

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده شیمی

گروه شیمی تجزیه

پایان نامه کارشناسی ارشد

پیش‌بینی اثرات بازداري بعضی ترکیبات دارویی با روش‌های QSAR

دانشجو:

یلدا قربان ابراهیمی

استاد راهنما:

دکتر ناصر گودرزی

استاد مشاور:

دکتر منصور عرب چم‌جنگلی

شهریور ماه ۱۳۹۱

تقدیم به:

مادره، که دستانش شکوفه پر شهد بهاریست

برای این زنبور کوچک.

و

برادر عزیزه

که قلبه شوق تپیدنش را

در وجود نازنینش یافته است

مقالات مستخرج از این پایان نامه:

۱- مطالعه رابطه کمی ساختار فعالیت (QSAR) بر روی فعالیت ضد روان‌پریشی برخی

از ترکیبات با استفاده از روش‌های خطی و غیر خطی، پانزدهمین سمینار شیمی ایران،

دانشگاه بوعلی سینا همدان، شهریور ۱۳۹۰.

۲- کاربرد الگوریتم ژنتیک- شبکه عصبی مصنوعی (GA-ANN) برای پیش‌بینی اثر بازدارندگی

مشتقات فنیل آلکیل آمین به عنوان داروهای ضد روان‌پریشی، پنجمین سمینار و کارگاه

ریاضی شیمی، دانشگاه پیام نور یزد، بهمن ۱۳۹۰.

چکیده

در بخش اول این پروژه، مطالعات کمی ساختار- فعالیت (QSAR) بر روی اثر بازداری (pIC50) ۴۰ ترکیب دارویی از مشتقات فنیل آلکیل آمین که به عنوان ترکیبات ضد روان‌پریشی عمل می‌کنند، انجام شد. برای انتخاب مهمترین توصیف‌گرها از روش انتخاب متغیر مرحله‌ای و الگوریتم ژنتیک استفاده شد. همچنین برای ایجاد مدل جهت پیش‌بینی اثر بازداری روش رگرسیون خطی چند گانه (MLR) به عنوان یک روش خطی، شبکه عصبی مصنوعی (ANN) به عنوان یک روش غیر خطی به کار برده شد. اعتبار این مدل‌ها توسط سری تست و تکنیک‌های حذف مرحله‌ای تک تک (LOO) و -Y تصادفی بررسی شد. نتایج نشان داد که ضریب تعیین (R^2) برای پیش‌بینی ثابت بازداری ترکیبات سری تست با مدل‌های SR-MLR، SR-ANN و GA-ANN به ترتیب برابر ۰/۹۱۸ ، ۰/۹۵۱ و ۰/۸۳۵ است.

در بخش دوم این پروژه برای انتخاب مهمترین توصیف‌گرها از روش انتخاب متغیر مرحله‌ای و الگوریتم ژنتیک برای مطالعه ارتباط کمی ساختار- خاصیت (QSPR) شاخص بازداری (RI) بعضی از ترکیبات آلی استفاده شد. همچنین برای ایجاد مدل جهت تعیین شاخص بازداری، روش رگرسیون خطی چند گانه (MLR) به عنوان یک روش خطی و شبکه عصبی مصنوعی (ANN) به عنوان یک روش غیر خطی به کار برده شد. مدل‌های بدست آمده به وسیله تکنیک توقف زود هنگام توسط سری تست و تکنیک‌های حذف مرحله‌ای گروهی (LGO) و -Y تصادفی اعتبار سنجی شدند، به طوری که مدل‌های بدست آمده توسط مدل‌های SR-MLR، SR-ANN و GA-ANN برای پیش‌بینی زمان بازداری ترکیبات سری تست که در مراحل ایجاد مدل شرکت نداشتند، به کار برده شد. ضرایب تعیین (R^2) برای آن دسته از ترکیبات به ترتیب برابر با ۰/۹۵۹ ، ۰/۹۹۹ و ۰/۹۹۹ بدست آمد.

کلمات کلیدی= ارتباط کمی ساختار-فعالیت، ارتباط کمی ساختار- خاصیت، اثر بازداری، شاخص بازداری، رگرسیون خطی چند گانه، شبکه عصبی مصنوعی، الگوریتم ژنتیک

فهرست مطالب

فصل اول

- ۱-۱- روان پریشی..... ۲
- ۲-۱- ضرورت تحقیق..... ۳
- ۳-۱- مروری بر کارهای انجام شده..... ۴

فصل دوم

- ۱-۲- کموتریکس..... ۷
- ۱-۱-۲- ارتباط کمی ساختار- فعالیت..... ۷
- ۲-۲-۲- ارتباط کمی ساختار- خاصیت..... ۸
- ۲-۲- رگرسیون خطی چندگانه..... ۸
- ۳-۲- شبکه عصبی..... ۹
- ۱-۳-۲- مدل ریاضی نرون مصنوعی..... ۱۰
- ۲-۳-۲- انواع شبکه‌های عصبی از نظر برگشت پذیری..... ۱۲
- ۳-۳-۲- مراحل آموزش در شبکه‌های پیشخور..... ۱۳
- ۴-۳-۲- بهبود عمومیت شبکه..... ۱۴
- ۴-۲- الگوریتم ژنتیک..... ۱۵
- ۱-۴-۲- مبانی الگوریتم ژنتیک..... ۱۶
- ۲-۴-۲- رمزبندی کروموزوم..... ۱۷
- ۳-۴-۲- عملگرهای GA..... ۱۸
- ۴-۴-۲- پیوند..... ۱۹
- ۵-۴-۲- جهش..... ۲۱
- ۶-۴-۲- پارامترهای GA..... ۲۱
- ۷-۴-۲- اجرای الگوریتم ژنتیک..... ۲۳
- ۸-۴-۲- الگوریتم ژنتیک و QSAR..... ۲۴
- ۵-۲- نرم افزارهای مورد استفاده..... ۲۴

فصل سوم

- ۱-۳- مدل سازی فعالیت دارویی مشتقات آریل الکوکسی فنیل آلکیل آمین به عنوان بازدارنده‌های روان پریشی..... ۲۸
- ۱-۱- انتخاب سری داده‌ها..... ۲۸
- ۲-۱-۳- رسم و بهینه سازی ساختار مولکول‌ها..... ۳۱
- ۳-۱-۳- محاسبه توصیفگرها..... ۳۱
- ۴-۱-۳- انتخاب توصیفگرهای مناسب..... ۳۲
- ۵-۱-۳- مدل سازی توسط روش رگرسیون خطی چندگانه..... ۳۴
- ۶-۱-۳- مدل سازی به روش شبکه عصبی مصنوعی..... ۳۵
- ۱-۶-۱-۳- انتخاب تعداد لایه‌های پنهان..... ۳۶
- ۱-۶-۲- انتخاب تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد نرون‌های لایه پنهان..... ۳۶

۳-۶-۱-۳	انتخاب تعداد دورهای آموزش.....	۴۲
۳-۶-۱-۴	بهینه کردن مقدار μ	۴۴
۳-۶-۱-۵	معماری شبکه عصبی بهینه شده.....	۴۵
۳-۷-۱-۳	انتخاب متغیر به روش الگوریتم ژنتیک.....	۴۶
۳-۸-۱-۸	مدل شبکه عصبی مصنوعی با استفاده از توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک.....	۴۹
۳-۸-۱-۱	انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع آموزش و نوع تابع انتقال.....	۵۰
۳-۸-۱-۲	انتخاب تعداد دورهای آموزش.....	۵۵
۳-۸-۱-۳	بهینه کردن مقدار μ	۵۶
۳-۸-۱-۴	معماری شبکه عصبی بهینه شده.....	۵۸
۳-۹-۱-۹	بررسی نتایج.....	۵۹
۳-۹-۱-۱	ارزیابی مدل‌ها با استفاده از سری ارزیابی.....	۵۹
۳-۹-۱-۲	ارزیابی مدل‌ها با استفاده از سری تست.....	۵۹
۳-۹-۱-۳	ارزیابی مدل‌های SR-MLR و SR-ANN و GA-ANN توسط روش رد مرحله‌ای تک تک.....	۶۳
۳-۹-۱-۴	ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری.....	۶۶
۳-۹-۱-۵	ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی.....	۶۷
۳-۱۰-۱-۱۰	بررسی ارتباط توصیفگرهای وارد شده در مدل با اثر بازدارندگی.....	۶۸
۳-۱۰-۱-۱۱	بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی.....	۷۱
۳-۱۱-۱-۱۱	نتیجه‌گیری نهایی.....	۷۲
۳-۱۱-۱-۱۲	آینده نگری.....	۷۳

فصل چهارم

۱-۴

مقدمه.....	75
۴-۱-۱-۱- پیشینه تحقیق.....	۷۶
۴-۲-۱-۱- انتخاب سری داده‌ها.....	۷۷
۴-۲-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها.....	۸۰
۴-۲-۱-۳- محاسبه توصیفگرها.....	۸۱
۴-۲-۱-۴- انتخاب توصیفگرهای مناسب.....	۸۱
۴-۲-۱-۵- مدل‌سازی توسط روش رگرسیون خطی چندگانه.....	۸۳
۴-۲-۱-۶- مدل‌سازی به روش شبکه عصبی مصنوعی.....	۸۴
۴-۲-۱-۷- انتخاب تعداد لایه‌های پنهان.....	۸۴
۴-۲-۱-۸- انتخاب تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد نرون‌های لایه پنهان.....	۸۵
۴-۲-۱-۹- انتخاب تعداد دورهای آموزش.....	۹۰
۴-۲-۱-۱۰- بهینه کردن مقدار μ	۹۰
۴-۲-۱-۱۱- معماری شبکه عصبی بهینه شده.....	۹۱
۴-۲-۱-۱۲- انتخاب متغیر به روش الگوریتم ژنتیک.....	۹۲
۴-۲-۱-۱۳- مدل شبکه عصبی مصنوعی با استفاده از توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک.....	۹۳
۴-۲-۱-۱۴- انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع آموزش و نوع تابع انتقال.....	۹۳

۹۹.....	انتخاب تعداد دوره‌های آموزش.....۲-۷-۲-۴
۹۹.....	بهینه کردن مقدار μ۳-۷-۲-۴
۱۰۰.....	معماری شبکه عصبی بهینه شده.....۴-۷-۲-۴
۱۰۱.....	بررسی نتایج.....۸-۲-۴
۱۰۱.....	ارزیابی مدل‌ها با استفاده از سری ارزیابی.....۱-۸-۲-۴
۱۰۵.....	ارزیابی مدل‌ها با استفاده از سری تست.....۲-۸-۲-۴
۱۰۸.....	ارزیابی مدل‌های SR-MLR و SR-ANN و GA-ANN توسط روش رد مرحله‌ای گروه.....۳-۸-۲-۴
۱۲۰.....	ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری.....۴-۸-۲-۴
۱۲۲.....	ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی.....۵-۸-۲-۴
۱۲۳.....	بررسی ارتباط توصیفگرهای وارد شده در مدل با اثر بازدارندگی.....۶-۸-۲-۴
۱۲۶.....	بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی.....۷-۸-۲-۴
۱۲۷.....	نتیجه‌گیری نهایی.....۹-۲-۴
۱۲۸.....	آینده نگری.....
۱۲۹.....	منابع.....

فهرست اشکال

- شکل (۱-۲) - مدل تک ورودی..... ۱۰
- شکل (۲-۲) - انواع توابع انتقال..... ۱۲
- شکل (۳-۲) - کروموزوم‌هایی با رمزبندی دودویی..... ۱۷
- شکل (۴-۲) - کروموزوم‌هایی با رمزبندی تبدیل..... ۱۸
- شکل (۵-۲) - تولید مثل بر اساس تقاطع یک نقطه‌ای..... ۱۹
- شکل (۶-۲) - تقاطع دو نقطه‌ای..... ۲۰
- شکل (۷-۲) - پیوند یکنواخت..... ۲۱
- شکل (۸-۲) - نحوه عملکرد عملگر جهش در سیستم دودویی..... ۲۱
- شکل (۹-۲) - فلوچارت یک الگوریتم ژنتیکی ساده..... ۲۴
- شکل (۱-۳) - ساختار آریل الکوکی فنیل آلکیل آمین..... ۲۸
- شکل (۲-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۳۸
- شکل (۳-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۳۹
- شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی..... ۴۰
- شکل (۵-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی..... ۴۱
- شکل (۶-۳) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دوره‌های آموزش متفاوت..... ۴۴
- شکل (۷-۳) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ ۴۴
- شکل (۸-۳) - ساختار شبکه عصبی بهینه شده..... ۴۶
- شکل (9-3) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۵۱
- شکل (۱۰-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۵۲
- شکل (۱۱-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزشی بایزین و تابع انتقال لگاریتم سیگموئیدی..... ۵۳
- شکل (۱۲-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزشی بایزین و تابع انتقال تانژانت سیگموئیدی..... ۵۴
- شکل (۱۳-۳) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش متفاوت..... ۵۶
- شکل (۱۴-۳) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد μ متفاوت..... ۵۷
- شکل (۱۵-۳) - ساختار شبکه عصبی بهینه شده..... ۵۸
- شکل (۱۶-۳) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری ارزیابی، الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۶۱
- شکل (۱۷-۳) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای تست الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۶۲
- شکل (۱۸-۳) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۶۵
- شکل (۱۹-۳) - نمودار باقی‌مانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۶۶
- شکل (۲۰-۳) - مشارکت توصیفگرها در شبکه عصبی بهینه..... ۷۲

شکل (۴-۱) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۸۶

شکل (۴-۲) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۸۷

شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی..... ۸۸

شکل (۴-۴) - نمودار سه بعدی شبکه‌هایی با تابع آموزش بایزین و تابع انتقال تانژانت سیگموئیدی..... ۸۹

شکل (۴-۵) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دوره‌های آموزش متفاوت..... ۹۱

شکل (۴-۶) - ساختار شبکه عصبی بهینه شده..... ۹۲

شکل (۴-۷) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۹۵

شکل (۴-۸) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۹۶

شکل (۴-۹) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۹۷

شکل (۴-۱۰) - نمودار سه بعدی شبکه‌هایی با تابع آموزش بایزین و تابع انتقال تانژانت سیگموئیدی..... ۹۸

شکل (۴-۱۱) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دوره‌های آموزش متفاوت..... ۱۰۰

شکل (۴-۱۲) - ساختار شبکه عصبی بهینه شده..... ۱۰۰

شکل (۴-۱۳) - نمودار مقادیر پیش‌بینی شده شاخص بازدارنده بر حسب مقادیر تجربی برای سری ارزیابی، الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۱۰۴

شکل (۴-۱۴) - نمودار مقادیر پیش‌بینی شده شاخص بازدارنده بر حسب مقادیر تجربی برای سری تست الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۱۰۷

شکل (۴-۱۵) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی شاخص بازدارنده به روش رد مرحله‌ای گروهی برای کل داده‌ها الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۱۱۹

شکل (۴-۱۶) - نمودار باقیمانده‌ها بر حسب مقدار تجربی شاخص بازدارنده برای کل داده‌ها الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN..... ۱۲۱

شکل (۴-۱۷) - مشارکت توصیفگرها در شبکه عصبی بهینه..... ۱۲۷

فهرست جداول

جدول (۳-۱) - ساختار و مقادیر $IC_{50}(nM)$ ترکیبات مورد مطالعه.....	۲۹
جدول (۳-۲) - فهرست توصیفگرهای محاسبه شده توسط نرم افزار Dragon.....	۳۲
جدول (۳-۳) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله.....	۳۳
جدول (۳-۴) - ماتریس همبستگی کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله.....	۳۴
جدول (۳-۵) - مقایسه آماره های چند مدل نهایی به دست آمده از روش MLR.....	۳۵
جدول (۳-۶) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی های مختلف و تعداد گره های متفاوت در لایه پنهان.....	۳۸
جدول (۳-۷) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی های مختلف و تعداد گره های متفاوت در لایه پنهان.....	۳۹
جدول (۳-۸) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۴۰
جدول (۳-۹) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۴۱
جدول (۳-۱۰) - توصیفگرهای انتخاب شده توسط شبکه عصبی.....	۴۲
جدول (۳-۱۱) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره های آموزش مختلف.....	۴۳
جدول (۳-۱۲) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ	۴۵
جدول (۳-۱۳) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....	۴۶
جدول (۳-۱۴) - کل توصیفگرهای انتخاب شده الگوریتم ژنتیک.....	۴۹
جدول (۳-۱۵) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۵۱
جدول (۳-۱۶) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۵۲
جدول (۳-۱۷) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۵۳

جدول (۳-۱۸) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۵۴
جدول (۳-۱۹) - مقادیر میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش متفاوت.....	۵۵
جدول (۳-۲۰) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ	۵۷
جدول (۳-۲۱) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....	۵۸
جدول (۳-۲۲) - پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری ارزیابی.....	۶۰
جدول (۳-۲۳) - پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری تست.....	۶۰
جدول (۳-۲۴) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-MLR و SR-ANN و GA-ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها.....	۶۳
جدول (۳-۲۵) - پارامترهای آماری برای مدل‌های برتر طراحی شده SR-MLR, SR-ANN و GA-ANN.....	۶۷
جدول (۳-۲۶) - الف: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش SR-ANN.....	۶۸
جدول (۳-۲۶) - ب: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش روش GA-ANN.....	۶۸
جدول (۳-۲۷) - ارتباط ساختاری مولکول‌ها با توصیفگر Mv.....	۶۹
جدول (۴-۱) - ساختار و مقادیر شاخص بازدارندگی ترکیبات مورد مطالعه در ۴ دمای متفاوت.....	۷۸
جدول (۴-۲) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله و معانی آن‌ها.....	۸۲
جدول (۶-۳) - ماتریس همبستگی کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله.....	۸۲
جدول (۴-۴) - مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR.....	۸۳
جدول (۴-۵) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۸۶
جدول (۴-۶) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۸۷
جدول (۴-۷) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۸۸
جدول (۴-۸) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۸۹
جدول (۴-۹) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف.....	۹۰
جدول (۴-۱۰) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....	۹۱
جدول (۴-۱۱) - کل توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک.....	۹۳

جدول (۴-۱۲) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۹۵
جدول (۴-۱۳) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۹۶
جدول (۴-۱۴) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۹۷
جدول (۴-۱۵) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان.....	۹۸
جدول (۴-۱۶) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دورهای آموزش مختلف.....	۹۹
جدول (۶-۱۷) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی.....	۱۰۱
جدول (۴-۱۸) - پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری ارزیابی.....	۱۰۲
جدول (۴-۱۹) - پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری تست.....	۱۰۵
جدول (۴-۲۰) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-MLR و SR-ANN و GA-ANN با استفاده از رد مرحله‌ای گروهی برای کل داده‌ها.....	۱۰۹
جدول (۴-۲۱) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط SR-MLR و SR-ANN و GA-ANN.....	۱۲۰
جدول (۴-۲۲) - مقادیر R^2 برای سری تست با استفاده از آزمون Y-تصادفی.....	۱۲۲
جدول (۴-۲۳) - اثر توصیفگر RDSUM بر شاخص بازداری.....	۱۲۴

فصل اول

مقدمه

۱-۱- روان پریشی

روان پریشی^۱ در اصطلاح روان پزشکی به وضعیت روانی گفته می شود که در آن فرد رابطه خود را با حقیقت از دست بدهد. روان پریشی همچنین شامل وضعیت اختلال روانی، هذیان و اختلال در بینش هم می شود. در افرادی که از جنون رنج می برند ممکن است هذیان، توهم^۲ و یا تغییرات شخصیتی^۳ بروز کند. این عارضه بسته به شدت آن ممکن است با رفتار غیرعادی همراه باشد و عملکرد فرد را در تعاملات اجتماعی و انجام فعالیت های روزمره دچار اختلال کند. داروهای ضد روان پریشی با هدف حداقل کردن این علائم مصرف می شوند. طیف گسترده ای از بیماری های سیستم عصبی مرکزی، از سموم خارجی و یا بیماری های فیزیولوژیک، می تواند نشانه های روان پریشی را تولید کند. تحقیقات نشان می دهد که علائم روان پریشی با تغییرات مواد شیمیایی در مغز همراه است. داروهای ضد روان پریشی کمک می کنند تا این مواد شیمیایی به میزان نرمال خود برگردند و به این وسیله علائم را کاهش داده یا از بین می برند. البته نباید فراموش کرد که این داروها درمان کننده نیستند [۱].

اولین داروهای ضد روان پریشی در سال ۱۹۵۰ کشف شدند اما استفاده از آنها از سال ۱۹۷۰ بصورت گسترده شروع شد. این داروها تمایل دارند که گیرنده های مسیرهای دوپامین^۴ مغز را غیر فعال کنند. اما هر یک از این داروها طیف وسیعی از گیرنده های هدف^۵ را در بر می گیرند [2].

¹- Psychosis

²- hallucinations

³- personality changes

⁴-dopamine pathways

⁵- receptor targets

۱-۲- ضرورت تحقیق

همان‌طور که گفته شد روان‌پریشی واژه‌ای گسترده است و طیف وسیعی از صدمات و بیماری‌های سیستم عصبی مرکزی را در بر می‌گیرد، که عوارض و نتایج همه آن‌ها از بین بردن عملکرد صحیح ذهن و مختل کردن کارکرد فرد به عنوان یک انسان سالم در اجتماع است. درمان دارویی می‌تواند در بیشتر موارد فرد بیمار را به وضعیت ذهنی و جسمی یک انسان سالم نزدیک کند و با صرف زمان و هزینه کمی چنین افرادی را به جامعه برگرداند. سنتز بیشتر داروهای شیمیایی شناخته شده از راه آزمون و خطا به انجام شده که این امر روند یافتن ترکیبات جدید و ساخت داروهای موثرتر را کند می‌کند و همچنین باعث صرف هزینه‌ای گزاف می‌شود. توسعه روش‌های تئوری به عنوان یک ابزار کارآمد برای رفع این محدودیت‌ها مورد توجه قرار گرفته است.

از جمله روش‌های تئوری، رابطه کمی ساختار- فعالیت (QSAR)^۱ است که می‌توان در آن از روش‌های مختلفی برای ایجاد روابط خطی و غیر خطی میان ساختار و فعالیت دارویی ترکیباتی که قبلاً فعالیت دارویی آن‌ها به صورت تجربی اندازه‌گیری شده‌اند، استفاده نمود. روش روابط کمی بهینه بین ساختار و فعالیت دارویی ترکیبات مورد بررسی را بدست می‌آورد. سپس امکان پیش‌بینی فعالیت دارویی برای ساختارهای مشابه اندازه‌گیری شده و یا ساختارهایی که فعالیت آن‌ها هنوز به صورت تجربی اندازه‌گیری نشده است فراهم می‌شود. بر این اساس به راحتی می‌تواند برای کمک به بهینه سازی سنتز داروها مورد استفاده قرار گیرد.

^۱-Quantitative Structure- Activity Relationship

۱-۳- مروری بر کارهای انجام شده

در سال ۲۰۰۳ هون یونگ کوه^۱ و همکارانش مطالعات QSAR را بر روی پایپروزیل آلکیل ایزوزایلول^۲ و مشتقاتش انجام دادند و ارتباط بین ساختار و فعالیت را با استفاده از مدل بدست آمده بررسی کردند، ضریب تعیین گزارش شده ۹۱۷/۰/ تعیین گردیده است [۳].

در سال ۲۰۰۳ توماس بیل^۳ و همکارانش مطالعات QSAR را بر روی ۳۷ ترکیب آدرنوسپتور^۴ از مشتقات سرتیندل^۵ تحقیق کردند. برای این ترکیبات با استفاده از روش 3D-QSAR مدل سازی شد. ضریب تعیین بدست آمده با این مدل ۰/۹۵ گزارش شد. این مدل قادر به پیش بینی فعالیت مشتقات سنتزی این سری از ترکیبات خواهد بود [۴].

در سال ۲۰۰۵ فرناندز و کامبا^۶ بازدارنده های CDK2 را با استفاده از شبکه ی عصبی مصنوعی مدل سازی نمودند که در آن دقت مدل برای سری آموزش ۸۷٪ بود [۵].

در سال ۲۰۰۶ نیم پا^۷ و همکارانش مطالعات QSAR را بر روی ۲۷۷۲ ترکیب از مشتقات دپامین ها و سرتونین ها انجام دادند. با استفاده از تجزیه مشخص کننده خطی (LDA)^۸ مدل سازی نرم افزاری مستقل و قیاس طبقاتی (SIMCA)^۹، پارتیشن بندی بازگشتی (RP)^{۱۰} و شبکه عصبی مصنوعی (ANN) توانستند فعالیت این سری ترکیبات پیش بینی کنند. ضریب تعیین گزارش شده ۷۳/۰ است [۶].

¹-Hun Yeong Koh

²- Piperazinylalkylisoxazole

³- Thomas Balle

⁴-adrenoceptor

⁵-sertindole

⁶-Fernandez and Camba

⁷- Nim Pae

⁸-Linear discriminant analysis

⁹- Soft independent modeling of class analogy

¹⁰- Recursive partitioning

جلال هروی در سال ۲۰۰۹ مدل سازی QSAR ۱- (۳و۳دی فنیل پروپیل)- پایپیریدینیل آمیدها^۱ را که به عنوان اصلاح کننده CCR5 در هنگام حمله ویروس HIV عمل می کنند، با استفاده از شبکه عصبی انجام داد. در این تحقیق هروی از الگوریتم ژنتیک برای انتخاب متغیرهای ورودی شبکه استفاده کرد و برتری این روش را نسبت به روش MLR اعلام کرد [۷].

عرب چم جنگلی در سال ۲۰۰۹ با استفاده از روش غیرخطی بر مبنای شبکه عصبی، مدلی مناسب برای فعالیت ضد HIV مشتقات پیرویل آریل سولفون ها^۲ ارائه داد به طوری که ضریب همبستگی برای سری تست ۰/۹۸۹ بود [۸]. در همان سال مطالعه دیگر بر اساس شبکه عصبی توسط آلمریکو^۳ برای مطالعه ضد ویروسی تعدادی از آنالوگ های سایکلوویر^۴ انجام شد [۹].

در سال ۲۰۱۰ پاتل^۵ و همکارانش مطالعه QSAR را بر روی ۴۶ مشتق دی هیدرو- آلکوکسی- بنزیل- اکسوپیریمیدین^۶ انجام دادند. مدل نهایی به دست آمده شامل ۴ توصیفگر بود و با توجه به علامت ضرائب این توصیفگرها، نتیجه گرفتند که استخلاف های آگریز و استخلاف های با شاخه بزرگتر، تأثیر مثبتی بر بازداری آنزیم نسخه بردار معکوس دارند [۱۰].

در تحقیق حاضر، ارتباط کمی ساختار- فعالیت برای پیش بینی فعالیت دارویی یکسری از ترکیبات ضد روان پریشی به کار گرفته شد. بر اساس دانش ما و تحقیقات صورت گرفته، برای ترکیبات بررسی شده در این پایان نامه قبلاً مدل سازی صورت نگرفته است و این کار به عنوان تحقیقی نوین در ارتباط با این ترکیبات است. در این پایان نامه از رگرسیون خطی و شبکه عصبی برای مدل سازی استفاده شده است زیرا توانایی مناسبی در برقراری ارتباط خطی و غیرخطی بین ساختار- فعالیت و نیز پیش بینی موارد مشابه دارند.

¹-1-(3,3-diphenylpropyl)-piperidinyl amides

²-1-[5-chlorophenyl) sulfonyl]-1H--pyrrole

³ - Almerico

⁴- Acyclovir

⁵ - Patel

⁶- dihydro-alkoxy-benzyl-oxypyrimidines

فصل دوم

کمو متریکس

فعالیت آن‌ها هنوز اندازه‌گیری نشده است، وجود دارد. بر این اساس QSAR به راحتی می‌تواند برای کمک به بهینه سازی سنتز داروها مورد استفاده قرار گیرد [۱۲ و ۱۳].

۲-۱-۲- ارتباط کمی ساختار - خاصیت

هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده مولکول انجام گیرد به آن ارتباط کمی ساختار-خاصیت می‌گویند، که در واقع رابط بین خواص فیزیکی و شیمیایی و ساختار شیمیایی ترکیبات است.

تدبیر اساسی QSPR به دست آوردن رابطه کمی بهینه ساختار مولکولی برای پیش‌بینی خواص میباشد و هنگامی که رابطه معتبری به دست آمد امکان پیش بینی فاکتورهای بازدارنده برای ساختارهای مشابه با ترکیبات اندازه گیری شده یا ساختارهای دیگر که هنوز اندازه‌گیری نشده و یا تهیه نشده‌اند، وجود دارد [14].

۲-۲- رگرسیون خطی چندگانه (MLR)

رگرسیون به مفهوم ایجاد یک رابطه بین یک متغیر از یک طرف و یک یا چند متغیر از طرف دیگر می‌باشد. اگر تنها یک متغیر مستقل داشته باشیم، رگرسیون را ساده و در غیر این صورت، رگرسیون را چندگانه می‌گویند [۱۵]. در بسیاری موارد نمی‌توان تغییرات یک متغیر را فقط به مقادیر یک متغیر دیگر مربوط کرد. به عبارت دیگر برای پیش بینی مقادیر یک متغیر، دانستن مقادیر دو یا چند متغیر دیگر لازم است در این حالت از روش‌های رگرسیون چندگانه استفاده می‌شود [۱۶].

ساده‌ترین و متداول‌ترین مدلی که برای ارتباط چند متغیر به کار می‌رود، مدل خطی چندگانه است. این مدل با رابطه خطی زیر تعریف می‌شود:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots + \beta_p X_p \quad (۱-۲)$$

که در آن Y متغیر وابسته (فعالیت دارویی)، X_i متغیر مستقل (توصیفگر) و β_i ضرایب رگرسیون نامیده می‌شوند. باید توجه کرد که مقدار ضریب متغیر مستقل در معادله، بیانی از میزان اهمیت آن متغیر در پیش بینی مقادیر متغیر وابسته است.

۲-۳- شبکه عصبی

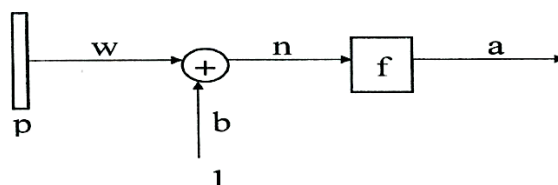
شبکه‌های عصبی چه در بعد آنالیز و توسعه ساختاری و چه در بعد پیاده سازی سخت‌افزاری، از نظر کمی، کیفی و توانایی، در حال رشد و پیشرفت می‌باشد و تکنیک‌های مختلف محاسبات عصبی از لحاظ تعداد همچنان در حال افزایش است. از مغز به عنوان یک سیستم پردازش اطلاعات با ساختار موازی و کاملاً پیچیده که دو درصد وزن بدن را تشکیل می‌دهد و بیش از بیست درصد کل اکسیژن بدن را مصرف می‌کند برای کلیه اعمال آگاهانه و بسیاری از رفتارهای ناخود آگاه استفاده می‌شود. مغز برای محاسبات خود، از ساختاری کاملاً مغایر با ساختاری کاملاً مغایر با ساختار کامپیوترهای متداول بر خوردار می‌باشد. تلاش برای فهم این موضوع خصوصاً از سال ۱۹۱۱ قوت گرفت، زمانی که برای نخستین بار شخصی به نام سگال^۱ اعلام کرد که مغز از عناصر اصلی ساختاری به نام نرون^۲ تشکیل یافته است [۱۷].

^۱.Segal

^۲.Neuron

۲-۳-۱- مدل ریاضی نرون مصنوعی

نرون کوچکترین واحد پردازشگر اطلاعات است، که اساس عملکرد شبکه‌های عصبی را تشکیل می‌دهد. شکل (۱-۲) ساختار یک نرون تک ورودی را نشان می‌دهد. اسکالرهایی p و a به ترتیب ورودی و خروجی می‌باشند [۱۷].



شکل (۱-۲)- مدل تک ورودی [۱۷]

میزان تأثیر p روی a به وسیله مقدار اسکالر w تعیین می‌شود. ورودی دیگر که مقدار ثابت ۱ است، در جمله بایاس b ضرب شده و سپس با wp جمع می‌شود، این حاصل جمع، ورودی خالص p' برای تابع محرک (یا تابع تبدیل) f خواهد بود. بدین ترتیب خروجی نرون با معادله زیر تعریف می‌شود:

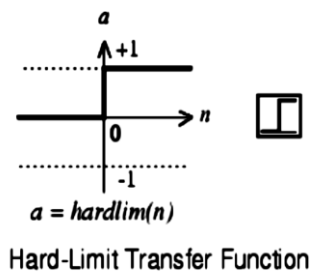
$$a = f(wp + b) \quad (۲-۲)$$

در مقایسه این مدل تک ورودی با یک نرون بیولوژیکی، معادل شدت سیناپس، مجموعه جمع کننده و تابع محرک، w معادل هسته سلول و a معادل سیگنال گذرنده از اکسون خواهد بود. نکته‌ای که باید به آن توجه شود اهمیت و تأثیر جمله بایاس b است. این جمله را می‌توان مانند وزنه w در نظر گرفت، با این تصور که میزان تأثیر ورودی ثابت ۱ را روی نرون منعکس می‌سازد. اهمیت جمله b به مرور توضیح داده خواهد شد.

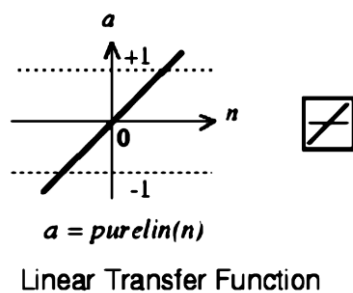
باید توجه داشت که پارامترهای w و b قابل تنظیم هستند و تابع محرک f نیز توسط طراح انتخاب می‌شود. بر اساس انتخاب f و نوع الگوریتم یادگیری، پارامترهای w و b تنظیم می‌شوند. یادگیری بدین معنی است که w و b طوری تغییر می‌کنند، که رابطه ورودی و خروجی نرون با هدف خاصی مطابقت نماید.

یکی از اجزای اصلی شبکه‌های عصبی تابع انتقال مورد استفاده می‌باشد که بر حسب کاربرد، می‌تواند انواع متفاوتی باشد. در برخی کاربردها، پاسخ مسئله از نوع دودویی است یعنی مسئله به گونه‌ای است که خروجی شبکه باید چیزی مانند آری یا نه باشد، در چنین مواردی نمی‌توان صرفاً بر جمع جبری سیگنال‌های ورودی تکیه کرد، بلکه ویژگی‌های خواسته شده در تابع انتقال گنجانده می‌شود. مثلاً اگر قرار باشد خروجی تابع فقط یکی از مقادیر صفر یا یک باشد، از تابع آستانه‌ای حدی^۱ استفاده می‌شود. بر اساس شکل (۲-۲ الف) در این تابع اگر آرگومان ورودی n (ورودی تابع انتقال) کوچکتر از صفر باشد، مقدار خروجی (a) این تابع صفر می‌شود و اگر آرگومان ورودی n بزرگتر یا مساوی از صفر باشد، خروجی برابر یک می‌شود. از توابع انتقال مشتق پذیری که مخصوصاً در شبکه‌های عصبی پس انتشار به کار می‌رود و در این تحقیق نیز مورد استفاده قرار گرفته، می‌توان به توابع انتقال خطی، لگاریتم سیگموئیدی و تانژانت سیگموئیدی اشاره کرد. تابع انتقال خطی شکل (۲-۲ ب) همان مقدار ورودی را به عنوان خروجی برمی‌گرداند ولی تابع لگاریتم سیگموئیدی ($\log \text{sig}$) مقادیر ورودی را در محدوده $-\infty$ تا $+\infty$ دریافت کرده و خروجی بین صفر و یک را تولید می‌کند و بر اساس شکل (۳-۴ د) تابع تانژانت سیگموئیدی (tansig) خروجی بین ۱ و -۱ تولید می‌کند [۱۷-۲۲].

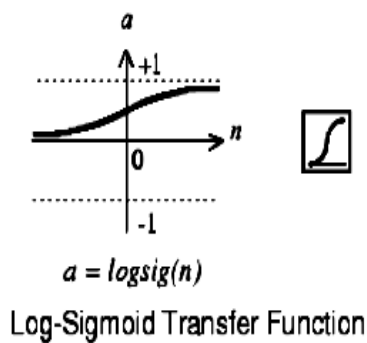
^۱ - Hard limit



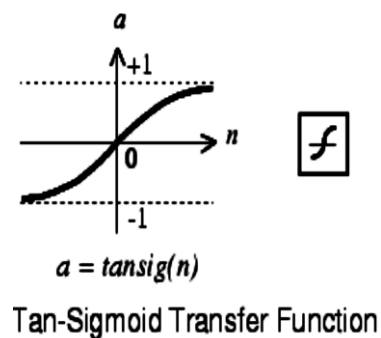
الف



ب



ج



د

شکل (۲-۲) - انواع توابع انتقال [۱۷]

۲-۳-۲- انواع شبکه‌های عصبی از نظر برگشت پذیری

• شبکه‌های پیشخور^۱

• شبکه‌های برگشتی^۲

شبکه‌های پیشخور، شبکه‌هایی هستند که مسیر پاسخ در آن‌ها همواره رو به جلو پردازش می‌شود و به نرون‌های لایه قبل باز نمی‌گردد. در این نوع شبکه‌ها به سیگنال اجازه داده می‌شود که در مسیر یکطرفه عبور کند یعنی از ورودی تا خروجی. بنابراین بازخوردی^۳ وجود ندارد بدین معنی که خروجی هر لایه تأثیری بر همان لایه و همچنین لایه‌های قبلی ندارد. در شبکه‌های برگشتی حداقل

^۱- Feed forward

^۲- Recurrent

^۳- Feedback

یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های دیگر همان لایه قبل وجود دارد. به عبارتی در این نوع شبکه‌ها بازخورد وجود دارد [۱۷و۲۲].

