سازی می‌کند. در این نوع شبکه عصبی، بیشتر رفتار شبکه‌ای مغز انسان و انتشار سیگنال در آن مد نظر بوده است و از این رو، گهگاه با نام شبکه‌های پیش‌خورد<sup>۲</sup> نیز خوانده می‌شوند. هر یک از سلول‌های عصبی مغز انسان، موسوم به نورون<sup>۳</sup>، پس از دریافت ورودی (از یک سلول

<sup>1</sup> Multi-Layer perceptron (MLP)

<sup>2</sup> Feedforward Networks

<sup>3</sup> Neuron

عصبی یا غیر عصبی دیگر)، پردازشی روی آن انجام می‌دهند و نتیجه را به یک سلول دیگر (عصبی یا غیر عصبی) انتقال می‌دهند. این رفتار تا حصول نتیجه‌ای مشخص ادامه دارد، که احتمالاً در نهایت منجر به یک تصمیم، پردازش، تفکر و یا حرکت خواهد شد.

### ۳. نتایج و بحث

#### ۳-۱. مدل‌سازی به روش رگرسیون خطی چندگانه (MLR)

پس از انتخاب مناسب‌ترین توصیف‌کننده‌ها توسط روش مرحله‌ای، مرحله بعدی، ایجاد مدل میان توصیف‌کننده‌های انتخاب شده و فعالیت‌های ضد میکروبی ترکیبات می‌باشد. بین توصیف‌کننده‌ها و فعالیت‌های ضد میکروبی مشتقات آنیلیدها برای سری آموزش با استفاده از روش MLR رابطه زیر به عنوان مدل خطی بدست آمد:

$$pMIC = 2.067 - 0.541 (GATS4v) + 0.021 (RDF085m) + 0.022 (RDF060e) - 0.067 (Mor09m) + 0.162 (Mor16e)$$

سپس از معادله بدست آمده برای پیش‌بینی فعالیت ضد میکروبی سری تست (پیش‌بینی) استفاده گردید.

مقادیر واقعی و پیش‌بینی شده فعالیت‌های ضد میکروبی (MIC) برای کلیه ترکیبات مجموعه آموزش و تست در جدول (۲) آورده شده است.

جدول ۲. مقادیر تجربی و محاسبه شده pMIC ترکیبات مختلف برای مجموعه‌های آموزشی و پیش‌بینی در مدل‌های SW-MLR, SW-SVM, SW-ANN

No.	X	R1	R2	R3	Exp.	SW-MLR	SW-ANN
1	C <sub>6</sub> H <sub>5</sub>	H	H	H	1.45	1.45	1.45
2	C <sub>6</sub> H <sub>5</sub>	Cl	H	H	1.32	1.36	1.32
3	C <sub>6</sub> H <sub>5</sub>	H	Cl	H	1.39	1.40	1.31
4	C <sub>6</sub> H <sub>5</sub>	H	H	Cl	1.40	1.41	1.39
5	C <sub>6</sub> H <sub>5</sub>	H	CH <sub>3</sub>	H	1.42	1.44	1.41
6*	C <sub>6</sub> H <sub>5</sub>	H	H	CH <sub>3</sub>	1.40	1.39	1.44
7	C <sub>6</sub> H <sub>5</sub>	OCH <sub>3</sub>	H	H	1.51	1.47	1.51
8	C <sub>6</sub> H <sub>5</sub>	H	H	OCH <sub>3</sub>	1.42	1.40	1.44
9	C <sub>6</sub> H <sub>5</sub>	H	NO <sub>2</sub>	H	1.44	1.49	1.44
10	C <sub>6</sub> H <sub>5</sub>	H	H	NO <sub>2</sub>	1.54	1.47	1.50
11*	C <sub>6</sub> H <sub>5</sub>	H	H	Br	1.34	1.47	1.32