شبکه‌های پیشخور معمولاً از یک یا چند لایه پنهان با تابع تبدیل سیگموئیدی و یک لایه خارجی با تابع تبدیل خطی تشکیل شده‌اند. از نمونه‌های بارز این دسته اول می‌توان به شبکه پرسپترون چند لایه^۱ و آدالین^۲ و از نمونه‌های شبکه‌های برگشتی به کوهونن^۳ و هاپفیلد اشاره نمود.

2-3-3- مراحل آموزش در شبکه‌های پیشخور

به طور کلی آموزش بر طبق مراحل زیر انجام می‌شود:

۱- انتشار ورودی‌ها از نرون‌های ورودی به سمت نرون‌های خروجی

۲- اختصاص ماتریس وزن‌های تصادفی به هریک از اتصالات

۳- مقایسه خروجی‌های شبکه با مقادیر واقعی (مقادیر هدف) و محاسبه‌ی خطای شبکه

۴- پس‌انتشار خطا از نرون‌های خروجی به سمت نرون‌های ورودی و اصلاح وزن‌ها

۵- ارزیابی عملکرد شبکه با توجه به تابع کارایی تعیین شده

مراحل فوق تا زمانی تکرار می‌شود که به حداکثر تکرار^۴ مجاز رسیده باشد یا مقدار تابع کارایی از مقداری که تعیین شده کمتر باشد.

بررسی قابلیت تعمیم شبکه‌ی عصبی آموزش‌دیده، آخرین مرحله در توسعه‌ی مدل شبکه‌ی عصبی مصنوعی است. برای یافتن شبکه‌ای با معماری مناسب به کمک الگوریتم‌های آموزشی، از رسم منحنی مقایسه‌ای مقادیر پیش‌بینی شده توسط مدل و مقادیر واقعی و نیز معیار مجذور خطای میانگین مربعات استفاده می‌شود [۱۷].

^۱- Multilayerperceptron(MLP)

^۲- Adaline

^۳- Kohonen

^۴- Epoch

۲-۳-۴- بهبود عمومیت شبکه^۱

یکی از مشکلات موجود در آموزش شبکه‌های عصبی بیش برآزش^۲ شبکه بر داده‌های آموزشی می‌باشد. منظور از بیش برآزش این است که پس از آموزش شبکه خطا بر روی مجموعه آموزشی به حداقل مقدار خود می‌رسد اما با ارائه داده‌های جدید به شبکه به عنوان ورودی خطا بسیار بالا می‌باشد. در واقع شبکه بر روی داده‌های آموزشی احاطه کامل می‌یابد اما نمی‌تواند در قبال داده‌های جدید عکس‌العمل مناسب را از خود نشان دهد. در اصطلاح می‌توان گفت شبکه عمومیت کافی ندارد [23].

دو راه حل برای افزایش عمومیت شبکه در جعبه ابزار شبکه‌های عصبی در نظر گرفته شده است: تنظیم^۳ و توقف زودرس^۴، در ادامه به بررسی این دو روش خواهیم پرداخت [۱۷].

• تنظیم

اولین روش برای بهبود عمومیت، تنظیم نام دارد. این روش شامل اصلاح توابع کارایی می‌باشد، که به صورت معمول این توابع کارایی مجموع مربعات خطاهای شبکه در ارتباط با مجموعه آموزشی در نظر گرفته می‌شوند. برای تنظیم دو روش وجود دارد، اصلاح توابع کارایی و تنظیم خودکار.

• توقف زودرس^۵

یک روش دیگر برای بهبود عمومیت توقف زودرس می‌باشد. از این روش به صورت پیش فرض در آموزش انواع شبکه‌های عصبی در این جعبه ابزار استفاده می‌شود. در این تکنیک داده‌های موجود به

¹- Improving Generalization

²- Overfitting

³- Regularization

⁴- Early stopping

⁵- Early stopping

سه زیر مجموعه تقسیم می‌شوند. زیر مجموعه اول همان آموزشی خواهد بود که برای محاسبه گرادیان و به روزرسانی وزن‌ها و بایاس‌ها از آن استفاده می‌شود.

زیر مجموعه دوم زیر جهت ارزیابی یا مجموعه معتبر سازی است. از خطای مجموعه ارزیابی در طول فرایند آموزش برای نظارت بر روال آموزش استفاده می‌شود. خطای مجموعه معتبرسازی در طول فرایند آموزش مانند خطای مجموعه آموزشی باید کاهش یابد. به هر حال زمانی که شبکه سعی در تطابق بیش از حد با داده‌ها دارد خطای مجموعه معتبرسازی شروع به افزایش می‌کند. زمانی که خطای معتبر سازی افزایش یابد آموزش متوقف می‌شود و مقدار وزن‌ها و بایاس‌ها با زمانی که این خطا حداقل بوده انطباق داده می‌شوند.

بخش سوم نیز زیر مجموعه آزمایشی می‌باشد که در طول فرایند آموزش کاربردی ندارد و از آن برای مقایسه مدل‌های مختلف استفاده می‌شود. همچنین این زیر مجموعه برای رسم خطای مجموعه آزمایشی در طول فرایند آموزش مفید است. اگر خطا در مجموعه آزمایشی به حداقل مقدار خود با تکرارهای متفاوتی نسبت به مجموعه معتبر سازی برسد این بدین معنی است که تقسیم‌بندی مناسبی انجام نشده است.

۲-۴- الگوریتم ژنتیک

الگوریتم ژنتیک، علی‌رغم نامش جزء روش‌های هوش مصنوعی غیر الگوریتمی به حساب می‌آید. بر خلاف روش‌های معمول آماری که دارای الگوریتم مشخص‌اند و در یک مسأله خاص همواره به یک جواب مشخص منتهی می‌شوند، الگوریتم ژنتیک همواره به یک جواب نمی‌رسد. الگوریتم ژنتیک را می‌توان روشی برای جست‌وجو در میان راه حل‌های مختلف مسأله دانست. به طور کلی روش‌های

جست و جو به سه دسته روش‌های مبتنی بر حساب دیفرانسیل^۱، روش‌های شمارشی^۲ و روش‌های تصادفی^۳ تقسیم‌بندی می‌شوند. روش الگوریتم ژنتیک از روش‌های تصادفی به حساب می‌آید [۲۴].

۲-۴-۱- مبانی الگوریتم ژنتیک

الگوریتم ژنتیک بر روی یک سری از جواب‌های مسأله، به امید بدست آوردن جواب‌های بهتر، قانون بقای بهترین را اعمال می‌کند. در هر نسل، به کمک فرآیند انتخابی متناسب با ارزش جواب‌ها و تولید مثل جواب‌های انتخاب شده به کمک عملگرهایی که از ژنتیک طبیعی تقلید شده‌اند، تقریب‌های بهتری از جواب نهایی بدست می‌آید. این فرایند باعث می‌شود که نسل‌های جدید با شرایط مسأله سازگارتر باشند.

الگوریتم با مجموعه‌ای از جواب‌ها (که به وسیله کروموزوم‌ها نمایش داده می‌شوند) به نام جمعیت^۴ آغاز می‌شود. جواب‌ها از یک جمعیت گرفته شده و برای ایجاد جمعیت جدیدی بکار می‌روند. هدف این است که جمعیت جدید از جمعیت پیشین بهتر باشد، از اینرو معیار انتخاب^۵ جواب‌هایی که برای ایجاد جواب‌های جدید (فرزندان^۶) انتخاب می‌شوند شایستگی^۷ آنها خواهد بود و به عبارتی هر چه شایستگی یک جواب بیشتر باشد، شانس بیشتری برای تولید نسل دارد. این روند تا زمانی که شرایط خاصی (به عنوان مثال تعداد جمعیت یا بهبود بهترین جواب) حکمفرما گردد تکرار می‌شود [۲۴].

^۱-Differential calculus

^۲-Numerical

^۳-Stochastic

^۴-Population

^۵-Selection

^۶-Offspring

^۷-Fitness

۲-۴-۲- رمزبندی کروموزوم

هر کدام از افراد جمعیت، که تقریب‌هایی از جواب نهایی‌اند، به صورت رشته‌هایی (کروموزوم) از حروف یا ارقام، کد گذاری (رمزبندی) می‌شوند. یک کروموزوم بایستی حاوی اطلاعاتی مطابق با جواب حاصل باشد. متداول‌ترین حالت، نمایش با ارقام صفر و یک است که به نام روش دودویی^۱ (باینری) شناخته می‌شود. در این روش هر کروموزوم به وسیله‌ی یک رشته دودویی نشان داده می‌شود. هر بیت در رشته (کروموزوم) می‌تواند ارائه دهنده برخی از مشخصات جواب باشد. حالت‌های دیگر، مثل استفاده از سه رقم، اعداد حقیقی و اعداد صحیح هم مورد استفاده قرار می‌گیرند [۲۴]. در ادامه رمزبندی دودویی و تبدیل^۲ معرفی خواهند شد.

• رمزبندی دودویی (باینری)

معمول‌ترین شیوه نمایش کروموزون‌ها در الگوریتم ژنتیک، شکل رشته‌های دودویی است. در این شیوه هر متغیر تصمیم‌گیری به صورت دودویی در آمده و سپس با در کنار هم قرار گرفتن این متغیرها، کروموزوم ایجاد می‌شود. رمزبندی دودویی اساس رایج‌ترین نوع رمزبندی است و این روش نسبتاً ساده است. در رمزبندی دودویی، هر کروموزوم یک رشته از بیت‌ها با مقدار صفر یا یک می‌باشد که مثالی از آن در شکل (۲-۳) آورده شده است [۲۴].

Chromosome A	10110010110010101110010101101
Chromosome B	11111110000011000001111010011

شکل (۲-۳)- کروموزوم‌هایی با رمزبندی دودویی [۲۴]

^۱-Binary string

^۲-Permutation encoding

رمزبندی دودویی، کروموزوم‌های ممکن زیادی را با کمک تعداد محدودی ژن ارائه می‌دهد. اما از طرفی این رمزبندی گاهی برای برخی مسائل طبیعی مناسب نیست و اصطلاحاتی در تعریف پیوند و جهش بایستی انجام گیرد.

• رمزبندی تبدیل

رمزبندی تبدیل را می‌توان در مسائل ترتیبی مانند مساله فروشنده دوره گرد بکار برد. در این رمزبندی، هر کروموزوم یک رشته از اعداد است که یک مکان را در یک توالی ارائه می‌کند. شکل (۲-۴) نمونه‌ای از این رمزبندی را نشان می‌دهد [۲۴].

Chromosome A	1 5 3 2 6 4 7 9 8 6 3 2 7
Chromosome B	2 4 5 2 7 8 4 3 9 7 5 7 2

شکل (۲-۴) - کروموزوم‌هایی با رمزبندی تبدیل [۲۴]

۲-۴-۳ - عملگرهای GA

همانطور که در طرح کلی الگوریتم ژنتیک دیدیم، پیوند و جهش از بخش‌های مهم الگوریتم ژنتیک هستند. عملکرد الگوریتم ژنتیک به طور اساسی تحت تاثیر این دو عملگر می‌باشد [۲۴].

۲-۴-۴- پیوند

وقتی که دو فرد از افراد یک نسل بر اساس شایستگی خود در مرحله انتخاب گزینش شوند، اجازه تولید مثل و تولید فرزندان جدید را خواهند داشت. عمل پیوند میان این دو فرد و تولید مثل نسل بعدی به وسیله عملگر تقاطع یا همان پیوند صورت می گیرد. به عبارتی، عملگر اصلی ایجاد نسل جدید در مرحله تکثیر، تقاطع و یا پیوند است. همانند کروموزومها در طبیعت، فرزندان حاصل از این عمل هر یک بخشی از اطلاعات روی کروموزومهای والد را دارند. عملگر تقاطع، انواع مختلفی دارد که بر اساس نوع مسأله و کدگذاری کروموزومها، برای ایجاد پیوند میان والدین یک نسل به کار برده می شود.

• پیوند یک نقطه ای^۱

عمل تکثیر در الگوریتم ژنتیک برای رد و بدل اطلاعات ژنتیکی بین یک جفت و یا تعداد بیش تری از افراد به کار می رود. ساده ترین نوع تکثیر، تقاطع یک نقطه ای است. اگر یک عدد صحیح از یک تا تعداد ارقام رشته منهای یک، انتخاب کنیم و اطلاعات دو رشته را در دو طرف این نقطه عوض کنیم، دو رشته جدید به وجود می آید که آنها را فرزند می خوانیم. نتیجه تقاطع یک نقطه ای به صورت شکل (۲-۵) در می آید.

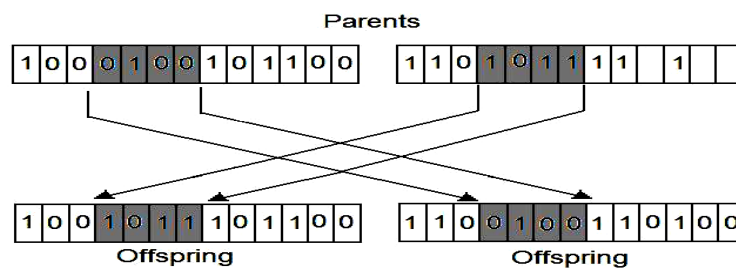
Chromosome 1	10001/00101100
Chromosome 2	11010/11110101
Offspring 1	10001/11110101
Offspring 2	11010/00101100

شکل (۲-۵) - تولید مثل بر اساس تقاطع یک نقطه ای [۲۴]

^۱-Single point crossover

• پیوند دو نقطه‌ای^۱

نوع دیگر عملگر تقاطع، تقاطع دو نقطه‌ای است. در این روش، تعدادی نقطه روی کروموزوم انتخاب شده و آن را به چند قسمت تقسیم می‌کند. سپس قسمت‌های مشابه از کروموزوم‌ها یکی در میان، با هم عوض می‌شوند. اولین قسمت هر دو کروموزوم بدون تغییر نگاهداشته می‌شود. در شکل (۶-۲) این شیوه برای تقاطع دو نقطه‌ای نشان داده شده است.



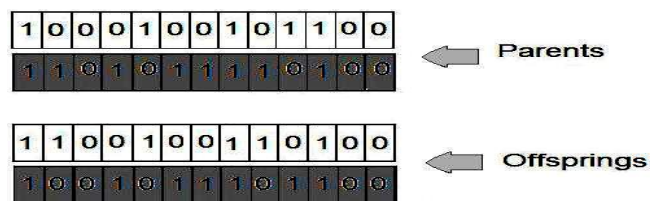
شکل (۶-۲) - تقاطع دو نقطه‌ای [۲۴]

• پیوند یکنواخت^۲

در تقاطع یک نقطه‌ای و چند نقطه‌ای، تعدادی نقطه تعریف می‌شد که کروموزوم در آن نقاط شکسته شده و اطلاعات آن با کروموزوم دیگری تعویض می‌شود. در روش تقاطع یکنواخت این حالت گسترش یافته است. یک رشته تصادفی از ارقام صفر و یک به طول کروموزوم ایجاد می‌شود. به بیان دیگر بیت‌ها به صورت اتفاقی از والدین کپی می‌شوند. شکل (۷-۲) پیوند یکنواخت را نشان می‌دهد.

^۱-Two point crossover

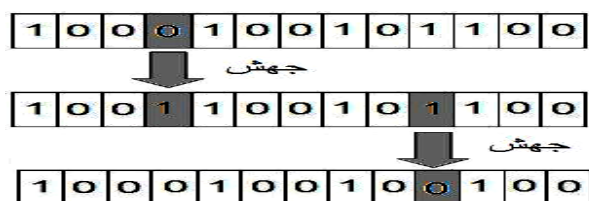
^۲-Uniform Crossover



شکل (۷-۲) - پیوند یکنواخت [۲۴]

۲-۴-۵ - جهش^۱

جهش به معنای تغییر مقدار یکی از خانه‌های رشته، از صفر به یک و یا از یک به صفر می‌باشد. به کمک این عملگر می‌توان امید داشت که کروموزوم‌های خوبی که در مراحل انتخاب و یا تکثیر حذف شده‌اند، دوباره احیا شوند [۲۴]. شکل (۸-۲) نحوه عملکرد عملگر جهش در سیستم دودویی را نشان می‌دهد.



شکل (۸-۲) - نحوه عملکرد عملگر جهش در سیستم دودویی [۲۴]

۲-۴-۶ - پارامترهای GA^۲

پارامترهای ژنتیک الگوریتم شامل احتمال (نرخ) پیوند، احتمال (نرخ) جهش و همچنین اندازه جمعیت می‌باشند. در ادامه به اختصار به این پارامترها خواهیم پرداخت [۲۵ و ۲۶].

^۱-Mutation

^۲-Parameters of GA

• احتمال (نرخ) پیوند^۱

اگر فرآیند پیوند نباشد، فرزندان دقیقاً همانند والدین خواهند بود و در صورت وجود پیوند، فرزندان از قسمت‌های مختلف کروموزوم والدین ساخته می‌شوند. اگر احتمال پیوند ۱۰۰٪ باشد در این صورت تمامی فرزندان به وسیله‌ی پیوند ساخته می‌شوند و اگر ۰٪ باشد تمام نسل جدید دقیقاً کپی کروموزوم‌های جمعیت پیشین خواهد بود. پیوند با این هدف ساخته شده است که کروموزوم‌های جدید بخش‌های خوب کروموزوم‌های قبلی را برداشته و از اینرو کروموزوم‌های جدید بهتر باشند. به هر حال مناسب است که برخی بخش‌های جمعیت پیشین را برای حضور در نسل بعدی حفظ کنیم.

• احتمال (نرخ) جهش^۲

اگر فرایند جهش نباشد، فرزندان بلافاصله پس از پیوند (یا کپی مستقیم) و بدون هیچ تغییری تشکیل می‌شوند و اگر جهش داشته باشیم یک بخش (یا بیشتر) از یک کروموزوم تغییر می‌کند. اگر احتمال جهش ۱۰۰٪ باشد تمام کروموزوم تغییر کرده و اگر ۰٪ باشد هیچ چیز تغییر نمی‌کند. به‌طور کلی جهش از سقوط به نقاط بهینه محلی جلوگیری می‌کند. جهش نبایستی زیاد اتفاق بیافتد زیرا در این صورت GA در اصل به شکل جستجوی اتفاقی در خواهد آمد.

• اندازه جمعیت

اندازه جمعیت یعنی اینکه چه تعداد کروموزوم در جمعیت (در یک نسل) وجود دارد. اگر تعداد کمی کروموزوم داشته باشیم، GA احتمال کمی برای اجرای پیوند دارد و تنها یک بخش کوچک از فضای جستجو مورد استفاده قرار می‌گیرد. از طرف دیگر اگر تعداد کروموزوم‌ها بسیار باشد، GA سرعت کمی

^۱-Crossover

^۲-Mutation

خواهد داشت. تحقیقات نشان می‌دهد که از یک محدوده‌ی خاص به بعد (که به روش رمزبندی و خود مسأله بستگی دارد) بکار بردن جمعیت‌های زیاد سودمند نیست، زیرا مسأله را سریع‌تر از حالت جمعیت‌های با اندازه متوسط حل نمی‌کند [۲۵].

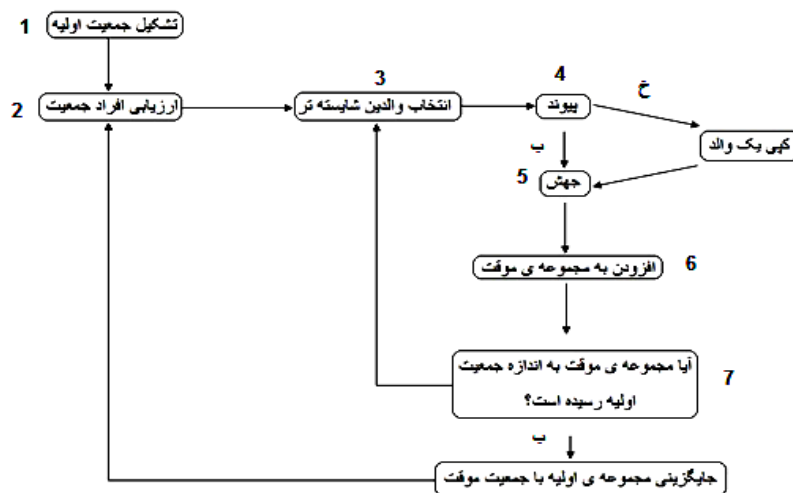
۲-۴-۷- اجرای الگوریتم ژنتیک

شکل (۲-۹) مراحل اجرای یک الگوریتم ژنتیک ساده را نشان می‌دهد که به‌طور خلاصه شرح داده می‌شود:

۱. در ابتدا، به صورت تصادفی جمعیتی از کروموزوم‌ها تشکیل می‌شوند. تعداد افراد جمعیت کاملاً انتخابی بوده و بستگی به نوع مسأله دارد.
۲. در این مرحله هر فرد به وسیله تابع شایستگی ارزیابی می‌گردد.
۳. با روش چرخه رولت افراد شایسته‌تر با احتمال بیشتری انتخاب می‌شوند. تعداد افراد انتخاب شده نیز امری انتخابی است و نخبه‌گرایی نام دارد. هر چه این تعداد بیشتر باشد، افراد شایسته جمعیت بهتر حفظ می‌شوند.
۴. افراد انتخاب شده با احتمال P_C با هم پیوند داده می‌شوند. برای این کار به صورت تصادفی یک عدد بین صفر و یک انتخاب می‌شود و اگر این عدد کمتر از P_C بود عمل تقاطع انجام شده، در غیر این صورت یکی از والدین به صورت تصادفی کپی می‌شود.
۵. عملگر جهش نیز با همان مکانیسم با احتمال P_M روی کروموزوم یا کروموزوم‌های ایجاد شده انجام می‌پذیرد.
۶. کروموزوم‌های ایجاد شده به یک مجموعه موقت از افراد می‌پیوندند.
۷. مراحل ۳ تا ۶ آنقدر تکرار می‌شوند تا تعداد افراد موجود در مجموعه موقت با تعداد افراد جمعیت ابتدایی برابر شود. وقتی این اتفاق افتاد، افراد مجموعه موقت جایگزین افراد جمعیت

ابتدایی شده و به این ترتیب نسل جدیدی تشکیل می‌شود. سپس، الگوریتم دوباره از مرحله ۲ تکرار می‌شود.

۸. مراحل ۱ تا ۷ آن قدر تکرار می‌شوند تا تعداد نسل‌ها به عدد تعیین شده برسد. شایسته‌ترین فرد نسل آخر جواب مسأله است [۲۴].



شکل (۲-۹) - فلوچارت یک الگوریتم ژنتیکی ساده [۲۴]

۲-۴-۸ - الگوریتم ژنتیک و QSAR

می‌توان از الگوریتم ژنتیک برای انتخاب متغیرها و مدل‌سازی با روش‌های دیگری همچون رگرسیون خطی چندگانه، حداقل مربعات جزئی و شبکه‌ی عصبی استفاده کرد. در روش GA-MLR، الگوریتم ژنتیک پارامترهای غیرخطی را بهینه کرده و پارامترهای خطی به وسیله رگرسیون خطی چندگانه در موازات یکدیگر محاسبه می‌شوند [۲۷].

۲-۵ - نرم‌افزارهای مورد استفاده

دانش کمومتریکس بسته‌های نرم‌افزاری متنوعی را برای انجام تمام مراحل مدل‌سازی بکار گرفته است که در ادامه به اختصار بسته‌های نرم‌افزاری استفاده شده در این تحقیق معرفی می‌شوند.

• بسته نرم‌افزاری Hyperchem

از بسته نرم‌افزاری Hyperchem [۲۸] برای رسم شکل مولکول‌ها و بهینه‌سازی ساختار با استفاده از روش‌های کوانتومی و مکانیکی، استفاده می‌شود. به کمک این برنامه می‌توان طول پیوند، زاویه پیوندی و زوایای پیچشی را در مولکول تعیین کرد. داده‌های حاصل از این نرم‌افزار را می‌توان به عنوان ورودی به سایر نرم‌افزارها معرفی نمود

• بسته نرم‌افزاری Dragon

بسته نرم‌افزاری برای محاسبه‌ی توصیف‌کننده‌های مولکولی می‌باشد که توسط گروه QSAR و کمومتریکس دانشگاه میلانو طراحی شده است [۲۹]. این توصیف‌گرهای مولکولی به منظور ارزیابی روابط ساختار-فعالیت یا ساختار-ویژگی مولکولی به کار می‌روند. اولین ویرایش آن به سال ۱۹۹۷ بر می‌گردد. به روز کردن‌ها و افزایش گنجایش توصیف‌کننده‌های مولکولی مرتباً به منظور تحقیقات پیشرفته‌تر در زمینه‌ی QSAR در این نرم‌افزار ایجاد شد. نرم‌افزار Dragon قادر به محاسبه‌ی بیش از ۱۴۹۷ توصیف‌کننده می‌باشد که به ۱۸ دسته اصلی تقسیم می‌شوند.

برای اجرای این نرم‌افزار کاربر نیاز به پرونده‌های ساختار مولکولی دارد که قبلاً توسط سایر نرم‌افزارهای مدل‌سازی مولکولی مانند Hyperchem تامین شده‌اند. برای استفاده‌ی بهتر از محاسبات این نرم‌افزار باید بهینه‌سازی سه بعدی ساختارها با هیدروژن‌هایشان به کار رود. Dragon یک پرونده‌ی خروجی کامل را که به سادگی توسط هر نرم‌افزار آنالیز همبستگی قابل کاربرد است، فراهم می‌سازد.

• بسته نرم‌افزاری SPSS^۱

SPSS [۳۰] که نخستین نسخه آن در سال ۱۹۷۰ توسط جمعی از فارغ التحصیلان دانشگاه استنفورد آمریکا ارائه شده، امکان تجزیه و تحلیل آماری داده‌ها را فراهم می‌آورد. برخی از قابلیت‌های این بسته نرم‌افزاری عبارتند از:

- ❖ تعیین تعداد فراوانی‌های هر یک از گروه‌ها در یک متغیر
- ❖ محاسبه میانگین ساده برای داده‌ها
- ❖ نمایش اطلاعات به صورت متنوع در قالب نمودار و جدول
- ❖ انجام رگرسیون تک متغیره و چند متغیره

• نرم‌افزار MATLAB^۲

نرم‌افزار MATLAB [۳۱] برای کسانی که با محاسبات عددی و بویژه جبر خطی سر و کار دارند تهیه شده است. MATLAB شامل یک سری از برنامه‌های کاربردی افزوده شده برای راه حل‌های مخصوص است که جعبه ابزار^۳ نامیده می‌شود که به کاربران اجازه یادگیری و بکار بردن تکنولوژی‌های خاص را می‌دهد و محیط آن را برای حل مسائل ویژه توسعه می‌بخشد. در این پژوهش از جعبه ابزارهای شبکه عصبی و الگوریتم ژنتیک برای مدل سازی غیر خطی فعالیت بیولوژیکی با ساختار ترکیبات استفاده شده است.

^۱ - Statistical Package for the Social Science

^۲ -Matrix Laboratory

^۳ -Toolbox

فصل سوم

مطالعه ارتباط کمی ساختار - فعالیت اثر

بازداری مشتقات فنیل آلکیل آمین با استفاده

روش‌های خطی و غیر خطی

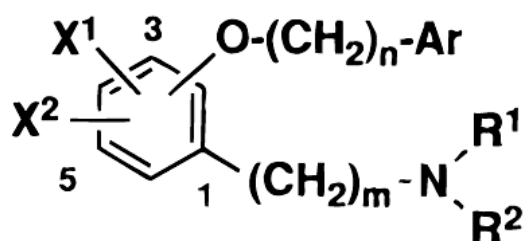
۳-۱- مدل سازی فعالیت دارویی مشتقات آریل الکوکسی فنیل آلکیل آمین^۱ به

عنوان بازدارنده های روان پریشی

به منظور مطالعه ارتباط کمی ساختار- فعالیت از شبکه عصبی مصنوعی و رگرسیون خطی چندگانه به عنوان روش های مدل سازی برای پیش بینی فعالیت ضد روان پریشی یک طبقه ی جدید از مشتقات آریل الکوکسی فنیل آلکیل آمین، استفاده شد و توانایی مدل های به دست آمده از هر دو روش در پیش بینی قدرت بازدارندگی این مشتقات، مورد ارزیابی قرار گرفت. به طور کلی بخش تجربی این تحقیق شامل انتخاب سری داده ها، رسم و بهینه سازی ساختار مولکول ها، محاسبه توصیفگرهای مولکولی، انتخاب توصیفگرهای مولکولی مناسب، مدل سازی و ارزیابی مدل برتر است.

۳- ۱- ۱- انتخاب سری داده ها

سری داده ها شامل فعالیت دارویی ۴۰ ترکیب مشتقات فنیل آلکیل آمین (شکل (۳-۱)) می - باشد و توسط آتسورو ناکازاتو^۲ و همکارانش گزارش شده است [۳۲].



شکل (۳-۱) - ساختار آریل الکوکسی فنیل آلکیل آمین [۳۲]

^۱- Arylalkoxyphenylalkylamine

^۲ - Atsuro Nakazato

در جدول (۱-۳) نیز جزئیات استخلاف‌ها و مقادیر عددی pIC₅₀ برای هر ترکیب آورده شده است.

IC₅₀^۱ غلظتی از دارو است که ۵۰٪ اثر بازدارندگی داشته باشد. ذکر این نکته ضروری است که هر چه

دارو قویتر باشد، غلظت کمتری از آن برای داشتن ۵۰٪ اثر بازدارندگی، نیاز است.

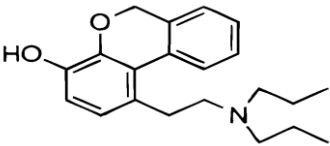
جدول (۱-۳) - ساختار و مقادیر pIC₅₀ ترکیبات مورد مطالعه [۳۲]

شماره ترکیب	X1	X2	-O(CH ₂) _n -Ar	m	-NR ₁ R ₂	pIC ₅₀
۱	4-MeO	H	3-OCH ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	7/55
۲ (t)	5-Cl	H	2-OCH ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	6/93
۳ (v)	4-MeO	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	8/18
۴	5-MeO	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	8/14
۵ (t)	4-MeO	H	3-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	8/88
۶	2-MeO	H	4-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	8/16
۷ (t)	3-MeO	H	4-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	7/92
۸ (t)	3-F	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	7/65
۹	5-F	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	7/88
۱۰ (v)	4-Cl	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	7/74
۱۱	3-Cl	5-Cl	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	6/55
۱۲	3-Br	5-Br	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	6/11
۱۳	5-Br	H	2-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr) ₂	۷/۳۵
۱۴ (t)	3-MeO	H	2-O(CH ₂) ₂ -Ph	۲	Pyrra	۷/۵۵
۱۵	4-MeO	H	3-O(CH ₂) ₂ -Ph	۲	Pyrra	۶/۹۷

^۱ - Half maximal inhibitory concentration

ادامه جدول (۱-۳)

شماره ترکیب	X1	X2	-O(CH ₂) _n -Ar	m	-NR ₁ R ₂	pIC ₅₀
۱۶	5-Cl	H	2-O(CH ₂) ₂ -Ph	۲	Morp	۶/۵۱
۱۷ (v)	3-MeO	H	2-O(CH ₂) ₂ -Ph	۲	Pipe-Ph	۶/۴۸
۱۸	4-MeO	H	3-O(CH ₂) ₂ -Ph	2	Pipe-Ph	۷/۴۸
19(v)	4-MeO	H	3-O(CH ₂) ₂ -Ph	2	Pipe-Pd-2	۶/۹۵
۲۰	4-MeO	H	3-O(CH ₂) ₂ -Ph	2	Pipe-Ph-OMe-2	۷/۶۹
۲۱	3-MeO	H	2-O(CH ₂) ₂ -Ph	3	N(<i>n</i> -Pr) ₂	۷/۸۸
۲۲	5-Br	H	2-O(CH ₂) ₂ -Ph	3	N(<i>n</i> -Pr) ₂	۷/۹۵
۲۳	5-Cl	H	2-O(CH ₂) ₂ -Ph	3	Pipe-Ph	۶/۸۴
۲۴ (v)	5-Cl	H	2-O(CH ₂) ₂ -Ph	3	Pipe-Pd-2	۶/۹۳
۲۵ (t)	5-Cl	H	2-O(CH ₂) ₂ -Ph	۳	Pipe-Pm-2	۶/۷۷
۲۶	5-Cl	H	2-O(CH ₂) ₂ -Ph	4	N(<i>n</i> -Pr) ₂	۷/۵۸
27(v)	4-HO	H	3-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr) ₂	۸/۵۱
۲۸	3-HO	H	4-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr) ₂	۷/۶۰
۲۹	5-Cl	H	2-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr) ₂	۸/۱۴
۳۰	4-MeO	H	3-O(CH ₂) ₂ -Ph-F-4	2	N(<i>n</i> -Pr) ₂	۹/۰
۳۱ (v)	4-MeO	H	3-O(CH ₂) ₂ -Ph-OMe-4	۲	N(<i>n</i> -Pr) ₂	۹/۰۰
۳۲	4-MeO	H	3-O(CH ₂) ₂ -Ph-(OMe) ₂ -3,4	2	N(<i>n</i> -Pr) ₂	۷/۱۷
۳۳ (t)	3-MeO	H	3-O(CH ₂) ₂ -Thi-	2	N(<i>n</i> -Pr) ₂	۸/۰۳
۳۴	3-MeO	H	3-O(CH ₂) ₃ -Ph	2	N(<i>n</i> -Pr) ₂	۸/۷۶
۳۵ (t)	4-MeO	H	3-O(CH ₂) ₃ -Ph	2	N(<i>n</i> -Pr) ₂	۸/۳۳
۳۶	4-MeO	H	3-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr)	۷/۴۶
۳۷	4-MeO	H	3-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr)(iso-Amy)	۸/۵۳
۳۸ (v)	4-MeO	H	3-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr)(<i>n</i> -Hex)	۸/۷۴
۳۹	4-MeO	H	3-O(CH ₂) ₂ -Ph	۲	N(<i>n</i> -Pr)((CH ₂) ₃ -OH)	۷/۷۲

شماره ترکیب	شکل ترکیب	p IC50
۴۰		6/25

۳-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، ساختار مولکولی هر ترکیب ابتدا در نرم افزار Hyperchem7.0 ترسیم شد. سپس با احتساب اتم‌های هیدروژن، ساختار سه بعدی ترکیبات با استفاده از روش نیمه تجربی کوانتومی AM1 بهینه شد و بهینه سازی تا زمانی ادامه یافت که جذر میانگین مربعات گرادیان^۱ انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد. همه این محاسبات در سطح هارتری- فاک محدود شده^۲ انجام شد.

۳-۱-۳- محاسبه توصیفگرها

برای ایجاد ارتباط ریاضی بین ساختار- فعالیت، به مقادیر عددی نیاز است پس باید ساختار مولکول‌ها به مقادیر کمی تبدیل شود. بدین منظور ساختار بهینه شده ترکیبات به نرم‌افزار Dragon منتقل شد و برای هریک از مولکول‌ها ۱۴۸۱ توصیفگر در ۱۸ گروه مختلف محاسبه گردید که مشخصات اصلی آنها در جدول (۳-۲) آمده است.

^۱ - Root Mean Square Gradient

^۲ - Restricted Hartee-Fock

جدول (۳-۲) - فهرست توصیفگرهای محاسبه شده توسط نرم افزار Dragon

ردیف	نام توصیفگر	ردیف	نام توصیفگر
۱	توصیفگر های زیر ساختاری	۱۰	توصیفگر های هندسی
۲	توصیفگرهای توپولوژیکی	۱۱	توصیفگرهای RDF
۳	شمارنده های مولکولی	۱۲	توصیفگرهای سه بعدی موریس
۴	توصیفگرهای BCUT	۱۳	توصیفگر های WHIM
۵	شاخص های بار توپولوژیکی Galvez	۱۴	توصیفگرهای GETAWAY
۶	خود ارتباطی های دوبعدی	۱۵	گروه های عاملی
۷	توصیفگرهای بار	۱۶	اجزای میان اتمی
۸	شاخص های آروماتیسیت	۱۷	توصیفگرهای تجربی
۹	پروفایل های مولکولی راندیک	۱۸	خصوصیات مولکولی

۳-۱-۴- انتخاب توصیفگرهای مناسب

مهمترین مرحله در مطالعات QSAR انتخاب توصیفگرهایی مناسب است که بتوانند فعالیت دارویی را به ساختار مولکول ارتباط دهند. با عنایت به این موضوع و با توجه به اینکه تعداد زیاد توصیفگر باعث پیچیدگی محاسبات می شود و همچنین تعدادی از توصیفگرها حاوی اطلاعات یکسان و در نتیجه اثر یکسان بر روی فعالیت مهارکنندگی هستند، لذا باید روش کاهش تعداد متغیرها به کار گرفته شود. به همین منظور، ابتدا توصیفگرهایی که دارای مقادیر یکسانی برای کل مولکولها بودند از فهرست توصیفگرها حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیفگری که ضریب همبستگی بزرگتر از ۰/۹ داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف و سپس رگرسیون مرحله به مرحله توسط نرم افزار SPSS مورد استفاده قرار گرفت. در روش

رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل می‌شوند و 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 داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. به این ترتیب داده‌های pIC_{50} به عنوان متغیر وابسته و توصیفگرها به عنوان متغیر مستقل در نظر گرفته شدند و تکنیک رگرسیون مرحله‌ای انجام شد که در این مرحله فقط ۱۱ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت دارویی دارند، شناخته شدند. این ۱۱ توصیفگر به همراه طبقه آنها در جدول (۳-۳) ارائه شده است.

جدول (۳-۳) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله

No	Symbol	Class	Meaning
۱	Mv	Constitutional descriptors	Mean atomic van der Waals volume
۲	BELm6	BCUT	Lowest eigenvalue n.6 of Burden
۳	GGI3	Charge indices	Topological charge index of order 3
۴	PJI2	Topological descriptors	2D Petitjean shape index
۵	JGI1	Galvez topol.charge indices	Mean topological charge index
۶	R5u	GETAWAY	R autocorrelation of lag 5/ unweighted
۷	Mor32m	3D-MoRSE descriptors	3D-MoRSE- signal32/weighted by atomic masses
۸	Mor30m	3D-MoRSE descriptors	3D-MoRSE- signal30/weighted by atomic masses
۹	RDF075u	RDF descriptors	Radial Distribution Function-7.5/unweighted
۱۰	RDF050m	RDF descriptors	Radial Distribution Function-5/weighted by atomic masses
۱۱	RDF050u	RDF descriptors	Radial Distribution Function-5/unweighted

همچنین ماتریس همبستگی بین این توصیفگرها در جدول (۳-۴) ارائه شده که این ماتریس عدم

همبستگی بین توصیفگرها را نشان می‌دهد.

جدول (۳-۴) - ماتریس همبستگی کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله

	Mv	PJI2	BELm6	GGI3	JGI1	R5u	Mor32m	Mor30m	RDF075u	RDF050m	RDF050u
Mv	۱										
PJI2	-0.050	۱									
GGI3	-۰/۱۲۶	۰/۶۳۳	۱								
PJI2	-۰/۱۵۴	۰/۴۳۲	۰/۴۳۲	۱							
JGI1	-۰/۳۵۹	-۰/۳۰۱	-۰/۳۴۸	-۰/۴۴۰	۱						
R5u	-۰/۳۸۹	۰/۵۵۶	۰/۵۰۶	۰/۳۵۵	-۰/۲۱۲	۱					
Mor32m	-۰/۱۰۴	-۰/۱۹۱	-۰/۴۹۵	-۰/۱۶۷	۰/۱۴۹	-۰/۱۰۶	۱				
Mor30m	-۰/۰۶۷	-۰/۴۲۶	۰/۰۲۲	۰/۲۴۸	۰/۰۴۰	۰/۱۴۰	۰/۰۰۳	۱			
RDF075u	۰/۲۴۵	۰/۵۱۱	۰/۰۴۷	۰/۲۲۲	-۰/۱۶۱	۰/۲۵۹	-۰/۰۰۶	-۰/۲۳۷	۱		
RDF050m	۰/۳۲۸	-۰/۵۶۰	-۰/۳۶۲	-۰/۲۷۲	۰/۰۶۴	-۰/۶۵۲	-۰/۱۱۳	۰/۴۰۲	-۰/۳۳۴	۱	
RDF050u	0/064	-0/070	0/533	0/560	0/008	0/87	-0/243	0/050	0/۵۶۴	0/760	۱

۳-۱-۵- مدل سازی توسط روش رگرسیون خطی چندگانه

برای ساختن مدل خطی مناسب که بیانگر ارتباط ساختاری ترکیبات مورد مطالعه با فعالیت بازدارندگی آنها باشد، مدل‌های مختلفی که توسط رگرسیون خطی چندگانه مرحله به مرحله به دست آمد، مورد بررسی قرار گرفتند. برای این کار داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری تست ۸ ترکیب، سری ارزیابی ۸ ترکیب و سری آموزش ۲۴ ترکیب را شامل می‌شوند. سپس، توسط سری آموزش، ابتدا با ۲ توصیفگر اول انتخاب شده، به روش اجباری مدل

ساخته شد و به وسیله‌ی معادله‌ی آن سری ارزیابی آزموده شد. در مرحله بعد با مدل ایجاد شده توسط ۳ توصیفگر، سری ارزیابی را آزموده و همین طور با معادلات حاصل از ۴ تا ۱۱ توصیفگر به طور مجزا سری ارزیابی مورد بررسی قرار گرفت. میانگین مربعات خطا (mse) و ضریب تعیین تصحیح شده (R_{adj}^2) محاسبه شده برای سری ارزیابی در جدول (۵-۳) آمده است.

جدول (۵-۳) - مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR

مدل پارامتر	2	3	4	5	6	7	8	9	10	11
MSE	0/202	0/141	0/074	0/069	0/057	0/047	0/044	0/061	0/038	0/049
R_{adj}^2 valid	0/587	0/805	0/904	0/934	0/95	0/941	0/918	0/866	0/909	0/907

همانطور که در جدول (۵-۳) مشاهده می‌شود مدل هشتم به خاطر داشتن خطای کمتر و R_{adj}^2 مناسب به عنوان مدل نهایی انتخاب شد.

به این ترتیب معادله خطی نهایی، با استفاده از سری آموزش به صورت زیر به دست آمد:

$$pIC50 = 10.27 - 24.77 Mv - 6.13 PJI2 + 11.42 BELm6 - 1.42 GGI3 + 10.75 JGI1 + 8.37 R5u + 1.25 Mor32m - 0.93 Mor30m$$

۳-۱-۶- مدل سازی به روش شبکه عصبی مصنوعی

بخاطر اینکه ممکن است مدل خطی بهترین مدل برای توصیف رفتار سری داده‌ها نباشد، شبکه عصبی مصنوعی برای مدل سازی غیر خطی داده‌ها مورد استفاده قرار گرفت. همانند روش MLR داده‌ها به صورت تصادفی و با نسبت ۶۰:۲۰:۲۰ به سه سری آموزش، ارزیابی و تست تقسیم شدند. سپس پارامترهای مؤثر بر آموزش موفق شبکه عصبی پس انتشار از قبیل تعداد ورودی‌ها، تعداد لایه‌های پنهان، تعداد گره‌های لایه پنهان، نوع تابع انتقال، نوع تابع آموزش، تعداد دوره‌های آموزش و پارامتر

mu با استفاده از سری آموزش و با روش ارزیابی تقاطعی بهینه شدند. در طی فرایند آموزش مقادیر وزن‌های ارتباطی تغییر می‌کند و معیار برای پایان دادن آموزش و بهینه شدن مقادیر وزن‌ها جهت‌پیش بینی پارامتر مورد بررسی، مقدار خطای شبکه است که طبق تابع کارایی تعیین شده، محاسبه می‌شود که در این روش، تابع کارایی، میانگین مربع خطا (MSE) در سری آموزش می‌باشد.

۳-۱-۶-۱- انتخاب تعداد لایه‌های پنهان

متأسفانه هیچ‌گونه نتایج تئوری قابل دسترس یا قوانین تجربی رضایت بخش که در تعیین تعداد لایه‌های پنهان به ما کمک کند، وجود ندارد با این وجود برای اغلب کاربردهای شبکه عصبی در شیمی، به نظر می‌رسد که یک لایه مخفی کافی باشد [۳۳]. به این ترتیب شبکه عصبی تشکیل شده، یک شبکه سه لایه شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است که تعداد ورودی‌ها (توصیفگرها) با تعداد نرون‌های لایه ورودی برابر است و لایه خروجی دارای یک نرون است که نشان دهنده pIC_{50} متناظر با هر ترکیب می‌باشد. ولی تعداد نرون‌های لایه پنهان باید بهینه شود.

۳-۱-۶-۲- انتخاب تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و

تعداد نرون‌های لایه پنهان

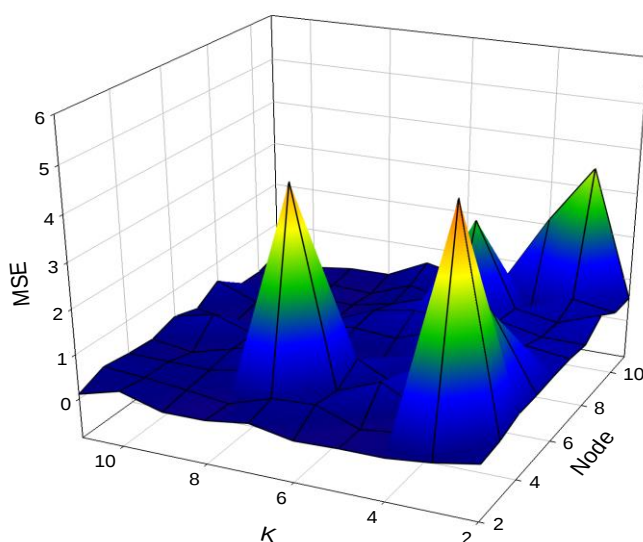
برای بهینه‌سازی تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد گره‌های لایه پنهان، شبکه با ورودی‌های از ۲ تا ۱۱ توصیفگر ایجاد شد و با الگوریتم‌های آموزشی تنظیم بایزین و لونبرگ-مارکوارت با تعداد متفاوت گره‌های لایه پنهان (۲ تا ۱۱ گره) آموزش داده شد. در حالی که برای انتخاب بهترین تابع انتقال در لایه پنهان از توابع تانژانت سیگموئیدی (tansig) و لگاریتم سیگموئیدی (logsig) به عنوان تابع انتقال استفاده شد. در تمامی این شبکه‌ها تعداد دورهای آموزش ثابت و برابر ۳۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده

گردید. در روند بهینه‌سازی پارامترهای فوق، به حداقل رساندن میانگین مربع خطاهای حاصل از بررسی سری ارزیابی به عنوان معیار انتخاب شد. نتایج بهینه‌سازی این پارامترها در جداول (۳-۶)، (۳-۷)، (۳-۸) و (۳-۹) گردآوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۳-۲)، (۳-۳)، (۳-۴) و (۳-۵) ارائه شده است. همانطور که نتایج نشان می‌دهد، میانگین مربع خطا (MSE) برای الگوریتم آموزشی تنظیم بایزین نسبت به الگوریتم آموزشی لونبرگ مارکوارت کمتر است. همین طور از قدرت تعمیم پذیری بالاتری نیز برخوردار است بدین معنی که دارای توانایی بیشتری برای برقراری تطابق میان ترکیباتی که در سری آموزش استفاده شده و آن‌هایی که در سری آموزش استفاده نشده‌اند، می‌باشد. همچنین تابع تانژانت سیگموئیدی نسبت به لگاریتم سیگموئیدی خطای کمتری دارد. در نتیجه تابع تانژانت سیگموئیدی به عنوان تابع انتقال لایه پنهان برگزیده شده است. طبق نتایج به دست آمده ۸ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط با فعالیت بازدارندگی دارند، انتخاب شدند. تعداد 6 گره نیز در لایه پنهان انتخاب شد.

جدول (۳-۶) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره‌های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	۲	۰/۳۵۴	۰/۱۷۳	۰/۰۶۶	۰/۰۶۶	۰/۰۳۵	۰/۲۲۳	۰/۰۸۹	۰/۰۸۱	۰/۳۵۹	۰/۱۰۶
	۳	۰/۴۱۱	۵/۱۵۵	۰/۱۶۹	۰/۰۵۸	۰/۴۱۱	۰/۳۶۱	۰/۰۴۴	۰/۱۷۴	۰/۲۶۷	۰/۳۳۹
	۴	۰/۵۷۰	۳/۰۴۹	۰/۱۸۹	۰/۷۰۰	۰/۳۱۷	۴/۶۳۴	۰/۴۱۰	۰/۲۳۴	۰/۲۴۸	۰/۲۶۰
	۵	۰/۵۶۴	۲/۱۵۸	۱/۰۹۶	۰/۱۵۷	۰/۶۴۴	۰/۲۵۳	۰/۵۸۱	۰/۳۳۹	۰/۰۸۲	۰/۲۱۷
	۶	۰/۵۹۲	۰/۷۲۱	۰/۱۱۸	۰/۴۹۸	۰/۰۵۸	۰/۱۴۸	۰/۳۶۷	۰/۳۱۶	۰/۶۱۰	۰/۳۹۲
	۷	۰/۵۹۳	۰/۷۲۰	۰/۸۳۵	۰/۷۱۲	۰/۳۵۶	۰/۴۳۷	۰/۳۲۷	۰/۰۸۵	۰/۰۷۹	۰/۲۲۳
	۸	۰/۵۳۰	۰/۵۴۳	۰/۷۴۷	۰/۶۲۲	۰/۶۵۵	۰/۴۹۳	۰/۰۸۵	۰/۰۴۲	۰/۱۹۰	۰/۵۵۵
	۹	۰/۸۵۶	۰/۵۰۹	۰/۶۷۷	۲/۵۷۸	۰/۶۱۹	۰/۲۹۴	۰/۰۹۸	۰/۱۲۰	۰/۲۴۲	۰/۰۹۷
	۱۰	۰/۵۵۲	۰/۷۱۲	۰/۲۵۲	۰/۴۱۱	۰/۱۹۵	۰/۳۹۳	۰/۰۳۲	۰/۰۸۲	۰/۰۴۷	۰/۰۹۲
	۱۱	۰/۵۴۵	۳/۴۴۵	۲/۱۳۲	۰/۵۷۹	۰/۴۰۹	۰/۶۱۶	۰/۱۷۰	۰/۱۶۰	۰/۰۳۹	۰/۲۵۶

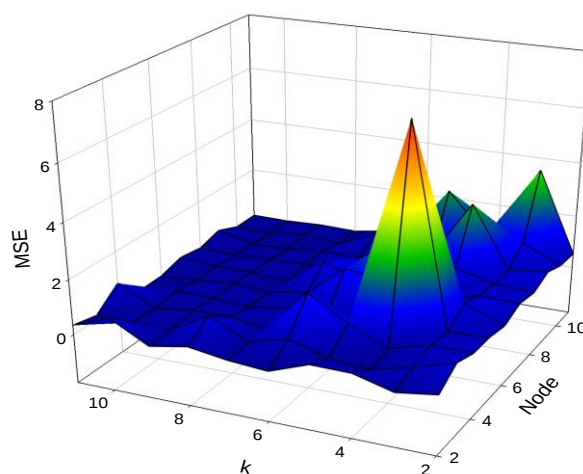


شکل (۳-۲) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (۷-۳) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره‌های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	2	۰/۳۵۴	۰/۱۴۵	۰/۵۳۱	۰/۵۵۱	۰/۰۳۵	۰/۱۴۷	۰/۳۶۲	۰/۱۱۰	۰/۸۶۷	۰/۳۴۷
	3	۰/۸۵۰	۰/۴۱۷	۰/۰۴۷	۰/۱۸۰	۰/۶۰۳	۰/۲۳۳	۰/۹۱۶	۰/۱۴۳	۰/۱۴۷	۰/۱۵۹
	4	۰/۵۸۷	۰/۶۰۷	۰/۲۰۵	۰/۸۵۴	۱/۶۹۵	۰/۰۴۱	۰/۳۳۳	۰/۳۰۹	۰/۱۲۶	۰/۸۳۲
	5	۰/۵۷۴	۰/۵۱۶	۷/۵۵۱	۰/۱۹۱	۰/۳۳۳	۰/۳۳۳	۰/۱۵۰	۰/۲۲۶	۰/۲۶۱	۰/۰۵۲
	6	۰/۴۳۰	۰/۹۸۶	۰/۱۳۶	۱/۳۹۹	۱/۵۲۶	۰/۳۹۲	۰/۱۷۱	۰/۲۶۱	۰/۱۲۹	۰/۰۷۸
	7	۰/۷۰۳	۰/۷۵۵	۰/۲۱۱	۲/۵۶۵	۰/۳۸۱	۱/۰۹۹	۰/۰۳۶	۰/۱۷۹	۰/۱۹۲	۰/۳۱۱
	8	۰/۵۷۲	۰/۶۷۸	۱/۰۳۶	۰/۲۹۱	۰/۹۸۲	۰/۳۱۸	۰/۲۲۱	۰/۱۸۳	۰/۱۳۲	۰/۱۹۱
	9	۰/۸۴۱	۰/۶۹۲	۲/۹۴۶	۰/۸۴۵	۰/۸۹۸	۰/۲۴۱	۰/۲۴۹	۰/۱۱۰	۰/۰۸۶	۰/۴۱۹
	10	۰/۵۸۱	۱/۲۱۱	۱/۲۶۹	۲/۷۹۱	۰/۳۱۵	۰/۶۷۴	۰/۰۷۳	۰/۲۲۵	۰/۰۴۳	۰/۰۹۹
	11	۰/۵۴۳	۳/۴۴۵	۱/۱۵۵	۰/۷۴۱	۰/۲۶۹	۰/۳۲۴	۰/۱۰۴	۰/۱۴۲	۰/۰۹۱	۰/۱۱۳

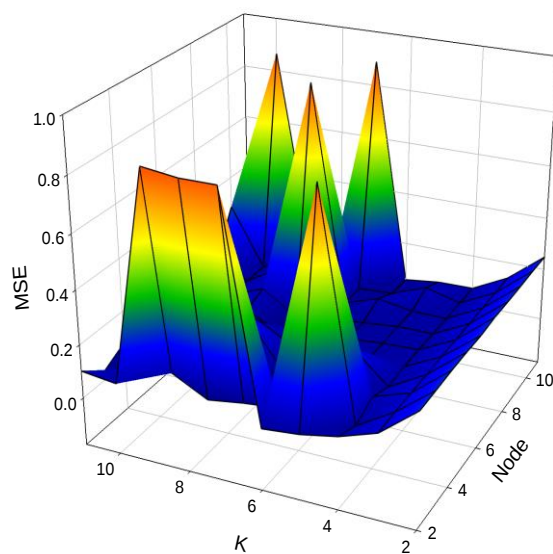


شکل (۳-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی

جدول (۸-۳) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

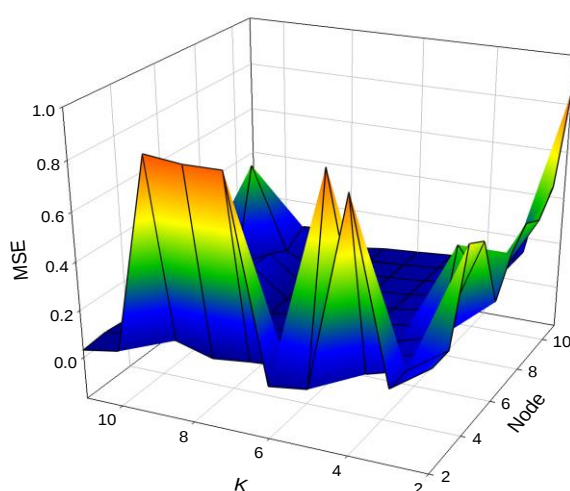
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	2	۰/۲۴۷	۰/۱۳۶	۰/۰۸۹	۰/۰۶۲	۰/۰۵۰	۰/۰۸۵۹	۰/۰۸۵۹	۰/۰۸۷۵	۰/۰۸۶	۰/۰۱۰۳
	3	۰/۲۴۸	۰/۱۳۸	۰/۰۵۶	۰/۰۸۶۷	۰/۰۵۰	۰/۰۴۶	۰/۰۲۶	۰/۰۹۱	۰/۰۷۱	۰/۰۳۰
	4	۰/۲۵۱	۰/۱۳۳	۰/۰۹۲	۰/۰۶۵	۰/۰۵۹	۰/۰۴۲	۰/۰۲۶	۰/۰۴۶	۰/۰۷۰	۰/۰۷۰
	5	۰/۲۵۰	۰/۱۳۱	۰/۰۶۰	۰/۰۱۷۸	۰/۰۵۳	۰/۰۴۸	۰/۰۲۱۷	۰/۰۴۹	۰/۰۵۰	۰/۰۳۹۸
	6	۰/۲۴۶	۰/۱۳۳	۰/۰۶۹	۰/۰۶۴	۰/۰۵۸	۰/۰۴۰	۰/۰۱۱۵	۰/۰۱۰۹	۰/۰۵۱	۰/۰۸۳
	7	۰/۲۵۱	۰/۱۳۸	۰/۰۶۹	۰/۰۶۸	۰/۰۵۱	۰/۰۴۵	۰/۰۲۴	۰/۰۷۱	۰/۰۵۸	۰/۰۱۲۴
	8	۰/۲۴۵	۰/۱۳۱	۰/۰۶۱	۰/۰۶۳	۰/۰۵۱	۰/۰۴۷	۰/۰۲۶	۰/۰۱۰۲	۰/۰۵۷	۰/۰۹۱
	9	۰/۲۴۹	۰/۱۳۱	۰/۰۶۳	۰/۰۶۸	۰/۰۸۵	۰/۰۴۴	۰/۰۸۷۴	۰/۰۱۲۲	۰/۰۵۶	۰/۰۱۰۵
	10	۰/۲۴۹	۰/۱۳۱	۰/۰۶۸	۰/۰۱۰۶	۰/۰۵۷	۰/۰۴۳	۰/۰۲۳	۰/۰۱۱۳	۰/۰۵۶	۰/۰۲۸۲
	11	۰/۲۴۸	۰/۱۳۷	۰/۰۶۸	۰/۰۶۶	۰/۰۵۰	۰/۰۸۸۹	۰/۰۲۲	۰/۰۱۲۳	۰/۰۸۶۸	۰/۰۱۱۳



شکل (۳-۴) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول (۳-۹) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	2	۰/۲۴۸	۰/۱۳۲	۰/۸۴۸	۰/۰۶۲	۰/۰۴۱	۰/۸۵۵	۰/۸۵۴	۰/۸۷۰	۰/۰۵۷	۰/۰۳۰
	3	۰/۲۴۸	۰/۱۳۳	۰/۰۸۶	۰/۸۶۴	۰/۰۵۱	۰/۰۳۹	۰/۰۲۱	۰/۰۶۶	۰/۰۵۴	۰/۰۴۱
	4	۰/۶۰۰	۰/۱۳۳	۰/۰۸۸	۰/۰۶۴	۰/۰۵۲	۰/۰۴۰	۰/۰۲۵	۰/۱۱۵	۰/۰۵۴	۰/۰۳۱
	5	۰/۵۵۲	۰/۱۳۳	۰/۰۸۹	۰/۰۶۴	۰/۰۵۱	۰/۰۴۰	۰/۰۲۱	۰/۱۴۹	۰/۰۵۴	۰/۰۶۰
	6	۰/۲۴۷	۰/۴۵۴	۰/۰۸۶	۰/۰۶۴	۰/۰۵۲	۰/۰۴۱	۰/۰۲۰	۰/۱۵۵	۰/۰۵۴	۰/۰۶۳
	7	۰/۳۸۸	۰/۱۳۳	۰/۰۸۹	۰/۰۶۵	۰/۰۵۴	۰/۰۴۱	۰/۰۲۰	۰/۱۳۸	۰/۰۵۴	۰/۰۴۱
	8	۰/۴۵۶	۰/۱۳۳	۰/۰۸۹	۰/۰۶۵	۰/۰۵۲	۰/۰۴۱	۰/۰۲۰	۰/۱۳۹	۰/۰۵۴	۰/۰۶۵
	9	۰/۴۱۴	۰/۱۳۳	۰/۰۸۷	۰/۰۶۵	۰/۰۵۴	۰/۰۴۱	۰/۰۲۰	۰/۰۹۳	۰/۴۵۸	۰/۰۹۲
	10	۰/۵۰۲	۰/۱۳۳	۰/۰۹۱	۰/۰۶۵	۰/۰۵۱	۰/۰۴۱	۰/۰۲۰	۰/۱۵۱	۰/۰۵۴	۰/۰۸۰
	11	۰/۸۵۳	۰/۱۳۳	۰/۰۸۹	۰/۰۶۵	۰/۰۵۳	۰/۰۴۲	۰/۰۲۰	۰/۰۸۹	۰/۰۵۴	۰/۰۷۷



شکل (۳-۵) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی

طبق نتایج به دست آمده ۸ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط با فعالیت بازدارندگی دارند، انتخاب شدند که جدول (۳-۱۰) لیست این توصیفگرها به همراه نام و گروه آن‌ها را نشان می‌دهد.

جدول (۳-۱۰) - توصیفگرهای انتخاب شده توسط شبکه عصبی

No	Symbol	Class	Meaning
۱	Mv	Constitutional descriptors	Mean atomic van der Waals volume
۲	BELm6	BCUT	Lowest eigenvalue n.6 of Burden
۳	GGI3	Charge indices	Topological charge index of order 3
۴	PJI2	Topological descriptors	2D Petitjean shape index
۵	JGI1	Galvez topol.charge indices	Mean topological charge index
۶	R5u	GETAWAY	R autocorrelation of lag 5/ unweighted
۷	Mor32m	3D-MoRSE descriptors	3D-MoRSE- signal32/weighted by atomic masses
۸	Mor30m	3D-MoRSE descriptors	3D-MoRSE- signal30/weighted by atomic masses

۳-۱-۶-۳- انتخاب تعداد دوره‌های آموزش

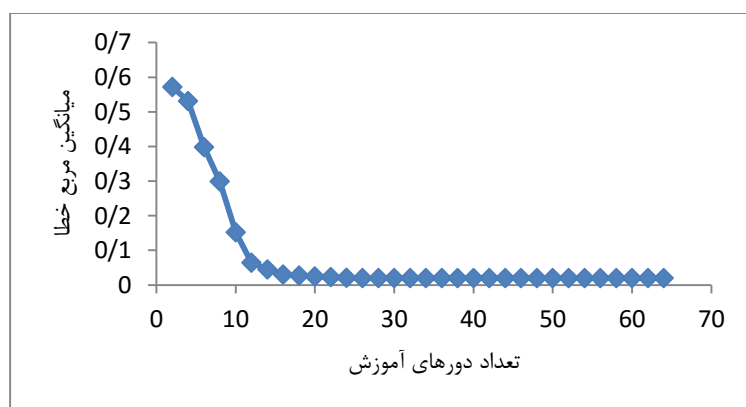
در طی آموزش شبکه، دقت تقریبی ANN برای داده‌های سری آموزش افزایش می‌یابد در حالی که توانایی پیش‌بینی برای ورودی‌های خارجی^۱ به شدت کاهش می‌یابد که این حالت را آموزش بیش از حد^۲ می‌گویند. برای جلوگیری از این حالت، از تکنیک توقف زودرس استفاده شد. بدین منظور با ۸ توصیفگر منتخب و قرار دادن ۶ گره در لایه پنهان، در الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی تعداد چرخه آموزش تغییر داده شد و برای هر چرخه میانگین مربع خطاهای سری ارزیابی برآورد شد (جدول (۳-۱۱)). سپس منحنی تعداد چرخه‌های آموزش بر حسب میانگین مربع خطاها بر طبق شکل (۳-۶) ترسیم شد. مقدار بهینه تعداد تکرار جایی است که میزان این خطا حداقل باشد. در نهایت مشاهده گردید تعداد دوره‌های آموزش ۲۶ دارای کمترین میانگین مربع خطا می‌باشد که به عنوان مقدار بهینه انتخاب شد.

جدول (۳-۱۱) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف

^۱ - External inputs

^۲ - Overtraining

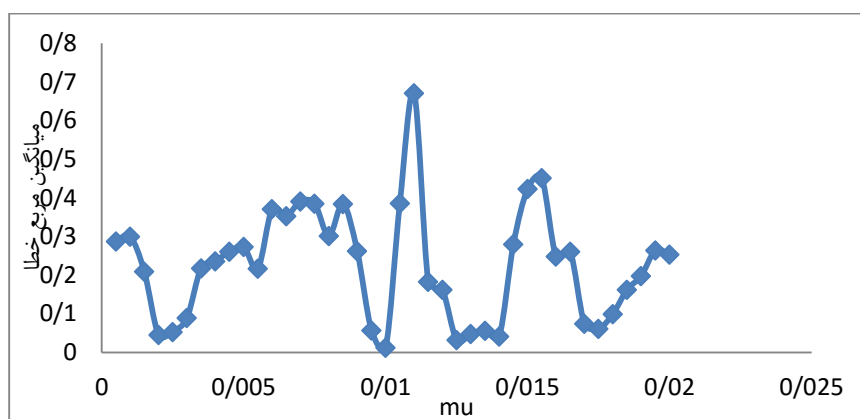
تعداد تکرار	میانگین مربع خطا	تعداد تکرار	میانگین مربع خطا
2	0/572	34	0/020
4	0/531	36	0/020
6	0/398	38	0/020
8	0/299	40	0/020
10	0/152	42	0/020
12	0/0645	44	0/020
14	0/0452	46	0/020
16	0/030	48	0/020
18	0/027	50	0/020
20	0/025	52	0/020
22	0/023	54	0/020
24	0/021	56	0/020
26	0/020	58	0/020
28	0/020	60	0/020
30	0/020	62	0/020
32	0/020	64	0/020



شکل (۳-۶) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دوره‌های آموزش متفاوت

۳-۱-۶-۴- بهینه کردن مقدار μ

جهت بهینه کردن مقدار μ ، ساختار شبکه با ۸ متغیر ورودی، ۶ گره در لایه پنهان و الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۲ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطای سری آموزش به روش ارزیابی تقاطعی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۳-۱۲) و شکل (۳-۷) مقدار بهینه ۰/۰۱ به دست آمد.



شکل (۳-۷) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ

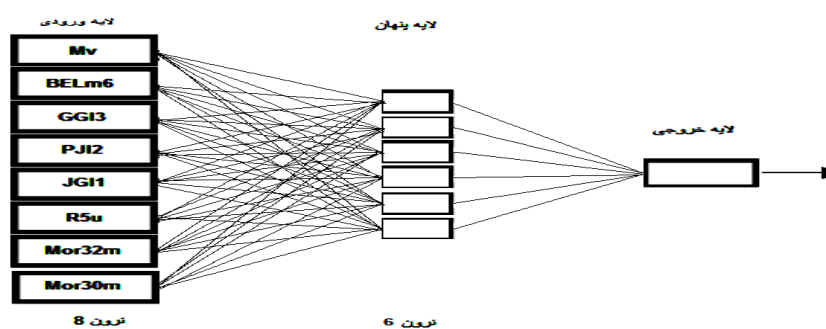
جدول (۳-۱۲) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ

μ	میانگین مربع خطا	μ	میانگین مربع خطا	μ	میانگین مربع خطا
0/0005	0/357	0/0075	0/385	0/0145	0/128
0/001	0/292	0/008	0/017	0/015	0/292
0/0015	0/175	0/0085	0/167	0/0155	0/289
0/002	0/032	0/009	0/182	0/016	0/290
0/0025	0/089	0/0095	0/049	0/0165	0/292
0/003	0/505	0/01	0/357	0/017	0/020
0/0035	0/027	0/0105	0/252	0/0175	0/055

0/004	0/298	0/011	0/346	0/018	0/316
0/0045	0/362	0/0115	0/303	0/0185	0/182
0/005	0/280	0/012	0/297	0/019	0/168
0/0055	0/280	0/0125	0/035	0/0195	0/173
0/006	0/434	0/013	0/036	0/02	0/175
0/0065	0/438	0/0135	0/037		
0/007	0/387	0/014	0/037		

۳-۱-۶-۵- معماری شبکه عصبی بهینه شده

با توجه به روند بهینه سازی که در قسمت های قبلی ذکر شد، شبکه عصبی مصنوعی که بتواند فعالیت ضد روان پریشی مشتقات ترکیب فنیل آلکیل آمین را پیش بینی کند، دارای ۸ توصیفگر، ۶ گره در لایه پنهان، تابع آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی است. شکل (۳-۸) ساختار شبکه عصبی بهینه نهایی را نشان می دهد. توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۳-۱۳) ارائه شده است.



شکل (۳-۸) - ساختار شبکه عصبی بهینه شده

جدول (۳-۱۳) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

تابع آموزش	Train br
------------	----------

Tansig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
۶	تعداد نرون لایه پنهان
۸	تعداد متغیرهای ورودی
۲۶	تعداد دوره‌های آموزش
0/01	mu پارامتر

۳-۱-۷- انتخاب متغیر به روش الگوریتم ژنتیک

در فصل دوم مفاهیم و اصول الگوریتم ژنتیک مطرح گردید. در اینجا مفاهیم کلی آن به اختصار شرح داده می‌شود.

الگوریتم ژنتیک برای حل یک مسئله مجموعه بسیار بزرگی از راه حل‌های ممکن را تولید می‌کند. هر یک از این راه حل‌ها با استفاده از یک “تابع تناسب” مورد ارزیابی قرار می‌گیرند. فضای جستجو در جهتی تکامل پیدا می‌کند که به راه حل مطلوب برسد. در صورت انتخاب صحیح پارامترها، این روش می‌تواند بسیار موثر عمل نماید.

در ابتدا یک جمعیت تصادفی از کروموزوم‌ها تولید می‌شود. سپس تناسب هر کروموزوم در جمعیت ارزیابی می‌گردد. مراحل ایجاد جمعیت جدید شامل موارد زیر است.

- انتخاب: انتخاب دو کروموزوم والد از جمعیت موجود بر طبق تناسب آن‌ها.
- پیوند: از ترکیب والدین یک فرزند جدید تولید می‌شود و اگر پیوندی صورت نگیرد فرزند دقیقاً کپی یکی از والدین می‌باشد.
- جهش: ممکن است با یک احتمال در هر یک از کروموزوم‌های فرزند جهش ایجاد شود.
- پذیرش: فرزند جدید در یک جمعیت جدید پذیرفته می‌شود.

- جایگزینی: جمعیت جدید تولید شده جایگزین جمعیت قبلی می شود تا الگوریتم بار دیگر تکرار شود.

- تست: اگر به شرط پایان الگوریتم رسیده باشیم اجرای برنامه متوقف می شود و بهترین راه حل از جمعیت جاری برگردانده می شود.

قابل ذکر است که الگوریتم ژنتیک هیچ چیزی درباره‌ی مسأله‌ای که حل می کند، نمی داند و برای اجرای جستجوی موثر فقط به مقادیر تابع هدف نیاز دارد، بنابراین شامل پیچیدگی‌های محاسباتی اضافی نمی شود، اما از آنجا که روشی غیر جبری است پاسخ دقیق مسأله را نیافته و حتی ممکن است برای حل یک مسأله‌ی مشخص با هر با بکارگیری، پاسخی متفاوت ارائه دهد، اگرچه تمامی این پاسخ-ها می توانند پاسخ‌هایی باشند که دقت مورد نظر را برآورده کنند. تاکنون هیچ روش استاندارد برای حل یک مسأله بخصوص در دسترس نبوده و با بکارگیری تجربیات موجود می توان الگوریتم ژنتیک را با احتیاجات خود منطبق کرد. بهینه کردن پارامترهای الگوریتم ژنتیک با حدس های اولیه‌ای که منجر به اصلاح بیشتر ساختار آن شود، آغاز شده و روش‌های مختلفی برای رسیدن به الگوریتم بهینه بکار گرفته می شود.

همانطور که پیشتر توضیح داده شد، هدف از اجرای الگوریتم حاضر شناسایی ترکیبی از توصیفگرهاست، به گونه‌ای که برای پیش‌بینی pIC50 مشتقات آریل الکوکسی فنیل آلکیل آمین بوسیله شبکه عصبی مصنوعی مورد استفاده قرار گیرند. در فرایند اجرای این الگوریتم، ابتدا به صورت اتفاقی یک جمعیت اولیه از توصیف کننده‌ها تولید می شود، برازندگی یا شایستگی تک تک اعضای هر نسل یعنی توصیف کننده‌ها، محاسبه و با توجه به شایستگی‌ها، نسل‌های بعدی با اعمال سه عملگر جمعیت، پیوند و جهش تولید یا باز ترکیب می شوند. این روند جستجو برای جمعیت‌های مختلف تا حصول ملاک خاتمه ادامه می یابد. لازم به ذکر است که برنامه استفاده شده در الگوریتم ژنتیک یک

برنامه بهینه شده می‌باشد. مقادیر بهینه شده‌ی پارامترهای جمعیت، پیوند و جهش به ترتیب ۳۰، ۵۰٪ و ۱٪ می‌باشد.

با هر بار اجرای برنامه ۱۰۰ کروموزوم که بهترین پاسخ را بدهند انتخاب می‌شوند.

تا اتمام روند الگوریتم ژنتیک n بار برنامه اجرا می‌شود. در نیمه‌ی اول اجرای برنامه بردار Y یا تابع متغیرهای وابسته همان مقادیر اصلی در نظر گرفته می‌شود و در نیمه دوم یک بردار Y به روش اتفاقی تولید می‌شود. فرض می‌کنیم برنامه ۴۰ بار اجرا شود. در پایان ۲ بردار خواهیم داشت که اولی شامل میانگین ۲۰ اجرای اول الگوریتم بر روی مقادیر اصلی، و دومی شامل میانگین ۲۰ اجرای دوم الگوریتم بر روی مقادیر تصادفی خواهد بود. به طور ساده می‌توان گفت که اجرای الگوریتم ژنتیک با بردار Y اصلی نشان دهنده توانایی الگوریتم مورد نظر برای مدل سازی "اطلاعات همراه با مقادیر اشتباه" است. و اجرای الگوریتم برای بردار Y تصادفی نیز نشان دهنده توانایی الگوریتم برای مدل سازی مقادیر "اشتباه" می‌باشد. حال تفاوت این دو^۳ می‌تواند توانایی الگوریتم ژنتیک را در ارتباط با مقادیر حقیقی نشان دهد. حسن عملکرد الگوریتم ژنتیک به وسیله ریشه متوسط مربع خطا (RMSE)^۴ در پیش‌بینی مقادیر محاسبه می‌گردد. RMSE چنین محاسبه می‌گردد.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

در فرمول ارائه شده n تعداد مقادیر موجود در سری مورد محاسبه است. توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک برای سری داده‌ها در جدول (۳-۱۴) آورده شده است.

جدول (۳-۱۴) - کل توصیفگرهای انتخاب شده الگوریتم ژنتیک

No	Symbol	Class	Meaning
----	--------	-------	---------

^۱-Information+noise

^۲-Noise

^۳-Maximum difference

^۴-Root mean square error

۱	BELm6	BCUT	Lowest einvalue n.6 of Burden
۲	Mv	Contitutionam descriptors	Mean atomic van der Waals volume
۳	X4sol	Topological descriptors	Salvation connectivity index chi-4
۴	PJI2	Topological descriptors	2D Petitjean shape index
۵	RDF030m	RDF descriptors	Radial Distribution Function-3.0/weighted by atomic masses
۶	GGI3	Charge indices	Topological charge index of order 3
7	RTu	GETAWAY	R total index / unweighted
8	ISH	GETAWAY	Standardized information content on the leverage equality
9	Gm	WHIM descriptors	Gtotal symmetry index / weighted by atomic masses

۳-۱-۸- مدل شبکه عصبی مصنوعی با استفاده از توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک

در این بخش نیز از شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی استفاده شد.