12	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	H	1.50	1.51	1.51
13	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	Cl	H	H	1.50	1.48	1.49
14	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	Cl	H	1.54	1.47	1.50
15	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	Cl	1.50	1.52	1.51
16	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	CH <sub>3</sub>	H	1.50	1.49	1.48
17*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	CH <sub>3</sub>	1.56	1.42	1.45
18	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	OCH <sub>3</sub>	H	H	1.44	1.42	1.43
19*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	OCH <sub>3</sub>	1.44	1.50	1.41
20	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	NO <sub>2</sub>	H	1.60	1.63	1.60
21	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	NO <sub>2</sub>	1.60	1.57	1.60
22	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	Br	1.48	1.50	1.51
23*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	H	1.48	1.50	1.49
24	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	Cl	H	H	1.51	1.47	1.49
25	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	Cl	H	1.43	1.45	1.42
26	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	Cl	1.49	1.53	1.55
27	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	CH <sub>3</sub>	H	1.46	1.48	1.47
28	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	CH <sub>3</sub>	1.46	1.46	1.46
29*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	OCH <sub>3</sub>	H	H	1.50	1.45	1.45
30	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	OCH <sub>3</sub>	150	1.52	1.51
31	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	NO <sub>2</sub>	H	160	1.62	1.60
32	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	NO <sub>2</sub>	1.45	1.50	1.45
33	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>11</sub> CH <sub>2</sub>	H	H	Br	1.49	1.52	1.52
34*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	H	1.51	1.51	1.52
35*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	Cl	H	H	1.60	1.51	1.34
36	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	Cl	H	1.50	1.50	1.48
37	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	Cl	1.64	1.57	1.59
38	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	CH <sub>3</sub>	H	1.48	1.52	1.46
39	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	CH <sub>3</sub>	1.48	1.48	1.52
40	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	OCH <sub>3</sub>	H	H	1.55	1.57	1.54
41*	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	OCH <sub>3</sub>	1.49	1.55	1.52
42	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	NO <sub>2</sub>	H	1.70	1.68	1.60
43	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	NO <sub>2</sub>	1.60	1.62	1.57
44	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>15</sub> CH <sub>2</sub>	H	H	Br	1.60	1.54	1.52

## ۲-۳. مدل سازی و پیش بینی توسط شبکه های عصبی مصنوعی

در قسمت دوم این کار، برای حصول نتایج بهتر، توصیف کننده های انتخاب شده توسط روش مرحله ایی، به شبکه عصبی مصنوعی وارد شدند. پردازش داده ها در محیط ویندوز ۱۰ و با استفاده از نرم افزار MATLAB انجام شد. یک شبکه سه لایه با تابع انتقال سیگموئیدی برای نورون ها طراحی شد. مقادیر اولیه وزن ها 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. بهینه سازی و بهنگام کردن وزنها و بایاس ها بوسیله الگوریتم BP<sup>۱</sup> انجام شده است. مجموعه داده ها به سه گروه تقسیم شده است: مجموعه آموزش، مجموعه ارزیابی و مجموعه تست. مجموعه آموزش (۵۰٪ داده ها) جهت آموزش دادن شبکه عصبی مصنوعی، مجموعه ارزیابی (۳۰٪ داده ها) برای ارزیابی مدل در طی آموزش دادن شبکه و ایجاد مدل مناسب و مجموعه پیش بینی (۲۰٪ داده ها) برای تست مدل ایجاد شده به کار رفت.

تعداد نورون ها در لایه ورودی با تعداد توصیف کننده های وارد شده به شبکه های عصبی مصنوعی برابر است. به ازای هر تعداد توصیف کننده وارد شده به شبکه عصبی، تعداد نورون ها در لایه مخفی بهینه شدند. بدین ترتیب که به ازای هر مدل ANN، تعداد نورون ها در لایه مخفی از ۱ تا ۱۰ تغییر داده شده و مقادیر RMSE برای مجموعه های آموزشی و پیش بینی محاسبه گردید. از رسم مقادیر RMSE بر حسب تعداد نورون ها در لایه مخفی، تعداد نورون های لایه مخفی بهینه شد. سپس بقیه پارامترها از جمله وزنها و بایاس ها، سرعت یادگیری و مومنتوم نیز بهینه گردید. جدول ۳ مشخصات شبکه عصبی مصنوعی بهینه شده را نشان می دهد.

جدول ۳. ساختار و مشخصات ANN تولید شده

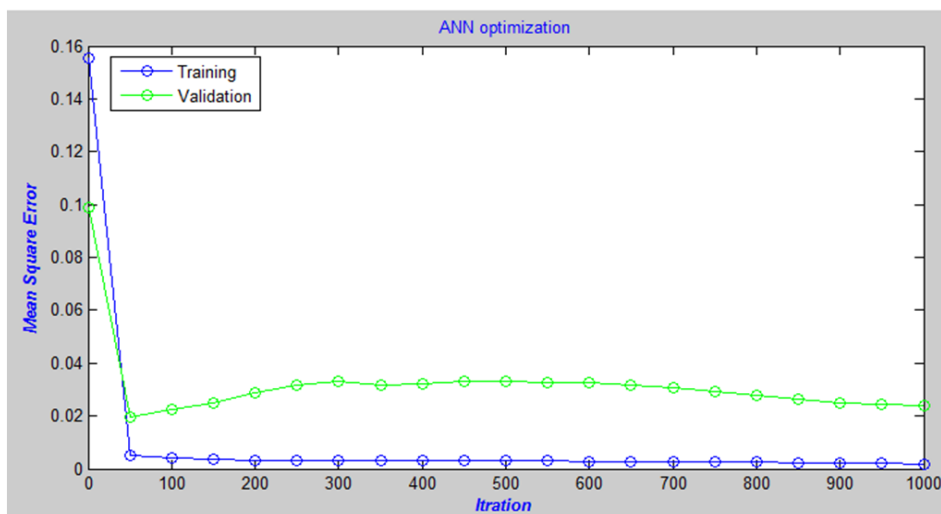
تعداد نورون ها در لایه ورودی	۵
تعداد نورون ها در لایه مخفی	۵
تعداد نورون ها در لایه خروجی	۱
سرعت آموزش	۰/۴
مومنتوم	۰/۳
تعداد چرخه های آموزشی	۱۰۰۰
تابع انتقال	Sigmoid

برای جلوگیری از Overfitting در طول آموزش مقادیر RMSE بعد از هر ۵۰ بار چرخه آموزشی، محاسبه و ثبت گردید. شکل ۱ نمودار میزان خطا بر حسب تعداد دورها برای این داده ها را نشان می دهد. همانطور که ملاحظه می شود میزان خطا

<sup>1</sup>Back-propagation

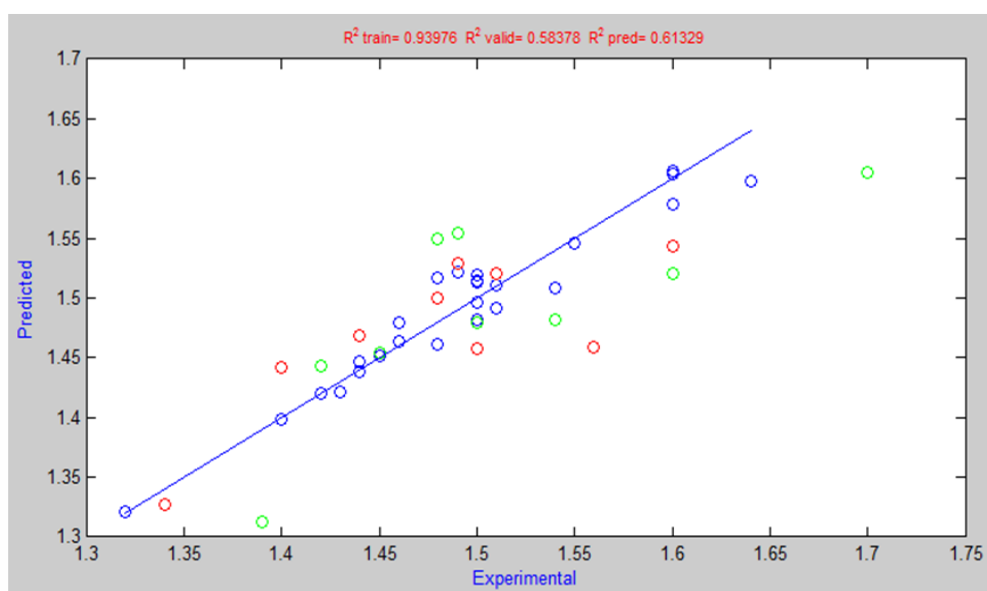


برای سری آموزش همواره در حال کاهش است اما برای سری ارزیابی در ۵۰ چرخه آموزش، کمترین خطا مشاهده میشود و بعد از آن افزایش می‌یابد. بنابراین این مقدار به عنوان مقدار بهینه تعداد چرخه‌های آموزش انتخاب شد.



شکل ۱. مقادیر RMSE برای مجموعه‌های آموزشی و ارزیابی برحسب تعداد چرخه‌های آموزش

با استفاده از مدل ANN بهینه شده مقادیر فعالیت‌های ضد میکروبی ترکیبات مورد نظر در مجموعه آموزش، ارزیابی و پیش‌بینی (تست) مورد محاسبه قرار گرفت و در جدول (۲) نشان داده شده است. در شکل (۲) مقادیر پیش‌بینی شده pMIC ترکیبات برحسب مقادیر تجربی در مجموعه‌های مختلف (آموزش، ارزیابی و تست) رسم شده‌اند. نزدیکی نتایج به خط راست، قدرت پیش‌بینی مدل را نشان می‌دهد.