۳-۱-۸-۱- انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع

آموزش و نوع تابع انتقال

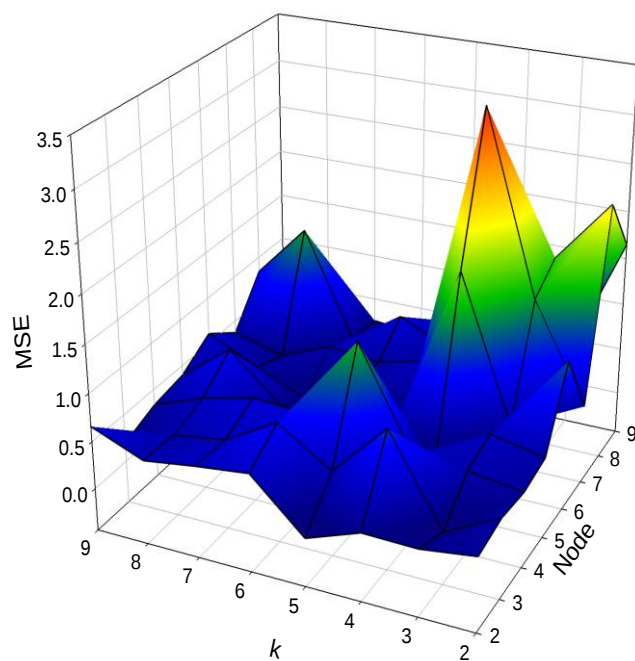
پس از طراحی اولیه شبکه، برای اینکه شبکه قدرت پیشگویی مناسبی داشته باشد باید به طور مناسب آموزش دیده باشد و با توجه به اینکه فرایند آموزش وقتگیر است، انتخاب الگوریتم آموزشی سریعتر امری مهم است. از میان الگوریتم‌های آموزشی متفاوت که در جعبه ابزار مطلب وجود دارد، الگوریتم لونیگ - مارکوارت و تنظیم بایزین دارای سرعت مناسب، قابلیت پیشگویی و تعمیم پذیری بهتری نسبت به سایر روش‌های معمول مانند روش نزول گرادیانی می‌باشند. بر این اساس شبکه با

ورودی‌های از ۲ تا ۹ توسط دو الگوریتم آموزشی لونبرگ - مارکوارت و تنظیم بایزین، با تعداد متفاوت گره در لایه پنهان از ۲ تا ۹ و همچنین توابع لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig)، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمامی این شبکه‌ها تعداد دوره‌های آموزش ثابت و برابر ۳۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده شد معیار نیز به حداقل رساندن میانگین مربع خطا (MSE) در نظر گرفته شد که نتایج این محاسبات در جداول (۳-۱۵)، (۳-۱۶)، (۳-۱۷) و (۳-۱۸) گرد آوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۳-۹)، (۳-۱۰)، (۳-۱۱) و (۳-۱۲) ارائه شده است.

با توجه به نتایج به دست آمده، الگوریتم آموزشی تنظیم بایزین نسبت به الگوریتم آموزشی لونبرگ- مارکوارت دارای میانگین مربع خطا (MSE) کمتری است. علاوه بر آن دارای قدرت تعمیم-پذیری بالاتری است. همچنین تابع تانژانت سیگموئیدی دارای خطای کمتری نسبت به لگاریتم سیگموئیدی است. بنابراین تابع انتقال تانژانت سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده ۵ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط با فعالیت بازدارندگی دارند، انتخاب شدند. تعداد ۴ گره نیز در لایه پنهان انتخاب شد.

جدول (۳-۱۵)- مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

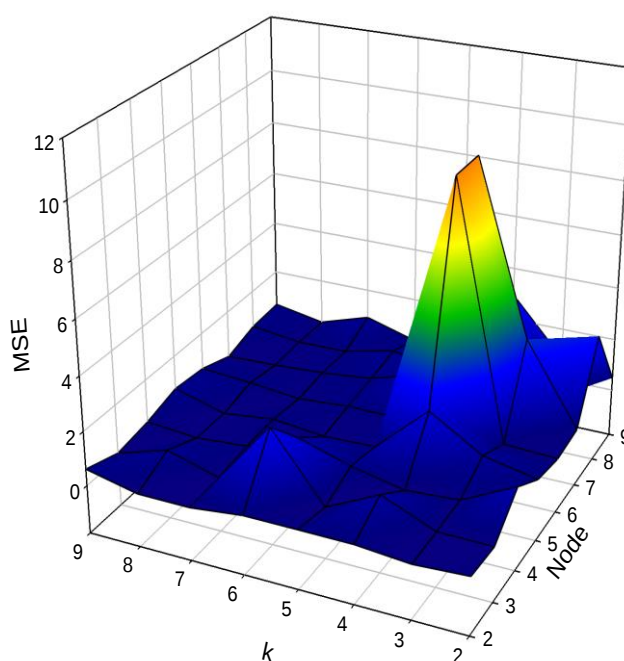
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره‌های لایه پنهان		2	3	4	5	6	7	8	9
	2	0/934	0/206	0/197	0/448	0/157	0/273	0/607	0/607
	3	1/093	0/274	1/718	0/634	0/791	3/606	0/426	0/442
	4	2/054	13/324	1/165	8/843	0/560	0/591	0/712	0/669
	5	1/341	0/202	4/808	0/712	0/4480	0/452	0/241	0/475
	6	1/877	2/969	0/668	0/795	0/664	0/373	0/848	0/211
	7	0/741	1/902	0/898	0/650	1/806	0/554	0/940	0/408
	8	7/863	2/903	3/6289	1/744	1/477	0/281	0/242	0/765
	9	0/496	5/219	0/139	0/084	3/8676	1/082	0/452	0/253



شکل (9-3) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (3-16) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

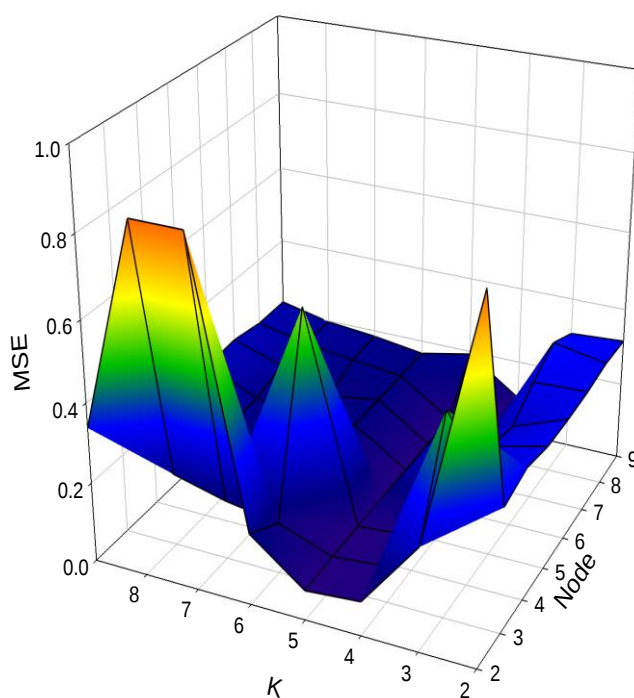
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	0/685	0/391	0/162	0/155	0/177	0/697	0/619	0/313
	3	0/688	0/294	0/214	0/907	0/800	0/670	0/403	0/777
	4	0/806	12/239	1/421	0/655	0/663	0/818	1/347	0/559
	5	0/610	0/208	3/028	0/395	1/190	0/242	0/579	0/218
	6	1/272	5/236	4/997	0/323	2/831	0/258	0/224	0/358
	7	0/580	1/276	1/100	1/244	0/843	0/515	0/851	0/216
	8	8/586	3/911	2/650	2/776	0/733	0/267	0/238	0/413
	9	1/024	3/278	0/205	0/343	4/099	6/261	0/244	0/211



شکل (3-10) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی

جدول (۳-۱۷) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

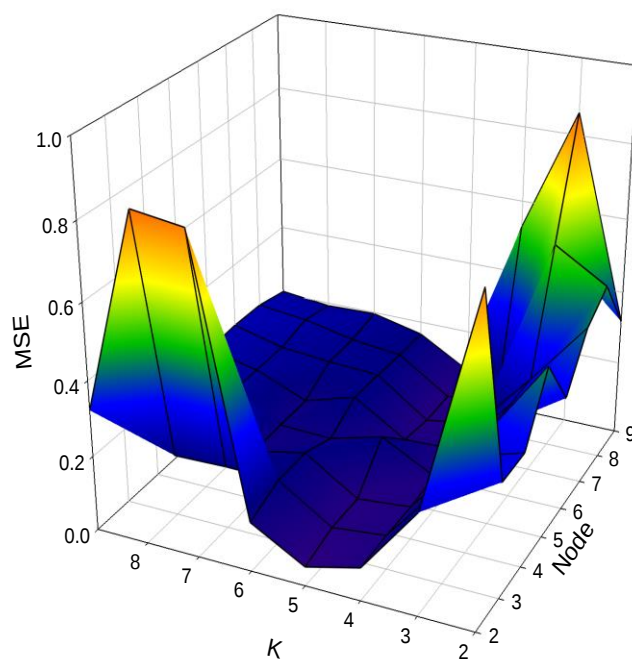
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	0/755	0/282	0/219	0/162	0/228	0/859	0/861	0/866
	3	0/752	0/297	0/219	0/157	0/245	0/384	0/352	0/331
	4	0/749	0/847	0/263	0/192	0/240	0/233	0/342	0/308
	5	0/778	0/464	0/346	0/179	0/247	0/223	0/372	0/309
	6	0/768	0/422	0/445	0/178	0/238	0/252	0/356	0/360
	7	0/744	0/408	0/353	0/177	0/236	0/261	0/350	0/359
	8	0/763	0/437	0/315	0/177	0/234	0/355	0/355	0/280
	9	0/862	0/427	0/298	0/176	0/232	0/359	0/401	0/278



شکل (۳-۱۱) - نمودار سه بعدی شبکه‌هایی با تابع آموزشی بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول (۳-۱۸) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	0/751	0/292	0/222	0/159	0/241	0/866	0/854	0/858
	3	0/746	0/313	0/221	0/157	0/238	0/520	0/346	0/312
	4	0/746	0/889	0/221	0/152	0/250	0/470	0/314	0/307
	5	0/750	0/387	0/408	0/161	0/232	0/366	0/270	0/244
	6	0/741	0/568	0/408	0/177	0/236	0/214	0/275	0/275
	7	0/859	0/416	0/544	0/177	0/230	0/216	0/269	0/278
	8	0/739	0/413	0/224	0/182	0/236	0/220	0/278	0/277
	9	0/866	0/412	0/378	0/174	0/226	0/229	0/266	0/281



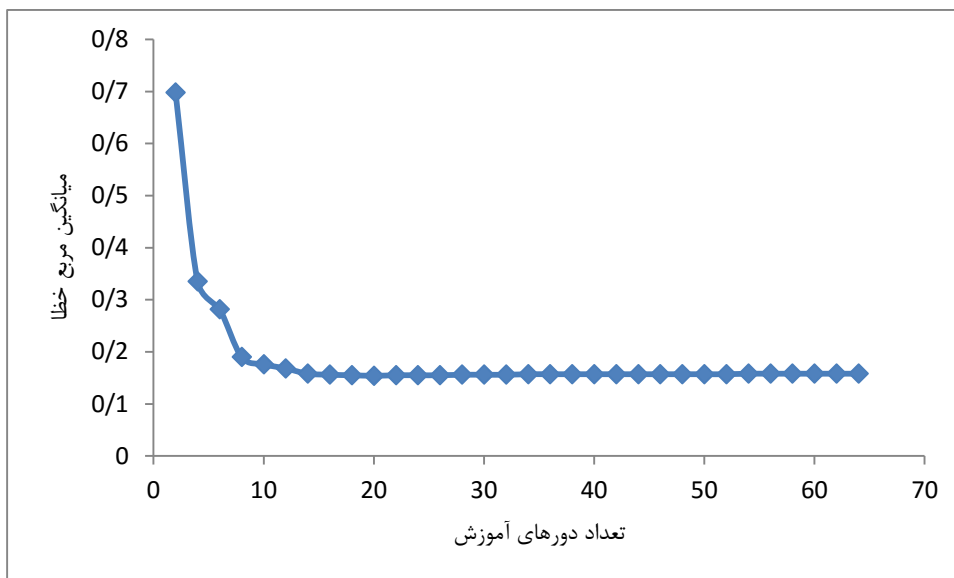
شکل (۳-۱۲)- نمودار سه بعدی شبکه‌هایی با تابع آموزشی بایزین و تابع انتقال تانژانت سیگموئیدی

۳-۱-۸-۲- انتخاب تعداد دورهای آموزش

در طی آموزش، شبکه با الگوریتم تنظیم بایزین، ۵ توصیفگر، ۴ گره در لایه پنهان و تابع انتقال تانژانت سیگموئیدی در تعداد دوره‌های متفاوت آموزش داده شد و میانگین مربع خطاها برای سری آموزش ثبت گردید. جدول (۳-۱۹) و شکل (۳-۱۳) نشان می‌دهد که مقدار میانگین مربع خطا (MSE) در تعداد دوره‌های آموزش ۲۰ به حداقل می‌رسد. بنابراین ۲۰ بار تکرار به عنوان تعداد تکرارهای بهینه برای یادگیری شبکه انتخاب شد.

جدول (۳-۱۹) – مقادیر میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش متفاوت

میانگین مربع خطا	تعداد دوره‌های آموزش	میانگین مربع خطا	تعداد دوره‌های آموزش
0/157	34	0/698	2
0/157	36	0/335	4
0/157	38	0/282	6
0/157	40	0/190	8
0/157	42	0/176	10
0/157	44	0/168	12
0/157	46	0/158	14
0/157	48	0/156	16
0/157	50	0/155	18
0/157	52	0/154	20
0/158	54	0/155	22
0/158	56	0/155	24
0/158	58	0/155	26
0/158	60	0/156	28
0/158	62	0/156	30
0/158	64	0/156	32



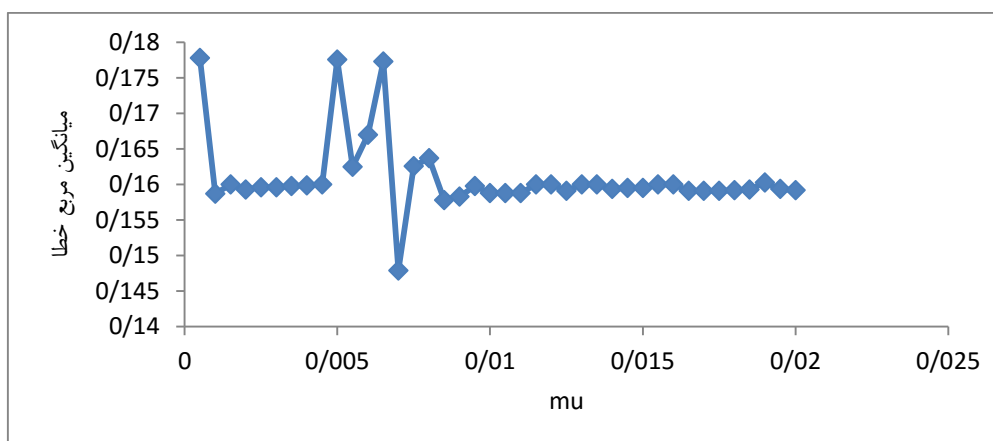
شکل (۳-۱۳) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش متفاوت

۳-۱-۸ - بهینه کردن مقدار μ

جهت بهینه کردن مقدار μ ، ساختار شبکه با ۵ متغیر ورودی، ۴ گره در لایه پنهان و الگوریتم آموزشی بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۵ تا ۰/۰۲ با گام‌های ۰/۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطای سری آموزش به روش ارزیابی تقاطعی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۳-۲۰) و شکل (۳-۱۴) مقدار بهینه ۰/۰۰۷ به دست آمد.

جدول (۳-۲۰) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ

μ	میانگین مربع خطا	μ	میانگین مربع خطا	μ	میانگین مربع خطا
0/0005	0/177	0/0075	0/162	0/0145	0/159
0/001	0/158	0/008	0/163	0/015	0/159
0/0015	0/160	0/0085	0/157	0/0155	0/160
0/002	0/159	0/009	0/158	0/016	0/160
0/0025	0/159	0/0095	0/159	0/0165	0/159
0/003	0/159	0/01	0/158	0/017	0/159
0/0035	0/159	0/0105	0/158	0/0175	0/159
0/004	0/159	0/011	0/158	0/018	0/159
0/0045	0/16	0/0115	0/16	0/0185	0/159
0/005	0/177	0/012	0/16	0/019	0/160
0/0055	0/162	0/0125	0/159	0/0195	0/159
0/006	0/167	0/013	0/16	0/02	0/159
0/0065	0/177	0/0135	0/16		
0/007	0/147	0/014	0/159		

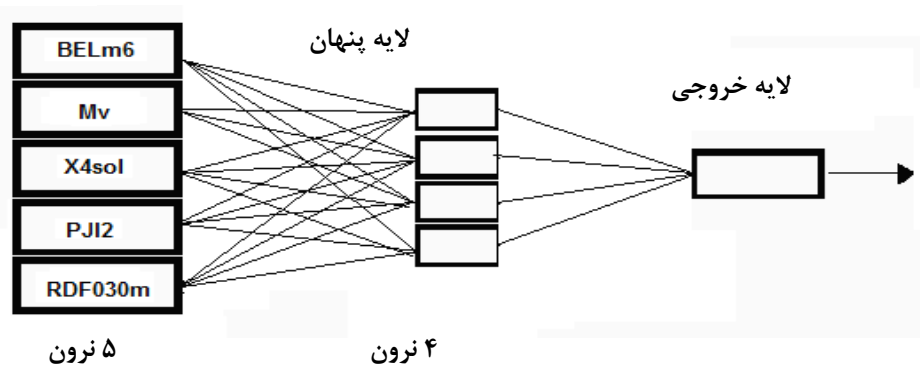


شکل (۳-۱۴) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد μ متفاوت

۳-۱-۸-۴- معماری شبکه عصبی بهینه شده

با توجه به روند بهینه سازی که در قسمت‌های قبلی ذکر شد، شبکه عصبی مصنوعی بهینه دارای ۵ توصیفگر، ۴ گره در لایه پنهان، تابع آموزشی بایزین و تابع انتقال تانژانت سیگموئیدی است. شکل (۳-۱۵) ساختار شبکه عصبی بهینه نهایی را نشان می‌دهد.

لایه ورودی



شکل (۳-۱۵) - ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۳-۲۱) ارائه شده است.

جدول (۳-۲۱) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

train br	تابع آموزش
tansig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
4	تعداد نرون لایه پنهان
۵	تعداد متغیرهای ورودی
۲۰	تعداد دوره‌های آموزش
0/007	پارامتر mu

۳-۱-۹- بررسی نتایج

اعتبار و اهمیت معادلات پیش‌بینی وقتی مشخص می‌گردد که خواص مولکول‌های مجهولی که در دسته‌ی آموزش موجود نیستند را پیش‌بینی کند. در اینجا چندین روش به منظور ارزیابی توانایی مدل‌های ارائه شده که دارای کمترین مقدار mse بودند، در پیش‌بینی داده‌های pIC50 ذکر شده است.

۳-۱-۹-۱- ارزیابی مدل‌ها با استفاده از سری ارزیابی

مدل‌های منتخب خطی و غیرخطی جهت پیش‌بینی فعالیت بازدارندگی سری ارزیابی که شامل ۸ ترکیب که بطور تصادفی انتخاب شده‌اند، به کارگرفته شد. پارامترهای آماری حاصل از مدل‌های خطی و غیر خطی با استفاده از شبکه عصبی بهینه شده و همین‌طور نتایج بدست آمده از الگوریتم ژنتیک و شبکه عصبی بهینه شده (GA-ANN) در جدول (۳-۲۲) آورده شده است. همچنین شکل (۳-۱۶)، ضریب تعیین مقدار پیش‌بینی شده بر حسب مقدار تجربی برای سری ارزیابی را با استفاده از هر سه مدل SR-MLR و SR-ANN و GA-ANN نشان می‌دهد. با توجه به نتایج بدست آمده می‌توان گفت روش SR-ANN دارای توانایی بیشتری برای پیش‌بینی دارد و خطای آن در مجموع کمتر از دو روش دیگر است. روش GA-MLR خطاهای بزرگتری نسبت به دو روش دیگر دارد. می‌توانیم نتیجه بگیریم برای این سری از داده‌ها روش خطی برای انتخاب متغیرها بهترین روش می‌باشد و برای مدل سازی نیز روش غیر خطی مناسبتر است.

۳-۱-۹-۲- ارزیابی مدل‌ها با استفاده از سری تست

بعد از پیش‌بینی فعالیت بازدارندگی سری ارزیابی، قدرت پیش‌بینی شبکه، بوسیله ۸ ترکیب که در سری تست قرار دارند مورد بررسی قرار گرفت. پارامترهای آماری حاصل از مدل‌های خطی و غیر خطی در جدول (۳-۲۳) آورده شده است. همچنین شکل (۳-۱۹)، ضریب تعیین مقدار پیش‌بینی

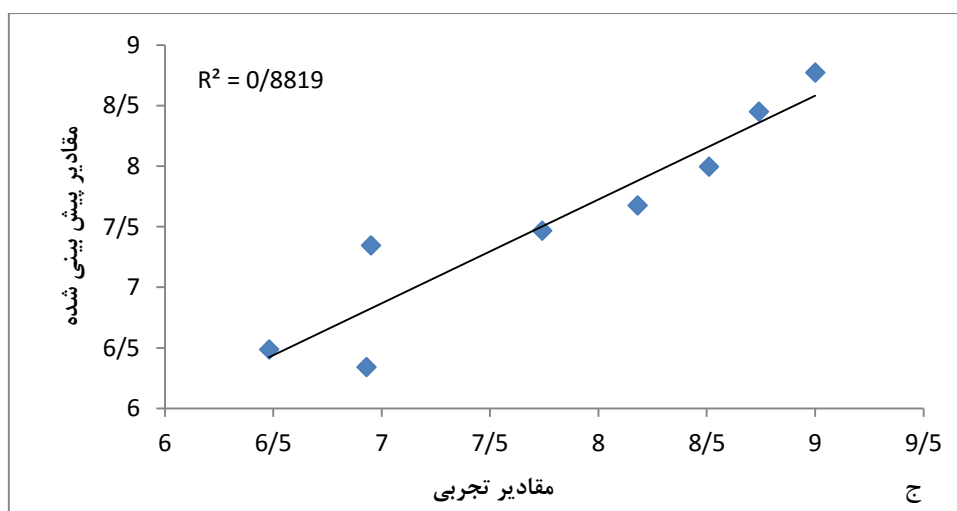
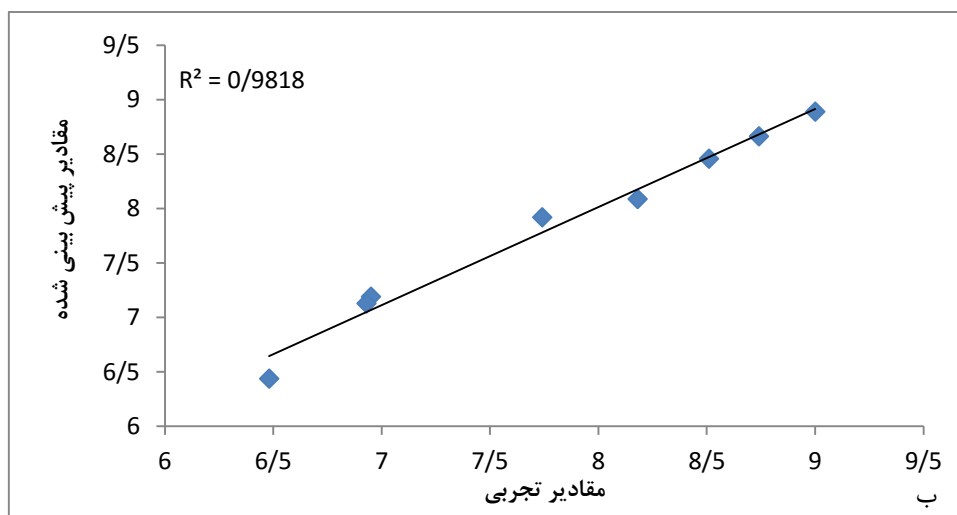
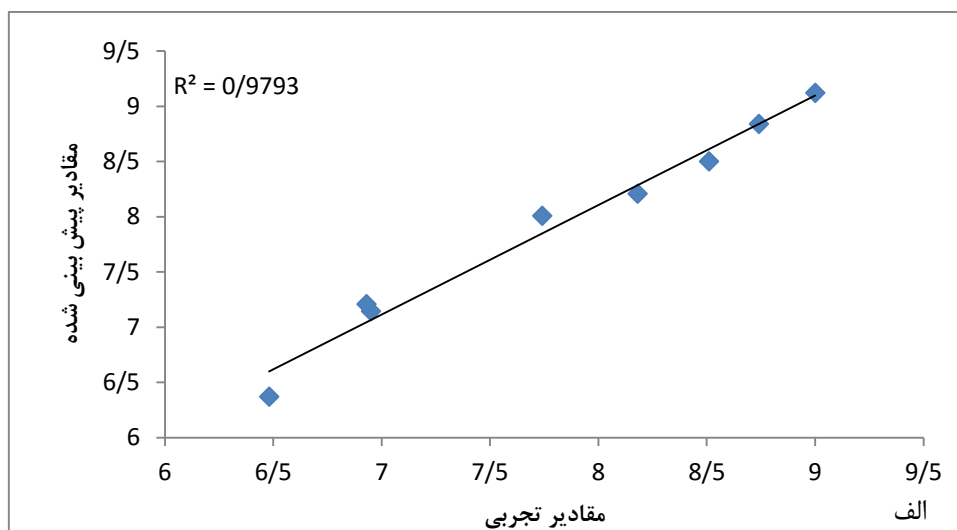
شده بر حسب مقدار تجربی برای سری تست را با استفاده از هر سه مدل SR-MLR و SR-ANN و GA-ANN نشان می‌دهد. در این مرحله نیز با توجه به نتایج سری تست فرض قبلی تایید می‌گردد.

جدول (۳-۲۲) - پارامترهای آماری مدل‌های SR-MLR، SR-ANN و GA-ANN برای سری ارزیابی

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده pIC50			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
3	8/18	8/21	8/09	7/67	-0/36	1/10	6/23
10	7/74	8/01	7/92	7/48	-3/48	-2/32	3/35
17	6/48	6/37	6/44	6/49	1/69	0/62	-0/15
19	6/95	7/14	7/19	7/35	-2/73	-3/45	-5/75
24	6/93	7/21	7/13	6/34	-4/04	-2/88	8/51
27	8/51	8/50	8/46	7/99	0/11	0/58	6/11
31	9/00	9/12	8/89	8/77	-1/33	1/22	2/55
38	8/74	8/84	8/66	8/45	-1/14	0/91	3/31

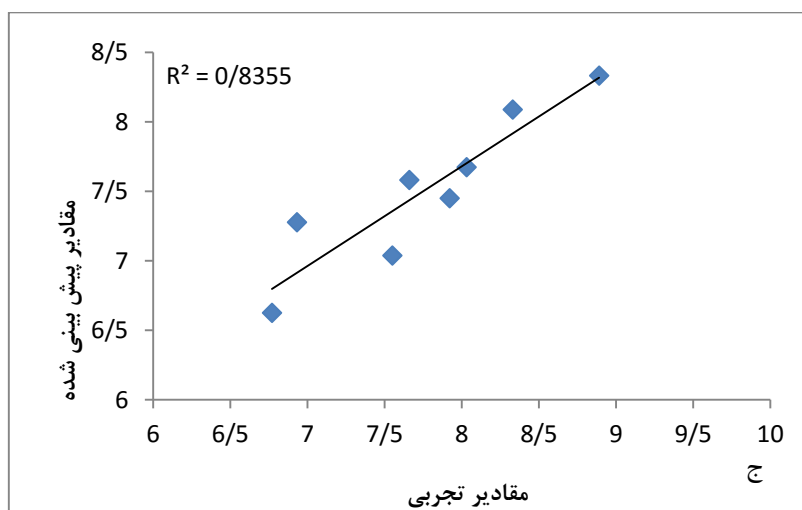
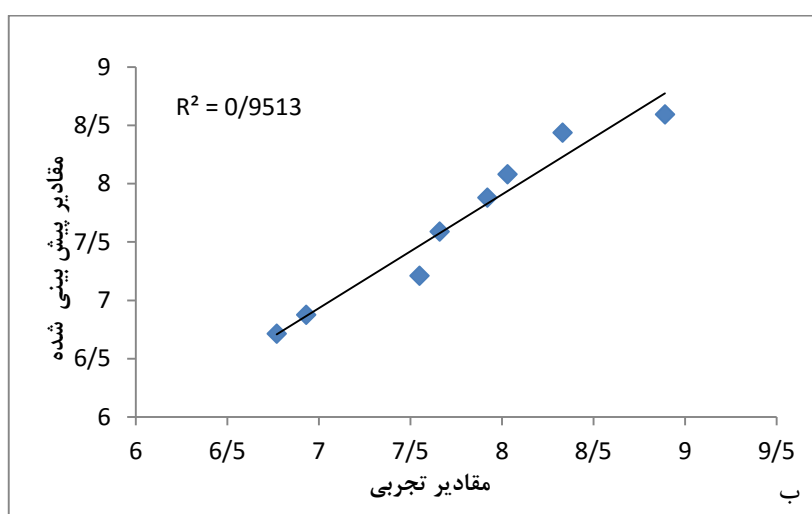
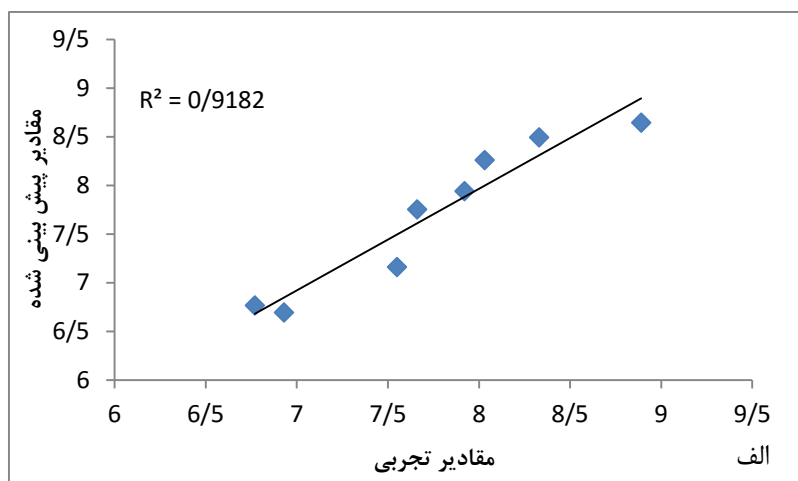
جدول (۳-۲۳) - پارامترهای آماری مدل‌های SR-MLR، SR-ANN و GA-ANN برای سری تست

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده pIC50			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
2	6/93	6/69	6/88	7/28	3/46	0/72	-5/05
5	8/89	8/64	8/59	8/33	2/81	3/37	6/29
7	7/92	7/94	7/88	7/45	-0/25	0/50	5/93
8	7/66	7/75	7/59	7/58	-1/17	0/91	1/04
14	7/55	7/16	7/21	7/04	5/16	4/50	6/75
25	6/77	6/76	6/71	6/62	0/15	0/89	2/22
33	8/03	8/26	8/08	7/67	-2/86	-0/62	4/48
35	8/33	8/49	8/44	8/09	-1/92	-1/32	2/88



شکل (۳-۱۶) - نمودار مقادیر پیش بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری ارزیابی، الف: با مدل SR-MLR

ب: با مدل SR-ANN ج: با مدل GA-ANN



شکل (۱۷-۳) - نمودار مقادیر پیش بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری تست الف: با مدل SR-MLR

ب: با مدل SR-ANN ج: با مدل GA-ANN

۳-۱-۹-۳- ارزیابی مدل‌های SR-MLR، SR-ANN و GA-ANN توسط روش رد

مرحله‌ای تک تک^۱

به منظور بررسی بیشتر قدرت پیش‌گویی مدل‌های خطی و غیرخطی تکنیک رد مرحله‌ای تک‌تک مورد استفاده قرار گرفت. در روش رد مرحله‌ای تک‌تک هر بار یکی از ترکیبات به طور کاملاً تصادفی از سری داده‌ها حذف شد و سپس با استفاده از مدل ساخته شده توسط بقیه ترکیبات، فعالیت دارویی ترکیب حذف شده، پیش‌بینی شد. این فرایند برای تمام اعضای سری داده‌ها تکرار شد. نتایج حاصل از رد مرحله‌ای در جدول (۳-۲۴) ارائه شده است. بررسی این نتایج نشان می‌دهد که روش SR-ANN در پیش‌بینی مقادیر pIC₅₀ این سری ترکیبات از دو روش دیگر موفق‌تر است. ضریب تعیین مشاهده شده بین مقادیر پیش‌بینی شده pIC₅₀ و مقادیر تجربی در شکل‌های (۳-۱۸)، نشان دهنده میزان نزدیکی فعالیت پیش‌بینی شده توسط مدل‌ها نسبت به مقدار تجربی آن است.