شکل ۲. مقادیر pMIC محاسبه شده برای مشتقات آنیلیدها براساس مدل ANN در مجموعه‌های آموزش (رنگ آبی)، ارزیابی (رنگ سبز) و تست (رنگ قرمز) برحسب مقادیر تجربی

### 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 ۴. پارامترهای آماری برای مدل‌های انتخاب شده

	Training set			Test set		
	R <sup>2</sup>	RMSE	F	R <sup>2</sup>	RMSE	F
MLR	0.808	0.033	24.420	0.073	0.077	0.230
ANN	0.940	0.017	299.35	0.613	0.021	0.952

### ۴. نتیجه گیری

در این تحقیق، مطالعه ارتباط کمی ساختار- فعالیت روی مشتقات آنیلید که به عنوان ترکیبات ضد میکروبی استفاده می‌شوند انجام شد. در ابتدا توصیف‌کننده‌های ساختاری ایجاد و مناسب‌ترین آن‌ها که بیشترین ارتباط را با فعالیت دارویی مشتقات مورد نظر داشتند به کمک رگرسیون مرحله‌ای انتخاب گردید و به روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) مدل مناسب ایجاد شد. مدل‌های ایجاد شده به روش ANN قابلیت پیش‌بینی بهتری ارائه نمود. مقدار بالای R<sup>2</sup> همچنین مقادیر پایین خطای پیش‌بینی نشان دهنده کارایی مطلوب مدل ANN در پیش‌بینی فعالیت ضد میکروبی ترکیبات آنیلیدها می‌باشد. همچنین در این کار توصیف‌کننده‌های GATS4v, RDF085m, RDF060e, Mor09m, Mor16e به عنوان توصیف‌کننده‌های موثر در قدرت بازدارندگی ترکیبات آنیلیدها آشکار گردید. با دانستن مقادیر این توصیف‌کننده‌ها برای هر ترکیب مجهول دیگر و با استفاده از مدل بدست آمده می‌توان مقادیر فعالیت ضد میکروبی آن ترکیب را بدست آورد. همچنین برای سنتز ترکیباتی با این خصوصیات باید به اهمیت این توصیف‌کننده‌ها توجه کرد. بنابراین از این روش‌ها می‌توان برای پیش‌بینی فعالیت ضد میکروبی ترکیبات مشابه استفاده نمود. همچنین نوع توصیف‌کننده‌ها نشان می‌دهد که ساختار توپولوژیکی و حجم مولکول، جرم مولکولی و الکترونگاتیوی، فاکتورهایی هستند که روی فعالیت این ترکیبات تاثیر می‌گذارند بنابراین با بهینه کردن این پارامترها می‌توان خصلت ضد میکروبی این ترکیبات را افزایش داد.