جدول (۳-۲۴)- نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-MLR، SR-ANN و GA-ANN با استفاده از رد

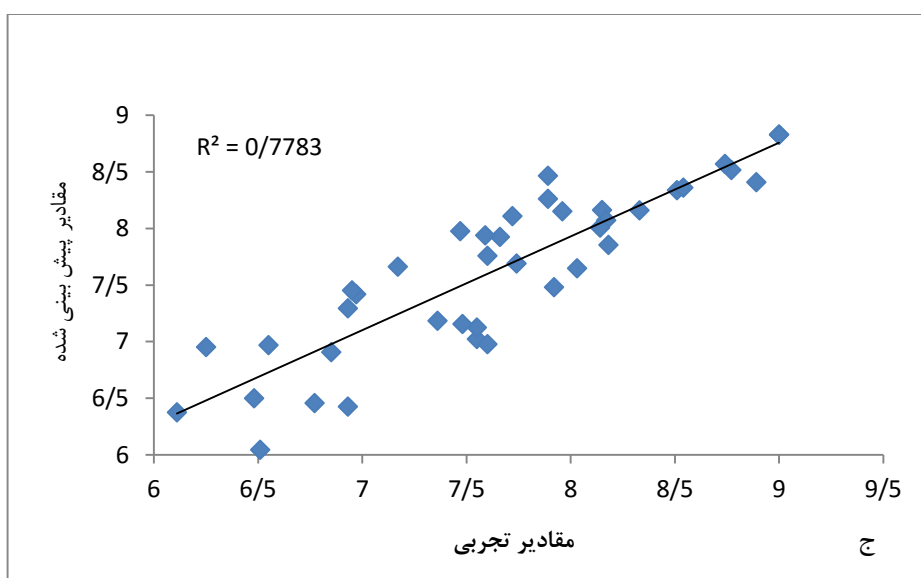
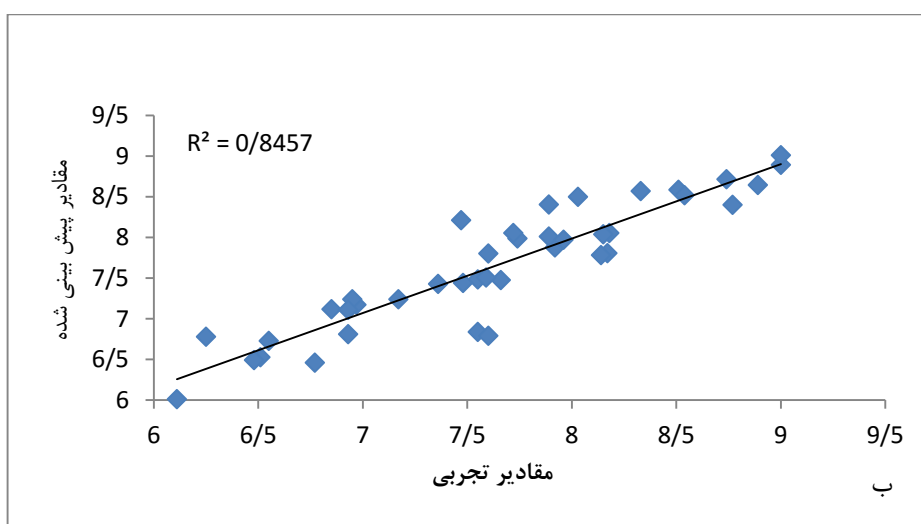
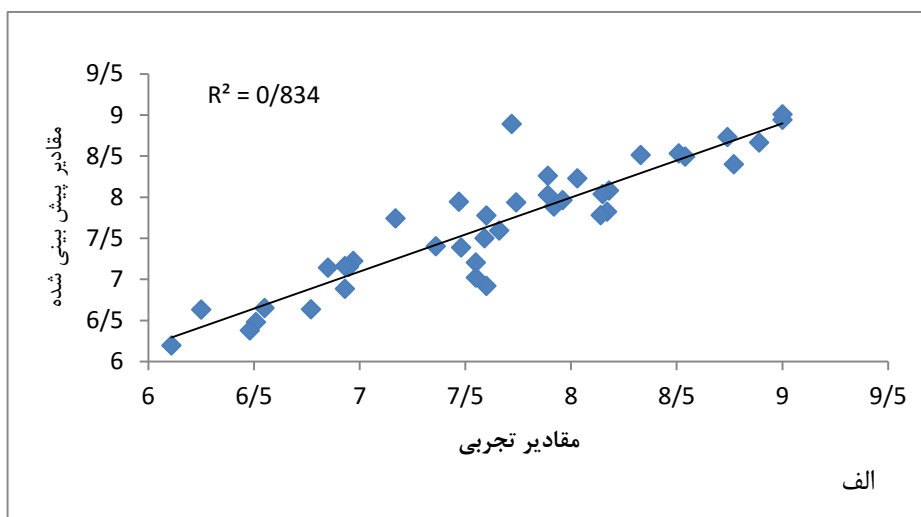
مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده pIC ₅₀			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
۱	7/55	7/02	6/84	7/02	7/02	-9/40	7/02
۲ (t)	6/93	6/88	6/81	7/29	0/72	-1/73	-5/19
۳ (v)	8/18	8/08	8/06	7/85	1/22	-1/47	4/03
۴	8/15	8/04	8/04	8/16	1/35	1/35	-0/12
۵ (t)	8/89	8/67	8/65	8/41	2/47	-2/70	5/40
۶	8/17	7/82	7/81	8/07	4/28	-4/40	1/22
۷ (t)	7/92	7/88	7/88	7/48	0/50	0/50	5/55
۸ (t)	7/66	7/59	7/47	7/92	0/91	2/48	-3/39
۹	7/89	8/02	8/01	8/26	-1/64	-1/52	-4/69
۱۰ (v)	7/74	7/93	7/99	7/69	-2/45	3/23	0/64
۱۱	6/55	6/65	6/73	6/97	-1/52	-2/74	-6/41
۱۲	6/11	6/19	6/01	6/38	-1/31	-1/64	-4/42

¹Leave-one-out

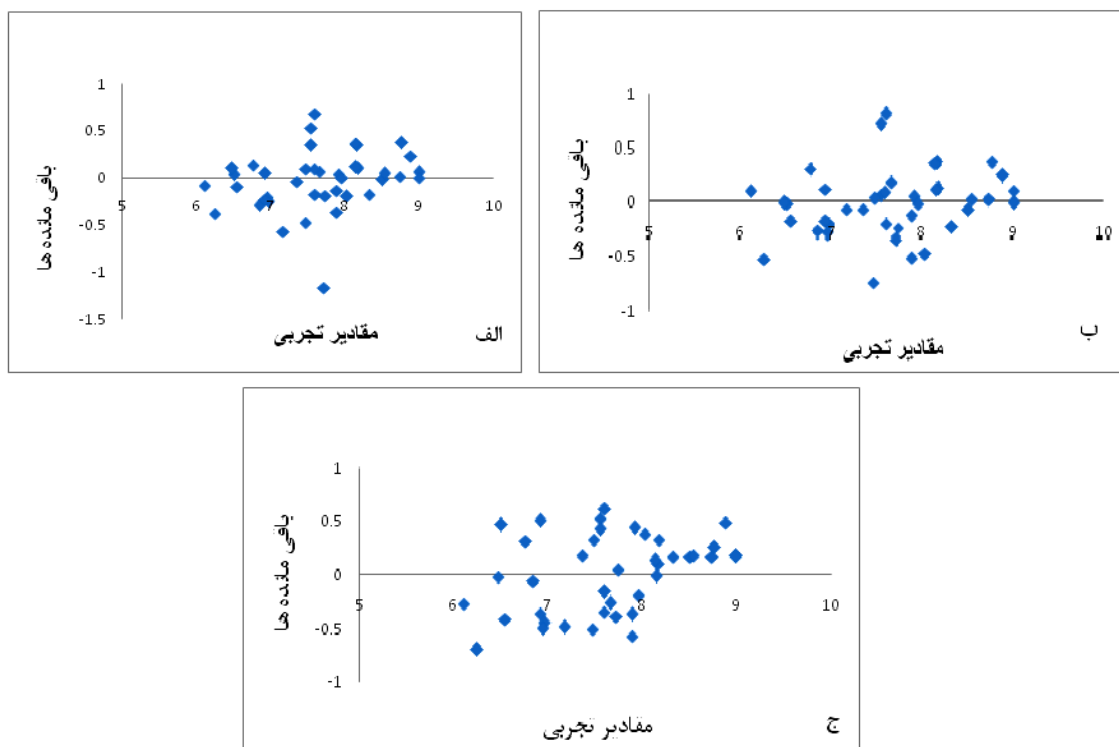
ادامه جدول (۳-۲۴)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده pIC50			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-ANN	SR-ANN	GA-ANN
۱۳	7/36	7/40	7/43	7/18	-0/54	-0/95	2/44
۱۴ (t)	7/55	7/20	7/49	7/125	4/63	0/79	5/69
15	6/97	7/22	7/17	7/42	-3/58	-2/87	-6/46
16	6/51	6/48	6/53	6/04	0/46	-0/31	7/22
۱۷ (v)	6/48	6/38	6/49	6/50	1/54	-0/15	-0/307
۱۸	7/48	7/39	7/44	7/15	1/20	0/53	4/41
19 (v)	6/95	7/15	7/24	7/45	-2/87	-4/17	-7/19
۲۰	7/69	6/92	6/79	6/98	10/01	11/70	9/23
۲۱	7/88	8/26	8/40	8/46	-4/82	-6/59	-7/36
۲۲	7/95	7/96	7/97	8/15	-0/12	-0/25	-2/51
۲۳	6/85	7/14	7/12	6/91	-4/23	-3/94	-0/87
۲۴ (v)	6/93	7/16	7/110	6/43	-3/32	-2/60	7/21
۲۵ (t)	6/77	6/63	6/46	6/46	2/06	4/58	4/58
۲۶	7/59	7/50	7/51	7/94	1/18	1/05	-4/61
27 (v)	8/51	8/53	8/59	8/34	-0/23	-0/94	2
۲۸	7/60	7/78	7/80	7/758	-2/37	-2/63	-2/10
۲۹	8/14	7/78	7/78	8/01	4/42	4/42	1/60
۳۰	9/00	8/94	8/89	8/83	0/67	1/22	1/89
۳۱ (v)	9/00	9/01	9/01	8/83	-0/11	-0/11	1/88
۳۲	7/17	7/74	7/24	7/66	-7/95	-0/98	-6/83
۳۳ (t)	8/03	8/23	8/50	7/65	-2/49	-5/85	4/73
۳۴	8/77	8/40	8/40	8/52	4/22	4/22	2/85
۳۵ (t)	8/33	8/51	8/57	8/16	-2/16	-2/88	2/04
۳۶	7/47	7/94	8/21	7/98	-6/29	-9/91	-6/83
۳۷	8/53	8/44	8/52	8/36	1/05	0/11	2/00
۳۸ (v)	8/74	8/73	8/71	8/57	0/11	0/34	1/94
۳۹	7/72	8/89	8/05	8/11	-15/15	-4/27	-5/05
40	6/25	6/63	6/78	6/95	-6/08	-8/48	-11/20



شکل (۳-۱۸) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده-ها
 الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN

نمودار باقی مانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد و اگر باقی مانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند، نشان می‌دهد که مدل مناسبی به دست آمده و هیچ خطای سیستماتیکی وجود ندارد. شکل (۳-۱۹) نمودار باقی مانده‌ها را که از تفاضل مقدار تجربی فعالیت دارویی از مقدار محاسبه شده توسط مدل‌های خطی و غیرخطی به دست می‌آید، بر حسب مقدار تجربی نشان می‌دهد. تقارن پراکندگی نقاط در دو طرف محور افقی عدم وجود خطای سیستماتیک را نشان می‌دهد.



شکل (۳-۱۹) - نمودار باقی مانده‌ها بر حسب مقدار تجربی IC_{50} برای کل داده‌ها: الف: با مدل SR-MLR؛ ب: با مدل SR-ANN؛ ج: با مدل GA-ANN

۳-۱-۹-۴- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای

آماري

مطابق جدول (۳-۲۵) هفت پارامتر آماری، جهت ارزیابی توانایی پیشگویی مدل‌های ساخته شده به روش SR-MLR، SR-ANN و GA-ANN به کار گرفته شد. نتایج رضایت بخش مدل SR-ANN

نشان می‌دهد که مشارکت بعضی توصیفگرها برای فعالیت دارویی (pIC50) ممکن است، غیرخطی باشد.

جدول (۳-۲۵) - پارامترهای آماری برای مدل‌های برتر طراحی شده SR-MLR، SR-ANN و GA-ANN

پارامتر	سری تست			کل داده‌ها		
	SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
MAE	2/21	۰/۱۳	۰/۳۴	۰/۲۲	۳/۱۱	4/00
MSE	۷/۴۴	۰/۰۳	۰/۱۴	۰/۱۰	۱۷/۵۰	24/45
Press	۵۹/۵۸	۰/۲۲	۱/۱۲	۴/۰۱	۷۰۰/۲۱	978/12
SEP	۲/۷۳	۰/۱۶	۰/۳۷	۰/۳۱	4/18	4/94
MRE	۲۸/۵۱	۰/۰۱	۰/۰۴	2/94	0/48	0/57
R ²	۰/۹۲	0/95	۰/۸۳	—	—	—
R ² (LOO)	—	—	—	0/83	۰/۸۳	۰/۷۸

۳-۱-۹-۵- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون

Y- تصادفی^۱

این تکنیک ارزیابی مدل با هدف بررسی هر گونه ارتباط تصادفی بین داده‌ها انجام شد. در این آزمون مقادیر تصادفی از متغیر وابسته تولید گردید. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. اگر مدل اصلی هیچ گونه ارتباط تصادفی نداشته باشد، تفاوت قابل توجهی بین مقدار ضریب تعیین مدل اصلی و مدل QSAR که با پاسخ‌های تصادفی توسعه یافته، وجود دارد. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۳-۲۶) نشان داده شده است. مقادیر کوچک ضریب تعیین (R²) بیانگر عدم وجود ارتباط شانس یا وابستگی ساختاری به سری آموزش در مدل توسعه یافته توسط شبکه می‌باشد.

^۱ - Y- randomization test

جدول (۳-۲۶) - الف: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش SR-ANN

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	0/002	۰/۰۱۲	۰/۰۰۱	۰/۰۰۹	۰/۰۰۳	۰/۱۸۰	۰/۰۲۴	۰/۰۰۲	۰/۰۲۰	۰/۰۴۵

جدول (۳-۲۶) - ب: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش GA-ANN

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	۰/۰۰۶	۰/۰۷۳	۰/۰۰۹	۰/۱۰۳	۰/۱۰۵	۰/۰۳۲	۰/۰۵۶	۰/۰۸۷	۰/۱۳۲	۰/۰۰۱

۳-۱- 10- بررسی ارتباط توصیفگرهای وارد شده در مدل با اثر بازدارندگی

در این قسمت با توجه به توصیفگرهای وارد شده در مدل SR-MLR، SR-ANN و GA-ANN یک بررسی اجمالی روی اثرات این متغیرها بر فعالیت دارویی ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۸ توصیفگر است که هر کدام بیانگر خصوصیت دو بعدی، هندسی و سه بعدی ترکیبات مورد بررسی می باشند.

• توصیفگرهای Constitutional

این دسته جزو ساده ترین توصیفگرها هستند که معمولاً مورد استفاده قرار می گیرند. نشان دهنده ترکیب یک جسم بدون توجه به هندسه ملکول مورد نظر هستند. مهمترین توصیف کننده های این دسته شامل عدد اتمی و تعداد پیوندها و همچنین مربوط به پیوندهای اتمی یگانه و چندگانه و پیوندهای آروماتیکی و همچنین شامل وزن مولکولی هستند. این توصیف کننده ها به هیچ نوع تغییر در صورتبندی و آرایش اتم ها نسبت به هم حساس نیستند. و بین ایزومرهای مختلف تفاوتی نمی کنند. یکی از توصیف کننده های این دسته Mv^1 است. این توصیفگر قادر است ارتباط بین حجم

¹ - Mean atomic van der Waals volume

واندروالسی با خصوصیات فیزیکی، شیمیایی و بیولوژیکی را فراهم سازد. حجم واندروالسی به وسیله شعاع واندروالسی (و برای مولکول ها ، فواصل بین اتمی و زوایا) اندازه گیری می شود. برای یک مولکول حجم واندروالسی کل همیشه از مجموع حجم های واندروالسی اتم های تشکیل دهنده آن کوچکتر است زیرا اتم ها در حین تشکیل پیوندهای شیمیایی همیشه همپوشانی انجام می دهند [۳۴]. با توجه به علامت منفی آن در مدل می توان گفت که با افزایش مقادیر این توصیف گر pIC_{50} کاهش می یابد. با توجه به جدول (۳-۲۷) با افزایش حجم مولکول ها مقدار پارامتر Mv کاهش یافته و برعکس، مقدار pIC_{50} کاهش می یابد.

جدول (۳-۲۷) - ارتباط ساختاری مولکول ها با توصیف گر Mv

شماره ترکیب	X1	X2	-O(CH ₂) _n -Ar	m	-NR ₁ R ₂	pIC_{50} (M)	پارامتر Mv
۳۷	4-MeO	H	3-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr)(iso-Amy)	8/54	0/58
4	5-MeO	H	2-O(CH ₂) ₂ -Ph	2	N(<i>n</i> -Pr) ₂	8/15	0/59
2	5-Cl	H	2-OCH ₂ -Ph	2	N(<i>n</i> -Pr) ₂	6/93	0/61

• توصیفگرهای BCUT

این توصیفگرها از دسته توصیفگرهای دو بعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده^۱ (ماتریس بوردن) می باشند. ماتریس بوردن یک گراف تهی از هیدروژن را ارائه می دهد که در آن B_{ii} ، یعنی عناصر قطری در ارتباط با خاصیت اتمی مانند الکترون گاتیویته، حجم واندروالس، جرم اتمی و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند [34 و 35]. اثر این توصیف کننده مثبت است بنابراین با افزایش آن فعالیت ضد روان پریشی مولکول ها

^۱ - Modified connectivity matrix

افزایش می‌یابد. هر چه جرم اتمی مولکول‌ها بیشتر می‌شود و تعداد حلقه‌های آروماتیک و باندهای دو گانه افزایش می‌یابند مقدار BEHm6¹ نیز افزایش می‌یابد.

• توصیفگرهای Chargeindex و Galvez

این توصیفگرها جهت ارزیابی و بررسی انتقال بار بین جفت اتم‌ها در یک مولکول و بنابراین انتقال بار در تمام مولکول پیشنهاد شدند. برای توضیح این توصیف کننده‌ها ابتدا ماتریس M را بصورت زیر تعریف می‌کنیم:

$$M = A \times D^2 \quad (1-3)$$

که در آن A ماتریس مجاورت² و D^2 ماتریس عکس مربع فاصله³ می‌باشد. توجه کنید که اعضای قطری می‌مانند. M ماتریس گالوز⁴ می‌باشد و یک ماتریس مربعی نامتقارن $A \times A$ می‌باشد که در آن A تعداد اتم‌های مولکول می‌باشد. M باعث یک ماتریس ترم بار⁵ نامتقارن، CT می‌شود که به صورت زیر تعریف می‌شود:

$$(2-3) \quad CT_{ij} = \delta(i) \text{ if } i=j$$

$$(3-3) \quad CT_{ij} = m_{ij} - m_{ji} \text{ if } i \neq j \quad (3-3)$$

که در آن m_{ij} عناصر M و $\delta(i)$ درجه‌ی بردار اتم i می‌باشد. ورودی‌های قطری CT ظرفیت توپولوژیکی اتم‌ها هستند.

برای هر فاصله‌ی طولی K می‌توان یک اندیس بار توپولوژیکی، G_k به صورت زیر تعریف کرد.

¹- Lowest einvalue n.6 of Burden

²-Adjacency

³-Reciprocal square distance

⁴-Galvez matrix

⁵-Charge term matrix

$$G_k = \frac{1}{2} \cdot \sum_i \sum_j |CT_{ij}| \cdot \delta(k, d_{ij}) \quad (4-3)$$

که در آن $\delta(k; d_y)$ دلتای کرونکر^۱ می‌باشد [۳۴].

توصیفگر G_{GI3} ^۲ از این گروه وارد مدل شده است که اثر آن بر روی فعالیت منفی می‌باشد.

۳-۱-۱۰-۱- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

میزان مشارکت توصیفگرهای منتخب به صورت زیر تعیین شد:

۱- توصیفگر مورد نظر به همراه اوزان مربوطه‌اش از شبکه بهینه شده حذف گردید.

۲- با استفاده از بقیه توصیفگرها مقدار متغیر وابسته (pIC_{50}) برای هر ترکیب سری آموزش به روش ارزیابی تقاطعی پیش‌بینی شد.

۳- میانگین خطای مطلق حاصل از ارزیابی تقاطعی ترکیبات سری آموزش محاسبه گردید.

۴- مراحل ۱ تا ۳ برای دیگر توصیفگرهای منتخب نیز تکرار شد.

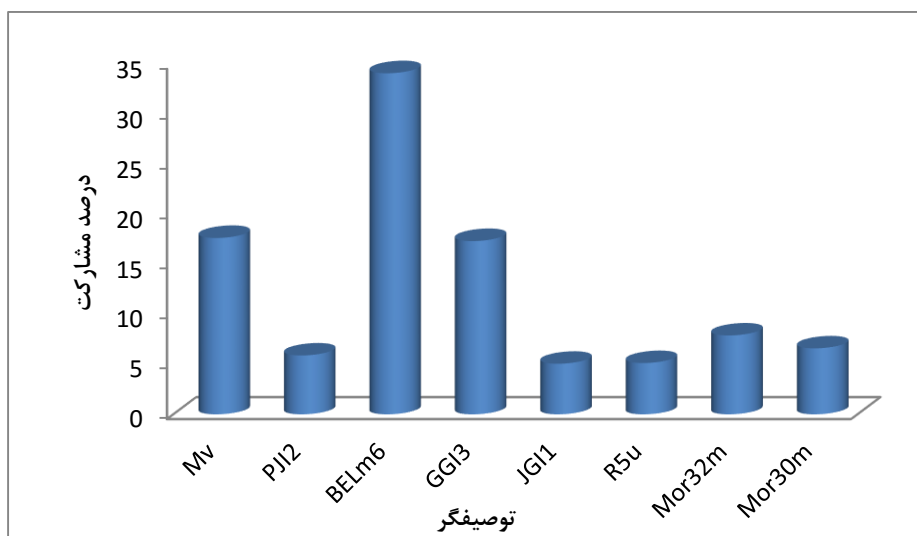
۵- سرانجام درصد مشارکت هر توصیفگر توسط رابطه (۳-۵) برآورد شد.

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (5-3)$$

در این رابطه c_i درصد مشارکت توصیفگر حذف شده Δm_i ، N تعداد توصیفگرهای مدل و Δm_i میانگین خطای مطلق حاصل از ارزیابی تقاطعی سری آموزش در غیاب توصیفگر Δm_i را نشان می‌دهد. که بر این اساس درصد مشارکت توصیفگرهای منتخب در ترکیبات مورد بررسی در شکل (۳-۲۰) ارائه شده است.

^۱-kroneker

^۲-Topological charge index of order 3



شکل (۳-۲۰) - مشارکت توصیفگرها در شبکه عصبی بهینه

بر اساس نمودار فوق توصیفگرهای BELm6 و Mv و GGI3 دارای بیشترین اثر مشارکت می‌باشند توصیفگر Mv از دسته توصیفگرهای بدون بعد است که ارتباط بین حجم و اندروالسی با خصوصیات فیزیکی، شیمیایی و بیولوژیکی و اثر بازدارندگی را نشان می‌دهد. توصیفگر BELm6 و GGI3 در برگیرنده اطلاعات دو بعدی مولکول‌ها هستند. بنابراین می‌توان گفت که فعالیت بازدارندگی این ترکیبات وابستگی قابل ملاحظه‌ای به توپولوژی ترکیب بازدارنده دارد. طبق شکل فوق توصیفگر Mor30m و MOR32m در برگیرنده اطلاعات سه بعدی مولکول هستند که این اطلاعات به صورتبندی دارو ارتباط دارد.

۳-۱۱- نتیجه‌گیری نهایی

همانطور که گفته شد هدف اصلی از کار حاضر، توسعه روش‌های خطی و غیرخطی جهت مطالعه ارتباط کمی بین ساختار و فعالیت ضد مشتقات فنیل آلکیل آمین به عنوان بازدارنده‌های روان‌پریشی می‌باشد. نتایج نشان می‌دهند که مدل‌های خطی و غیرخطی توسعه یافته می‌توانند به عنوان روش-

های موفق برای مدل سازی و پیشگویی فعالیت بازدارندگی ترکیبات مورد مطالعه باشند و این مدل ها می توانند به محققان در طراحی داروهایی با فعالیت بازدارندگی قویتر کمک کنند.

آینده نگری

❖ پژوهشگران می توانند نتایج حاصل از این مطالعات را در طراحی ترکیبات (داروها) جدید با

فعالیت بازدارندگی قویتر به کار برند.

❖ از آن جایی که در تحقیق حاضر توصیفگرهای به دست آمده از روش رگرسیون خطی

چندگانه و الگوریتم ژنتیک به عنوان ورودی شبکه عصبی مصنوعی قرار گرفته است ، می توان

از روش هایی همچون SVM¹ و LS-SVM² به جای شبکه عصبی به استفاده کرد.

❖ می توان از روش های خطی دیگر مانند حداقل مربعات جزئی و آنالیز اجزای اصلی را جایگزین

روش رگرسیون خطی چندگانه کرد و نتایج به دست آمده را با هم مقایسه کرد.

از روش های نوین مانند CoMFA³ و CoMSIA⁴ برای پیش بینی فعالیت بازدارندگی ترکیبات مورد

مطالعه استفاده کرد.

¹ - Supported Vector Machine

² - Least Square Supported Vector Machine

³ - Comparative Molecular Field Analysis

⁴ - Comparative Molecular Similarity Indices Analysis

فصل چهارم

مطالعه ارتباط کمی ساختار- شاخص بازداری

مجموعه‌ای از ترکیبات آلی با استفاده از روش-

های خطی و غیر خطی

اندازه‌گیری شاخص بازداری ترکیبات آلی به طور معمول با استفاده از تکنیک‌های مختلف کروماتوگرافی انجام می‌شود. وجود ستون‌های کروماتوگرافی مختلف برای اندازه‌گیری مواد گوناگون و پیشرفت‌هایی که از گذشته تا به حال در زمینه کروماتوگرافی پدید آمده‌اند اندازه‌گیری شاخص بازداری تعداد زیادی از ترکیبات را امکان‌پذیر کرده است و بسیاری از مسائل مربوط به جداسازی و شناسایی را حل نموده است. اما در کنار این مزایا، تکنیک‌های کروماتوگرافی محدودیت‌هایی نیز دارند که از آن جمله می‌توان به عدم توانایی آن‌ها در اندازه‌گیری پاره‌ای از ترکیبات که تهیه و یا آماده سازی آن‌ها بسیار مشکل است یا ترکیباتی که در ستون‌های کروماتوگرافی ناپایدارند، اشاره نمود. هم چنین تکنیک‌های کروماتوگرافی نیاز به آماده سازی‌های مختلف، انجام عملیات ویژه بر روی نمونه و ستون‌های کروماتوگرافی و در مواردی بهینه سازی پارامترها دارند که ممکن است با صرف زمان و هزینه‌های بسیار زیاد همراه باشند. بنابراین این محدودیت‌ها موجب می‌شوند که در کنار روش‌های آزمایشگاهی تجربی، روش‌های تئوری نیز برای استفاده از نتایج تجربی توسعه یابند. از جمله این روش‌ها مطالعات کمومتریکس به ویژه در زمینه‌ی QSPR می‌باشد که همان گونه که قبلاً ذکر شد در آن می‌توان از تکنیک‌های مختلفی برای ایجاد روابط خطی و یا غیر خطی میان خواص ساختاری و فاکتور بازداری ترکیباتی که شاخص بازداری آن‌ها با استفاده از روش‌های تجربی اندازه‌گیری شده‌اند، استفاده نمود. تدبیر اساسی QSPR به دست آوردن رابطه‌ی کمی بهینه‌ای بین ویژگی‌های ساختار مولکولی از ساختار مولکولی و خواص مختلف مولکول‌ها می‌باشد و هنگامی که رابطه‌ی معتبری به دست آمد امکان پیش‌بینی یک خاصیت مانند فاکتورهای بازداری برای ساختارهای مشابه با ترکیبات اندازه‌گیری شده یا ساختارهای دیگر که هنوز اندازه‌گیری نشده و یا تهیه نشده‌اند، وجود دارد.

۴-۱-۱- پیشینه تحقیق

در سال ۲۰۰۷ مطالعه‌ای بر اساس الگوریتم ژنتیک GA و شبکه‌ی عصبی ANN برای مدل سازی و بررسی رفتار بازداري برخی ترکیبات آلی در HPLC فاز معکوس توسط یک گروه تحقیقاتی ایتالیایی با سرپرستی کارلوسی^۱ انجام شد [۳۶].

در سال ۲۰۰۸ مهدی ستاری و فرهاد قراقیزی ضریب پخش ۳۷۸ ترکیب خالص در هوا را با استفاده از الگوریتم ژنتیک بر پایه رگرسیون خطی چندگانه (GA-MLR) و شبکه عصبی پیشخور بررسی کردند که در آن GA-MLR برای انتخاب توصیف‌گرها به عنوان ورودی به عنوان ورودی‌های شبکه استفاده شد. ضریب همبستگی مدل خطی چندگانه شامل هفت توصیف کننده، ۰/۹۳۳ به دست آمد و همان مقدار برای شبکه برابر ۰/۹۶۴ به دست آمد [۳۷].

لوان و همکارانش^۲ در سال ۲۰۰۸ با استفاده از QSPR مدل‌هایی را برای پیش‌بینی ضریب تقسیم خون به هوا برای ترکیبات آلی فرار با استفاده از روش‌های خطی و غیر خطی بر پایه‌ی شبکه‌های عصبی مصنوعی محاسبه نمودند. هر دو روش خطی و غیر خطی نتایج مناسبی را ارائه دادند به طوری که مقدار R برای دو روش خطی و غیر خطی به ترتیب برابر ۰/۹۶۴ و ۰/۹۷۹ بود [۳۸].

همچنین لووان با همکاری گروه دیگری در همان سال برای پیش‌بینی فاکتورهای کالبراسیون کمی در کروماتوگرافی گازی برای تعداد ۲۰۸ ترکیب آلی از تکنیک QSPR استفاده نمود، بدین صورت که از دو روش ANN و MLR مدل‌هایی به دست آمد و مقایسه‌ی ویژگی‌های آماری این دو مدل با $R^2=0/927$ برای MLR و $R^2=0/959$ برای ANN باز هم برتری مدل و استفاده از این مدل را برا پیش‌بینی فاکتورهای کالبراسیون کمی برای سایر ترکیبات نشان می‌دهد [39].

۱-G. Carlucci

۲-F.Luan et al

بهرام همتی نژاد و همکارانش نیز در سال ۲۰۰۸ در چارچوب مطالعات برای ایجاد مدل‌هایی که ارتباط دهنده حلالیت با ساختار مشتقات آنتراکینون، آنترون و زانتون در دی اکسید کربن فوق بحرانی باشند، کار کردند. در این کار ۲۹ مولکول در دماها و فشارهای مختلف که مجموعاً ۱۱۹۰ داده حلالیت را ایجاد می‌نمود، در دسترس قرار داشت. مدل سازی و استفاده از توصیف‌گرهای مولکولی انتخاب شده و داده‌های حلالیت توسط روش‌های خطی MLR و غیر خطی ANN انجام شد که هر دوی روش‌های خطی و غیر خطی منجر به پیش بینی‌های خوبی شدند [۴۰].

برای ترکیبات مورد بررسی در این پایان‌نامه قبلاً مدل‌سازی صورت نگرفته است و این کار به عنوان تحقیقی جدید در ارتباط با این ترکیبات آلی می‌باشد که در این پایان‌نامه از رگرسیون خطی چندگانه شبکه‌ی عصبی مصنوعی همراه با الگوریتم ژنتیک برای مدل سازی استفاده شده است.

۴-۲- انتخاب سری داده‌ها

سری داده‌ها از نتایج تعیین شاخص‌های بازداری ۶۰ مولکول با کروماتوگرافی گازی مؤئینه با فاز ساکن، انجام شده توسط یک گروه تحقیقاتی اسپانیایی، بدست آمد [۴۱]. این گروه تحقیقاتی داده‌های این آزمایش را برای چهار دمای مختلف به دست آوردند و به دلیل حضور ۶۰ مولکول در ستونی با چهار دمای مختلف مجموعاً ۲۴۰ شاخص بازداری خواهیم داشت. در جدول (۴-۱) نیز مقادیر عددی شاخص بازداری برای هر ترکیب در چهار دمای متفاوت آورده شده است.

جدول (۴-۱)- ساختار و مقادیر شاخص بازداري تركيبات مورد مطالعه در ۴ دمای متفاوت [۴۱]

شماره تركيب	نام تركيب	شاخص بازداري			
		353K	373K	393K	413K
۱	Benzene	658/1	662/3	668/4	672/10
۲	Toluene	760/8	765/6	771/1	775/7
۳	Ethylbenzene	852/8	858/1	863/9	869/1
۴	n-Propylbenzene	943	948/6	954/6	960/4
۵	n-Butylbenzene	۱۰۴۱/۹	1047/6	1053/6	1059/5
۶	Propanone	۴۷۳/۳	471/2	479/9	472
۷	2-Butanone	۵۷۵/۷	575/2	580/5	576/8
۸	2-Pentanone	۶۶۶/۲	666/6	670	668/9
۹	2-Hexanone	۷۶۷/۸	768/6	771/1	771
۱۰	2-Heptanone	۸۶۸/۵	869/3	871/3	871/8
۱۱	2-Octanone	۹۶۸/۸	969/8	971/7	972/6
۱۲	2-Nonanone	۱۰۶۹/۶	1070/7	1072/4	1073/6
۱۳	Cyclobutanone	۶۳۸/۶	641/1	644/8	647
۱۴	Cyclopentanone	۷۶۴/۲	768/9	774/6	779/5
۱۵	Cyclohexanone	866/8	873/9	882	889/6
۱۶	Cycloheptanone	980/7	989/2	998/7	1008
۱۷	Cyclooctanone	1082/9	1093	1103/9	1115
۱۸	n-Propanol	537/3	536	533/3	530/4
۱۹	n-Butanol	642/6	641/1	639/4	638/1
۲۰	n-Pentanol	746/9	745/2	744/1	743/5
۲۱	n-Hexanol	849/5	848/1	847/2	846/6
۲۲	n-Heptanol	951/6	950/2	949/4	949/2
۲۳	n-Octanol	1053/1	1051/9	1051/3	1051/2
24	25 Cyclohexanol	865/7	869/9	875/1	880/6

ادامه جدول (۴-۱)

شماره ترکیب	نام ترکیب	شاخص بازداری			
		353K	373K	393K	413K
۲۵	2-Methyl-2-pentanol	716/8	717/2	718/3	718/1
۲۶	Methyl acetate	511/9	510/2	510/8	507/3
۲۷	28 Ethyl acetate	593/9	591/1	589/6	586/8
۲۸	<i>n</i> -Propyl acetate	693/5	691/2	689/7	688/1
۲۹	<i>n</i> -Butyl acetate	793/8	791/7	790/4	788/4
۳۰	<i>n</i> -Propylamine	526/4	526/6	528/4	523/8
۳۱	<i>n</i> -Butylamine	626/6	628/1	627/9	628/3
۳۲	<i>n</i> -Pentylamine	729/4	730/6	731	730/4
۳۳	<i>n</i> -Hexylamine	830/2	831/4	832/4	833/2
۳۴	<i>n</i> -Heptylamine	931/8	932/8	934/1	934/8
۳۵	Cyclohexylamine	847/8	855/2	862/4	870/2
۳۶	Aniline	947/3	952/1	957/9	964/3
۳۷	<i>n</i> -Butanenitrile	638/5	639/8	640/6	643/1
۳۸	<i>n</i> -Pentanenitrile	740/7	742/6	744/2	747/2
۳۹	<i>n</i> -Hexanenitrile	842/8	844/8	847	849/9
۴۰	Benzonitrile	947/6	954/4	961/5	969/5
۴۱	1-Nitropropane	706/5	708/6	710/8	714/7
۴۲	Nitrobenzene	1047/1	1056/2	1065/9	1076/2
۴۳	<i>n</i> -Butyl iodide	806/9	813/2	819/6	825/9
۴۴	<i>n</i> -Propyl chloride	539/8	541/5	542	544/5
۴۵	<i>n</i> -Butyl chloride	641/3	643/6	645/6	647/7
۴۶	<i>n</i> -Pentyl chloride	743/5	745/9	748/3	751/4
۴۷	<i>n</i> -Hexyl chloride	844/8	847/4	850	853/1
48	Chlorobenzene	835/2	۹۰/۸	848/9	856/3

ادامه جدول (۴-۱)

شماره ترکیب	نام ترکیب	شاخص بازداري			
		353K	373K	393K	413K
۴۹	1,4-Dioxane	689/9	692/7	696	699/1
۵۰	Tetrahydrofuran	619/6	622/5	625/9	628/6
۵۱	Pyridine	726	731/8	735/9	741/7
۵۲	<i>N,N</i> -Dimethylformamide	744/9	747	750/3	753/3
۵۳	2-Octyne	865/4	864/1	862/8	861/4
۵۴	<i>n</i> -Hexane	600	599/9	600	599/9
۵۵	<i>n</i> -Heptane	700	700/1	700/1	700/2
۵۶	<i>n</i> -Octane	800	800	800	800
۵۷	<i>n</i> -Nonane	900	900	900	899/9
۵۸	<i>n</i> -Decane	1000	1000	1000	1000
۵۹	<i>n</i> -Undecane	1100	1100	1100	1100
۶۰	Cyclohexane	668/5	673/4	679/5	684/6

۴-۲-۱- رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، ساختار مولکولی هر ترکیب ابتدا در نرم افزار Hyperchem7/0 و سپس با احتساب اتم‌های هیدروژن، ساختار سه بعدی ترکیبات با استفاده از روش نیمه تجربی کوانتومی AM1 بهینه شد. البته بهینه‌سازی تا زمانی ادامه یافت که جذر میانگین مربعات گرادیان^۱ انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد. همه این محاسبات در سطح هارتری-فاک^۲ محدود شده انجام شد.

^۱ - Root Mean Square Gradient

^۲ - Restricted Hartee-Fock

۴-۲-۲- محاسبه توصیفگرها

برای ایجاد ارتباط ریاضی بین ساختار-خاصیت، به مقادیر عددی نیاز است و لذا باید ویژگی‌های ساختاری مولکول‌ها به مقادیر کمی تبدیل شود. بدین منظور ساختار بهینه شده ترکیبات به نرم افزار Dragon منتقل و برای هریک از مولکول‌ها ۱۴۸۱ توصیفگر در ۱۸ دسته مختلف محاسبه گردید.

۴-۲-۳- انتخاب توصیفگرهای مناسب

مهمترین مرحله در مطالعات QSPR انتخاب توصیفگرهای مناسب است که بتوانند خاصیت شیمیایی را به ساختار مولکول ارتباط دهند. با عنایت به این موضوع و با توجه به اینکه تعداد زیاد توصیفگر، باعث پیچیدگی محاسبات می‌شود و همچنین تعدادی از توصیفگرها حاوی اطلاعات یکسان و در نتیجه اثر یکسان بر روی فعالیت شیمیایی هستند، لذا باید روش کاهش تعداد متغیرها به کار گرفته شود. به همین منظور، ابتدا توصیفگرهایی که دارای مقادیر یکسانی برای کل مولکول‌ها بودند از فهرست توصیفگرها حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیفگری که ضریب همبستگی بزرگتر از ۰/۹ داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف شد سپس رگرسیون مرحله به مرحله توسط نرم افزار SPSS مورد استفاده قرار گرفت. در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل می‌شوند و 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 داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. به علاوه با افزودن کلید واژه‌های مناسب توصیف کننده دما (T) نیز از جدول استخراج شده و به مجموعه توصیفگرهای قبلی اضافه شد. به این ترتیب داده‌های شاخص بازداری^۱ (RI)، به عنوان متغیر وابسته و توصیفگرها به عنوان متغیر مستقل در نظر گرفته

^۱-Retention index

شدند و تکنیک رگرسیون مرحله‌ای انجام شد که در این مرحله فقط ۱۰ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت دارویی دارند، انتخاب شدند. نام و طبقه‌ی این ۱۰ توصیفگر در جدول (۲-۴) ارائه شده است. همچنین ماتریس همبستگی بین این توصیفگرها در جدول (۳-۴) ارائه شده که این ماتریس عدم همبستگی بین توصیفگرها را نشان می‌دهد.

جدول (۲-۴) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله و معانی آنها

تعداد	نماد	طبقه	معانی
۱	RDSUM	Topological	reciprocal distance Wiener-type index
۲	GATS3e	2D autocorrelations	Geary autocorrelation-lag 3
۳	GATS2e	2D autocorrelations	Geary autocorrelation-lag 2
۴	GATS3m	2D autocorrelations	Geary autocorrelation-lag 3
۵	X0Av	topological	Average valence connectivity index
۶	piPC07	topological	molecular multiple path count of order 07
۷	HATS4m	GETAWAY	leverage-weighted autocorrelation of lag 4
۸	Mor12u	3D-MoRSE	3D-MoRSE-signal12
۹	Pol	topological	polarity number
۱۰	T		

جدول (۳-۴) - ماتریس همبستگی کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله

	RDSUM	GATS3e	GATS2e	GATS3m	X0Av	piPC07	HATS4m	Mor12u	Pol	T
RDSUM	1									
GATS3e	-0/13	1								
GATS2e	0/02	-0/20	1							
GATS3m	-0/31	0/29	0/16	1						
X0Av	-0/27	-0/09	0/29	0/80	1					
piPC07	0/56	-0/20	-0/11	-0/32	0/35	1				
HATS4m	-0/14	-0/09	-0/13	0/67	0/47	0/12	1			
Mor12u	-0/60	0/18	-0/20	0/07	0/00	-0/10	0/17	1		
Pol	0/95	-0/19	-0/07	-0/32	-0/26	0/62	-0/03	-0/57	1	
T	0/00	0/00	0/00	0/00	0/00	0/00	0/00	0/00	0/00	1

۴-۲-۴- مدل سازی توسط روش رگرسیون خطی چندگانه

برای ساختن مدل خطی مناسب که بیانگر ارتباط ساختاری ترکیبات مورد مطالعه با شاخص بازداري آنها باشد، مدل‌های مختلفی که توسط رگرسیون خطی چندگانه مرحله به مرحله به دست آمد، مورد بررسی قرار گرفتند. برای این کار داده‌ها به طور تصادفی به سه سری آموزش و ارزیابی و تست تقسیم شدند. سری تست ۵۰ ترکیب، سری ارزیابی ۵۰ ترکیب و سری آموزش نیز ۱۴۰ ترکیب را شامل می‌شوند. سپس توسط سری آموزش، ابتدا با ۲ توصیفگر اول انتخاب شده، به روش اجباری مدل ساخته شد و به وسیله‌ی معادله‌ی آن، سری ارزیابی آزموده شد. در مرحله بعد با مدل ایجاد شده توسط ۳ توصیفگر، سری ارزیابی را آزموده و همین طور با معادلات حاصل از ۴ تا ۱۰ توصیفگر به طور مجزا سری ارزیابی مورد بررسی قرار گرفت. میانگین مربعات خطا (mse) و ضریب تعیین تصحیح شده (R_{adj}^2) محاسبه شده برای سری ارزیابی در جدول (۵-۵) آمده است.

جدول (۴-۴) - مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR

مدل پارامتر	2	3	4	5	6	7	7	8	9	10
mse	2587/20	1936/80	906/89	1064/80	1148/20	854/98	1303/30	1571/40	1627/70	1571/40
R_{adj}^2	0/89	0/92	0/97	0/96	0/97	0/97	0/95	0/94	0/94	0/94

همانطور که در جدول (۴-۴) مشاهده می‌شود مدل هفتم به خاطر داشتن تعداد توصیفگر و خطای کمتر بر سایر مدل‌ها برتری داشته و به عنوان مدل نهایی انتخاب شد. به این ترتیب معادله خطی نهایی، با استفاده از سری آموزش به صورت زیر به دست آمد:

$$RI = 809.266 + 0.128 T + 35.507 RDSUM - 87.301 GATS3e - 51.824 GATS2e + 8267.080 GATS3m - 508.949 X0Av - 3.055 piPC07$$

۴-۲-۵ - مدل سازی به روش شبکه عصبی مصنوعی

بخاطر اینکه ممکن است مدل خطی بهترین مدل برای توصیف رفتار سری داده‌ها نباشد، شبکه عصبی مصنوعی برای مدل‌سازی غیر خطی داده‌ها مورد استفاده قرار گرفت. همانند روش MLR داده‌ها به صورت تصادفی و با نسبت ۶۰:۲۰:۲۰ به سه سری آموزش، ارزیابی و تست تقسیم شدند. سپس پارامترهای مؤثر بر آموزش موفق شبکه عصبی پس انتشار از قبیل تعداد ورودی‌ها، تعداد لایه‌های پنهان، تعداد گره‌های لایه پنهان، نوع تابع انتقال، نوع تابع آموزش، تعداد دوره‌های آموزش و پارامتر μ با استفاده از سری آموزش و با روش ارزیابی تقاطعی بهینه شدند. در طی فرایند آموزش مقادیر وزن‌های ارتباطی تغییر می‌کند و معیار برای پایان دادن آموزش و بهینه شدن مقادیر وزن‌ها جهت پیش‌بینی پارامتر مورد بررسی، مقدار خطای شبکه است که طبق تابع کارایی تعیین شده، محاسبه می‌شود که در این روش، تابع کارایی، میانگین مربع خطا (MSE) در سری آموزش می‌باشد.

۴-۱-۵-۱- انتخاب تعداد لایه‌های پنهان

متأسفانه هیچ‌گونه نتایج تئوری قابل دسترس یا قوانین تجربی رضایت بخش که در تعیین تعداد لایه‌های پنهان به ما کمک کند، وجود ندارد با این وجود برای اغلب کاربردهای شبکه عصبی در شیمی، به نظر می‌رسد که یک لایه مخفی کافی باشد. به این ترتیب شبکه عصبی تشکیل شده، یک شبکه سه لایه شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است که تعداد ورودی‌ها (توصیفگرها) با تعداد نرون‌های لایه ورودی برابر است و لایه خروجی دارای یک نرون است که نشان دهنده‌ی شاخص بازداري متناظر با هر ترکیب می‌باشد. ولی تعداد نرون‌های لایه پنهان باید بهینه شود.

۴-۲-۵-۲- انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع

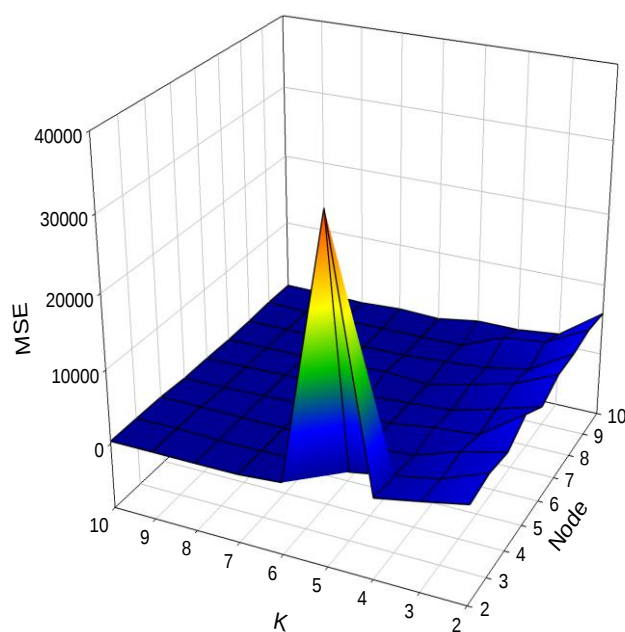
آموزش و نوع تابع انتقال

پس از طراحی اولیه شبکه، برای اینکه شبکه قدرت پیش‌بینی مناسبی داشته باشد باید به طور مناسب آموزش دیده باشد و با توجه به اینکه فرایند آموزش وقتگیر است، انتخاب الگوریتم آموزشی سریعتر امری مهم است. شبکه با ورودی‌های از ۲ تا ۱۰ توسط دو الگوریتم آموزشی لونبرگ - مارکوارت و تنظیم بایزین، با تعداد متفاوت گره در لایه پنهان از ۲ تا ۱۰ و همچنین توابع لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig)، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمامی این شبکه‌ها تعداد دورهای آموزش ثابت و برابر ۳۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده شد معیار نیز به حداقل رساندن میانگین مربع خطا (MSE) در نظر گرفته شد که نتایج این محاسبات در جداول (۴-۵)، (۴-۶)، (۴-۷) و (۴-۸) گرد آوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۴-۱)، (۴-۲)، (۴-۳) و (۴-۴) ارائه شده است.

با توجه به نتایج به دست آمده، الگوریتم آموزشی لونبرگ مارکوارت نسبت به الگوریتم آموزشی تنظیم بایزین دارای میانگین مربع خطا (MSE) کمتری است. علاوه بر آن دارای قدرت تعمیم‌پذیری بالاتری است. همچنین تابع تانژانت سیگموئیدی دارای خطای کمتری نسبت به لگاریتم سیگموئیدی است. بنابراین تابع انتقال تانژانت سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده ۹ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط با فعالیت بازدارندگی دارند، انتخاب شدند. تعداد ۹ گره نیز در لایه پنهان انتخاب شد.

جدول (۴-۵) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)										
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9	10
	2	4283/6	3064/2	1927/7	36124/8	819/4	359/2	333/5	351/01	342/3
	3	4923/8	2640/1	1491/8	455/9	444/3	260/9	251/1	390/3	350/3
	4	4367/9	2984/1	944/7	588/3	219/7	229/8	274/2	352/0	349/2
	5	6272/9	2057/7	1585/1	703/4	696/7	463/9	301/4	109/0	81/5
	6	4230/1	2502/3	901/3	484/7	78/4	546/9	314/9	162/9	94/8
	7	5659/1	3019/5	1524/6	134/6	629/7	295/5	140/7	120/2	20/5
	8	5709/1	2162/9	1018/1	549/8	186/2	110/5	154/3	67/4	102/9
	9	6250/9	3234/8	2538/7	136/0	96/6	61/3	26/8	26/0	342/8
	10	5887/7	1618/0	974/1	996/7	107/5	338/3	46/2	256/9	322/7

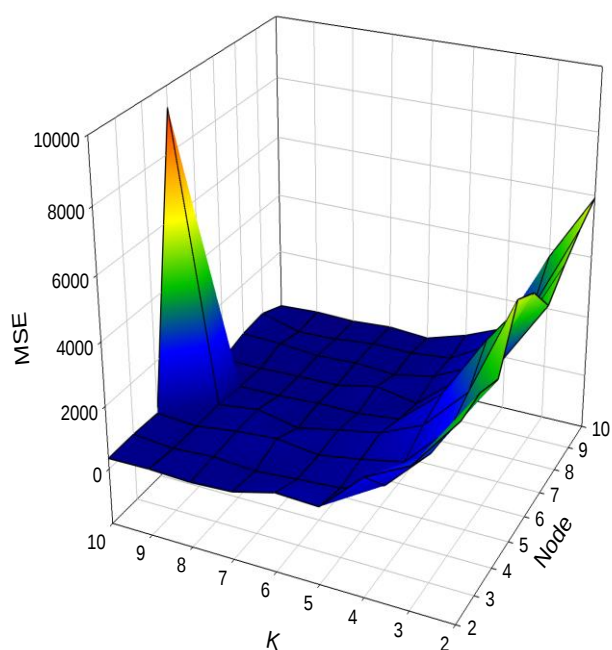


شکل (۴-۱) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (۴-۶) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

لونیگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)										
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9	10
	2	۴۲۸۳/۲	۲۶۴۰/۸	۱۹۴۰/۰	۶۵۹/۸	۷۱۶/۵	۳۷۲/۱	۳۰۲/۴	۳۵۸/۵	۳۳۹/۷
	3	۴۳۶۸/۵	۲۲۴۳/۲	۸۸۸/۷	۴۳۲/۸	۳۴۵/۹	۲۳۶/1	۲۳۸/1	۲۵۲/۴	۳۶۳/۷
	4	۴۰۶۱/۶	۲۲۱۲/3	۸۳۶/۳	۵۶۴/۲۲	۳۴۰/۹	۵۲۲/4	۴۰۹/۹	۲۳۶/۵	۳۲۲/۳
	5	۵۷۳۷/۱	۲۶۲۹/۹	۱۲۵۳/۱	۶۵۹/۴	۴۰۴/۸	۱۲۲/۴	۴۰۰/۴	۲۳۴/۷	۹۳۵۰/۳
	6	۵۳۱۲/۷	۳۵۶۹/۹	۱۵۲۱/۵	۳۶۶/۰	۲۲۴/5	۲۲۲/۵	۱۵۱/۲	۱۳۱/۰	۴۷/۱
	7	۴۲۹۹/۲	۲۲۵۰/۱	۱۲۸۹/۷	۱۶۷/۸	۴۱۱/۱	۱۳۰/۸	۱۰۶/۳	۶۴/۱	۷۷/۷۲
	8	۴۹۴۳/۹	۲۳۶۸/۲	۱۰۵۱/۷	۳۸۷/۱	۱۱۸/۳	۲۰۷/۸	۱۴۵/۵	۲۵۹/۵	۳۰۵/۱
	9	۵۴۸۳/۶	۲۷۱۹/۱	۱۰۸۰/۶	۳۲۷/3	۱۹۵/۳	۲۰۴/1	۶۰/۹	۱۷/۷	۳۸۱/۵
	10	۵۹۱۳/۹	۳۷۴۸/۷	۱۰۲۱/۴	۴۶۱/2	۵۳/۱۰	۱۴۷/۷	۵۸/۳	۹۱/۲	۲۵/۴



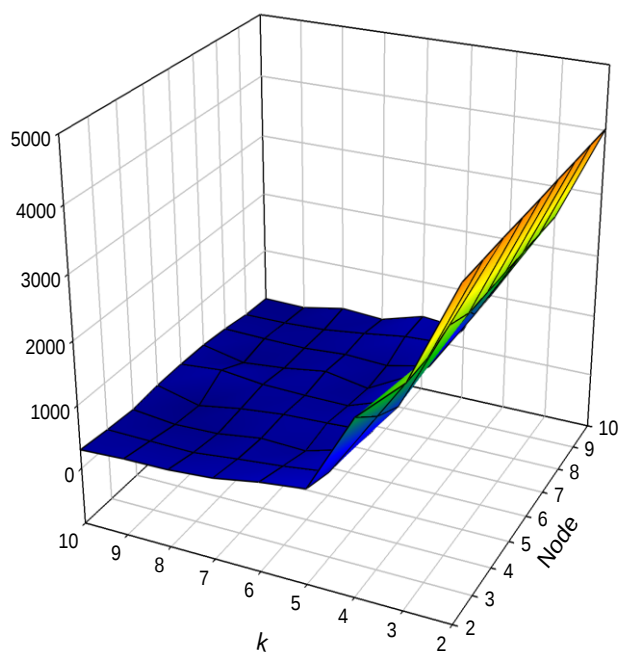
شکل (۴-۲) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی

جدول (۴-۷) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی بایزین و

تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)

تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9	10
	2	4024/8	2503/6	1842/8	597/2	515/6	392/9	341/8	333/5	299/8\
	3	4027/1	2524/2	1492/3	462/8	371/3	303/4	227/8	236/3	248/2\
	4	4032/1	2649/5	1281/6	432/7	227/5	256/3	156/7	135/5	212/2
	5	4027/2	2508/3	1430/8	343/5	304/8	276/4	201/2	84/9	260/9
	6	4024/5	2508/2	1486/8	478/9	226/6	232/3	252/6	334/1	273/5
	7	4029/1	2508/5	1842/9	541/3	161/0	287/7	184/3	117/4	266/8
	8	4026/6	2508/25	1376/4	351/7	203/1	238/1	190/2	201/5	222/0
	9	4027/1	2508/5	1394/31	404/7	370/1	270/1	185/5	153/3	150/8
	10	4026/3	2507/7	1806/9	326/6	396/7	195/6	241/2	109/4	124/9



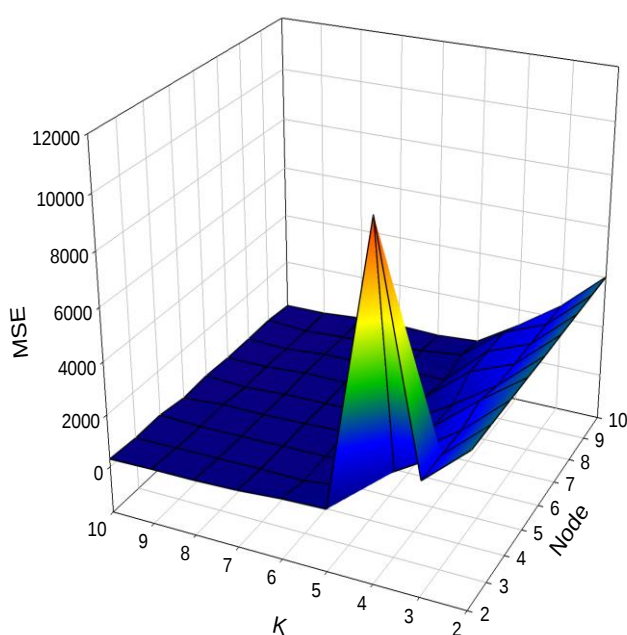
شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول (۴-۸) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)										
۹		2	3	4	5	6	7	8	9	10
	2	۴۰۲۲/۶	۲۵۰۷/۹	۱۱۱۰/۹	۵۷۶/۷	۵۱۹/۱	۳۹۵/۳	۳۴۱/۶	۲۳۲/۴	۲۹۹/۰

3	۴۰۲۰/۳	۲۵۰۹/۴	۱۴۷۴/۰	۳۳۷/۵	۲۵۲/۷	۲۷۷/۴	۲۵۵/۷	۲۳۹/۹	۲۲۵/۵
4	۴۰۲۱/۲	۲۵۰۸/۵	۱۳۵/۳	۴۳۶/۶	۲۷۳/۳	۲۵۵/۱	۲۲۳/۸	۲۱۰/۴	۲۶۷/۵
5	۴۰۲۰/۶	۲۷۸۲/۸	۱۳۹۸/۳	۳۹۷/۹	۳۳۰/۵	۶۳۴/۳	۲۲۲/۴	۲۳۰/۳	۱۰۹/۷
6	۴۰۳/۹	۲۵۰۸/۳	۱۳۵۶/۱	۳۱۱/۶	۲۴۳/۳	۲۲۱/۰	۲۰۱/۸	۱۰۸/۵	۱۹۴/۸
7	۴۰۱۷/۸	۲۵۰۸/۸	۱۸۴۰/۹	۴۸۹/۹	۲۹۹/۴	۴۲/۲	۲۵۴/۴	۱۳۲/۹	۱۴۱/۶
8	۴۰۱۸/۹	۲۵۰۸/۴	۱۸۳۶/۸	۲۸۵/۹	۳۰۳/۹	۲۷۸/۲	۲۳۰/۴	۹۱/۴	۱۵۳/۳
9	۴۰۱۸/۴	۲۵۰۸/۳	۱۸۳۵/۹	۵۲۰/۶	۳۳۸/۱	۷۶۶/۲	۱۷۳/۷	۸۵/۸	۱۸۹/۶
10	۴۰۱۸/۸	۲۵۰۷/۶	۱۳۴۸/۷	۳۱۷/۹	۲۰/۷	۸۸۷/۳	۲۲۹/۴	۹۹/۹	۸۵/۷



شکل (۴-۴) - نمودار سه بعدی شبکه‌هایی با تابع آموزش بایزین و تابع انتقال تانژانت سیگموئیدی

۴-۲-۵-۳- انتخاب تعداد دوره‌های آموزش

در طی آموزش، شبکه با الگوریتم آموزشی لونبرگ مارکوارت، ۹ توصیفگر، ۹ گره در لایه پنهان و تابع انتقال تانژانت سیگموئیدی در تعداد دوره‌های متفاوت، آموزش داده شد و میانگین مربع خطاها برای سری آموزش ثبت گردید. برای هر چرخه میانگین مربع خطاهای سری آموزش برآورد شد

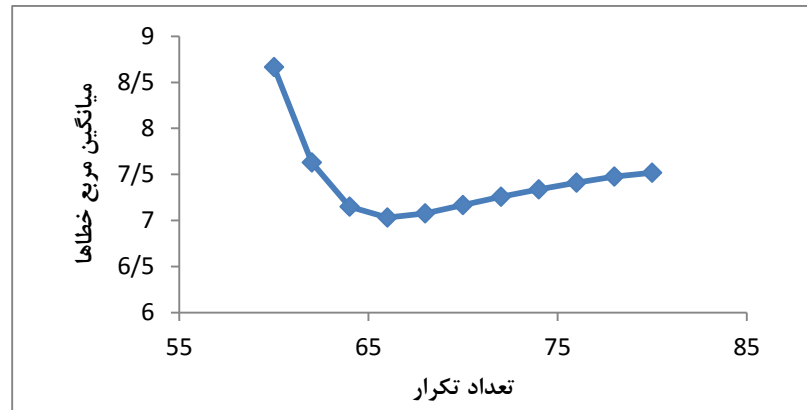
(جدول (4-9)). سپس منحنی تعداد چرخه‌های آموزش بر حسب میانگین مربع خطاها بر طبق شکل (4-5) ترسیم شد. مقدار بهینه تعداد تکرار جایی است که میزان این خطا حداقل باشد. کمترین خطا مربوط به تعداد دور آموزش ۶۶ می‌باشد.

۴-۲-۴-۵-۴- بهینه کردن مقدار μ

جهت یافتن مقدار بهینه μ ، ساختار شبکه با ۹ متغیر ورودی، ۹ گره در لایه پنهان، تابع انتقال تانژانت سیگموئیدی و الگوریتم آموزشی لوبز-مارکوارت در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۱ تا ۱ با گام‌های ۰/۰۵ تغییر داده شد و آنگاه برای هر مورد، مقدار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش محاسبه شد. مشاهده گردید که در این سری از داده‌ها با تغییر مقدار μ ، تغییری در mse ایجاد نمی‌شود. به همین دلیل از همان مقدار پیش فرض آن یعنی ۰/۰۵ استفاده شد.

جدول (4-9) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف

میانگین مربع خطا	تعداد دوره‌های آموزش	میانگین مربع خطا	تعداد دوره‌های آموزش
7/2	70	8/7	60
7/3	72	7/6	62
7/3	74	7/1	64
7/4	76	7/0	66
7/5	78	7/1	68
7/5	80	7/2	70



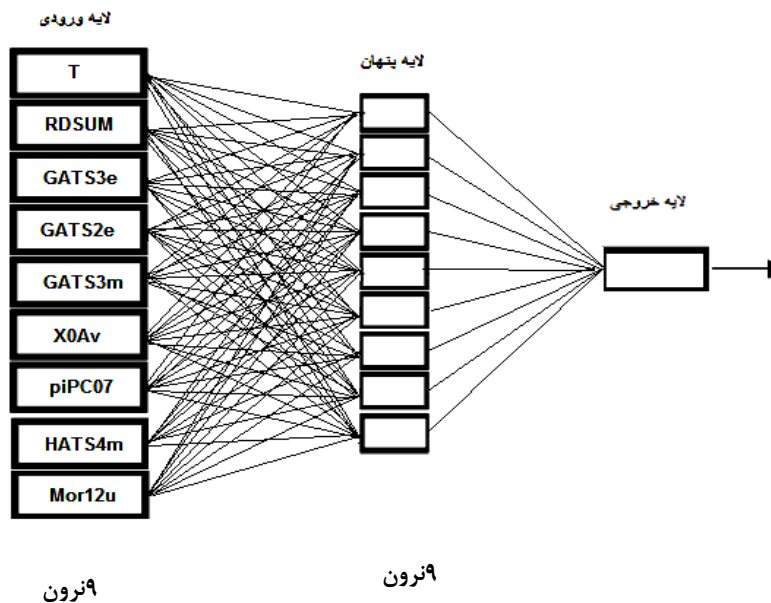
شکل (۴-۵) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دورهای آموزش متفاوت

۴-۲-۵-۵ - معماری شبکه عصبی بهینه شده

با توجه به روند بهینه سازی که در قسمت‌های قبلی ذکر شد، شبکه عصبی مصنوعی بهینه دارای ۹ توصیفگر، ۹ گره در لایه پنهان، تابع آموزشی لونیگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی است. مقادیر بهینه شده در جدول (۴-۱۰) آمده است. شکل (۴-۶) ساختار شبکه عصبی بهینه نهایی را نشان می‌دهد.

جدول (۴-۱۰) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

train lm	تابع آموزش
tansig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
۹	تعداد نرون لایه پنهان
۹	تعداد متغیرهای ورودی
۶۶	تعداد دورهای آموزش
۰/۰۵	پارامتر μ



شکل (۴-۶) - ساختار شبکه عصبی بهینه شده

۴-۲-۶ - انتخاب متغیر به روش الگوریتم ژنتیک

همانطور که پیشتر توضیح داده شد، هدف از اجرای الگوریتم حاضر شناسایی ترکیبی از توصیفگرهاست، به گونه‌ای که برای پیش‌بینی شاخص بازداري ترکیبات بوسیله شبکه عصبی مصنوعی مورد استفاده قرار گیرند. در فرایند اجرای این الگوریتم، ابتدا به صورت اتفاقی یک جمعیت اولیه از توصیف کننده‌ها تولید می‌شود، برازندگی یا شایستگی تک تک اعضای هر نسل یعنی توصیف کننده‌ها، محاسبه و با توجه به شایستگی‌ها، نسل‌های بعدی با اعمال سه عملگر جمعیت، پیوند و جهش تولید یا باز ترکیب می‌شوند. این روند جستجو برای جمعیت‌های مختلف تا حصول ملاک خاتمه ادامه می‌یابد. لازم به ذکر است که برنامه استفاده شده در الگوریتم ژنتیک یک برنامه بهینه شده می‌باشد. مقادیر بهینه شده‌ی پارامترهای جمعیت، پیوند و جهش به ترتیب ۳۰، ۵۰٪ و ۱٪ می‌باشد.

توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک برای سری داده‌ها در جدول (۴-۱۱) آورده شده است.

جدول (۴-۱۱) - کل توصیفگرهای انتخاب شده توسط الگوریتم ژنتیک

No	Symbol	Class	Meaning
۱	T		teperature
۲	BEHv3	BCUT	highest eigenvalue n.3 of Burden matrix
۳	GATS3e	2D autocorrelations	Geary autocorrelation-lag 3
۴	GATS2e	2D autocorrelations	Geary autocorrelation-lag 2
۵	MPC06	topological	Rmolecular path count of order
۶	Mor26v	3D-MorSE	3D-MorSE-signal26
7	BEHe1	BCUT	highest eigenvalue n.1 of Burden
8	Ss	constitutional	Sum of the Kier-Hall electrotopological
9	HATS4m	GETAWAY	leverage-weighted autocorrelation of lag 4

۴-۲-۷- مدل شبکه عصبی مصنوعی با استفاده از توصیفگرهای انتخاب شده

توسط الگوریتم ژنتیک

در این بخش نیز از شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی استفاده شد.

۴-۲-۷-۱- انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع

آموزش و نوع تابع انتقال

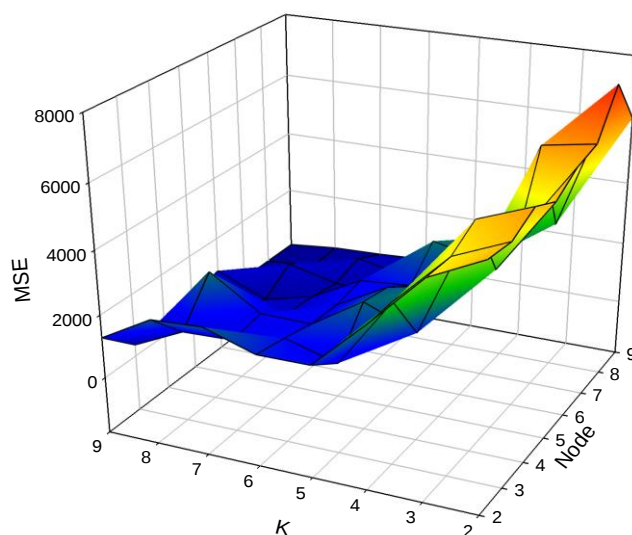
شبکه با ورودی‌های از ۲ تا ۹ توسط دو الگوریتم آموزشی لونیگ - مارکوارت و تنظیم بایزین، با تعداد متفاوت گره در لایه پنهان از ۲ تا ۹ و همچنین توابع لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig)، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمامی این شبکه‌ها تعداد دوره‌های آموزش ثابت و برابر ۳۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده شد معیار نیز به حداقل رساندن میانگین مربع خطا (MSE) در نظر گرفته شد

که نتایج این محاسبات در جداول (۱۲-۴)، (۱۳-۴)، (۱۴-۴) و (۱۵-۴) گرد آوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۷-۴)، (۸-۴)، (۹-۴) و (۱۰-۴) ارائه شده است.

الگوریتم آموزشی لونبرگ - مارکوارت نسبت به الگوریتم آموزشی بایزین دارای میانگین مربع خطا (MSE) کمتری است. همچنین تابع تانژانت سیگموئیدی دارای خطای کمتری نسبت به لگاریتم سیگموئیدی است. بنابراین تابع انتقال تانژانت سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده ۹ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط شاخص بازداری دارند، انتخاب شدند. تعداد ۸ گره نیز در لایه پنهان انتخاب شد.

جدول (۴-۱۲) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	۵۷۰۷/۵	۴۹۱۹/۰	۳۲۷۶/۹	۱۷۴۳/۳۰	۱۷۳۴/۰	۲۲۵۹/۵	۲۱۷۰/۲	۱۲۸۴/۴
	3	۵۵۹۶/۶۵	۴۸۹۹/۸	۳۲۸۹/۶	۱۰۸۶/۳	۱۵۴۳/۸	۱۱۶۳/۶	۱۱۱۰/۹	۳۶۴/۳
	4	۶۰۰۵/۴	۵۳۸۰/۶	۱۶۶۵/۹	۲۳۲۷/۹	۱۳۱۷/۹	۱۳۴۰/۷	۲۳۹۳/۷	۵۷۸/۰
	5	۵۵۹۳/۱	۳۲۹۰/۷	۲۴۸۲/۱	۲۱۴۸/۲	۱۲۷۵/۹	۶۷۸/۲	۷۱۱/۶	۶۱۲/۲
	6	۶۰۶۴/۱	۳۷۱۴/۶	۱۷۳۳/۸	۱۵۶۶/۳	۱۴۳۷/۵	۶۰۳/۹	۱۹۵/۰	۹۰۰/۹
	7	۶۲۴۷/۸	۵۹۸۲/۴	۲۰۵۵/۶	۲۴۲۷/۲	۲۹۹/۸	۴۲۱/۱	۸۰۱/۱	۱۷/۸
	8	۷۵۷۷/۴	۲۹۵۵/۶	۲۲۶۴/۱	۱۱۲۱/۸	۳۵۸/۵	۶۶۸/۷	۱۲۲/۸	۲۷/۴
	9	۶۰۸۱/۱	۴۱۲۳/۹	۱۶۸۰/۶	۵۳۰/۹	۳۱۳/۲	۲۲۶/۳	۱۷۶/۵	۲۳/۹

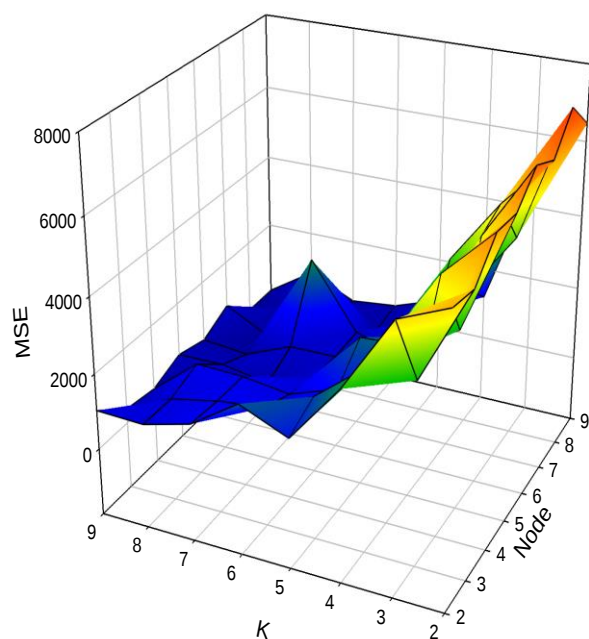


شکل (۴-۷) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (۴-۱۳) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی

لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	5702/7	5154/4	3192/4	1631/99	2180/0	1331/9	992/4	1035/4
	3	5534/2	3038/8	3023/0	2001/1	1393/8	1241/5	976/8	382/2
	4	6296/6	5035/5	2499/9	1509/3	854/0	1223/4	1259/4	257/3
	5	6204/4	3020/5	2316/8	854/0	769/8	362/6	593/0	567/9
	6	6850/9	4857/9	1974/6	769/8	818/8	748/5	236/1	353/5
	7	6485/5	5143/0	3421/6	818/8	1331/9	2611/6	90/4	737/9
	8	7341/9	3708/2	1970/6	894/1	247/4	469/7	332/8	11/5
	9	6478/0	4749/7	1202/7	813/2	372/6	237/7	817/2	11/6

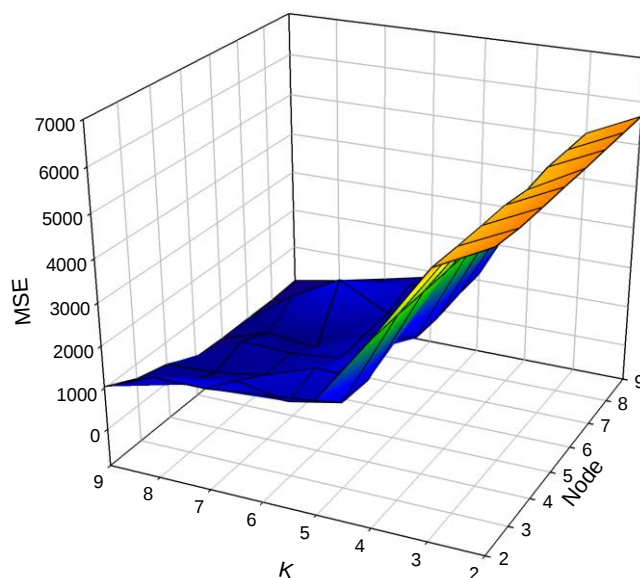


شکل (۴-۸) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی

جدول (۴-۱۴) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم

بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

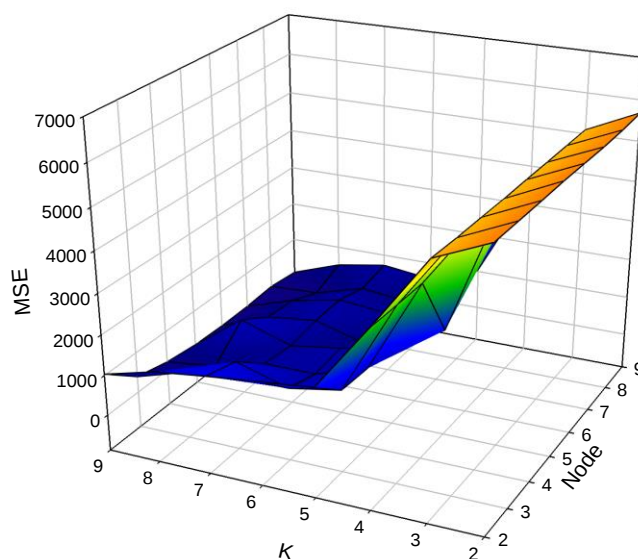
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	۵۷۸۴/۹	5147/6	3399/9	1787/0	1652/0	1523/2	1522/2	13038/7
	3	۵۷۰۵/۱	۵۱۱۵/۴	۳۴۱۵/8	۱۱۶۰/۲	۹۱۱/۸	۱۱۳۸/۹	۷۷۹/۷	۶۵۸/۱
	4	۵۶۸۸/۶	۵۰۹۶/۵	۳۴۰۹/6	۱۱۱۳/۸	۱۱۴۸/۴	۷۱۲/۷	۵۱۴/۶	۴۵۹/۵
	5	۵۶۸۶/۹	۵۰۹۴/۲	۳۴۰۷/۵	۱۵۹۰/9	۸۲۸/2	۷۰۳/۳	۷۴۶/8	۸۲/۳
	6	۵۶۷۰/۱	۴۹۹۷/۷	۳۴۱۱/۱	۱۰۴۳/۵	۸۹۲/۹	۲۸۴/۷۴	۴۳۶/۴	۴۱/۱
	7	۵۶۷۴/۱	۵۰۹۹/۳	۳۴۰۴/۸	۱۰۳۶/۴	۸۹۸/۷	۱۵۳۴/۹	۸۸۲/۲	۳۶/6
	8	۵۶۶۹/۶	۵۰۹۵/۸	۳۳۵۹/8	۱۱۲۷/6	۱۱۴۲/۰	۸۵۰/8	۳۲۲/۱	۲۷/۴
	9	۵۶۶۷/۵	۵۰۹۱/۵	۳۳۷۹/۴	۱۱۶۰/۶	۸۵۰/۹	۵۶۸/۷	۲۹۹/۹	۷۰/۹



شکل (۴-۹) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونیگ-ماکوارت و تابع انتقال تانزانت سیگموئیدی

جدول (۴-۱۵) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)									
تعداد گره های لایه پنهان		2	3	4	5	6	7	8	9
	2	5727/8	5172/9	3435/1	1789/3	1655/0	1518/0	1358/9	1034/5
	3	5714/8	5138/1	3419/9	1141/9	915/2	1305/3	774/9	397/6
	4	5705/8	5128/1	3412/6	1102/2	693/9	691/4	361/4	213/3
	5	5702/7	5109/5	1758/9	1015/8	790/0	601/4	1022/3	113/6
	6	5678/4	5111/9	3403/9	1038/8	350/6	725/2	576/3	87/7
	7	5654/5	5104/4	3403/5	1033/9	634/1	757/9	505/3	100/2
	8	5657/2	5102/2	3398/7	1036/9	415/1	561/5	203/7	153/1
	9	5689/3	5097/9	3397/7	1054/3	562/4	716/9	499/9	61/1



شکل (۴-۱۰) - نمودار سه بعدی شبکه‌هایی با تابع آموزش بایزین و تابع انتقال تانژانت سیگموئیدی

۴-۲ - ۷-۲ - انتخاب تعداد دوره‌های آموزش

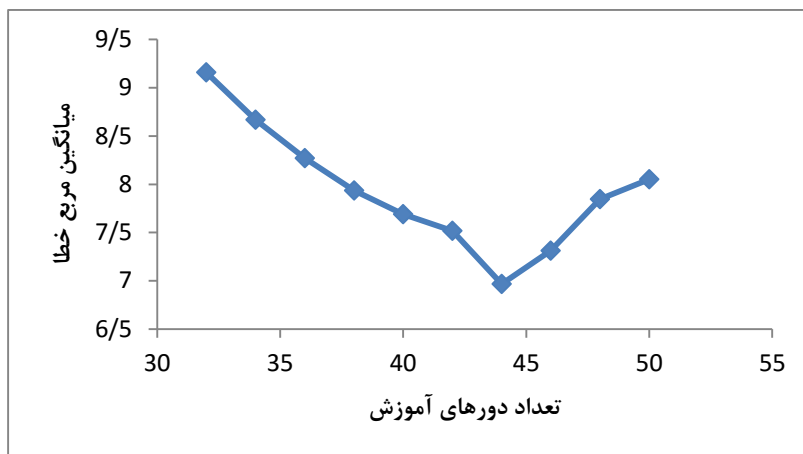
با ۹ توصیفگر منتخب، الگوریتم آموزشی لونبرگ مارکوارت و تابع انتقال تانژانت سیگموئیدی و قرار دادن ۸ گره در لایه پنهان، تعداد چرخه آموزش تغییر داده شد و برای هر چرخه میانگین مربع خطاهای سری آموزش به روش ارزیابی تقاطعی برآورد شد (جدول (۴-۱۶)) سپس منحنی تعداد چرخه‌های آموزش بر حسب میانگین مربع خطاها بر طبق شکل (۴-۱۱) ترسیم شد. مقدار بهینه تعداد تکرار جایی است که میزان این خطا حداقل باشد. مشاهده می‌کنیم که کمترین مقدار مربع خطا مربوط به تعداد دور آموزش ۴۴ می‌باشد.