## ۵. مراجع

- [1] Kim, Y. C., Ji, X. D., Melman, N., Linden, J., & Jacobson, K. A. (2000). Anilide derivatives of an 8-phenylxanthine carboxylic congener are highly potent and selective antagonists at human A2B adenosine receptors. *Journal of medicinal chemistry*, 43(6), 1165-1172.
- [2] Kitagawa, O., Yoshikawa, M., Tanabe, H., Morita, T., Takahashi, M., Dobashi, Y., & Taguchi, T. (2006). Highly enantioselective synthesis of atropisomeric anilide derivatives through catalytic asymmetric N-arylation: conformational analysis and application to asymmetric enolate chemistry. *Journal of the American Chemical Society*, 128(39), 12923-12931.
- [3] Legoabe, L., Kruger, J., Petzer, A., Bergh, J. J., & Petzer, J. P. (2011). Monoamine oxidase inhibition by selected anilide derivatives. *European journal of medicinal chemistry*, 46(10), 5162-5174.
- [4] Ma, Y., Yang, X., Han, H., Wen, Z., Yang, M., Zhang, Y., ... & Yang, Y. (2021). Design, synthesis and biological evaluation of anilide (dicarboxylic acid) shikonin esters as antitumor agents through targeting PI3K/Akt/mTOR signaling pathway. *Bioorganic Chemistry*, 111, 104872.
- [5] Yuan, C., Zhu, L., Chen, C., Chen, X., Yang, Y., Lan, Y., & Zhao, Y. (2018). Ruthenium (II)-enabled para-selective C-H difluoromethylation of anilides and their derivatives. *Nature communications*, 9(1), 1-10.
- [6] Li, M., Yu, H., Wang, Y., Li, J., Ma, G., & Wei, X. (2020). QSPR models for predicting the adsorption capacity for microplastics of polyethylene, polypropylene and polystyrene. *Scientific reports*, 10(1), 1-11.
- [7] Sepehri, B. (2020). A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view. *Journal of Molecular Liquids*, 297, 112013.
- [8] Borhani, T. N., García-Muñoz, S., Luciani, C. V., Galindo, A., & Adjiman, C. S. (2019). Hybrid QSPR models for the prediction of the free energy of solvation of organic solute/solvent pairs. *Physical Chemistry Chemical Physics*, 21(25), 13706-13720.
- [9] Zhu, T., Gu, L., Chen, M., & Sun, F. (2021). Exploring QSPR models for predicting PUF-air partition coefficients of organic compounds with linear and nonlinear approaches. *Chemosphere*, 266, 128962.
- [10] Yan, F., Shi, Y., Wang, Y., Jia, Q., Wang, Q., & Xia, S. (2020). QSPR models for the properties of ionic liquids at variable temperatures based on norm descriptors. *Chemical Engineering Science*, 217, 115540.
- [11] Haiying, Y., Meiping, L., & Junsheng, H. (2020). A Comparative QSPR Study on Photolysis Half-lives of Polychlorinated Dibenzofurans by MLP-ANN and SVM. *Asian Journal of Ecotoxicology*, (4), 240-247.
- [12] Moussaoui, M., Laidi, M., Hanini, S., Abdallah, A. E. H., & Hentabli, M. (2021). Critical Properties and Acentric Factors of Pure Compounds Modelling Based on QSPR-SVM with Dragonfly Algorithm. *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske*, 70(7-8), 375-386.
- [13] Zarei, K., Atabati, M. and Ebrahimi, M., 2007. Quantitative structure-property relationship study of the solvent polarity using wavelet neural networks. *Analytical Sciences*, 23(8), pp.937-942.

- [14] Quang, N. M., Mau, T. X., Nhung, N. T. A., An, T. N. M., & Van Tat, P. (2019). Novel QSPR modeling of stability constants of metal-thiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN. *Journal of Molecular Structure*, 1195, 95-109.
- [15] Maouz, H., Khaouane, L., Hanini, S., Ammi, Y., Hamadache, M., & Laidi, M. (2020). QSPR studies of carbonyl, hydroxyl, polyene indices, and viscosity average molecular weight of polymers under photostabilization using ANN and MLR approaches. *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske*, 69(1-2), 1-16.
- [16] Cleiton A. Nunes, Matheus P. Freitas., (2013). aug-MIA-QSAR modeling of antimicrobial activities and design of multi-target anilide derivatives. *Journal of Microbiological Methods*, 94, 217–220.

## Application of multivariate linear regression and artificial neural networks to predict the antimicrobial activity of some anilide derivatives by quantitative structure-activity relationship (QSAR) method

Mehdi Nekoei<sup>\*1</sup>, Syed Hamed Mosavi<sup>2</sup>, Parastoo Pourali<sup>2</sup>

<sup>1</sup>Department of Chemistry, Shahrood Branch, Islamic Azad University, Shahrood, Iran

<sup>2</sup>Department of Biology, Shahrood Branch, Islamic Azad University, Shahrood, Iran

Submitted: 16 September 2021, Revised: 15 December 2021, Accepted: 26 December 2021

### Abstract

Quantitative structure-activity relationship (QSAR) study was performed to predict the antimicrobial activity of some anilide derivatives using multivariate linear regression (MLR) and artificial neural networks (ANN). First, the structure of the compounds, the drawing and the appropriate group of descriptors were calculated. Then, the step selection method was used to obtain the best descriptors that were most associated with the antimicrobial activity of the compounds. With this method, 5 descriptors were selected and first the linear MLR model was constructed. Then, artificial neural network was used to obtain better results. The values of coefficient of determination ( $R^2$ ) and root mean square error (RMSE) for the test series were 0.07 and 0.073 for the MLR linear model and 0.613 and 0.021 for the nonlinear ANN model, respectively. Statistical data show the superiority of ANN method over MLR method.

**Keywords:** *Quantitative structure-activity relationship, aniline 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