۴-۲-۷-۳- بهینه کردن مقدار μ

جهت یافتن مقدار بهینه μ ، ساختار شبکه با ۹ متغیر ورودی، ۸ گره در لایه پنهان و الگوریتم آموزشی لونبرگ-مارکوارت در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۱ تا ۱ با گام‌های ۰/۰۵ تغییر داده شد و آنگاه برای هر مورد، مقدار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش محاسبه شد. مشاهده گردید که در این سری از داده‌ها با تغییر مقدار μ ، تغییری در $lmse$ ایجاد نمی‌شود. به همین دلیل از همان مقدار پیش فرض آن یعنی ۰/۰۵ استفاده شد.

جدول (۴-۱۶)- مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف

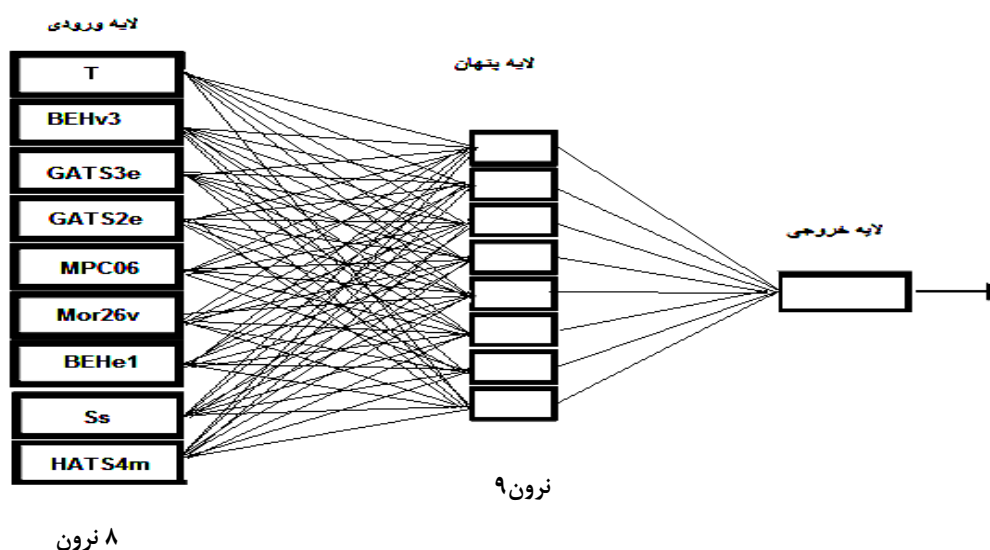
میانگین مربع خطا	تعداد دوره‌های آموزش	میانگین مربع خطا	تعداد دوره‌های آموزش
7/5	42	9/2	32
7/0	44	8/7	34
7/3	46	8/3	36
7/8	48	7/9	38
8/0	50	7/7	40



شکل (۴-۱۱) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دورهای آموزش متفاوت

۴-۲-۷- معماری شبکه عصبی بهینه شده

با توجه به روند بهینه سازی که در قسمت های قبلی ذکر شد، شبکه عصبی مصنوعی بهینه دارای ۹ توصیفگر، ۸ گره در لایه پنهان، تابع آموزشی لونیگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی است. شکل (۴-۱۲) ساختار شبکه عصبی بهینه نهایی را نشان می دهد.



شکل (۴-۱۲) - ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۴-۱۷) ارائه شده است.

جدول (۴-۱۷) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

train lm	تابع آموزش
tansig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
۹	تعداد نرون لایه پنهان
۸	تعداد متغیرهای ورودی
۴۴	تعداد دوره‌های آموزش
۰/۰۵	پارامتر μ

۴-۲-۸- بررسی نتایج

اعتبار و اهمیت معادلات پیشگویی وقتی مشخص می‌گردد که خواص مولکول‌های مجهولی که در دسته‌ی آموزش موجود نیستند را پیشگویی کند. در اینجا چندین روش به منظور ارزیابی توانایی مدل‌های ارائه شده (که دارای کمترین مقدار mse بودند) در پیش‌بینی داده‌های شاخص بازدارندگی شده است.

۴-۲-۸-۱- ارزیابی مدل‌ها با استفاده از سری ارزیابی

مدل‌های منتخب خطی و غیرخطی جهت پیش‌بینی شاخص بازدارندگی سری ارزیابی که شامل ۵۰ ترکیب که بطور تصادفی انتخاب شده‌اند، به کارگرفته شد. پارامترهای آماری حاصل از مدل‌های خطی و غیر خطی با استفاده از شبکه عصبی بهینه شده و همین‌طور نتایج بدست آمده از الگوریتم ژنتیک و شبکه عصبی بهینه شده (GA-ANN) در جدول (۴-۱۸) آورده شده است. همچنین شکل (۴-۱۳)،

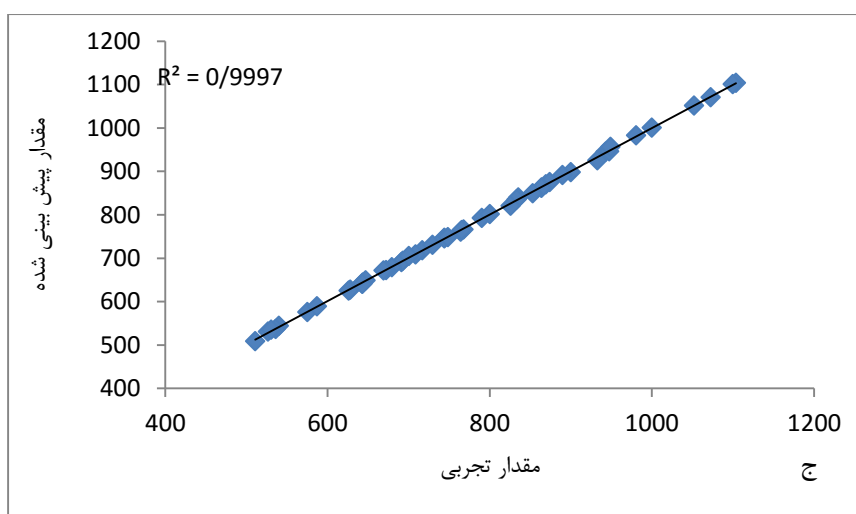
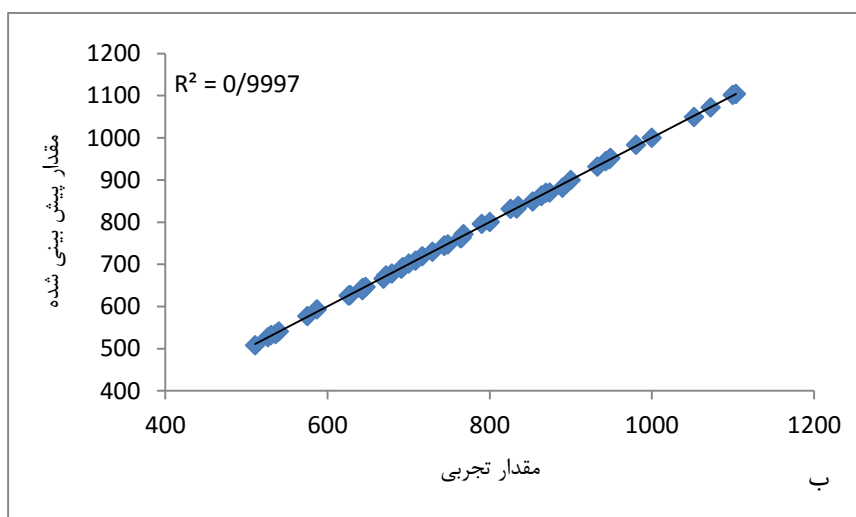
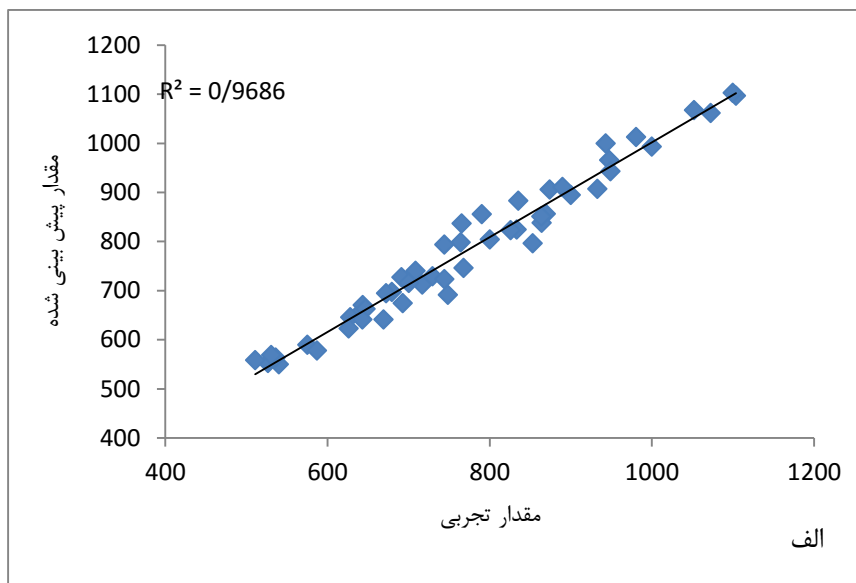
ضریب تعیین مقدار پیش بینی شده بر حسب مقدار تجربی برای سری ارزیابی را با استفاده از هر سه مدل SR-MLR و SR-ANN و GA-ANN نشان می‌دهد. نتایج نشان می‌دهد هر سه روش از دقت قابل قبولی برای پیش بینی شاخص بازدارندگی ترکیبات مورد نظر برخوردارند اما نتایج حاصل از روش SR-ANN به مقادیر داده‌های تجربی نزدیکتر می‌باشد.

جدول (۴-۱۸) - پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری ارزیابی

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازدارندگی			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
۴	943/0	999/8	944/9	945/7	-6/0	-0/2	-0/3
۹	767/8	746/2	771/3	765/7	2/8	-0/4	0/3
۱۴	764/2	798/6	762/1	761/0	-4/5	0/3	0/4
۱۶	980/7	1012/9	983/8	983/1	-3/3	-0/3	-0/2
۱۹	642/6	659/6	641/2	641/6	-2/6	0/2	0/1
۲۵	716/8	712/5	719/6	718/3	0/6	-0/4	-0/2
۳۰	526/4	552/3	526/8	530/3	-4/9	-0/1	-0/7
۳۲	729/4	729/4	729/5	730/5	0/01	-0/01	-0/1
۴۰	947/6	966/1	949/5	946/4	-1/9	-0/2	0/1
۴۴	539/8	550/1	541/0	544/1	-1/9	-0/2	-0/8
۴۸	835/2	882/8	838/7	839/8	-5/7	-0/4	-0/5
۵۷	900/0	894/9	899/6	898/2	0/6	0/04	0/2
۶۲	765/6	837/0	763/4	764/3	-9/3	0/3	0/2
۶۷	575/2	589/9	577/4	575/9	-2/6	-0/4	-0/1
۷۰	869/3	856/1	869/7	869/9	1/5	-0/05	-0/1
۷۵	873/9	905/9	870/3	875/6	-3/7	0/4	-0/2
۷۸	536/0	564/0	534/5	536/8	-5/2	0/3	-0/1
۸۳	1051/9	1067/4	1049/8	1051/8	-1/5	0/20	0/01
۸۸	691/2	727/5	689/5	690/9	-5/2	0/2	0/04
۹۴	932/8	907/5	931/6	925/8	2/7	0/1	0/7

ادامه جدول (۴-۱۸)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
۹۴	932/8	907/5	931/6	925/8	2/7	0/1	0/7
۱۰۱	708/6	740/2	708/6	708/8	-4/5	0/0	-0/03
۱۰۵	643/6	670/7	643/9	643/3	-4/2	-0/05	0/05
۱۰۹	692/7	674/3	693/1	693/6	2/7	-0/06	-0/1
۱۱۳	864/1	838/0	863/1	862/7	3/0	0/1	0/1
۱۱۸	1000/0	993/0	1000/2	1000/4	0/7	-0/02	-0/04
۱۲۳	863/9	851/3	862/1	862/3	1/4	0/2	0/2
۱۳۲	1072/4	1061/4	1072/5	1070/9	1/0	-0/01	0/1
۱۳۷	1103/9	1096/8	1103/8	1104/1	0/6	0/01	-0/02
۱۴۰	744/1	793/7	743/9	746/9	-6/7	0/0	-0/4
۱۴۶	510/8	558/5	508/5	508/4	-9/3	0/4	0/5
۱۴۹	790/4	855/7	796/1	792/8	-8/3	-0/7	-0/3
۱۵۱	627/9	645/9	627/7	627/1	-2/9	0/03	0/1
۱۵۸	744/2	723/3	744/1	745/2	2/8	0/01	-0/1
۱۶۶	748/3	691/5	747/5	748/7	7/6	0/1	-0/05
۱۷۰	625/9	622/7	625/5	625/3	0/53	0/06	0/1
۱۷۶	800/0	804/2	800/2	801/2	-0/5	-0/02	-0/1
۱۸۰	679/5	697/3	678/1	679/3	-2/6	0/2	0/03
۱۸۱	672/1	694/9	673/8	672/2	-3/4	-0/2	-0/01
۱۸۸	668/9	641/0	665/7	671/3	4/2	0/5	-0/4
۱۹۳	647/0	662/7	646/3	648/6	-2/4	0/1	-0/2
۱۹۵	889/6	911/0	881/9	891/4	-2/4	0/9	-0/2
۱۹۸	530/4	569/1	532/3	534/9	-7/3	-0/4	-0/8
۲۰۲	949/2	942/9	952/6	956/7	0/7	-0/4	-0/8
۲۰۷	586/8	578/3	593/2	589/0	1/4	-1/1	-0/4
۲۱۳	833/2	824/7	832/1	835/5	1/0	0/1	-0/3
۲۱۷	643/1	641/0	638/4	640/6	0/3	0/7	0/4
۲۲۳	825/9	823/0	831/9	820/4	0/3	-0/7	0/7
۲۲۷	853/1	796/5	848/7	849/1	6/6	0/5	0/5
۲۳۵	700/2	715/6	701/7	704/6	-2/2	-0/2	-0/6



شکل (۴-۱۳) - نمودار مقادیر پیش‌بینی شده شاخص بازداري بر حسب مقادير تجربی برای سری ارزیابی، الف: با

مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN

۴-۲-۸-۲- ارزیابی مدل‌ها با استفاده از سری تست

بعد از پیش‌بینی شاخص بازداري سری ارزیابی، قدرت پیش‌بینی شبکه، بوسیله ۵۰ ترکیب که در سری تست قرار دارند مورد بررسی قرار گرفت. پارامترهای آماری حاصل از مدل‌های خطی و غیر خطی با استفاده از شبکه عصبی بهینه شده و همین‌طور نتایج بدست آمده از الگوریتم ژنتیک و شبکه عصبی بهینه شده (GA-ANN) در جدول (۴-۱۹) آورده شده است. همچنین شکل (۴-۱۴)، ضریب تعیین مقدار پیش‌بینی شده بر حسب مقدار تجربی برای سری تست را با استفاده از هر سه مدل SR-MLR و SR-ANN و GA-ANN نشان می‌دهد. نتایج بدست آمده فرضیه قبل را تایید می‌کند.

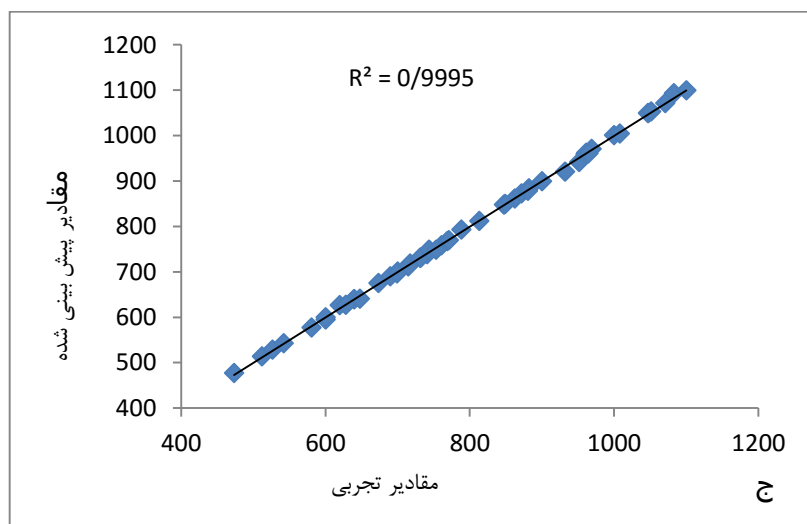
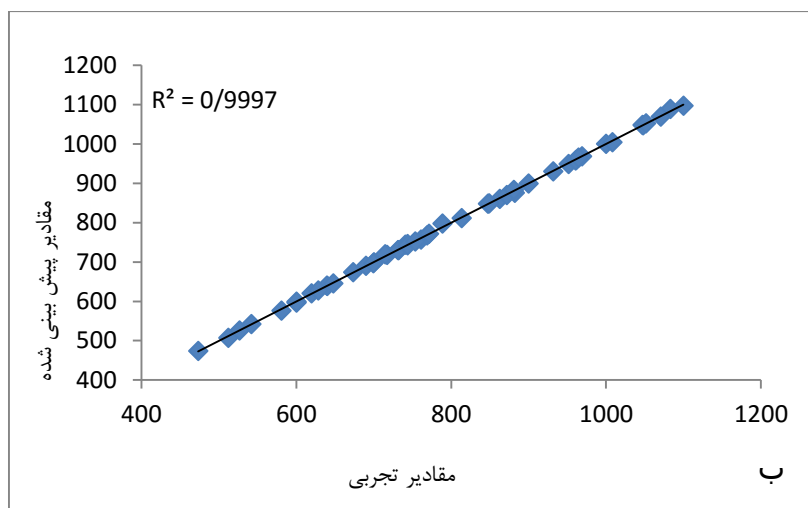
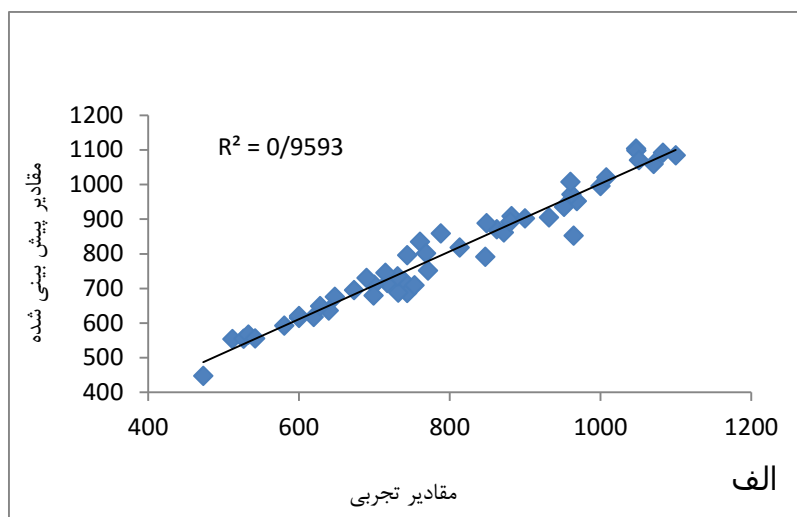
جدول (۴-۱۹)- پارامترهای آماری مدل‌های SR-MLR و SR-ANN و GA-ANN سری تست

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداري			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
2	760/8	834/4	757/2	758/4	-9/7	0/5	0/3
6	473/3	447/9	474/7	476/7	5/4	-0/3	-0/7
11	968/8	951/7	968/8	970/7	1/8	0.0	-0/2
17	1082/9	1091/7	1089/3	1093/4	-0/8	-0/6	-1/0
22	951/6	935/2	949/5	941/3	1/7	0/2	1/1
26	511/9	553/4	507/7	513/4	-8/1	0/8	-0/3
34	931/8	904/9	930/8	920/4	2/9	0/1	1/2
38	740/7	718/1	744/6	738/6	3/0	-0/5	0/3
42	1047/1	1103/4	1049/1	1049/6	-5/4	-0/2	-0/2
46	743/5	686/4	746/0	748/6	7/7	-0/3	-0/7
50	619/6	617/6	620/9	626/8	0/3	-0/2	-1/2
54	600/0	614/9	597/1	594/7	-2/5	0/5	0/9
59	1047/6	1097/7	1097/7	1099/3	-4/8	-4/8	-4/9
65	1100/0	1084/1	1048/1	1049/5	1/4	4/7	4/6
72	1070/7	1058/8	1070/2	1071/0	1/1	0/05	-0/03
74	768/9	801/1	767/8	767/3	-4/2	0/1	0/2
85	717/2	715/1	718/9	718/1	0/3	-0/2	-0/1

ادامه جدول (۴-۱۹)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداري			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN

90	526/6	554/9	526/8	528/6	-5/4	-0/047	-0/4
97	639/8	635/9	639/9	639/2	0/6	-0/01	0/1
103	813/2	817/9	812/0	811/7	-0/6	0/1	0/2
107	847/4	791/4	848/4	847/5	6/6	-0/1	-0/01
111	731/8	687/3	731/0	731/1	6/1	0/1	0/1
115	700/1	710/4	699/9	700/2	-1/5	0/03	-0/01
120	673/4	694/8	674/6	675/0	-3/2	-0/2	-0/2
127	580/5	592/5	576/6	576/80	-2/1	0/7	0/6
129	771/1	751/3	772/1	769/5	2/6	-0/1	0/2
135	882/0	908/4	876/0	884/1	-3/0	0/7	-0/2
139	533/3	566/6	640/4	639/6	-6/2	-20/1	-19/9
143	1051/3	1070/0	1053/3	1052/8	-1/8	-0/2	-0/1
148	689/7	730/1	690/9	690/2	-5/9	-0/2	-0/1
152	731/0	734/5	730/9	730/6	-0/5	0/01	0/05
155	862/4	871/0	861/2	861/1	-1/0	0/1	0/1
160	961/50	971/2	960/4	961/7	-1/0	0/1	-0/02
164	542/0	555/12	543/0	542/9	-2/4	-0/2	-0/2
168	848/9	887/9	849/2	848/9	-4/6	-0/03	0/0
174	600/0	620/0	600/0	600/0	-3/3	0/0	0/0
178	1000/0	995/6	1000/8	1000/7	0/4	-0/1	-0/1
184	960/4	1007/5	958/6	959/3	-4/9	0/2	0/1
190	871/8	861/3	871/1	872/6	1/2	0/1	-0/1
196	1008/0	1020/6	1004/9	1004/2	-1/2	0/3	0/4
200	743/5	796/3	743/2	747/4	-7/1	0/04	-0/5
204	880/6	892/4	883/4	877/9	-1/3	-0/3	0/3
209	788/4	858/3	798/4	792/9	-8/9	-1/3	-0/6
211	628/3	648/5	628/1	627/0	-3/2	0/03	0/2
216	964/3	852/6	966/2	960/9	11/6	-0/2	0/3
221	714/7	745/4	720/4	712/1	-4/3	-0/8	0/4
225	647/7	675/8	646/2	640/7	-4/3	0/2	1/1
229	699/1	679/4	697/70	696/1	2/8	0/2	0/4
232	753/3	709/1	752/3	748/5	5/9	0/1	0/6
237	899/9	902/6	900/4	899/4	-0/3	-0/05	0/05



شکل (۴-۱۴) - نمودار مقادیر پیش‌بینی شده شاخص بازدارندگی بر حسب مقادیر تجربی برای سری تست الف: با مدل

ج: با مدل GA-ANN ب: با مدل SR-ANN الف: با مدل SR-MLR

۴-۲-۸-۳- ارزیابی مدل‌های SR-MLR و SR-ANN و GA-ANN توسط روش

رد مرحله‌ای گروهی^۱

به منظور بررسی بیشتر قدرت پیش‌گویی مدل‌های خطی و غیرخطی روش رد مرحله‌ای گروهی نیز مورد استفاده قرار گرفت. در روش رد مرحله‌ای گروهی هر بار دسته‌ای از ترکیبات به طور کاملاً تصادفی از سری داده‌ها حذف شدند (در اینجا هر بار ۲۵٪ ترکیبات حذف شدند) و سپس با استفاده از مدل ساخته شده توسط بقیه ترکیبات، فعالیت دارویی ترکیبات حذف شده، پیش‌بینی شد. این فرایند برای تمام اعضای سری داده‌ها تکرار شد. نتایج حاصل از رد مرحله‌ای گروهی در جدول (۴-۲۰) ارائه شده است. این نتایج تعمیم‌پذیری بالای مدل خطی و غیرخطی طراحی شده را بیان می‌کند. ضریب تعیین مشاهده شده بین مقادیر پیش‌بینی شده شاخص بازداری (RI) و مقادیر تجربی در شکل (۴-۱۵) نشان دهنده میزان نزدیکی فعالیت پیش‌بینی شده توسط مدل‌ها نسبت به مقدار تجربی آن است.

^۱ - Leave-group-out

جدول (۴-۲۰) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-MLR و SR-ANN,GA-ANN با استفاده از رد

مرحله‌ای گروهی برای کل داده‌ها

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
1	658/0	705/9	658/0	658/0	-7/3	-0/9	0/015
2(t)	760/0	802/4	763/0	763/0	-5/5	-0/8	-0/3
3	852/0	883/1	849/0	849/0	-3/6	-0/7	0/3
4(v)	943/0	964/8	939/0	939/0	-2/3	-0/4	0/3
5	104/10	1047/1	103/7	103/7	-0/5	-0/3	0/4
6(t)	473/0	424/9	472/0	472/0	10/2	-0/1	0/2
7	577/0	539/0	572/0	572/0	6/4	-0/2	0/6
8	666/0	654/3	670/0	670/0	1/8	0/1	-0/6
9(v)	767/0	760/2	761/0	761/0	1/0	-0/2	0/9
10	868/0	861/5	864/0	864/0	0/8	0/5	0/4
11(t)	968/0	960/9	959/0	959/0	0/8	0/2	0/9
12	106/9	1066/9	1063/0	1063/0	0/2	0/2	0/6
13	638/0	620/0	641/0	641/0	2/9	-0/05	-0/5
14(v)	764/2	777/6	765/5	765/5	-1/8	0/6	-0/2
15	866/8	879/1	870/7	870/7	-1/4	-0/6	-0/4
16(v)	980/7	986/240	984/9	984/9	-0/6	-0/3	-0/4
17(t)	1082/9	1067/5	1087/4	1087/4	1/4	0/3	-0/4
18	537/3	537/9	534/1	534/1	-0/1	0/6	0/6
19(v)	642/6	632/4	638/1	638/1	1/6	0/4	0/70
20	746/9	744/6	745/4	745/4	0/3	0/1	0/2
21	849/5	845/7	843/4	843/4	0/4	0/6	0/7
22(t)	951/6	919/5	947/1	947/1	3/4	0/1	0/5
23	1053/1	1040/3	1048/4	1048/4	1/2	0/5	0/4
24	865/7	883/6	861/2	861/2	-2/1	0/3	0/5

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
25(v)	716/8	744/1	718/90	717/0	-3/8	-0/3	-0/03

26(t)	511/9	514/3	507/9	510/6	-0/5	0/8	0/2
27	593/9	582/3	587/3	588/7	1/9	1/1	0/9
28	693/5	724/4	688/5	690/6	-4/4	0/7	0/4
29	793/8	841/04	785/9	787/6	-5/9	1/0	0/8
30(v)	526/4	536/2	527/2	526/5	-1/8	-0/1	0/0
31	626/6	631/8	641/2	627/3	-0/8	-2/3	-0/1
32(v)	729/4	719/5	736/1	723/3	1/4	-0/9	0/8
33	830/2	809/5	827/4	837/6	2/5	0/3	-0/9
34(t)	931/8	899/9	939/4	931/2	3/4	-0/8	0/1
35	847/8	873/8	849/5	846/0	-3/1	-0/2	0/2
36	947/3	879/2	941/8	941/7	7/2	0/6	0/6
37	638/5	653/9	632/2	638/3	-2/4	1/0	0/03
38(t)	740/7	750/8	745/3	739/0	-1/4	-0/6	0/2
39	842/8	846/9	852/5	845/2	-0/5	-1/1	-0/3
40(v)	947/6	952/4	955/2	954/3	-0/5	-0/8	-0/7
41	706/5	759/4	698/3	704/6	-7/5	1/2	0/3
42(t)	1047/1	1075/1	1046/9	1047/0	-2/7	0/02	0/01
43	806/9	832/3	807/3	808/6	-3/1	-0/05	-0/2
44(v)	539/8	533/4	542/8	542/7	1/2	-0/5	-0/5
45	641/3	632/9	644/6	646/5	1/3	-0/5	-0/8
46(t)	743/5	720/6	750/6	748/1	3/077	-0/9	-0/6
47	844/8	813/8	850/5	847/9	3/7	-0/7	-0/4
48(v)	835/2	875/6	837/8	834/8	-4/8	-0/3	0/047

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازاری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
49	689/9	675/4	684/5	691/0	2/1	0/8	-0/3
50(t)	619/6	628/3	620/3	620/0	-1/4	-0/1	-0/1
51	726/0	695/9	730/9	728/0	4/1	-0/7	-0/4
52	744/9	719/9	744/1	750/0	3/3	0/1	-0/7
53	865/4	812/5	857/7	863/0	6/1	0/9	0/2
54(t)	600/0	617/8	594/9	599/0	-3/0	0/8	0/1
55	700/0	708/2	697/5	698/0	-1/1	0/4	0/3

56	800/0	800/3	796/3	799/0	-0/0	0/5	0/02
57(v)	900/0	896/1	893/8	899/0	0/4	0/7	0/03
58	1000/0	995/3	1000/1	100/1	0/5	-0/01	-0/190
59(t)	1100/0	1098/5	1105/8	110/1	0/1	-0/5	-0/1
60	668/5	694/4	670/3	669/0	-3/9	-0/3	-0/164
61	662/3	708/1	663/8	663/0	-6/9	-0/2	-0/2
62(v)	765/6	804/5	768/3	765/0	-5/1	-0/3	0/1
63	858/1	885/3	858/6	858/0	-3/2	-0/1	-0/01
64	948/6	966/9	947/0	948/0	-1/9	0/2	0/05
65(t)	1047/6	1049/3	1046/7	104/7	-0/2	0/1	0/02
66	471/2	427/0	474/5	475/0	9/4	-0/7	-0/8
67(v)	575/2	541/1	577/4	578/0	5/9	-0/4	-0/5
68	666/6	656/5	668/8	668/0	1/5	-0/3	-0/2
69	768/6	762/3	770/1	769/0	0/8	-0/2	-0/1
70(v)	869/3	863/6	870/9	869/0	0/6	-0/2	-0/1
71	969/8	963/0	969/5	970/0	0/7	0/030	-0/1
72(t)	1070/7	1069/0	1070/5	107/2	0/1	0/018	-0/1
73	641/1	622/2	642/3	641/0	2/9	-0/2	-0/1

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجری	مقدار پیش‌بینی شده شاخص بازداري			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
74(t)	768/9	779/8	768/3	769/0	-1/4	0/1	-0/03
75(v)	873/9	881/3	874/7	875/0	-0/8	-0/1	-0/2
76	989/2	988/4	990/4	989/0	0/1	-0/1	-0/05
77	1093/0	1069/7	1091/9	109/3	2/1	0/1	-0/04
78(v)	536/0	540/1	531/5	534/0	-0/8	0/9	0/3
79	641/1	634/5	640/6	641/0	1/0	0/1	0/01
80	745/2	746/7	741/5	746/0	-0/2	0/5	-0/2
81	848/1	847/8	848/2	843/0	0/03	-0/01	0/5
82	950/2	921/6	946/7	948/0	3/0	0/4	0/2
83(v)	1051/9	1042/4	1052/4	103/8	0/9	0/0	1/2
84	869/9	885/7	872/5	878/0	-1/8	-0/3	-1/0
85(t)	717/200	746/204	717/800	709/0	-4/0	-0/1	1/1
86	510/2	516/5	510/7	539/0	-1/2	-0/1	-5/8
87	591/1	584/4	591/3	582/0	1/1	-0/03	1/5

88(v)	691/2	726/5	689/2	703/0	-5/1	0/3	-1/8
89	791/7	843/1	792/1	792/0	-6/5	-0/05	-0/1
90(t)	526/6	538/3	528/6	523/0	-2/2	-0/4	0/6
91	628/1	633/9	630/3	623/0	-0/9	-0/3	0/8
92	730/6	721/6	732/5	725/0	1/2	-0/3	0/7
93	831/4	811/6	831/8	839/0	2/4	-0/05	-0/9
94(v)	932/8	902/0	932/7	935/0	3/3	0/01	-0/3
95	855/2	875/9	854/6	841/0	-2/4	0/1	1/6
96	952/1	881/4	952/0	951/0	7/4	0/01	0/1
97(t)	639/8	656/1	640/2	645/0	-2/5	-0/1	-0/9

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجری	مقدار پیش‌بینی شده شاخص بازاری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
98	742/6	752/9	744/3	736/0	-1/4	-0/2	0/9
99	844/8	849/1	843/6	857/0	-0/5	0/1	-1/4
100	954/4	954/6	954/1	958/0	0/0	0/0	-0/4
101(v)	708/6	761/5	709/2	708/0	-7/5	-0/1	-0/01
102	1056/2	1077/3	1056/4	105/7	-2/0	-0/02	-0/1
103(t)	813/2	834/5	813/4	815/0	-2/6	0/0	-0/2
104	541/500	535/594	542/000	538/000	1/1	-0/1	0/6
105(v)	643/600	635/026	645/200	638/000	1/3	-0/2	0/7
106	745/900	722/757	744/900	750/000	3/1	0/1	-0/6
107(t)	847/400	815/963	848/400	836/000	3/7	-0/1	1/2
108	841/9	877/5	845/0	843/0	-4/2	-0/4	-0/2
109(v)	692/7	677/5	692/5	688/0	2/2	0/03	0/6
110	622/5	630/5	622/6	623/0	-1/3	-0/02	-0/2
111(t)	731/8	698/0	730/9	741/0	4/6	0/1	-1/3
112	747/	722/1	748/2	759/0	3/3	-0/2	-1/7
113(v)	864/1	814/6	862/1	861/0	5/7	0/2	0/3
114	599/9	619/9	600/1	600/0	-3/3	-0/03	-0/016
115(t)	700/1	710/3	700/5	695/9	-1/5	-0/1	0/6
116	800/0	802/5	802/6	801/8	-0/3	-0/3	-0/2

117	900/0	898/2	898/9	887/0	0/2	0/1	1/4
118(v)	1000/0	997/4	1001/4	1006/0	0/3	-0/1	-0/6
119	1100/0	1100/6	1099/4	1083/3	-0/1	0/05	1/5
120(t)	673/4	696/6	676/7	677/0	-3/4	-0/5	-0/6
121	668/4	710/2	666/9	682/0	-6/2	0/2	-2/1

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازاری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
122	771/1	806/6	769/3	777/8	-4/6	0/2	-0/9
123(v)	863/9	887/4	863/6	863/2	-2/7	0/0	0/1
124	954/6	969/0	952/5	954/1	-1/5	0/2	0/05
125	1053/6	1051/4	1054/3	1049/6	0/2	-0/1	0/4
126	479/9	429/2	471/9	478/5	10/6	1/7	0/3
127(t)	580/5	543/3	575/2	545/1	6/4	0/9	6/1
128	670/0	658/6	668/2	690/7	1/7	0/3	-3/1
129(t)	771/1	764/5	771/0	744/7	0/9	0/01	3/4
130	871/3	865/8	870/0	873/4	0/6	0/1	-0/2
131	971/7	965/2	969/9	980/0	0/7	0/2	-0/8
132(v)	1072/4	1071/2	1072/9	1060/3	0/1	-0/05	1/1
133	644/8	624/3	643/6	638/5	3/2	0/2	1/0
134	774/6	781/9	775/0	774/6	-1/0	-0/05	0/0
135(t)	882/0	883/4	879/1	882/8	-0/1	0/3	-0/1
136	998/7	990/5	997/7	1012/8	0/8	0/1	-1/4
137(v)	1103/9	1071/8	1102/7	1095/5	3/0	0/1	0/8
138	533/3	542/2	532/4	520/1	-1/7	0/2	2/5
139(t)	639/4	636/7	639/1	643/2	0/4	0/05	-0/6
140(v)	744/1	748/9	746/0	752/2	-0/6	-0/2	-1/1
141	847/2	849/9	847/2	840/9	-0/3	0/0	0/7
142	949/4	923/8	953/9	951/7	2/7	-0/5	-0/2
143(t)	1051/3	1044/6	1051/5	1049/1	0/6	-0/02	0/2
144	875/1	887/9	876/3	872/8	-1/5	-0/1	0/3
145	718/3	748/3	715/5	716/5	-4/2	0/4	0/2

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازاری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
146(v)	510/8	518/6	511/5	507/2	-1/5	-0/1	0/7
147	589/6	586/6	590/8	583/0	0/5	-0/2	1/1
148(t)	689/7	728/7	691/4	687/7	-5/6	-0/2	0/3
149(v)	790/4	845/3	792/4	789/4	-6/9	-0/2	0/1
150	528/4	540/4	525/4	522/8	-2/3	0/6	1/1
151(v)	627/9	636/1	628/6	627/8	-1/3	-0/1	0/01
152(t)	731/0	723/7	729/4	721/7	1/0	0/2	1/3
153	832/4	813/8	831/3	839/0	2/2	0/1	-0/8
154	934/1	904/2	931/6	925/6	3/2	0/3	0/9
155(t)	862/4	878/0	864/8	857/3	-1/8	-0/3	0/6
156	957/9	883/5	959/3	954/4	7/8	-0/1	0/4
157	640/6	658/2	639/7	638/9	-2/8	0/1	0/3
158(v)	744/2	755/1	745/1	742/5	-1/5	-0/120	0/2
159	847/0	851/3	846/6	842/1	-0/5	0/05	0/6
160(t)	961/5	956/7	962/0	969/5	0/5	-0/05	-0/8
161	710/8	763/6	712/4	702/9	-7/4	-0/2	1/1
162	1065/9	1079/4	1066/3	1065/9	-1/3	-0/04	0/0
163	819/6	836/6	819/7	827/5	-2/1	-0/01	-0/9
164(t)	542/0	537/7	543/4	540/9	0/8	-0/2	0/2
165	645/6	637/2	645/9	643/3	1/3	-0/05	0/4
166(v)	748/3	724/9	748/4	747/3	3/1	-0/01	0/1
167	850/0	818/1	846/7	847/6	3/7	0/4	0/3
168(t)	848/9	879/9	849/0	847/9	-3/6	-0/01	0/1
169	696/0	679/6	697/2	688/8	2/3	-0/2	1/0
170(v)	625/9	632/6	627/7	619/8	-1/0	-0/3	0/9

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداري			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
171	735/9	700/2	736/5	730/0	4/8	-0/1	0/8
172	750/3	724/2	749/3	743/0	3/5	0/1	1/0
173	862/8	816/8	864/1	856/0	5/3	-0/1	0/8
174(t)	600/0	622/1	599/4	600/0	-3/7	0/1	0/0
175	700/1	712/5	701/6	697/0	-1/8	-0/2	0/4
176(v)	800/0	804/6	802/5	797/0	-0/6	-0/3	0/3
177	900/0	900/4	898/1	895/0	-0/04	0/2	0/5
178(t)	1000/0	999/6	1003/0	1002/0	0/04	-0/3	-0/3
179	1100/0	1102/8	1100/5	1095/0	-0/2	-0/04	0/5
180(v)	679/5	698/7	679/2	676/0	-2/8	0/04	0/5
181(v)	672/1	712/3	669/7	680/9	-6/0	0/4	-1/3
182	775/7	808/8	771/4	785/9	-4/3	0/5	-1/3
183	869/1	889/5	866/8	875/6	-2/3	0/2	-0/7
184(t)	960/4	971/2	957/0	964/3	-1/1	0/3	-0/4
185	1059/5	1053/5	1055/3	1058/3	0/6	0/4	0/113
186	472/0	431/3	476/3	470/0	8/6	-0/9	0/4
187	576/8	545/4	583/2	577/3	5/4	-1/1	-0/1
188(v)	668/9	660/8	673/9	671/8	1/2	-0/7	-0/4
189	771/0	766/6	773/9	769/7	0/6	-0/376	0/168
190(t)	871/8	867/9	873/2	876/0	0/4	-0/1	-0/5
191	972/6	967/3	975/4	965/9	0/5	-0/2	0/7
192	1073/6	1073/3	1074/3	1075/8	0/03	-0/1	-0/2
193(v)	647/0	626/4	652/7	661/3	3/2	-0/9	-2/2
194	779/5	784/1	777/20	787/6	-0/6	0/3	-1/0

ادامه جدول (۴-۲۰)

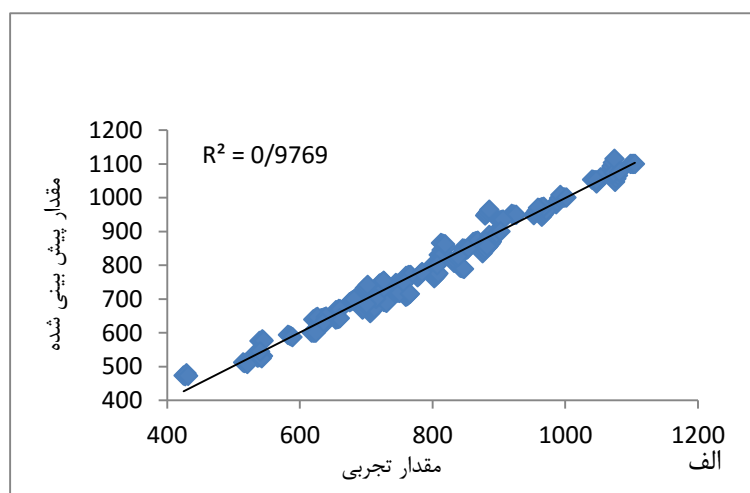
شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده شاخص بازداري			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
195(v)	889/0	885/6	885/5	892/0	0/4	0/4	-0/3
196(t)	1008/0	992/7	1004/5	1001/0	1/5	0/3	0/7
197	1115/0	1073/9	1110/2	1079/1	3/7	0/4	3/2

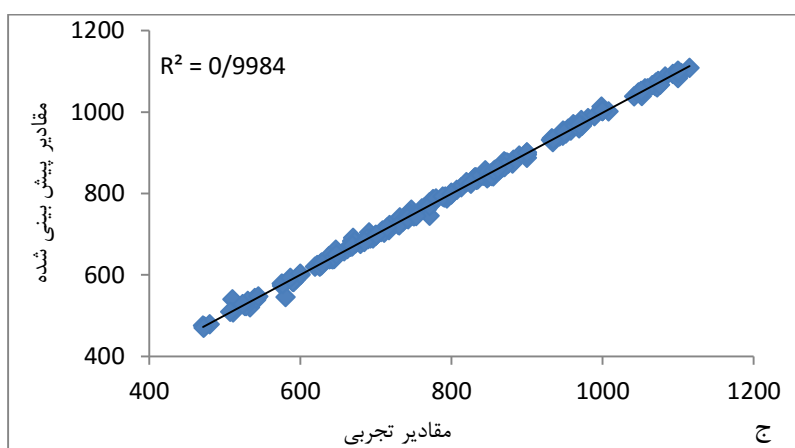
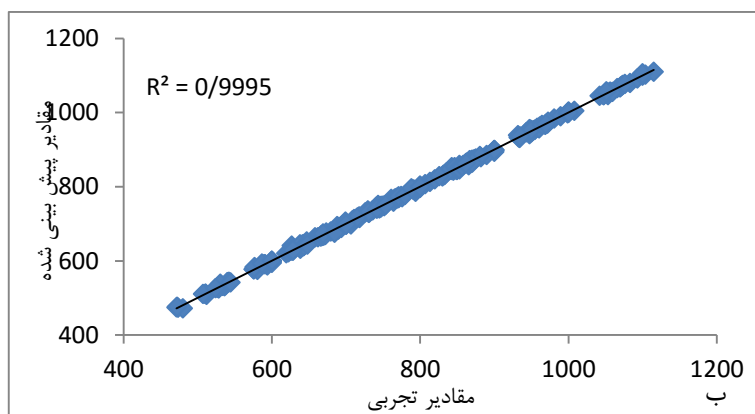
198(v)	530/0	544/4	538/6	536/1	-2/7	-1/6	-1/6
199	638/0	638/8	644/3	645/0	-0/1	-0/9	-1/1
200(t)	743/0	751/0	749/3	746/0	-1/1	-0/8	-0/4
201	846/0	852/1	852/4	849/0	-0/7	-0/7	-0/3
202(v)	949/0	925/9	951/9	948/0	2/4	-0/3	0/1
203	1051/0	1046/7	1057/7	1051/0	0/4	-0/6	0/0
204(t)	880/0	890/0	883/3	873/0	-1/1	-0/4	0/8
205	718/0	750/4	719/5	722/0	-4/5	-0/2	-0/5
206	507/0	520/7	510/7	509/1	-2/7	-0/7	-0/4
207(v)	586/8	588/7	593/2	592/0	-0/3	-1/1	-0/9
208	688/1	730/8	694/1	688/0	-6/2	-0/9	0/01
209(t)	788/0	847/5	795/3	791/0	-7/5	-0/9	-0/4
210	523/0	542/6	525/9	527/0	-3/7	-0/5	-0/8
211(t)	628/0	638/2	627/2	630/0	-1/6	0/1	-0/3
212	730/0	725/9	730/8	723/0	0/6	-0/1	0/9
213	833/0	815/9	831/6	830/0	2/0	0/2	0/3
214	934/0	906/3	935/4	931/0	3/0	-0/1	0/3
215	870/0	880/2	864/6	864/0	-1/2	0/6	0/7
216(t)	964/0	885/7	967/1	965/0	8/1	-0/3	-0/1
217(v)	641/0	660/3	640/2	642/0	-3/0	0/1	0/1
218	747/0	757/3	743/8	744/0	-1/4	0/4	0/4
219	849/0	853/4	847/0	841/0	-0/412	0/2	0/9

ادامه جدول (۴-۲۰)

شماره ترکیب	مقدار تجربی	مقدار پیش‌بینی شده‌شخص بازاری			درصد خطا		
		SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
220	969/0	958/8	969/7	968/0	1/0	-0/1	0/1
221(t)	714/0	765/8	716/6	715/8	-7/2	-0/4	-0/2
222	1076/0	1081/6	1077/6	1066/7	-0/5	-0/1	0/9
223(v)	825/0	838/7	828/6	823/6	-1/7	-0/4	0/2
224	544/0	539/9	541/5	546/8	0/7	0/4	-0/5
225(t)	647/0	639/3	643/5	650/1	1/2	0/5	-0/5
226	751/4	727/0	747/0	747/7	3/2	0/6	0/5
227(v)	853/0	820/2	847/1	842/4	3/8	0/7	1/2
228	856/0	882/0	855/6	855/5	-3/0	0/05	0/1

229(t)	699/0	681/8	705/5	695/0	2/5	-0/9	0/6
230	628/6	634/8	628/8	627/2	-1/0	-0/03	0/2
231	741/7	702/3	740/8	738/0	5/3	0/1	0/5
232(t)	753/3	726/4	750/5	744/0	3/6	0/4	1/2
233	861/4	818/9	863/4	860/4	4/9	-0/2	0/1
234	599/0	624/3	600/4	602/3	-4/2	-0/2	-0/5
235(v)	700/2	714/6	697/9	697/5	-2/1	0/3	0/4
236	800/0	806/7	799/7	797/0	-0/8	0/04	0/4
237(t)	899/0	902/5	899/5	900/8	-0/4	-0/05	-0/2
238	1000/0	1001/7	999/9	1002/0	-0/2	0/01	-0/2
239(v)	1100/0	1104/9	1104/5	1100/4	-0/4	-0/4	-0/0
240	684/0	700/8	675/7	680/7	-2/4	1/2	0/5





شکل (۴-۱۵) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی شاخص بازداري به روش رد مرحله‌ای گروهی برای کل داده‌ها الف: با مدل SR-MLR ب: با مدل SR-ANN ج: با مدل GA-ANN

نمودار باقی‌مانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد و اگر باقی‌مانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند، نشان می‌دهد که مدل مناسبی به دست آمده و هیچ خطای سیستماتیکی وجود ندارد. شکل (۴-۱۶) نمودار باقی‌مانده‌ها را که از تفاضل مقدار تجربی فعالیت دارویی از مقدار محاسبه شده توسط مدل‌های خطی و غیرخطی به دست می‌آید، بر حسب مقدار تجربی نشان می‌دهد. تقارن پراکندگی نقاط در دو طرف محور افقی عدم وجود خطای سیستماتیک را نشان می‌دهد.

۴-۲-۸-۴- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای

آماري

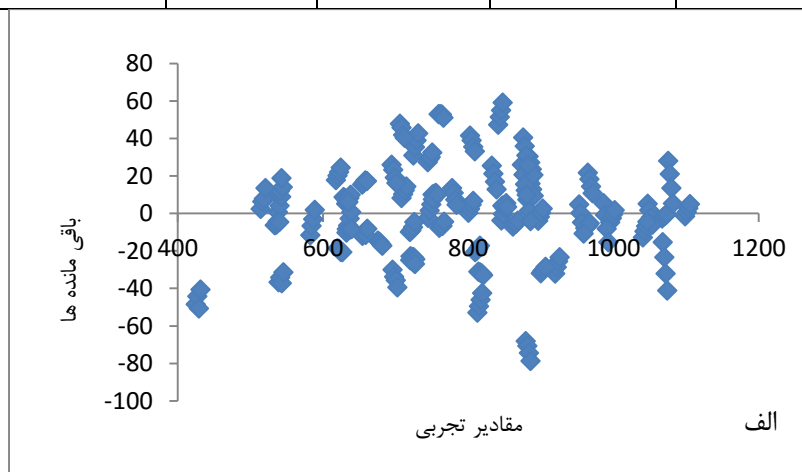
مطابق جدول (۴-۲۱) هفت پارامتر آماری، جهت ارزیابی توانایی پیشگویی مدل‌های ساخته شده به

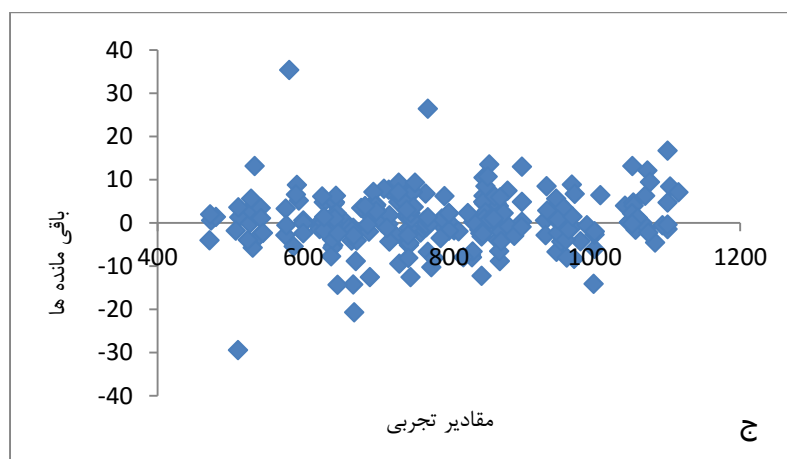
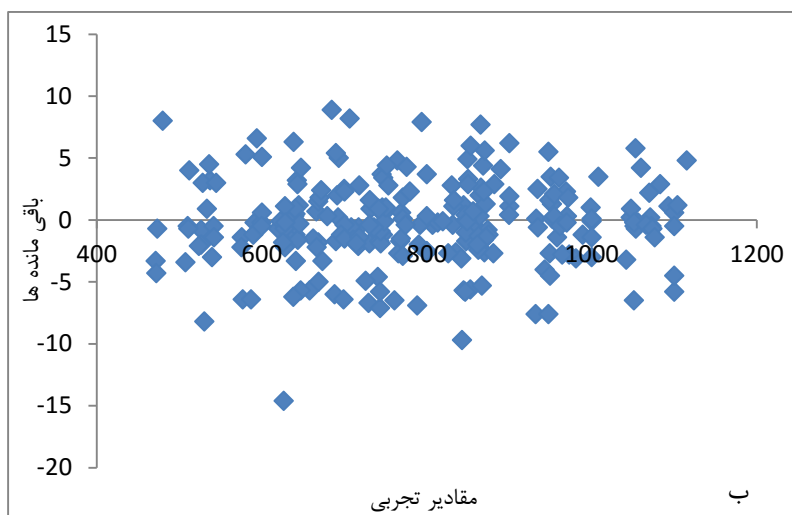
روش SR-MLR, SR-ANN و GA-ANN به کار گرفته شد. نتایج رضایت بخش مدل SR-ANN

نشان می‌دهد که مشارکت بعضی توصیفگرها برای شاخص بازداري ممکن است، غیرخطی باشد.

جدول (۴-۲۱) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط SR-MLR, GA-ANN و SR-ANN

پارامتر	سری تست			کل داده‌ها		
	SR-MLR	SR-ANN	GA-ANN	SR-MLR	SR-ANN	GA-ANN
MAE	۲۷/۱	۱/۸	۲/۴	۱۸/۳	۲/۵	۴/۳
MSE	۱۲۱۳/۲	۷/۱	۱۲/۶	۵۸۹/۷	۱۱/۷	۴۰/۲
Press	۶۰۶۶۰/۴	۳۵۴/۹	۶۳۱/۸	۱۴۱۵۳۷/۸	۲۸۱۰/۳	۹۶۶۰/۵
SEP	۳۴/۸	۲/۷	۳/۵	۲۴/۳	۳/۴	۶/۳
MRE	۳/۵	۰/۲	۳۱/۳	۲/۴	۰/۳	۰/۶
R ²	۰/۹۵۹	۰/۹۹۹	۰/۹۹۹	-	-	-
R ² (LGO)	-	-	-	۰/۹۷۶	۰/۹۹۹	۰/۹۹۸





شکل (۴-۱۶) - نمودار باقی مانده ها بر حسب مقدار تجربی شاخص بازداري برای کل داده ها الف: با مدل SR-MLR ب: با

مدل SR-ANN ج: مدل GA-ANN

۴-۲-۸-۵- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون

Y- تصادفی^۱

این تکنیک ارزیابی مدل با هدف بررسی هر گونه ارتباط تصادفی بین داده ها انجام شد. در این آزمون مقادیر تصادفی از متغیر وابسته تولید گردید. مدل QSPR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. اگر مدل اصلی هیچ گونه

^۱ - Y- randomization test

ارتباط تصادفی نداشته باشد، تفاوت قابل توجهی بین مقدار ضریب تعیین مدل اصلی و مدل QSPR که با پاسخهای تصادفی ایجاد شده، وجود دارد. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۲۲-۴) نشان داده شده است. مقادیر کوچک ضریب تعیین (R^2) بیانگر عدم وجود ارتباط شانس یا وابستگی ساختاری به سری آموزش در مدل توسعه یافته توسط شبکه می باشد.

جدول (۲۲-۴) - الف: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش SR-ANN

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	0/03	۰/01	۰/02	۰/2	۰/۳	۰/01	۰/1	۰/۱	۰/0	۰/1

جدول (۲۲-۴) - ب: مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی روش GA-ANN

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	۰/۰۱	۰/۱	۰/۱	۰/۰۵	۰/۱	۰/۰۲	۰/۱	۰/۲	۰/۰۱	۰/02

۴-۲-۸-6- بررسی ارتباط توصیفگرهای وارد شده در مدل با فعالیت

بازدارندگی

در این قسمت با توجه به توصیفگرهای وارد شده در مدل SR-MLR، SR-ANN و GA-ANN یک بررسی اجمالی روی اثرات این متغیرها بر شاخص بازدارندگی ترکیبات مورد مطالعه صورت خواهد گرفت.

• توصیفگرهای Topological

این دسته توصیفگرها که از گراف مولکولی تهی از هیدروژن محاسبه می‌شوند، از ساده‌ترین توصیفگرهای مولکولی بوده و به ساختار فضایی مولکول ارتباطی ندارند. توصیفگرهای توپولوژیکی نشان‌دهنده‌ی نوع اتم، نوع و تعداد پیوند و چگونگی ارتباط با یکدیگر بوده و به سادگی از روی ساختمان دوبعدی مولکول قابل محاسبه‌اند [34 و 35]. از میان این توصیفگرها، متوسط مربع شاخص فاصله متقابل وینر (RDSUM)¹ در مدل ظاهر شده است. شاخص وینر که با W ، نشان داده می‌شود، به عنوان مجموع تمام فاصله‌های توپولوژیکی در گراف هیدروژن تهی تعریف می‌شود. با نگاهی به ساختار و مقادیر توصیفگر RDSUM در می‌یابیم که با افزایش اندازه مولکول‌ها رابطه مستقیم دارد. اما تغییر در مرتبه پیوندها تأثیری در افزایش یا کاهش این توصیفگر بر روی مولکول‌هایی با تعداد کربن-های یکسان ندارد. همچنین در مولکول‌های قطبی تر افزایش می‌یابد. چند مثال از اثر این توصیفگر در جدول (۴-۲۳) آورده شده است.

جدول (۴-۲۳)-الف: اثر توصیفگر RDSUM بر شاخص بازداری

شاخص بازداری	RDSUM	نام ترکیب
537/3	۴/۴	n-Propanol
849/5	11/1	n-Hexanol
951/6	۷۴۳/۱	n-heptanol

جدول (۴-۲۳)-ب: اثر توصیفگر RDSUM بر شاخص بازداری

شاخص بازداری	RDSUM	نام ترکیب
638/5	6/4	n-Butanenitrile

¹-reciprocal distance Wiener-type index

<i>n</i> -Butylamine	6/4	626/6
<i>n</i> -Hexanenitrile	11/1	842/8
<i>n</i> -Hexylamine	11/1	830/2

همانطور که در جدول ((۴-۲۳) الف) می‌بینیم با افزایش اندازه مولکول‌ها، RDSUM و نیز شاخص بازداري افزایش می‌یابند. با توجه به جدول ((۴-۲۳) ب) با تغییر گروه‌های عاملی تغییر چندانی در RDSUM مولکول‌هایی با تعداد کربن‌های (R) یکسان ایجاد نمی‌شود.

• توصیفگرهای D-autocorrelation

توصیفگرهای دوبعدی از گراف مولکولی و از طریق محاسبه مجموع اوزان اتم‌های انتهایی کل مسیرها با طول مسیر مورد نظر (lag)، به دست می‌آیند. نوعی از این توصیفگرها گروه Autocorrelation Geary است که ضریب گری نام دارد و بدین صورت محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (۴-۱)$$

که w یک ویژگی اتم، $\bar{w}$ میانگین مقدار آن ویژگی روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} نیز که به تابع کرونکر^۲ معروف است، در حالتی که $d=d_{ij}$ باشد مقدار این تابع یک است، و در غیر اینصورت صفر است و Δ نیز مجموعه‌ی δ هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است. از میان این توصیفگرها، GATS2e و GATS3e وارد مدل برتر شدند که نقش خاصیت فیزیکی و الکترونگاتیویته‌ی ساندerson را بیان می‌کنند. هر دو

^۱ -Reciprocal distance Wiener-type index

^۲ - Kronecker

توصیف‌گرها دارای اثر متوسط منفی بوده، بدین معنی که افزایش مقادیر این توصیف‌گرها باعث کاهش شاخص بازداری می‌شود [34و35].

• توصیف‌گرهای 3D - MORSE

توصیف‌گرهای 3D - Morse (نمایش سه بعدی ساختار مولکول براساس تفرق الکترون) از طریق معادله تبدیلی که در پراش الکترون استفاده می‌شود، محاسبه می‌گردند:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (۴-۲)$$

I شدت الکترون پراکنده شده، A_i و A_j خاصیت اتمی اتم i و j، S زاویه پراکندگی، r_{ij} فاصله بین اتم‌های i و j و N نیز تعداد کل اتم‌ها را نشان می‌دهد. این روش باعث می‌شود که ساختار سه بعدی مولکول به یک کد ثابت تبدیل شود [34و35].

برخی از این توصیف‌گرهای سه بعدی مانند Mor12u^۲ در مدل برتر دیده شود. چون این توصیف‌گرها آرایش سه بعدی اتم‌ها را بدون اینکه به اندازه مولکول ارتباط داشته باشند، بیان می‌کنند بنابراین برای تعداد زیادی مولکول با تفاوت‌های ساختاری زیاد قابل کاربرد است. با افزایش Mor12u شاخص بازداری نیز افزایش می‌یابد.

• دما

دما معیاری از سردی و گرمی مواد را مشخص می‌کند. در حقیقت دما خاصیتی است که جهت جریان انرژی را مشخص می‌کند. انرژی به صورت گرما از ناحیه‌ای با دمای بیشتر به ناحیه‌ای با دمای

^۱ - 3D- Molecular Representation of Structure based on Electron diffraction

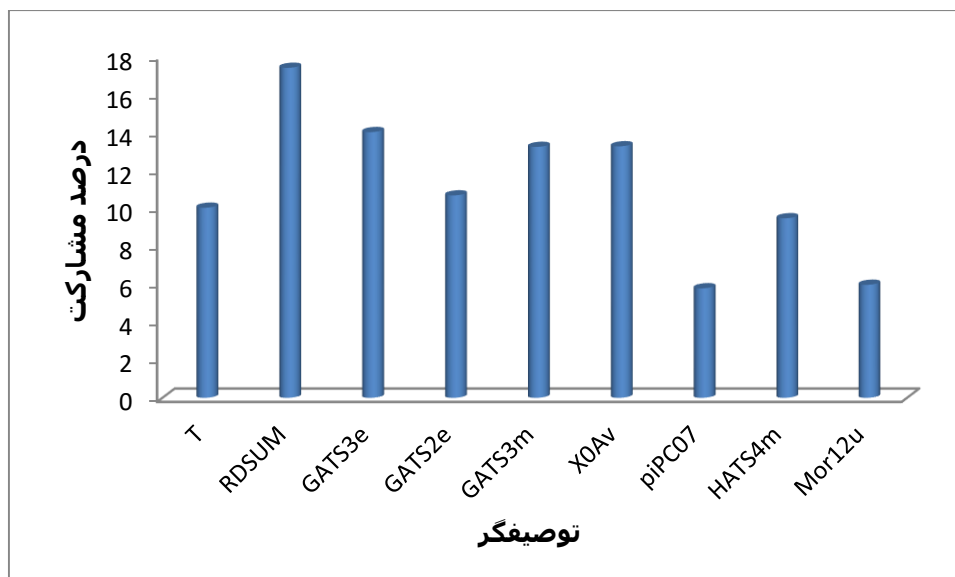
^۲ -3D-MoRSE- signal 12/ unweighted

کمتر جریان می‌یابد. با افزایش دما فاصله مولکول‌ها از یکدیگر افزایش یافته در نتیجه، حجم مواد زیاد می‌شود، اما جرم آن‌ها ثابت می‌ماند، پس با افزایش دما حجم افزایش و دانسیته کاهش می‌یابد.

در مدل خطی حاصل شده نیز، دما علامت مثبت دارد و رابطه‌ی مستقیم با شاخص بازداری نشان می‌دهد. یعنی با افزایش دما، میزان شاخص بازداری افزایش می‌یابد.

۷-۸-۲-۴ - بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

درصد مشارکت توصیفگرهای منتخب در ترکیبات مورد بررسی در شکل (۴-۱۷) نشان داده شده است. بر اساس این نمودار توصیفگرهای RDSUM، GAT3e، GAT3m و X0AV دارای بیشترین اثر مشارکت می‌باشند. توصیفگرهای GAT3e و GAT3m در برگرنده اطلاعات دو بعدی مولکول هستند. RDSUM و X0AV توصیفگرهای توپولوژی مولکول‌ها هستند. بنابراین می‌توان گفت که شاخص بازداری این ترکیبات وابستگی قابل ملاحظه‌ای به توپولوژی ترکیب دارد.



شکل (۴-۱۷) - مشارکت توصیفگرها در شبکه عصبی بهینه

4-2-9- نتیجه‌گیری نهایی

در این تحقیق برای اولین بار از روش‌های SR-ANN، SR-MLR و GA-ANN برای پیش‌بینی شاخص بازدارندگی یکسری از ترکیبات (آلکان‌ها، الکل‌ها، آلدئیدها و ...) در دماهای متفاوت به کار گرفته شد. ارزیابی مدل بویژه با سری تست نشان دهنده صحت و اعتبار مدل‌های پیشنهادی است بنابراین بر اساس این مدل‌ها می‌توان شاخص بازدارندگی را برای همین ترکیبات که هنوز سنتز نشده‌اند، پیش‌بینی کرد. در واقع این روش به محققان این امکان را می‌دهد که قبل از اینکه بخواهند یک ترکیب را سنتز کنند، ابتدا با استفاده از مدل معتبر شاخص بازدارندگی آن را پیش‌بینی کنند و در صورتی که دارای شاخص بازدارندگی مناسبی باشند، ترکیب سنتز شود و بدین ترتیب در وقت و هزینه صرفه جویی خواهد شد.

آینده نگری

❖ از آن جایی که در تحقیق حاضر توصیفگرهای به دست آمده از روش رگرسیون خطی چندگانه و الگوریتم ژنتیک به عنوان ورودی شبکه عصبی مصنوعی قرار گرفته است، می‌توان از روش‌هایی همچون SVM¹ و LS-SVM² به جای شبکه عصبی استفاده کرد.

❖ می‌توان روش‌های خطی دیگر مانند حداقل مربعات جزئی و آنالیز اجزای اصلی را جایگزین روش رگرسیون خطی چندگانه کرد و نتایج به دست آمده را با هم مقایسه کرد.

¹ - Supported Vector Machine

² -Least Square Supported Vector Machine

همچنین می‌توان از روش‌های نوین مانند ¹CoMFA و ²CoMSIA برای پیش‌بینی شاخص بازدارندگی ترکیبات مورد مطالعه استفاده کرد.

¹-Comparative Molecular Field Analysis

² - Comparative Molecular Similarity Indices Analysis

Reference:

- [1]-Gelder, Michael (2005). "Psychiatry", P. 12. Oxford University Press Inc., New York.
- [2]- Jaspers, Karl (1997-11-27) [1963]. *Allgemeine Psychopathologie (General Psychopathology)*. Translated by J. Hoenig & M.W. Hamilton from German (Reprint ed.). Baltimore, Maryland: Johns Hopkins University Press. ISBN 0-8018-5775-9.
- [3]- Hun Yeong Koh (2003). "QSAR Studies on Piperazinylalkylisoxazole Analogues Selectively Acting on Dopamine D3 Receptor by HQSAR and CoMFA". *Bioorganic & Medicinal Chemistry, Volume 11, 2003, Pages 1293–1298*.
- [4]- Thomas Balle, Kim Andersen, Karina Krojer Soby, Tommy Liljefors (2003). "α₁ Adrenoceptor subtype selectivity: 3D-QSAR models for a new class of α₁ adrenoceptor antagonists derived from the novel antipsychotic sertindole". *Journal of Molecular Graphics and Modelling, Volume 21, Issue 6, June 2003, Pages 523-534*.
- [5]- M. Fernandez, A. Tundidor- Camba, J Caballero, (2005) "Modelling of cyclindependent Kinase inhibition by 1H-pyrazolo[3,4-d] pyrimidine derivatives using artificial neural network ensembles" *J. Chem. Inf. Model.* 45, 6, pp 1884-1895.
- [6]- Ae Nim Pae. (2006) "Classification of dopamine, serotonin, and dual antagonists by decision trees." *Bioorganic & Medicinal Chemistry, Volume 14, June 2006, Pages 2763–2770*.
- [7]-Jalali-Heravi M., Mani-Varnosfaderani A., (2009),"QSAR modeling of 1-(3,3-diphenylpropyl)-piperidinyl amides as CCR₅ modulators using multivariate adaptive regression spline and Bayesian regularization genetic neural networks", *J. QSAR Comb. Sci.*, 9, pp 946-958.
- [8]- Arab Chamjangali M., (2009), "Modeling of cytotoxicity data (CC₅₀) of anti-HIV 1-[5-chlorophenyl) sulfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg- Marquardt artificial neural network", *J.Chem. Bio. Drug. Des.*, 73, pp 456-465.
- [9]- Almerico A.M, Tutone M., Lauria A., (2009),"A QSAR study investigating the potential anti-HIV-I effect of some Acyclovir and Ganciclovir analogs", *ARKIVOC*, pp 85-94.
- [10]- Patel A., Gupta L., Karthikeyan C., Trivedi P.(2010), "QSAR studies on dihydro-alkoxy-benzyl-oxopyrimidines(DABOs) derivatives, as new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors" *J. Current Pharmaceutical Research* , 1, pp 19-25.
- [11] -Horwery, D. G., Hirsch, R.F., 1983, *Chemometrics in the chemistry curriculum*, *j. Chem*, Ed. 60, pp: 656-659.
- [12]-He,L.; JRS, P.C.J. *Mol*, 2005,Graph. Model.23, pp: 503.
- [13]-Introduction to molecular descriptors, <http://www.qsar.pe.kr.descriptors>.
- [14]-Charton M., Charton B. (2002) "Advances in Quantitative Structure-Property Relationship", University of Bristol, UK, 3, pp 140.

[15] - فرشادفر ع، (۱۳۸۰)، "اصول و روشهای پیشرفته آماری (تجزیه رگرسیون)"، چاپ دوم، انتشارات طاق بستان.

[16] - مرکز انفورماتیک و توسعه، (۱۳۷۳)، "راهنمای نرم افزار SPSS/PC+"، چاپ اول، انتشارات راهبرد.

[17]- منهای م، (۱۳۸۷) " مبانی شبکه‌های عصبی (هوش محاسباتی)" جلد اول، مرکز نشر دانشگاه صنعتی امیر کبیر، تهران.

[18]- McCulloch W.W. and Pitts W. (1943) "A Logical Calculus of the Ideas Imminent in Nervous Activity", *Bull. Math. Biophys.*, 5, pp 115-133.

[19]- Rosenblatt F. (1958) "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Cornell Aeronautical Lab, Psychological Review, 65, 6, pp 386-408.

[20]- Rumelhart D.E. and McClelland J.L., (1986), "Parallel Distributed Processing: Exploration in the Microstructure of Cognition", 1, Cambridge, MA:MIT Press, Chapter 1.

[21]- هاگان. ت. ه.، دیموث. ه.، بیل م.، "طراحی شبکه‌های مصنوعی"، کی.م. انتشارات کیان رایانه سبز.

[22]- کی.م. (۱۳۸۷) "شبکه‌های عصبی در MATLAB"، چاپ دوم، انتشارات کیان رایانه سبز.

[23]- Caballero J.; Fernandez M.; (2006), "Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks"; *J. Mol. Model*; 12, pp 168-181.

[24]- باوی. ا.، صالحی. م. (۱۳۸۷) "الگوریتم ژنتیک و بهینه‌سازی سازه‌های سازه‌های مرکب"، انتشارات عابد.

[25]- Chen D. (1997) "Using Genetic Algorithms for the optimal design of structural systems", Arizona state university, pp 70-93.

[26]- Imam H. and Al-Shihri M. A. (2000) "A primitive crossover for improving the reliability of genetic algorithms for structural optimization, in Computational Engineering Using Metaphors from Nature", Civil-Comp press, Edit. B.H.V. Topping, pp 91-97.

[27]- Erbatur F., Hasancebi O., Tutuncu I. Kilic H. (2000) "Optimal design of planar and space structures with Genetic algorithms", *Computers and Structures*, 75, pp 209-224.

[28]- HyperChem 7.0 Toronto, Canada: HyperCube Inc, [http:// www.hyper.com](http://www.hyper.com).

[29]- Todeschini R. Milano Chemometrics and QSPR Group, <http://www.disat.unimib.it.vhtml>.

[30]- SPSS for windows Statistical package for IBM PC, SPSS Inc, <http://www.spss.com>.

[31]- MATLAB 7.8, the Math Work, Inc., Natick, MA, USA.

[32]- Atsuro Nakazato, Kohmei Ohta, Yoshinori Sekiguchi, Shigeru Okuyama, Shigeyuki Chaki, Yutaka Kawashima, and Katsuo Hatayama, (1998), "Design, Synthesis, Structure-Activity Relationships, and Biological Characterization of Novel Arylalkoxyphenylalkylamine ó Ligands as Potential Antipsychotic Drugs" *J. Med. Chem.* volume 42, issues, pp 1076-1087.

[33]- Zupan J.J., Gasteiger J., (1993), "Neural networks for chemists an introduction", VCH Publishers, Weinheim, Germany.

[34]- Todeschini R., Consonni V., (2000), "Handbook of molecular descriptors", Wiley-VCH, Weinheim, Germany.

- [35]- Marjanovic M., Kralj M., Supek F., Frkanec L., Piantunida I., Smuc T., Tusek-Bozic L., (2007), "*Antitumor potential of crown ether: structure-activity relationship, cell cycle disturbances, and cell death studies of a series of Ionophores*", J. Med. Chem., 50, pp 1007-1018.
- [36]- Carluci G., D'Archivio A.A., Maggi M.A., Mazzeo P., Ruggieri F., (2007)" *Investigation of retention behaviour of non-steroidal anti-inflammatory drugs in high performance liquid chromatography by using quantitative structure-retention relationships*", Anal.Chem.Acta, 601, pp:68-76.
- [37]- Sattari M., Gharagheizi F. (2008)"*Prediction of molecular diffusivity of pure components into air: A QSPR approach*", chemosphere, 72, pp:1298-1302.
- [38]- Luan F., Liu H.T., Ma W.P., Fan B.T.,(2008) "*QSPR analysis of air-to-blood distribution of volatile organic compounds*", Ecotoxicology and Environmental Safety, 71, pp:731-739.
- [39]- Luan F., Liu H.T., Wen Y., Zhang X., (2008)" *Quantitative structure- property relationship study for estimation of quantitative calibration factors of some organic compounds in gas chromatography*", Anal.Chem.Acta, 612, pp:126-135.
- [40]-Hemmateenejad B., Shamsipur M., Miri P., Elyasi M., Foroghnia F., Shaghi H.,(2008)Linear and nonlinear quantitative structure- property relationship models for solubility of some anthraquinone, anthrone and xanthone derivatives in supercritical carbon dioxide, Anal.Chem.Acta, 610, pp:25-34.
- [41]-Rosa Lebrón-Aguilar, Jesús Eduardo Quintanilla-López, Ana María Tello, José María Santiuste,(2007), "*Isothermal retention indices on poly(3,3,3-trifluoropropylmethylsiloxane) stationary phases*", Journal of Chromatography A, Volume 1160, Issues 1-2, 10 August, pp 276–288.

Astract

In the first section of this project, quantitative structure-activity relationship (QSAR) study was conducted on the inhibition effect (pIC₅₀) of 40 drug compounds of Arylalkoxyphenylalkylamine derivatives as inhibition of schizophrenia. The stepwise multiple linear regression and genetic algorithm methods were used to select the most important descriptors. Multiple linear regression (MLR) as a linear method and artificial neural network (ANN) as nonlinear method were used to predicting the inhibition effect of these compounds. The validation of the MLR and ANN models was performed using test set, leav-one-out and Y-randomization techniques. The obtained results are shown, determination coefficients for prediction of inhibition effect of the test set by SR-MLR, SR-ANN and GA-ANN models were 0/918, 0/951 and 0/835 respectively.

In the second part of this project, quantitative structure-property relationship (QSPR) study was developed to predict of the retention index (RI) of some organic compounds. The stepwise multiple linear regression and genetic algorithm methods were used to select the most important descriptors. Multiple linear regression (MLR) as a linear method and artificial neural network (ANN) as nonlinear method used to predicting the retention index. The validation of the MLR and ANN models was performed using test set, leave-group-out and Y-randomization techniques. The obtained results are shown, determination coefficients for prediction of retention index of the test set by SR-MLR, SR-ANN and GA-ANN models were 0/959, 0/999 and 0/999 respectively.

Keywords: Quantitative structure-activity relationship, Quantitative structure-property relationship ,inhibition effect, retention index, multiple linear regression, genetic algorithm, artificial neural network



Shahrood university of technology

Faculty of chemistry

**Prediction of Inhibition effects of some drug
compounds using QSAR methods**

Yalda Ghorban Ebrahimi

Supervisor:

Dr. N. Goudarzi

September 2012